

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

3/2021

Enabling Power Beacons and Wireless Power Transfers for Non-Orthogonal Multiple Access Networks

C.-B. Le and N. D. Nguyen

Paper

1

Evaluation of Radio Channel Utility using Epsilon-Greedy Action Selection

K. Malon

Paper

10

Developing RF Power Sensor Calibration Station in Direct Comparison Transfer System using Vector Network Analyzer

J. Szatkowski

Paper

18

Network Traffic Classification in an NFV Environment using Supervised ML Algorithms

G. Ilievski and P. Latkoski

Paper

23

A Shared Cybersecurity Awareness Platform

M. Amanowicz

Paper

32

Markov Decision Process based Model for Performance Analysis an Intrusion Detection System in IoT Networks

G. Kalnoor and Gowrishankar S

Paper

42

Linear and Planar Array Pattern Nulling via Compressed Sensing

J. R. Mohammed, R. H. Thaher, and A. J. Abdulqader

Paper

50

High Temperature Effects in Fused Silica Optical Fibers

K. Borzycki, M. Jaworski, and T. Kossek

Paper

56

(Contents Continued on Back Cover)

Editor-in-Chief

Adrian Kliks, *Poznan University of Technology, Poland*

Steering Editor

Jordi Mongay Batalla, *National Institute of Telecommunications, Poland*

Editorial Advisory Board

Hovik Baghdasaryan, *National Polytechnic University of Armenia, Armenia*

Naveen Chilamkurti, *LaTrobe University, Australia*

Luis M. Correia, *Instituto Superior Técnico, Universidade de Lisboa, Portugal*

Luca De Nardis, *DIET Department, University of Rome La Sapienza, Italy*

Nikolaos Dimitriou, *NCSR "Demokritos", Greece*

Ciprian Dobre, *Politechnic University of Bucharest, Romania*

Filip Idzikowski, *Poznan University of Technology, Poland*

Andrzej Jajszczyk, *AGH University of Science and Technology, Poland*

Albert Levi, *Sabancı University, Turkey*

Marian Marciniak, *National Institute of Telecommunications, Poland*

George Mastorakis, *Technological Educational Institute of Crete, Greece*

Constandinos Mavromoustakis, *University of Nicosia, Cyprus*

Klaus Mößner, *Technische Universität Chemnitz, Germany*

Imran Muhammad, *King Saud University, Saudi Arabia*

Mjumo Mzyece, *University of the Witwatersrand, South Africa*

Daniel Negru, *University of Bordeaux, France*

Ewa Orłowska, *National Institute of Telecommunications, Poland*

Jordi Perez-Romero, *UPC, Spain*

Michał Pióro, *Warsaw University of Technology, Poland*

Konstantinos Psannis, *University of Macedonia, Greece*

Salvatore Signorello, *University of Lisboa, Portugal*

Adam Wolisz, *Technische Universität Berlin, Germany*

Tadeusz A. Wysocki, *University of Nebraska, USA*

Publications Staff

Content Editor: **Robert Magdziak**

Managing Editor: **Ewa Kapuściarek**

Technical Editor: **Julia Malińska**

on-line: ISSN 1899-8852

© Copyright by National Institute of Telecommunications, Warsaw 2021

Enabling Power Beacons and Wireless Power Transfers for Non-Orthogonal Multiple Access Networks

Chi-Bao Le¹ and Nhan Duc Nguyen²

¹ Faculty of Electronics Technology, Industrial University of Ho Chi Minh City (IUH), Ho Chi Minh City, Vietnam

² Innovation Center, Van Lang University, Ho Chi Minh City, Vietnam

<https://doi.org/10.26636/jtit.2021.152421>

Abstract—This paper studies downlink cellular networks relying on non-orthogonal multiple access (NOMA). Specifically, the access point (AP) is able to harvest wireless power from the power beacon (PB). In the context of an AP facilitated with multiple antennas, the transmit antenna selection procedure is performed to process the downlink signal, with the transmission guaranteed by energy harvesting. Therefore, a wireless power transfer-based network is introduced to overcome power outages at the AP. In particular, an energy-constrained AP harvests energy from the radio frequency signals transmitted by the PB in order to assist in transmitting user data. Outage performance and ergodic capacity are evaluated with the use of closed-form expressions. In order to highlight some insights, approximate computations are provided. Finally, numerical simulations are performed to confirm the benefits of combining the downlink NOMA transmission and the transmit power scheme at the AP in order to serve a multitude of users.

Keywords—ergodic capacity, NOMA, outage probability, power beacon.

1. Introduction

The main requirements faced by 5G systems include the following: high demand for data-intensive services and improved bandwidth availability for cellular networks [1], [2]. For multiple accesses (MA) techniques in 5G, some non-orthogonal techniques such as power domain non-orthogonal multiple accesses (PD-NOMA) [3] and sparse code multiple accesses (SCMA) [4], are introduced. In the context of NOMA, PD-NOMA allocates a sub-carrier to multiple users at the same time, by employing superimposed coding on the transmitter side, while user signals are detected thanks to the successive interference cancellation (SIC) method adopted at the receiver side. In SCMA, each sub-carrier may be implemented by allocating different codebooks on the transmitter side, while simultaneously applying the message passing algorithm (MPA) for the receiver side to detect user signals.

PD-NOMA and SCMA have been recently considered, in numerous works, as appropriate candidates enabling to de-

ploy the MA technique in a 5G context, [5]–[9]. Ding *et al.* [5] studied PD-NOMA-based systems to evaluate their user pairing-related abilities. By pairing users who enjoy a good channel situation with those suffering from poor channel conditions, overall throughput of the system may be improved. Hanif *et al.* [6] proposed a multi-user, multiple-input multiple-output (MIMO) PD-NOMA-based system by relying the joint power allocation and precoding design. They proposed a method for achieving the maximum system sum rate. It has been reported in many recent works that wireless transfers of power from natural sources may be achieved, and numerous advancements concerning this technique have been reported [10]. For example, multi-user communication scenarios were recommended in emerging communication systems in order to implement the Internet of Things (IoT) and fifth-generation (5G) networks [11], [12].

The benefits of energy harvesting were reported in [13]–[20]. In paper [13], benefits for the operation of wireless networks, stemming from the wireless power transfer method were described. This work proposed a new expression for achieving optimal throughput in energy-aware cooperative systems with a general time-power energy harvesting protocol, namely time-power switching-based relaying (TPSR). In particular, the impact that relay node and destination node hardware imperfections exert on two-way relaying networks (TWRN) was shown. Interestingly, to maximize system throughput, an optimized policy for joint wireless information and energy transfers was determined by identifying optimal time switching and power splitting fractions.

In [16], the authors presented a small-cell network operating in the context of heterogeneous cellular networks, for both downlink (DL) and uplink (UL) scenarios, relying on three techniques, namely energy harvesting, full-duplex transmission mode, and the power domain-based NOMA scheme. Compared to the conventional half-duplex orthogonal multiple accesses (OMA) scheme that has been widely implemented in current wireless communication systems, the full-duplex (FD) NOMA relying on an energy harvest-

ing scheme offers great potential in terms of a further enhancement of the system's performance and has additional advantages, such as spectral efficiency, connectivity-related capabilities, and outage-related performance. In [18], the authors studied a simultaneous wireless information and power transfer for a NOMA network, with the relay being an energy-constrained device. The relay harvests energy from source radio frequency (RF) signals using the time-switching protocol. Both imperfect channel state information (ICSI) and residual hardware impairments (RHIs) were considered. To characterize these effects in the network under consideration, outage probability (OP) and throughput-related expressions were designed.

Motivated by the recent publications [13]–[20], we consider, in this study, a power beacon (PB) helping the access point (AP) transmit signal at the downlink portion of a NOMA system.

The remaining part of this paper is organized as follows. Section 2 introduces the principle of a downlink NOMA and describes how the signal may be processed and detected at each of the receivers. Section 3 provides an analysis of outage probability and some useful insights. Section 4 presents the results of simulations with two users, thus allowing us to confirm some of the comparisons made. A summary is provided in Section 5.

Table 1
Key parameters of the system model

Symbols	Description
a_i	Power allocation coefficient
P_S	Transmit power at AP
P_P	Transmit power at PB
$\bar{x}_i$	Information of U_i
R_i	Target rate at U_i
T	Total time used for energy harvesting and information processing
η	Energy harvesting efficiency
θ	Time switching factor

2. System Model

In Fig. 1, the AP is equipped with multiple antennas, i.e. N to guarantee the operation of such an AP on the downlink, while the PB transfers wireless energy to the AP. The AP transmits the superimposed signal to destinations U_1 , U_2 which will receive signal $x_S = \sqrt{a_1 P_S} x_1 + \sqrt{a_2 P_S} x_2$, in which a_1 and a_2 are power allocation factors. The condition applying to the factors is that $a_1 + a_2 = 1$, with $a_2 > a_1$. During the energy harvesting phase, the index of the best antenna may be determined and the energy received from the signal at the AP may be expressed as follows:

$$n^* = \arg \max_{n=1,2,\dots,N} \bar{X}, \quad \bar{X} \in \left\{ |h_{0,n}|^2, |h_{n,1}|^2, |h_{n,2}|^2 \right\}, \quad (1)$$

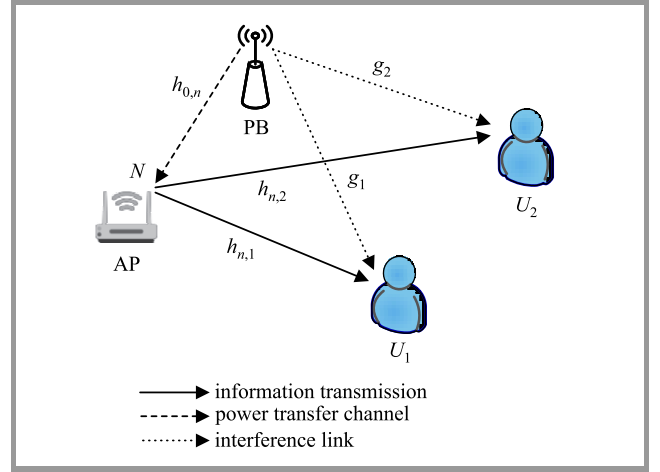


Fig. 1. Model of a power beacon-based downlink NOMA system.

and

$$y_S^n = \sqrt{P_P} h_{0,n} x_S + \omega_S^n, \quad (2)$$

where P_P is the transmit power at the PB, $h_{0,n}$ denotes the power transfer channel with $h_{0,n} \sim CN(0, \lambda_0)$, x_S is the energy signal vector satisfying the total power constraint $\mathbb{E}\{x_1^2\} = \mathbb{E}\{x_2^2\} = 1$ in which $\mathbb{E}\{\cdot\}$ is the expectation operator and ω_S^n denotes the additive white Gaussian noise (AWGN) with $\omega_S^n \sim CN(0, N_0)$.

In this model, the power-splitting energy harvesting technique is employed [13]. Hence, the total harvested energy at the end of the first phase is:

$$E = \eta P_P \theta T |h_{0,n}|^2, \quad (3)$$

where T is the block time in which a certain amount of information is transmitted from the AP node to two user nodes, θ is the fraction of the block time in which the AP harvests energy from the PB's information signal and η ($0 < \eta < 1$) denotes the energy conversion efficiency. In line with reports from several previous papers, we assume that all of the harvested energy is used during the information transmission phase. Hence, the transmit power of the source is:

$$P_S = \frac{E}{(1 - \theta)T} = \frac{\eta P_P \theta |h_{0,n}|^2}{1 - \theta}. \quad (4)$$

During the wireless information transfer (WIT) phase, the AP employs the harvested energy to convey independent signals, serving multiple users through the NOMA paradigm.

In this case, the PB uses the beamforming technique in order to enhance performance. As such, the signal received by two users may be written as:

$$y_{U_i}^n = h_{n,i} (\sqrt{a_1 P_S} \bar{x}_1 + \sqrt{a_2 P_S} \bar{x}_2) + \sqrt{P_P} g_i x_S + \omega_{U_i}^n, \quad (5)$$

where $h_{n,i}$, $i \in \{1, 2\}$ denotes the main channel from AP to U_i in which $h_{n,1} \sim CN(0, \lambda_1)$ and with $h_{n,2} \sim CN(0, \lambda_2)$, while g_i denotes interference from the PB to two users.

The elements of $h_{n,i}$ and g_i are independent and identically distributed zero-mean complex Gaussian random variables with variance λ_i and Ω_i , respectively. $\bar{x}_i$ denotes the source symbol with unit power with $\mathbb{E}\{\bar{x}_i^2\} = 1$. x_S is the jamming signal with unit power. Here, $\omega_{U_i}^*$ is the AWGN at user with variance N_0 . Therefore, the received signal-to-interference-plus-noise ratio (SINR) at U_1 to detect U_2 message x_2 is given by:

$$\begin{aligned} \gamma_{U_2 \rightarrow U_1} &\simeq \frac{a_2 P_S |h_{n^*,1}|^2}{a_1 P_S |h_{1,n^*}|^2 + P_P \Omega_1 + N_0} \\ &\simeq \frac{\varphi_2 |h_{0,n^*}|^2 |h_{n^*,1}|^2}{\varphi_1 |h_{0,n^*}|^2 |h_{n^*,1}|^2 + \Theta_1}, \end{aligned} \quad (6)$$

where $\rho = \frac{P_P}{N_0}$ is the transmit signal-to-noise ratio (SNR) at the source. $\varphi_1 = \frac{a_1 \eta \rho \theta}{(1-\theta)}$, $\varphi_2 = \frac{a_2 \eta \rho \theta}{(1-\theta)}$ and $\Theta_1 = \rho \Omega_1 + 1$.

We assume that $\mathbb{E}\{|g_i|^2\} \approx \Omega_i, i \in \{1, 2\}$.

After SIC, the received SINR at U_1 detecting its own message x_1 is:

$$\gamma_{U_1} \simeq \frac{\varphi_1 |h_{0,n^*}|^2 |h_{n^*,1}|^2}{\Theta_1}. \quad (7)$$

The received signal at U_2 is y_{U_2} in Eq. (5), and the SINR at U_2 is represented by:

$$\gamma_{U_2} \simeq \frac{\varphi_2 |h_{0,n^*}|^2 |h_{n^*,2}|^2}{\varphi_1 |h_{0,n^*}|^2 |h_{n^*,2}|^2 + \Theta_2}, \quad (8)$$

where $\Theta_2 = \rho \Omega_2 + 1$.

3. Outage Probability Analysis

3.1. Outage Probability of U_1

According to the NOMA protocol, the complementary events of an outage at U_1 can be explained by the fact that U_1 may detect x_2 as well as its own message x_1 . From the above description, the outage probability of U_1 can be defined as:

$$\begin{aligned} \mathcal{OP}_1 &= \Pr(\gamma_{U_2 \rightarrow U_1} < \gamma_{h2} \cup \gamma_{U_1} < \gamma_{h1}) \\ &= 1 - \Pr(\gamma_{U_2 \rightarrow U_1} > \gamma_{h2}, \gamma_{U_1} > \gamma_{h1}) \\ &= 1 - \Pr(|h_{0,n^*}|^2 |h_{n^*,1}|^2 > \chi), \end{aligned} \quad (9)$$

where the threshold SNRs are $\gamma_{h1} = 2^{\frac{R_1}{1-\theta}} - 1$, $\gamma_{h2} = 2^{\frac{R_2}{1-\theta}} - 1$ in which R_i is the target rate at U_i , $\delta_1 = \frac{\Theta_1 \gamma_{h1}}{\varphi_1}$, $\delta_2 = \frac{\gamma_{h2} \Theta_1}{\varphi_2 - \varphi_1 \gamma_{h2}}$ and $\chi = \max(\delta_1, \delta_2)$.

Before computing the outage probability, we assume that all channel models are followed by Rayleigh fading. The probability density function (PDF) of channel gains $\bar{X}$, $\bar{X} \in \{|h_{0,n}|^2, |h_{n,1}|^2, |h_{n,2}|^2\}$, i.e. $f_{\bar{X}}(x)$ is given by [22, Eq. (6)]:

$$f_{\bar{X}}(x) = \sum_{n=1}^N \binom{N}{n} \frac{n(-1)^{n-1}}{\lambda_{\bar{X}}} e^{-\frac{nx}{\lambda_{\bar{X}}}}. \quad (10)$$

Proposition 1: The following PDF of $f_{|h_{0,n^*}|^2 |h_{n^*,i}|^2}$, $i \in \{1, 2\}$ is calculated as:

$$\begin{aligned} f_{|h_{0,n^*}|^2 |h_{n^*,i}|^2}(x) &= \sum_{n_0=1}^N \sum_{n_i=1}^N \binom{N}{n_i} \binom{N}{n_0} \\ &\times \frac{n_i(-1)^{n_i+n_0-2}}{\lambda_i} \left[\frac{\lambda_i}{n_i} - 2\sqrt{\frac{n_0 \lambda_i x}{\lambda_0 n_i}} K_1 \left(2\sqrt{\frac{n_0 n_i x}{\lambda_0 \lambda_i}} \right) \right], \end{aligned} \quad (11)$$

where $K_n(x)$ is the first-order modified Bessel function of the second kind.

Proof 1: See Appendix A.

Proposition 2: The closed-form expression of approximated $\mathcal{OP}_1$ can be represented as:

$$\begin{aligned} \mathcal{OP}_1 &\approx 1 - \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} (-1)^{n_1+n_0-2} \\ &+ \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} \frac{n_1(-1)^{n_1+n_0-2}}{\lambda_1} \\ &\times \frac{\pi^2}{2K} \sum_{k=1}^K \sqrt{1 - \xi_k^2} \sec^2 \left(\frac{(\xi_k + 1)\pi}{4} \right) \Upsilon(\xi_k), \end{aligned} \quad (12)$$

where $\xi_k = \cos \left(\frac{\pi(2k-1)}{2K} \right)$, $\Xi(a) = \tan \left(\frac{(a+1)\pi}{4} \right) + \chi$ and $\Upsilon(a) = \sqrt{\frac{n_0 \lambda_1 \Xi(a)}{\lambda_0 n_1}} K_1 \left(2\sqrt{\frac{n_0 n_1 \Xi(a)}{\lambda_0 \lambda_1}} \right)$.

Proof 2: See Appendix B.

3.2. Outage Probability of U_2

In a similar way, the outage probability at the second user is computed as:

$$\begin{aligned} \mathcal{OP}_2 &= 1 - \Pr(\gamma_{U_2} > \gamma_{h2}) \\ &= 1 - \Pr(|h_{0,n^*}|^2 |h_{n^*,2}|^2 > \bar{\delta}_2), \end{aligned} \quad (13)$$

where $\bar{\delta}_2 = \frac{\gamma_{h2} \Theta_2}{\varphi_2 - \varphi_1 \gamma_{h2}}$.

Similarly to the manner in which $\mathcal{OP}_1$ was solved, the close-form expression of approximated $\mathcal{OP}_2$ may be computed as:

$$\begin{aligned} \mathcal{OP}_2 &\approx 1 - \sum_{n_0=1}^N \sum_{n_2=1}^N \binom{N}{n_2} \binom{N}{n_0} (-1)^{n_2+n_0-2} \\ &+ \sum_{n_0=1}^N \sum_{n_2=1}^N \sum_{q=1}^Q \binom{N}{n_2} \binom{N}{n_0} \frac{\sqrt{1 - \xi_q^2} n_2 \pi^2}{2Q \lambda_2} \\ &\times (-1)^{n_2+n_0-2} \Theta(\xi_q) \sec^2 \left(\frac{(\xi_q + 1)\pi}{4} \right), \end{aligned} \quad (14)$$

where $\xi_q = \cos \left(\frac{\pi(2q-1)}{2Q} \right)$, $\Phi(a) = \left[\tan \left(\frac{(a+1)\pi}{4} \right) + \bar{\delta}_2 \right]$ and $\Theta(a) = \sqrt{\frac{n_0 \lambda_2 \Phi(a)}{\lambda_0 n_2}} K_1 \left(2\sqrt{\frac{n_0 n_2 \Phi(a)}{\lambda_0 \lambda_2}} \right)$.

3.3. Approximate Expression of Outage Probability for Two Users

Based on the analytical results presented in Eq. (14) and Eq. (12), when $\rho \rightarrow \infty$, the approximate outage probabilities of two users with $\chi \approx 0$ and $\bar{\delta}_2 \approx 0$ are given as:

$$\begin{aligned} \mathcal{OP}_1^\infty &\approx 1 - \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} (-1)^{n_1+n_0-2} \\ &\times \left[1 - \frac{n_1 \pi^2}{2\lambda_1 K} \sum_{k=1}^K \sqrt{1 - \xi_k^2} \sec^2 \left(\frac{(\xi_k + 1)\pi}{4} \right) \right. \\ &\times \left. \sqrt{\frac{n_0 \lambda_1 \hat{\Xi}(\xi_k)}{\lambda_0 n_1}} K_1 \left(2 \sqrt{\frac{n_0 n_1 \hat{\Xi}(\xi_k)}{\lambda_0 \lambda_1}} \right) \right], \end{aligned} \quad (15)$$

and

$$\begin{aligned} \mathcal{OP}_2^\infty &\approx 1 - \sum_{n_0=1}^N \sum_{n_2=1}^N \binom{N}{n_2} \binom{N}{n_0} (-1)^{n_2+n_0-2} \\ &\times \left[1 - \frac{n_2 \pi^2}{2\lambda_2 Q} \sum_{q=1}^Q \sqrt{1 - \xi_q^2} \sec^2 \left(\frac{(\xi_q + 1)\pi}{4} \right) \right. \\ &\times \left. \sqrt{\frac{n_0 \lambda_2 \hat{\Xi}(\xi_q)}{\lambda_0 n_2}} K_1 \left(2 \sqrt{\frac{n_0 n_2 \hat{\Xi}(\xi_q)}{\lambda_0 \lambda_2}} \right) \right], \end{aligned} \quad (16)$$

where $\hat{\Xi}(a) = \lceil \tan((\alpha + 1)\frac{\pi}{4}) + 1 \rceil$.

Next, in order to obtain more insights, diversity orders for U_1 and U_2 are investigated. In particular, they may be defined as [15, Eq. (18)]

$$d_a = - \lim_{\rho \rightarrow \infty} \frac{\log(\mathcal{OP}_a^\infty)}{\log(\rho)}, a \in \{1, 2\}. \quad (17)$$

By substituting Eqs. (16) and (15) into Eq. (17), the diversity orders of U_1 and U_2 are given as:

$$d_1 = d_2 = 0. \quad (18)$$

Remark. From Eq. (18) at high SNR, the outage probability of U_1 and U_2 approaches a fixed non-zero constant, indicating that outage performance error floors exist.

3.4. Throughput for Two Users

Based on achievable outage probability, throughput in the delay-limited transmission mode may be evaluated as [16, Eq. (42)]:

$$\tau_\star = (1 - \mathcal{OP}_\star) R_\star, \star \in \{1, 2\}. \quad (19)$$

4. Ergodic Capacity for Two Users

In this section, the ergodic capacity is further metric evaluated at U_2 as [23]:

$$\mathcal{C}_2 = (1 - \theta) \mathbb{E}[\log_2(1 + \gamma_{U_2})]. \quad (20)$$

Proposition 3: The closed-form expression for ergodic rate of $\mathcal{C}_2$ is given by:

$$\begin{aligned} \mathcal{C}_2 &= \sum_{n_0=1}^N \sum_{n_2=1}^N \binom{N}{n_0} \binom{N}{n_2} \frac{(1-\theta)(-1)^{n_0+n_2-2}}{\ln 2} \\ &\times \left[G_{1,3}^{3,1} \left(\tilde{\Psi} \middle| \begin{matrix} 0 \\ 0, 1, 0 \end{matrix} \right) - G_{1,3}^{3,1} \left(\tilde{\Psi} \middle| \begin{matrix} 0 \\ 0, 1, 0 \end{matrix} \right) \right], \end{aligned} \quad (21)$$

where $G_{p,q}^{m,n}[\cdot]$ is the Meijer G-function given in [21, Eq. (9.301)], $\tilde{\Psi} = \frac{n_2 n_0 \Theta_2}{\lambda_2 \lambda_0 (\varphi_2 + \varphi_1)}$ and $\tilde{\Psi} = \frac{n_2 n_0 \Theta_2}{\lambda_2 \lambda_0 \varphi_1}$.

Proof 3: See Appendix C.

Next, the ergodic capacity at U_1 is given as:

$$\begin{aligned} \mathcal{C}_1 &= (1 - \theta) \mathbb{E}[\log_2(1 + \gamma_{U_1})] \\ &= (1 - \theta) \mathbb{E} \left[\log_2 \left(1 + \underbrace{\frac{\varphi_1 |h_{0,n^*}|^2 |h_{n^*,1}|^2}{\Theta_1}}_X \right) \right] \\ &= \frac{(1 - \theta)}{\ln 2} \int_0^\infty \frac{1 - F_X(x)}{1 + x} dx. \end{aligned} \quad (22)$$

Based on the formula (31) from appendix C, $F_X(x)$ is:

$$\begin{aligned} F_X(x) &= 1 - \Pr \left(|h_{0,n^*}|^2 |h_{n^*,1}|^2 > \frac{\Theta_1 x}{\varphi_1} \right) \\ &= 1 - 2 \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_0} \binom{N}{n_1} (-1)^{n_0+n_1-2} \\ &\times \sqrt{\frac{n_2 n_0 \Theta_1 x}{\lambda_2 \lambda_0 \varphi_1}} K_1 \left(2 \sqrt{\frac{n_2 n_0 \Theta_1 x}{\lambda_2 \lambda_0 \varphi_1}} \right). \end{aligned} \quad (23)$$

By replacing Eq. (23) with Eq. (22) and using [21, Eq. (7.811.5), (9.34.3)], $\mathcal{C}_1$ may be defined as:

$$\begin{aligned} \mathcal{C}_1 &= \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_0} \binom{N}{n_1} \frac{(1 - \theta)(-1)^{n_0+n_1-2}}{\ln 2} \\ &\times G_{1,3}^{3,1} \left(\frac{n_2 n_0 \Theta_1}{\lambda_2 \lambda_0 \varphi_1} \middle| \begin{matrix} 0 \\ 0, 1, 0 \end{matrix} \right). \end{aligned} \quad (24)$$

5. Numerical Results

In this section, we simulate some theoretical results illustrated by specific figures in order to show the outage performance of the system. The key parameters are presented in Table 2. Additionally, the Gauss-Chebyshev parameter is selected as $K = Q = 100$ to yield a close approximation. Figure 2 shows an improvement in outage performance at high transmit SNR. One may notice that lower outage probability rates may be achieved at SNR greater than 40 dB. The outage performance of the second user outperforms that of the first user. The higher the number of antennas at the AP, the better the outage behavior. In particular, performance is the best for the second user. It is further confirmed that the approximated outage performance is very close to the exact value at high SNR. The differences concerning

Table 2
Definition of system parameters [25]

Parameter	Notation	Values
Power splitting factors	a_1, a_2	0.1, 0.9
Target SINR rates to decode x_1 and x_2	$R_1 = R_2$	0.5
Energy conversion efficiency	η	0.8
Fraction of the block time	θ	0.2
Exponential distribution random variables	λ_0	1
	λ_1	0.4
	λ_2	0.6
	$\Omega_1 = \Omega_2$	0.01

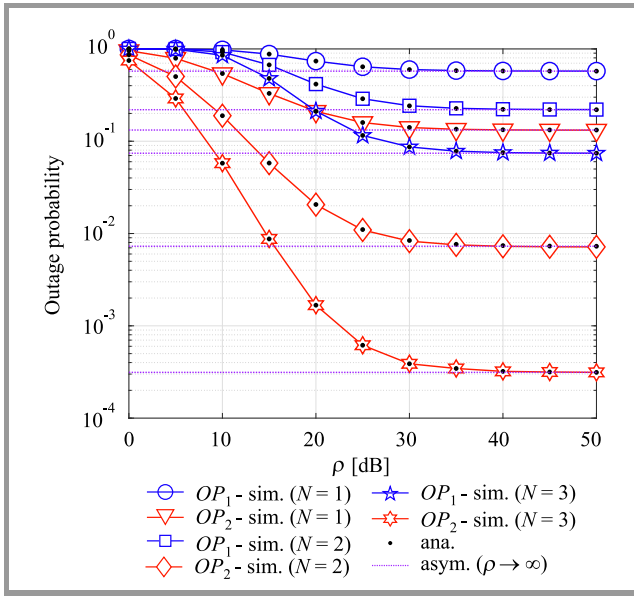


Fig. 2. Outage probability versus transmit SNR at AP with varying number of transmit antennas.

performance of two users exist throughout the entire SNR range. This is caused primarily by the fact that different power allocation factors are assigned to the individual users. In order to consider the impact that interference from the PB has on the two users, the outage probability of NOMA in the downlink mode is shown in Fig. 3. The performance of NOMA improves remarkably with the different power levels of interference channels. This is a promising result, as lower interference boosts outage performance. These outage trends for two users are similar to the trend shown in Fig. 2. Consequently, limitation of the impact of interference channels contributes to enhancing the performance of a NOMA system.

As can be seen from Fig. 4, with the increase of SNR, the outage performance of NOMA is still better than that of OMA as in Fig. 4. Interestingly, the saturation happens

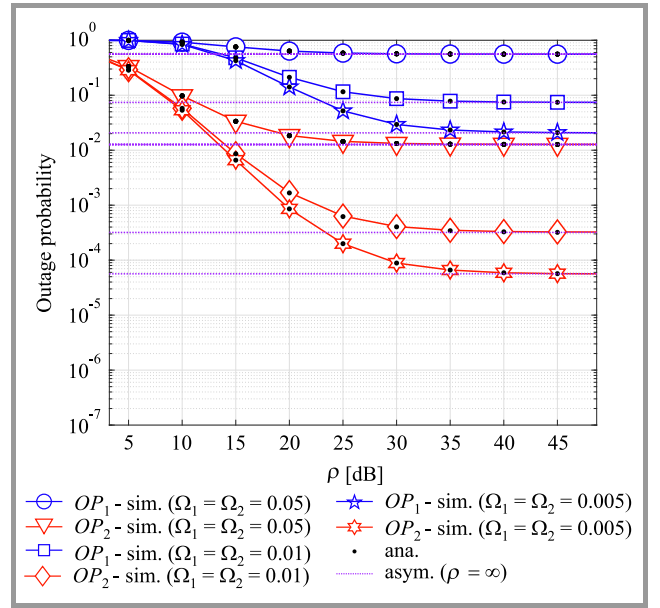


Fig. 3. Outage performance of two users with different levels of interference and $N = 3$.

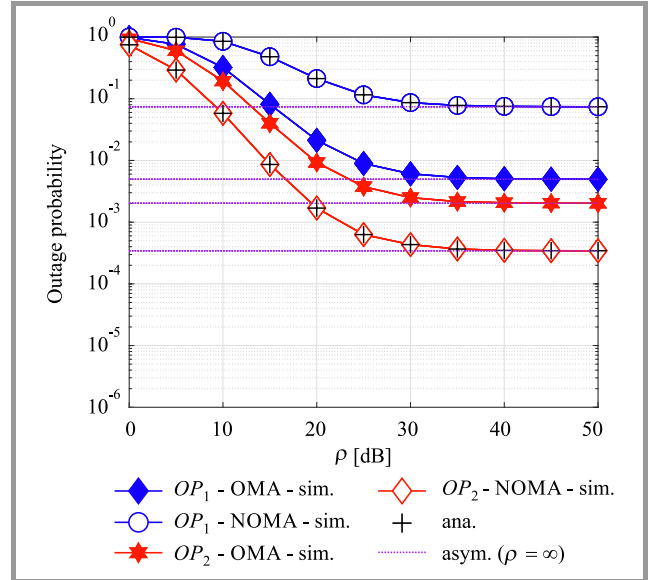


Fig. 4. Comparison between OMA and NOMA with $N = 3$.

at high SNR. This result indicates that NOMA shows its superiority compared with OMA.

In Fig. 5, one may observe that a large amount of energy harvested influences the outage performance of two users. At a high region of θ , higher harvested energy leads to higher SINR, then outage performance can be improved significantly. Specifically, outage performance may be changed significantly in the case of $N = 3$. Therefore, a power beacon plays an important role in improving system performance.

As seen from Fig. 6, the throughput of two users increases significantly along with the increase in data rates. There are specific optimal points along these throughput performance

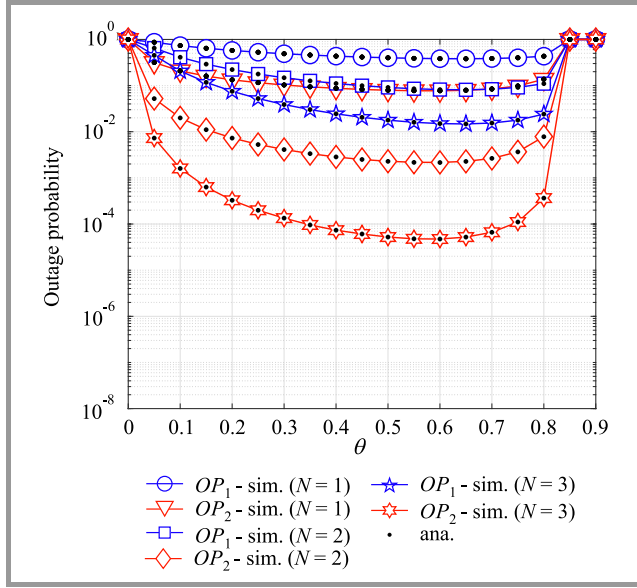


Fig. 5. Outage performance of two users vs. θ , with $\rho = 40$ dB.

curves. The throughput-related result is consistent with outage performance of two users shown in the previous figures.

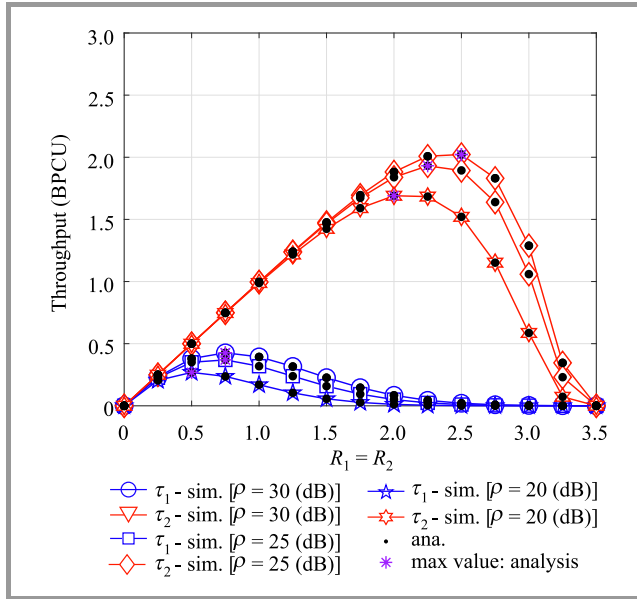


Fig. 6. Throughput performance of two users, with $N = 3$, $a_1 = 0.05$, $a_2 = 0.95$ and $\rho = 40$ dB.

Specific ergodic capacity trends may be noticed with SNR increasing from 0 to 50, as illustrated in Fig. 7. It is obvious that high SNR rates lead to better ergodic performance, but ergodic capacity increases significantly in the low SNR range only, as demonstrated by the fact that the lines remain unchanged at high SNR rates. In the considered cases, the highest ergodic capacity may be observed when the AP is facilitated with $N = 3$ antennas. This is caused by the fact that the selected antennas contribute to the improvement in ergodic capacity.

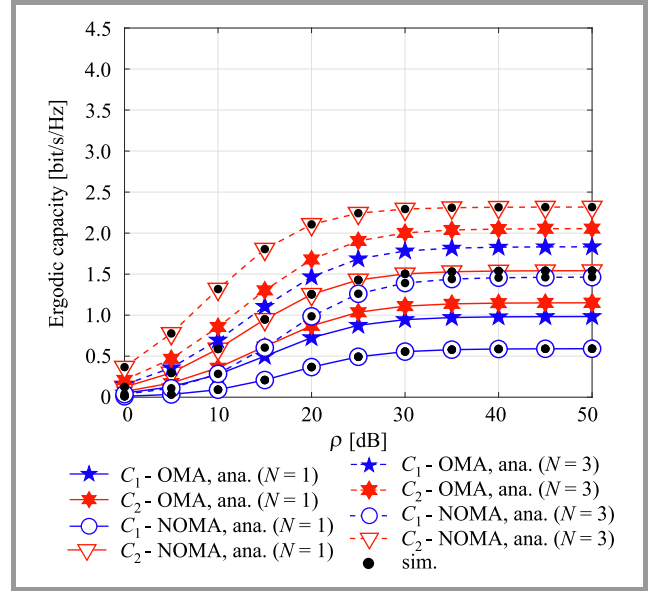


Fig. 7. Ergodic capacity versus transmit SNR.

Variations in the power splitting coefficient θ affect the ergodic capacity as well as illustrated in Fig. 8. It is worth noting that user U_2 in the NOMA mode shows superior performance for two cases $N = 1, 3$. This is caused mainly by the fact that SINR depends on the level of power used to transmit signals at the AP. In particular, the power splitting coefficient of θ results in higher transmit power at the AP.

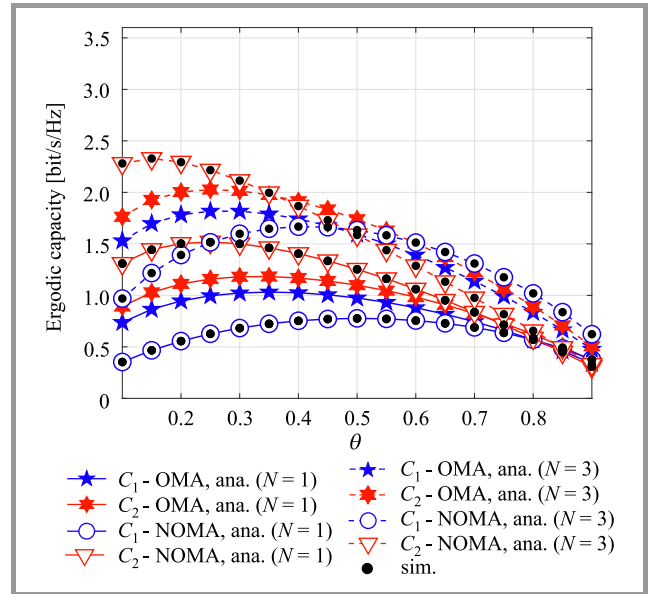


Fig. 8. Ergodic capacity versus power splitting coefficient θ , with $\rho = 30$ dB.

Similarly to Fig. 8, ergodic capacity may be improved once the number of transmit antennas at the AP is increased, as shown in Fig. 9. It is noteworthy that the ergodic capacity of user U_1 in the NOMA mode is likely affected by varying N .

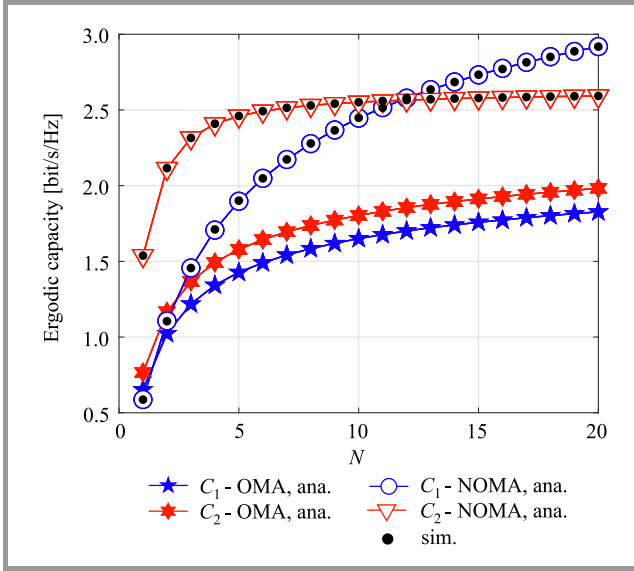


Fig. 9. Ergodic capacity versus N , with $\rho = 30$ dB.

6. Conclusion

In this study, a downlink of NOMA relying on the WIT scheme was considered. Specifically, by employing energy from the PB, the AP assists in transmitting the signal to two users in the downlink mode. Performance of the system depends on the amount of power harvested, meaning that larger numbers of AP antennas contribute to improvement in performance. Furthermore, performance of the system may be boosted by limiting the impact of interference. Finally, we have analyzed performance of the NOMA network and derived closed-form expressions for outage probability, throughput, and ergodic capacity. It has been shown that the theoretical approximations align perfectly with the simulation results. Additionally, the impact of the WIT strategy is analyzed. For example, a significant improvement is achieved in terms of outage behavior, throughput and ergodic capacity.

Appendix A

From formula (10), we have $f_{|h_{0,n^*}|^2|h_{n^*,i}|^2}(x)$ are the PDF of two random variables (RVs) $|h_{0,n^*}|^2|h_{n^*,i}|^2$ with $i \in \{1, 2\}$ is calculated by:

$$\begin{aligned} f_{|h_{0,n^*}|^2|h_{n^*,i}|^2}(x) &= \Pr\left(|h_{0,n^*}|^2 < \frac{x}{|h_{n^*,i}|^2}\right) \\ &= \int_0^\infty f_{|h_{n^*,i}|^2}(y) \int_0^\infty f_{|h_{0,n^*}|^2}(z) dy dz \\ &= \sum_{n_0=1}^N \sum_{n_i=1}^N \binom{N}{n_i} \binom{N}{n_0} \frac{n_i(-1)^{n_i+n_0-2}}{\lambda_i} \end{aligned}$$

$$\begin{aligned} &\times \int_0^\infty e^{-\frac{n_i y}{\lambda_i}} \left[1 - e^{-\frac{n_0 x}{\lambda_0 y}}\right] dy \\ &= \sum_{n_0=1}^N \sum_{n_i=1}^N \binom{N}{n_i} \binom{N}{n_0} \frac{n_i(-1)^{n_i+n_0-2}}{\lambda_i} \\ &\times \left[\int_0^\infty e^{-\frac{n_i y}{\lambda_i}} dy - \int_0^\infty e^{-\frac{n_i y}{\lambda_i} - \frac{n_0 x}{\lambda_0 y}} dy \right]. \end{aligned} \quad (25)$$

Based on [14, Eq. (3.324.1)] and by applying some polynomial expansion manipulations, we obtain the statistic function of two RVs $|h_{0,n^*}|^2|h_{n^*,i}|^2$ as in Proposition 1.

Appendix B

Using the PDF from Eq. (11), the outage probability $\mathcal{OP}_1$ may be expressed by:

$$\begin{aligned} \mathcal{OP}_1 &= 1 - \Pr(|h_{0,n^*}|^2|h_{n^*,1}|^2 > \chi) \\ &= 1 - \int_\chi^\infty f_{|h_{0,n^*}|^2|h_{n^*,1}|^2}(x) dx \\ &= 1 - \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} \frac{n_1(-1)^{n_1+n_0-2}}{\lambda_1} \\ &\times \int_\chi^\infty \left[\frac{\lambda_1}{n_1} - 2\sqrt{\frac{n_0 \lambda_1 x}{\lambda_0 n_1}} K_1\left(2\sqrt{\frac{n_0 \lambda_1 x}{\lambda_0 \lambda_1}}\right) \right] dx \\ &= 1 - \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} (-1)^{n_1+n_0-2} \\ &+ 2 \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} \frac{n_1(-1)^{n_1+n_0-2}}{\lambda_1} \\ &\times \underbrace{\int_\chi^\infty \sqrt{\frac{n_0 \lambda_1 x}{\lambda_0 n_1}} K_1\left(2\sqrt{\frac{n_0 \lambda_1 x}{\lambda_0 \lambda_1}}\right) dx}_{\mathcal{A}}. \end{aligned} \quad (26)$$

Unfortunately, identifying a closed-form expression for $\mathcal{A}$ is a tough task, but an accurate approximation may be obtained instead. With variable $x = \tan[(t+1)\pi/4] + \chi \Rightarrow dx = \pi \sec^2[(t+1)\pi/4]/4$ and the Gaussian-Chebyshev quadrature from [24, Eq. (25.4.38)], $\mathcal{A}$ can be represented as:

$$\begin{aligned} \mathcal{A} &= \frac{\pi}{4} \int_{-1}^1 \sec^2\left[\frac{(t+1)\pi}{4}\right] \sqrt{\frac{n_0 \lambda_1 (\tan[(t+1)\pi/4] + \chi)}{\lambda_0 n_1}} \\ &\times K_1\left(2\sqrt{\frac{n_0 \lambda_1 (\tan[(t+1)\pi/4] + \chi)}{\lambda_0 \lambda_1}}\right) dt \\ &\approx \frac{\pi^2}{K4} \sum_{k=1}^K \sqrt{1 - \xi_k^2} \sec^2\left[\frac{(\xi_k + 1)\pi}{4}\right] \Upsilon(\xi_k), \end{aligned} \quad (27)$$

where $\sec^2(t) = \frac{1}{\cos^2(t)}$, $\xi_k = \cos\left(\frac{\pi(2k-1)}{2K}\right)$, $\Xi(t) = \tan\left[\frac{(t+1)\pi}{4}\right] + \chi$ and $\Upsilon(t) = \sqrt{\frac{n_0\lambda_1\Xi(t)}{\lambda_0n_1}} K_1\left[2\sqrt{\frac{n_0n_1\Xi(t)}{\lambda_0\lambda_1}}\right]$. Combining Eq. (27) into Eq. (26), $\mathcal{OP}_1$ can be identified by:

$$\begin{aligned} \mathcal{OP}_1 \approx & 1 - \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} (-1)^{n_1+n_0-2} \\ & + \sum_{n_0=1}^N \sum_{n_1=1}^N \binom{N}{n_1} \binom{N}{n_0} \frac{n_1(-1)^{n_1+n_0-2}}{\lambda_1} \\ & \times \frac{\pi^2}{2K} \sum_{k=1}^K \sqrt{1 - \xi_k^2} \sec^2\left[\frac{(\xi_k+1)\pi}{4}\right] \Upsilon(\xi_k). \end{aligned} \quad (28)$$

Proof 2 is completed.

Appendix C

By the definition of the expectation operator and after integration-by-part, $\mathcal{C}_2$ can then be evaluated as [23], [26]:

$$\begin{aligned} \mathcal{C}_2 & \triangleq (1-\theta) E[\log_2(1 + \gamma_{U_2})] \\ & = \frac{1-\theta}{\ln 2} \int_0^\infty \frac{1}{1+x} \left[1 - F_{|h_{0,n^*}|^2 |h_{n^*,2}|^2}(x)\right] dx \\ & = \frac{1-\theta}{\ln 2} \int_0^{\frac{\varphi_2}{\varphi_1}} \frac{1}{1+x} \bar{F}_{|h_{0,n^*}|^2 |h_{n^*,2}|^2} \left(\frac{x\Theta_2}{\varphi_2 - x\varphi_1}\right) dx, \end{aligned} \quad (29)$$

where $\bar{F}_{|h_{0,n^*}|^2 |h_{n^*,2}|^2}(x) = 1 - F_{|h_{0,n^*}|^2 |h_{n^*,2}|^2}(x)$. Note that $\mathcal{C}_2$ is derived on the condition of $\varphi_2 - x\varphi_1 > 0$. By changing variable $t = \frac{x\Theta_2}{\varphi_2 - x\varphi_1}$ and by performing a series of calculations, Eq. (29) can be further derived as [27]:

$$\begin{aligned} \mathcal{C}_2 & = \frac{1-\theta}{\ln 2} \int_0^\infty \left(\frac{1}{t + \Theta_2(\varphi_2 + \varphi_1)^{-1}} - \frac{1}{t + \Theta_2\varphi_1^{-1}} \right) \\ & \times \bar{F}_{|h_{0,n^*}|^2 |h_{n^*,2}|^2} \left(\frac{x\Theta_2}{\varphi_2 - x\varphi_1} \right) dt. \end{aligned} \quad (30)$$

In Eq. (29), $F_{|h_{0,n^*}|^2 |h_{n^*,2}|^2}(x)$ is calculated as:

$$\begin{aligned} F_{|h_{0,n^*}|^2 |h_{n^*,2}|^2}(x) & = 1 - \Pr\left(|h_{0,n^*}|^2 |h_{n^*,2}|^2 > t\right) \\ & = 1 - \sum_{n_0=1}^N \sum_{n_2=1}^N \binom{N}{n_0} \binom{N}{n_2} \frac{n_0(-1)^{n_0+n_2-2}}{\lambda_0} \\ & \times \int_0^\infty e^{-\frac{n_2 t}{\lambda_2 y} - \frac{n_0 y}{\lambda_0}} dy \\ & = 1 - 2 \sum_{n_0=1}^N \sum_{n_2=1}^N \binom{N}{n_0} \binom{N}{n_2} (-1)^{n_0+n_2-2} \\ & \times \sqrt{\frac{n_2 n_0 t}{\lambda_2 \lambda_0}} K_1\left(2\sqrt{\frac{n_2 n_0 t}{\lambda_2 \lambda_0}}\right). \end{aligned} \quad (31)$$

By substituting Eq. (31) into Eq. (30), $\mathcal{C}_2$ can be given by:

$$\begin{aligned} \mathcal{C}_2 & = \sum_{n_0=1}^N \sum_{n_2=1}^N \binom{N}{n_0} \binom{N}{n_2} \frac{(1-\theta)(-1)^{n_0+n_2-2}}{\ln 2} \\ & \times \int_0^\infty \left(\frac{1}{t + \Theta_2(\varphi_2 + \varphi_1)^{-1}} - \frac{1}{t + \Theta_2\varphi_1^{-1}} \right) \\ & \times 2\sqrt{\frac{n_2 n_0 t}{\lambda_2 \lambda_0}} K_1\left(2\sqrt{\frac{n_2 n_0 t}{\lambda_2 \lambda_0}}\right) dt. \end{aligned} \quad (32)$$

Finally, with the help of [21, Eq. (7.811.5), (9.34.3)] and after some manipulations, we can obtain (21).

Proof 3 is completed.

References

- [1] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies", *IEEE Vehicular Technol. Mag.*, vol. 9, no. 1, pp. 71–78, 2014 (DOI: 10.1109/MVT.2013.2295070).
- [2] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System level performance evaluation of downlink non-orthogonal multiple access (NOMA)", in *Proc. IEEE 24th Annual Int. Symp. on Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, London, UK, 2013, pp. 611–615 (DOI: 10.1109/PIMRC.2013.6666209).
- [3] D.-T. Do, A.-T. Le, C.-B. Le, and B. M. Lee, "On exact outage and throughput performance of cognitive radio based non-orthogonal multiple access networks with and without D2D link", *Sensors*, vol. 19, no. 15, pp. 3314, 2019 (DOI: 10.3390/s19153314).
- [4] D.-T. Do and M.-S. Van Nguyen, "Device-to-device transmission modes in NOMA network with and without wireless power transfer", *Computer Commun.*, vol. 139, pp. 67–77, 2019 (DOI: 10.1016/j.comcom.2019.04.003).
- [5] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions", *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, 2016 (DOI: 10.1109/TVT.2015.2480766).
- [6] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems", *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, 2016 (DOI: 10.1109/TSP.2015.2480042).
- [7] D.-T. Do, M. Vaezi, and T.-L. Nguyen, "Wireless powered cooperative relaying using NOMA with imperfect CSI", in *Proc. of IEEE Globecom Workshops*, Abu Dhabi, UAE, pp. 1–6, 2018 (DOI: 10.1109/GLOCOMW.2018.8644154).
- [8] D.-T. Do and A.-T. Le, "NOMA based cognitive relaying: Transceiver hardware impairments, relay selection policies and outage performance comparison", *Computer Commun.*, vol. 146, pp. 144–154, 2019 (DOI: 10.1016/j.comcom.2019.07.023).
- [9] D.-T. Do, A.-T. Le, and B.-M. Lee, "On performance analysis of underlay cognitive radio-aware hybrid OMA/NOMA networks with imperfect CSI", *Electronics*, vol. 8, no. 7, pp. 819, 2019 (DOI: 10.3390/electronics8070819).
- [10] P.-M. Nam, D.-T. Do, T.-T. Nguyen, and P. T. Tin, "Energy harvesting assisted cognitive radio: random location-based transceivers scheme and performance analysis", *Telecommun. Systems*, vol. 67, no. 1, pp. 123–132, 2018 (DOI: 10.1007/s11235-017-0325-0).
- [11] T.-L. Nguyen and Dinh-Thuan Do, "Exploiting impacts of intercell interference on SWIPT-assisted non-orthogonal multiple access", *Wireless Commun. and Mobile Comput.*, vol. 2018, 2018 (DOI: 10.1155/2018/2525492).
- [12] D.-T. Do, M.-S. Van Nguyen, T.-A. Hoang, and M. Voznak, "NOMA-assisted multiple access scheme for IoT deployment: Relay selection model and secrecy performance improvement", *Sensors*, vol. 19, no. 3, pp. 736, 2019 (DOI: 10.3390/s19030736).

- [13] D.-T. Do, "Energy-aware two-way relaying networks under imperfect hardware: optimal throughput design and analysis", *Telecommun. Systems*, vol. 62, no. 2, pp. 449–459, 2016 (DOI: 10.1007/s11235-015-0085-7).
- [14] X. Li *et al.*, "Effective rate of MISO systems over $\kappa - \mu$ shadowed fading channels", in *IEEE Access*, vol. 5, pp. 10605–10611, 2017 (DOI: 10.1109/ACCESS.2017.2705018).
- [15] X. Yue *et al.*, "Exploiting full/half-duplex user relaying in NOMA systems", *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 560–575, 2018 (DOI: 10.1109/TCOMM.2017.2749400).
- [16] D.-T. Do, C.-B. Le, and F. Afghah, "Enabling full-duplex and energy harvesting in uplink and downlink of small-cell network relying on power domain based multiple access", *IEEE Access*, vol. 8, pp. 142772–142784, 2020 (DOI: 10.1109/ACCESS.2020.3013912).
- [17] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual transceiver hardware impairments on cooperative NOMA networks", *IEEE Transac. on Wireless Commun.*, vol. 19, no. 1, pp. 680–695, 2020 (DOI: 10.1109/TWC.2019.2947670).
- [18] X. Li, J. Li, and L. Li, "Performance analysis of impaired SWIPT NOMA relaying networks over imperfect Weibull channels", in *IEEE Systems J.*, vol. 14, no. 1, 2020, pp. 669–672 (DOI: 10.1109/JSYST.2019.2919654).
- [19] X. Li, M. Liu, C. Deng, P. T. Mathiopoulos, Z. Ding, and Y. Liu, "Full-duplex cooperative NOMA relaying systems with I/Q imbalance and imperfect SIC", *IEEE Wireless Commun. Letters*, vol. 9, no. 1, pp. 17–20, 2020 (DOI: 10.1109/LWC.2019.2939309).
- [20] Xingwang Li *et al.*, "A unified framework for HS-UAV NOMA network: performance analysis and location optimization", *IEEE Access*, vol. 8, pp. 13329–13340, 2020 (DOI: 10.1109/ACCESS.2020.2964730).
- [21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York, NY, USA: Academic Press, 2000 (ISBN: 9780080542225).
- [22] C.-B. Le and D.-T. Do, "On outage performance of backscatter NOMA relaying systems equipping with multiple antennas", *Electronics Letters*, vol. 55, no. 19, pp. 1066–1067, 2019 (DOI: 10.1049/el.2019.1390).
- [23] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying protocols for wireless energy harvesting and information processing", *IEEE Transac. on Wireless Commun.*, vol. 12, no. 7, pp. 3622–3636, 2013 (DOI: 10.1109/TWC.2013.062413.122042).
- [24] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, NY, USA: Dover, 1972 (ISBN: 9780486612720).
- [25] P. Yan, J. Yang, M. Liu, J. Sun, and G. Gui, "Secrecy outage analysis of transmit antenna selection assisted with wireless power beacon", *IEEE Transac. on Vehicular Technol.*, vol. 69, no. 7, pp. 7473–7482, 2020 (DOI: 10.1109/TVT.2020.2992766).
- [26] X. Wang, M. Jia, I. W.-H. Ho, Q. Guo, and F. C. M. Lau, "Exploiting full-duplex two-way relay cooperative non-orthogonal multiple access", *IEEE Transac. Commun.*, vol. 67, no. 4, pp. 2716–2729, 2019 (DOI: 10.1109/TCOMM.2018.2890264).
- [27] T.-L. Nguyen, C.-B. Le, and D.-T. Do, "Performance analysis of multi-user NOMA over $\alpha - \kappa - \mu$ shadowed fading", *Electronics Letters*, vol. 56, no. 15, pp. 771–773, 2020 (DOI: 10.1049/el.2019.4265).



Chi-Bao Le has been working closely with Dr. Thuan of the Wireless Communications and Signal Processing Research Group at Industrial University of Ho Chi Minh City, Vietnam. He is currently pursuing his M.Sc. degree in wireless communications. His research interests include electronic design, signal processing in wireless communications network, non-orthogonal multiple access, UAV, backscatter communication, physical layer security and reconfigurable intelligent surfaces.

E-mail: lechibao@iuh.edu.vn

Faculty of Electronics Technology

Industrial University of Ho Chi Minh City

Ho Chi Minh City

Vietnam



Nhan Duc Nguyen received his M.Eng. in electronic materials from International Training Institute for Materials Science (ITIMS), Hanoi University of Technology in 1998, and his Ph.D. degree in electrical and computer systems engineering from Monash University, Australia in 2011. He joined the Faculty of Telecommunications,

Postal and Telecommunications Institute of Technology in Vietnam, as a lecturer, in 1999. He served as the Head of the Signals and Systems Department at Postal and Telecommunications Institute of Technology from 2014 to 2020. He is currently serving as a Systems Engineering Director at the Innovation Center, Van Lang University. His research interests focus on optical communications, numerical modeling and analysis, signal processing, as well as sensor data processing in machine learning.

E-mail: nhan.nd@vlu.edu.vn

Innovation Center

Van Lang University

Ho Chi Minh City

Vietnam

Evaluation of Radio Channel Utility using Epsilon-Greedy Action Selection

Krzysztof Malon

Military University of Technology, Warsaw, Poland

<https://doi.org/10.26636/jtit.2021.153621>

Abstract—This paper presents an algorithm that supports the dynamic spectrum access process in cognitive radio networks by generating a sorted list of best radio channels or by identifying those frequency ranges that are not in use temporarily. The concept is based on the reinforcement learning technique named Q-learning. To evaluate the utility of individual radio channels, spectrum monitoring is performed. In the presented solution, the epsilon-greedy action selection method is used to indicate which channel should be monitored next. The article includes a description of the proposed algorithm, scenarios, metrics, and simulation results showing the correct operation of the approach relied upon to evaluate the utility of radio channels and the epsilon-greedy action selection method. Based on the performed tests, it is possible to determine algorithm parameters that should be used in this proposed deployment. The paper also presents a comparison of the results with two other action selection methods.

Keywords—cognitive radio, dynamic spectrum access, spectrum monitoring, machine learning, Q-learning.

1. Introduction

With the dynamic development of wireless communications systems, spectrum scarcity has become an increasingly important problem. The vast majority of radio frequency bands are assigned to licensed (primary) users on an exclusive basis. At the same time, by analyzing the use of frequency resources over time [1]–[3], one can identify the so-called spectrum holes. This term refers to frequency bands that are not in use temporarily and may be utilized for transmission by secondary (unlicensed) users. This approach is referred to as dynamic spectrum access (DSA). To apply this concept, cognitive radio (CR) technology is proposed. The functionalities of CR include the ability to receive various information from the surrounding environment, analyze it, make decisions, and perform specific actions. The ability to learn (improve functionality) from previous reactions and from the results obtained is another essential feature of CR.

Implementation of DSA requires constant monitoring of the available radio resources and means that the usefulness of individual frequency ranges (i.e. radio channels) needs to

be determined. For secondary users, channels with low occupancy ratios (activity of other users) are the most important ones.

2. Evolution of Radio Channel Utility

The algorithm proposed in this paper for evaluating the utility of radio channels supports DSA and is based on the machine learning method named Q-learning. This technique belongs to the class of reinforcement learning methods in which learning takes place through experimentation. In addition to reinforcement learning, two primary groups of machine learning methods may be distinguished, namely supervised and unsupervised learning [4], [5]. Reinforcement learning is considered to be useful in terms of CR, especially in monitoring and accessing the spectrum in a dynamically changing environment [4]. In the considered solution, reinforcement learning of the single state [6], [7] or stateless type [8], [9] is analyzed. The proposed algorithm does not require any knowledge of the radio environment. It recognizes and learns spectrum usability-related information by relying on the trial and error method [6], [10]. On the other hand, if the state of several frequency channels is known, it may also be used during the initialization step or during the algorithm's operation.

Figure 1 depicts the general scheme of the proposed algorithm that consists of four primary stages repeated as the system is operated. Before the algorithm starts, the Q matrix should be initialized. This matrix consists of channels a and their estimated qualities $Q(a)$.

The first step of the algorithm is the selection of action (a). During this stage, a new channel for sensing (i.e. radio spectrum monitoring) is indicated. Random and cyclic algorithms are the most straightforward and the most popular solution, as they search the entire action space with equal probabilities. In the first case, the action (channel) is selected randomly. The second solution assumes that channels are specified in sequence over a repeated cycle. Another proposal is the epsilon-greedy strategy, which is a greedy policy variant (see Fig. 2). By using this approach, one may exploit (use) the best actions and reduce the effort

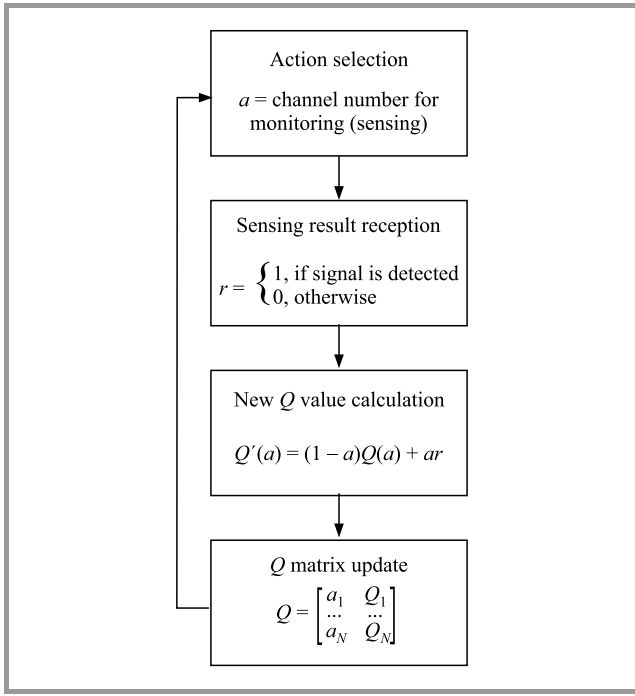


Fig. 1. Radio channel utility evaluation algorithm.

required to explore (search) for others. According to this principle, the following action is chosen:

- randomly with low probability ϵ ,
- or according to the current policy of maximizing rewards with a probability of $1 - \epsilon$.

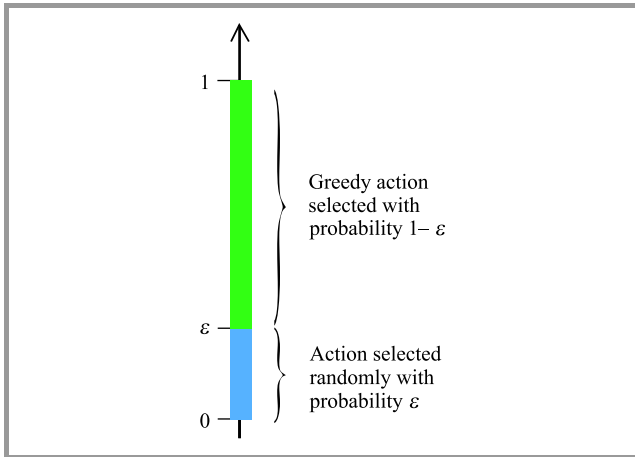


Fig. 2. Epsilon-greedy action selection method.

The ϵ value determines the probability of taking the greedy action, selecting the best channel for sensing. Otherwise, it also defines the probability $1 - \epsilon$ of performing a random action. This makes it possible to find other channels with good qualities. As shown in Fig. 2, by changing the ϵ value, a trade-off between exploration and exploitation is reached. An example algorithm for the epsilon-greedy action selection method is shown in Fig. 3. Firstly, a random number

p ($p = 0, \dots, 1$) is generated. Then, its value is compared with the defined ϵ . Depending on the result, a random action is performed or the best channel is selected.

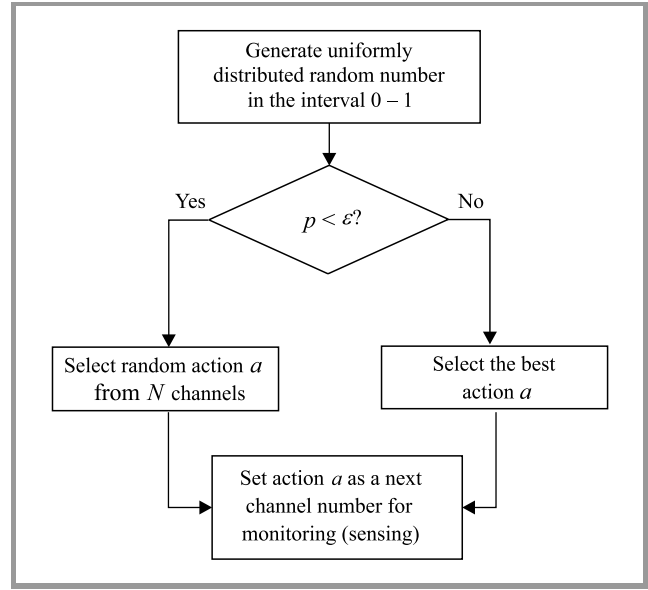


Fig. 3. Epsilon-greedy action selection algorithm.

In the next step, the radio channel utility evaluation algorithm is executed on the selected frequency channel. Two basic sensing approaches may be used in the proposed solution: local spectrum monitoring or cooperative sensing of spatially distributed radio network nodes. The problem of optimizing the placement of sensing elements is considered, inter alia, in [11] and [12]. The use of cooperative spectrum monitoring allows to reduce the severity of the problem of the so-called hidden nodes [13]. In such cases, it is necessary to apply a certain data fusion method, e.g. the Dempster-Shaffer theory [14].

The spectrum monitoring result r is passed to the following step of the algorithm in which the calculation of a new $Q'(a)$ value for the analyzed channel is performed. Both the newly obtained r result and the previous value $Q(a)$ are considered. The significance of new and historical data is defined by the learning rate α . The calculation of the new value $Q'(a)$ is performed using the following relationship:

$$Q'(a) = (1 - \alpha)Q(a) + \alpha r, \quad (1)$$

where:

- a – selected action,
- $Q(a)$ – Q value for the selected action,
- $Q'(a)$ – new (updated) Q value for the selected action,
- $\alpha \in (0, 1)$ – learning rate,
- r – reward.

In the next step, the determined value $Q'(a)$ is used to update the Q matrix, and then the algorithm cycle repeats.

3. Simulations and Results

This section of the paper presents the scenarios, metrics, and simulation results of the proposed algorithm using the epsilon-greedy action selection approach.

To evaluate the proposed algorithm, two base scenarios were prepared. Each scenario consists of radio channels with their occupancy defined over time. To generate spectrum occupancy figures, a statistical approach based on the On-Off model (see Fig. 4) was used [15]. The traffic generated may be interpreted as originating from one or more users. This model assumes two possible states: occupied (On) and not-occupied (Off).

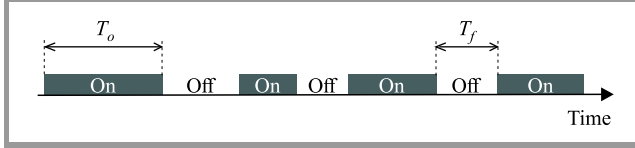


Fig. 4. On-Off spectral occupancy model [15].

In the selected Poisson-exponential model, the arrival procedure is modeled by a Poisson process. The time between successive spectral occupancies T_f (Off periods) is defined by an exponentially distributed random process:

$$f(T_f) = \frac{1}{\bar{T}_f} e^{-\frac{T_f}{\bar{T}_f}}, \quad (2)$$

with a mean interarrival time of:

$$E[T_f] = \bar{T}_f. \quad (3)$$

The duration of the occupancy time T_o (On periods) is also modeled as an exponentially distributed random process given by:

$$f(T_o) = \frac{1}{\bar{T}_o} e^{-\frac{T_o}{\bar{T}_o}}, \quad (4)$$

with a mean occupancy time of:

$$E[T_o] = \bar{T}_o. \quad (5)$$

According to the ITU-R report [16], spectrum resource occupancy SRO is the ratio of the number of channels in use (occupied) to the total number of channels in the entire frequency band. SRO for multiple channels within a specific time, called the integration time, is calculated as follows:

$$SRO = \frac{N_0}{N}, \quad (6)$$

where: N_0 – number of samples on any channel with a level above the threshold and N – total number of samples taken on all channels during the integration time.

In this case, the integration time is equated with the scenario time. Spectrum resource occupancy SRO may be interpreted as the average occupancy of the channels.

Table 1
Scenario 1 parameters

Parameter name	Parameter value		
	Even-numbered channels (2, 4, 6, 8, 10, 12)	Odd-numbered channels (1, 3, 5, 7, 9, 11)	All channels
Simulation time T	10,000		
Number of channels M	6	6	12
Average On time $\bar{T}_o$	10	10	-
Average Off time $\bar{T}_f$	10	30	-
Spectrum resource occupancy SRO	0.5	0.25	0.375

Table 2
Scenario 2 parameters

Parameter name	Parameter value		
	Even-numbered channels (2, 4, 6, 8, 10, 12)	Odd-numbered channels (1, 3, 5, 7, 9, 11)	All channels
Simulation time T	10,000		
Number of channels M	6	6	12
Average On time $\bar{T}_o$	40	40	-
Average Off time $\bar{T}_f$	40	120	-
Spectrum resource occupancy SRO	0.5	0.25	0.375

For evaluation purposes, two scenarios are considered. Both consist of twelve radio channels for which 10,000 states are defined (Table 1 and Table 2). The channels are divided into two groups with different parameters. Spectrum resource occupancy SRO for even-numbered channels is about 0.5, whereas for odd-numbered channels $SRO \approx 0.25$. This means that the overall spectrum occupancy for both scenarios (all channels) equals 0.375. The difference between scenario 1 and scenario 2 is in average On and Off times. Channel state changes in scenario 2 are

four times slower compared to scenario 1, e.g. $\overline{T}_o$ and $\overline{T}_f$ for even-numbered channels in scenario 1 are equal to 10, while for the scenario 2 these parameters are equal to 40.

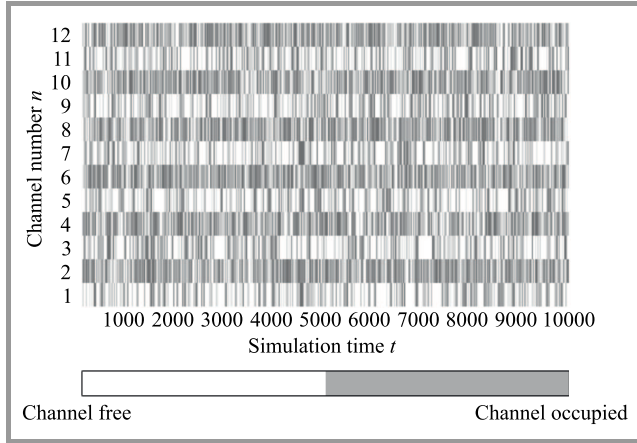


Fig. 5. Radio channel occupancy for base scenario 1.

The generated radio channel occupancy rates are shown in Figs. 5 and 6 for scenario 1 and scenario 2, respectively. The results are consistent with the Poisson-exponential model. White color represents not-occupied states (free), whereas gray is used to identify occupied channels.

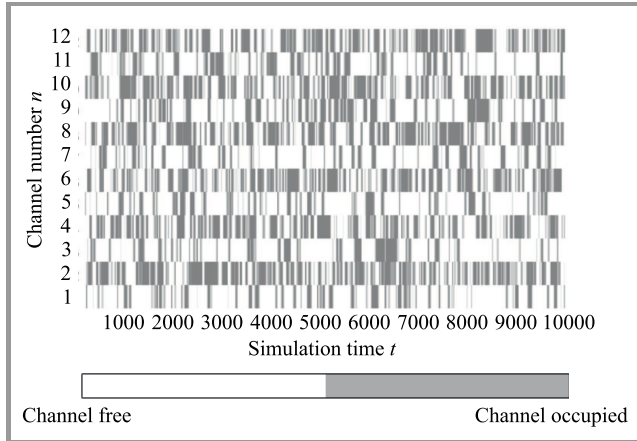


Fig. 6. Radio channel occupancy for base scenario 2.

The two presented scenarios serve as a basis for the simulations (base scenarios). They are used for the first iteration of the simulations performed according to the algorithm illustrated in Fig. 7.

Each scenario is simulated for different ϵ values, and then the relevant metrics are calculated. After analyzing the obtained metrics, ϵ value rendering the best results is selected. Consequently, there is a Q matrix and the best radio channel for each time step. In the next stage, information about temporally best channels is used to modify the input scenario, and then the simulations are performed relying on this updated data. That allows the scenarios used in subsequent iterations to be defined based on an increasing spectrum resource occupancy rate.

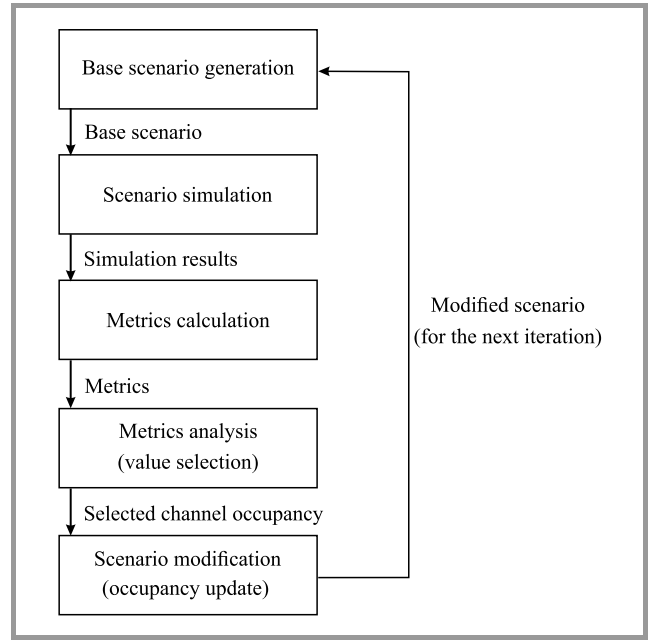


Fig. 7. Simulation algorithm – successive iterations.

3.1. Metrics Used

To evaluate the proposed solution, specific metrics are proposed. The first one is channel utility Utl , defined as:

$$Utl = \frac{N_f}{T}, \quad (7)$$

where: N_f – number of samples on the selected channel with a level below threshold (channel not-occupied) and T – total number of samples taken on the selected channel during the scenario time.

The Utl value can vary from 0 to 1. The second metric is spectrum resource occupancy gain SRO_{gain} defined as:

$$SRO_{gain} = SRO_2 - SRO_1, \quad (8)$$

where: SRO_1 – reference spectrum resource occupancy value (occupancy of the scenario prepared for simulation) and SRO_2 – spectrum resource occupancy after simulation (including the occupancy resulting from the use of the best channel determined by the algorithm).

The goal is to obtain the highest SRO_{gain} and therefore the greatest Utl value, as radio resources are then used more efficiently. SRO_1 may be defined in the same way as in Eq. (6):

$$SRO_1 = \frac{N_0}{N}. \quad (9)$$

The use of the spectrum, when the system is using the best channels indicated by the proposed algorithm, increases proportionally to the Utl value. In such a case, the not-occupied states of the selected channel N_f change their status to occupied and increase the utilization of frequency resources. Accordingly, SRO_2 may be defined as:

$$SRO_2 = \frac{N_o + N_f}{N}. \quad (10)$$

The total number of samples N taken on all channels during the integration time is expressed by:

$$N = MT, \quad (11)$$

where M is the number of channels.

The SRO_{gain} may be defined as:

$$SRO_{gain} = \frac{N_o + N_f}{N} - \frac{N_0}{N} = \frac{N_f}{MT} = \frac{Utl}{M}. \quad (12)$$

Considering the range of variability of the parameter Utl , the maximum SRO_{gain} is:

$$SRO_{gainMax} = \frac{1}{M}. \quad (13)$$

3.2. Results

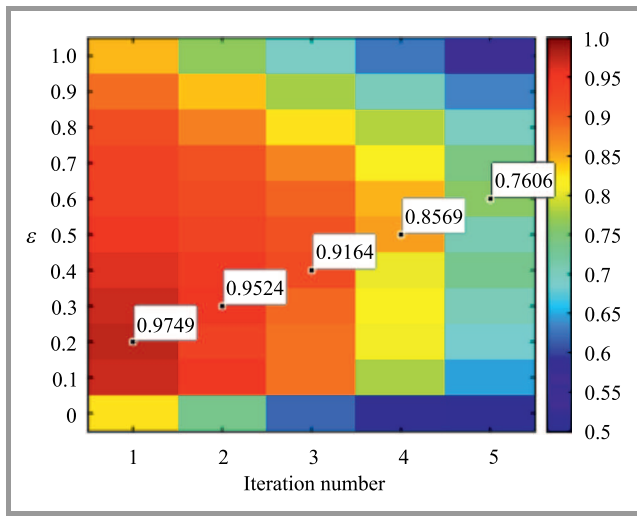


Fig. 8. Utility values for different epsilon ϵ values in successive iterations of base scenario 1. (see the digital version for color images)

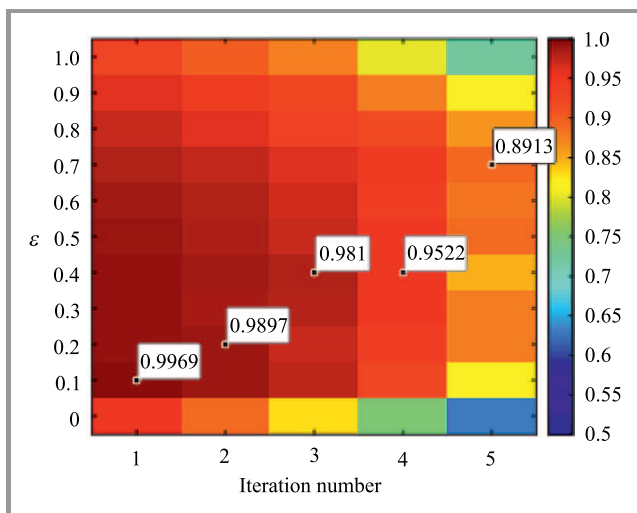


Fig. 9. Utility values for different epsilon ϵ values in successive iterations of base scenario 2.

Figures 8–11 show Utl , SRO , and SRO_{gain} in successive iterations for both scenarios. Utility values presented in Figs. 8 and 9 are calculated for the first channel in the Q matrix – the best radio channel in each simulation step.

Better results are obtained for scenario 2. Here, higher utility values are obtained compared to those for scenario 1 for the same spectrum resource occupancy (iteration number). It is so because of the different channel state changes dynamics. In scenario 1, shorter On and Off times cause frequent channel state changes. The larger the iteration number, the greater value of ϵ provides the best utility values. It means that for a higher spectrum resource occupancy rate, the epsilon-greedy action selection method should increase exploration.

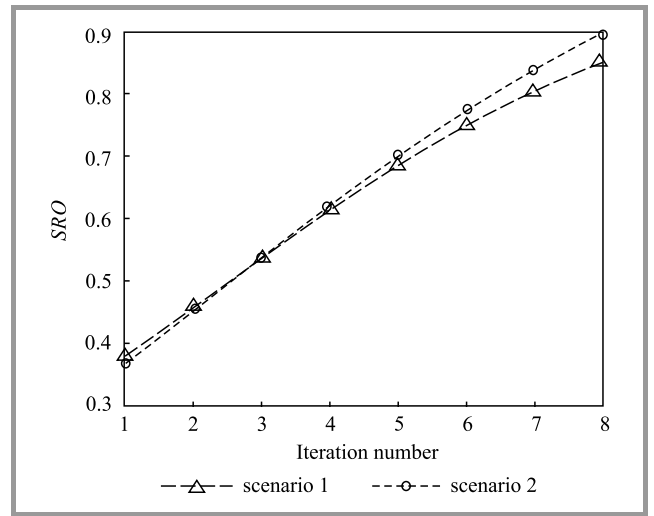


Fig. 10. Spectrum resource occupancy SRO in successive iterations for both scenarios.

Figures 10 and 11 show spectrum resource occupancy and SRO_{gain} for both scenarios. For the first three iterations,

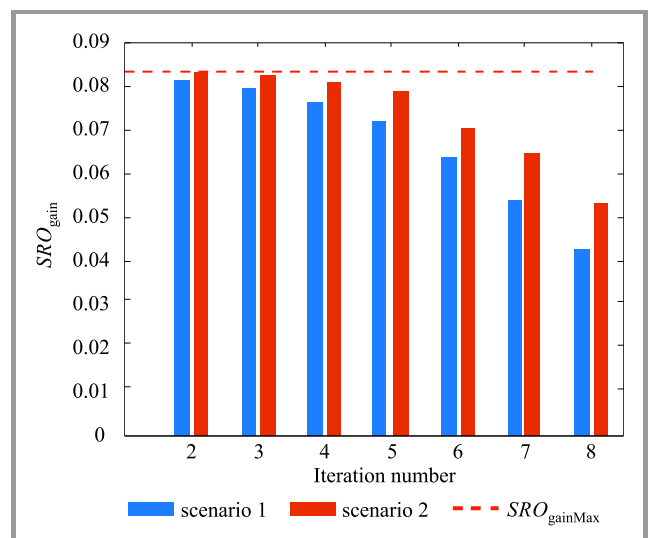


Fig. 11. Spectrum resource occupancy gain SRO_{gain} in successive iterations of both scenarios.

SRO_{gain} values are close to their maximum $SRO_{gainMax}$ (red dashed line in Fig. 11). As mentioned before, better results (higher SRO_{gain}) can be obtained for scenario 2 due to lower channel state changes dynamics. As the iteration number increases, the spectrum occupancy grows, and thus the chances of finding free radio resources (not-occupied channel) decreases. Therefore, the channel utility Utl values shown in Figs. 8 and 9 and the spectrum resource occupancy gain SRO_{gain} presented in Fig. 11 take a lower values with the increase in the iteration number.

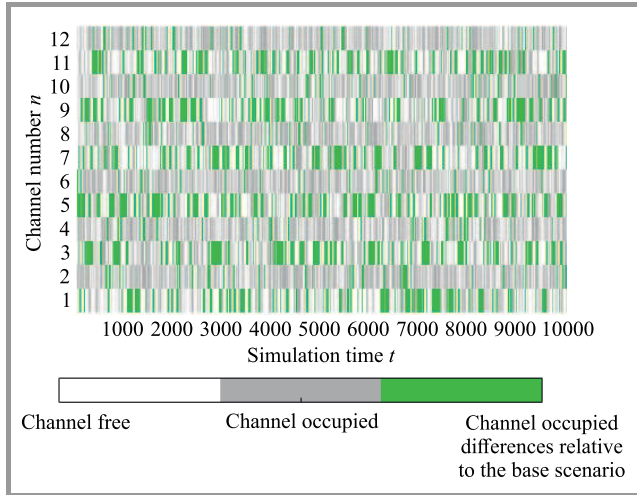


Fig. 12. Radio channel occupancy for the third iteration of base scenario 1.

Figures 12 and 13 show the growth in radio channel occupancy after two iterations, compared to the base scenarios. White and gray colors represent free and occupied states in the base scenario. Green indicates new occupied states resulting from including the temporarily best channels selected by the proposed algorithm. As one may notice, odd-numbered channels are chosen more often because their spectrum resource occupancy is lower. Please refer

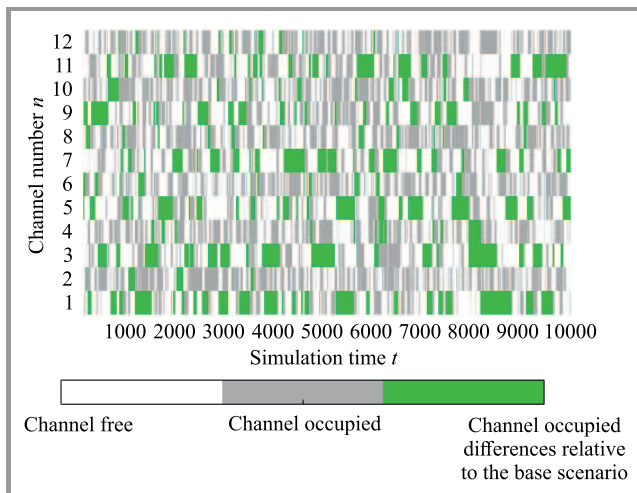


Fig. 13. Radio channel occupancy for the third iteration of base scenario 2.

to Tables 1 and 2 for detailed parameters. This behavior confirms the correct operation of the algorithm identifying temporarily free channels.

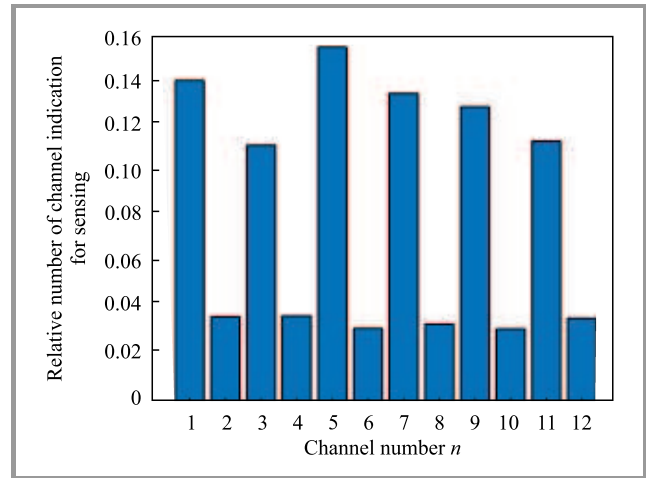


Fig. 14. Radio channels selected by the epsilon-greedy method ($\epsilon = 0.2$) in the first iteration of base scenario 1.

Figures 14 and 15 depict the effect of the epsilon-greedy action selection method. They show how often particular channels are selected for sensing (monitoring). There are two example results from the scenario 1 simulations. The first one concerns the base scenario (first iteration), when SRO is approx. 0.375. In this case the ϵ value is set to 0.2, which means that greedy actions are taken with the probability of 0.8 (please refer to Fig. 2). This results in more frequent selection of less busy channels (odd-numbered channels). Figure 15 presents the behavior of the epsilon-greedy action selection method in the fourth iteration. In this situation, the ϵ value that allows to obtain the best Utl is 0.5 (see Fig. 8). As the spectrum occupancy grows it is needed to increase the exploration to find other free channels. Probabilities of the selection of individual channels are equalized.

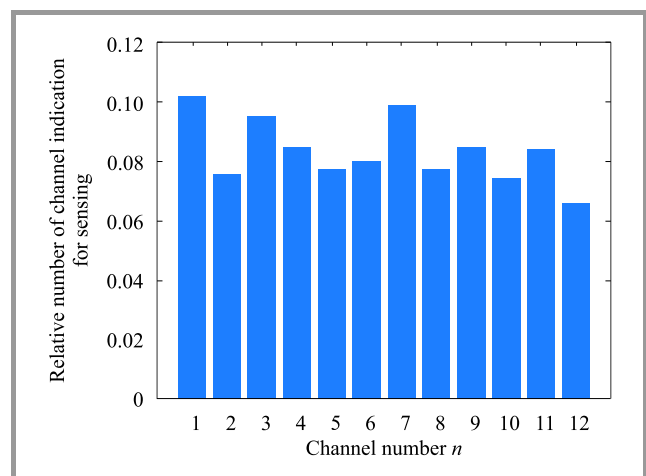


Fig. 15. Radio channels selected by the epsilon-greedy method ($\epsilon = 0.5$) in the fourth iteration of base scenario 1.

Radio channels utilities in the first and fourth iteration of base scenario 1 are presented in Figs. 16 and 17, respectively. These results compare three action selection methods: epsilon-greedy, random and cyclic. Analysis of the first channel index in the Q matrix (the best one) shows a significant advantage that the epsilon-greedy algorithm has over the other two methods.

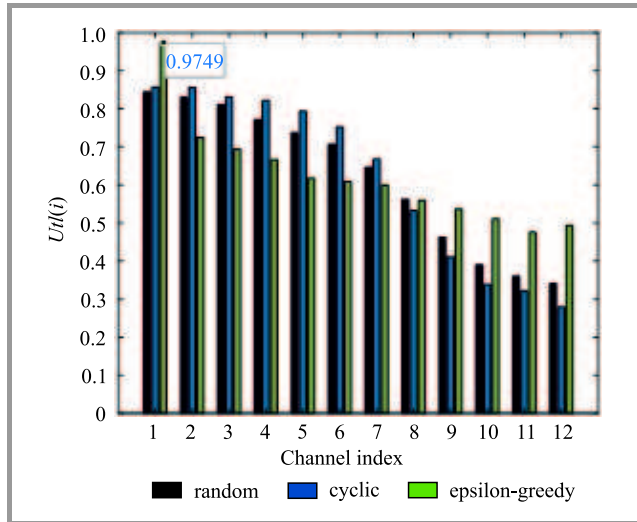


Fig. 16. Radio channel utility rate in the first iteration of base scenario 1.

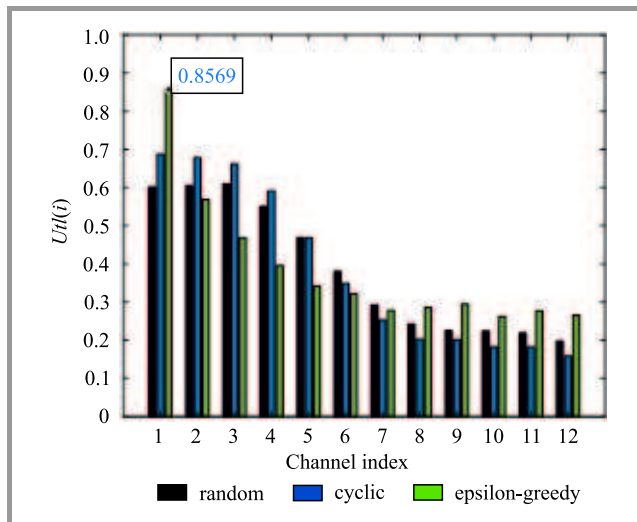


Fig. 17. Radio channel utility rate in the fourth iteration of base scenario 1.

4. Conclusions

This paper presents an algorithm for evaluating the usefulness of radio channels based on a machine learning technique named Q-learning. The proposed algorithm identifies radio channels for sensing using the epsilon-greedy action selection method. This process aims to reach a trade-off

between exploration and exploitation of the available radio channels. Based on the results from the process of monitoring frequency resources, individual radio channels are evaluated. As a result, a sorted list of radio channels capable of supporting DSA is generated. An essential feature of the proposed concept is that it does not need to be initialized, meaning it may work in an unknown electromagnetic environment, gradually building its situational awareness. The presented scenarios, metrics, and simulation results show the algorithm's correct operation and the proper choice of the action selection method. The tests performed have identified the crucial ϵ values that allow to reach the maximum spectrum utilization rate under specific conditions. The epsilon-greedy action selection method is also compared with two other approaches: random and cyclic. It has been shown that the channel utilization rates obtained using the epsilon-greedy approach are much better.

5. Acknowledgment

This work was financed by the Military University of Technology under research project no. UGB/22-854/2021/WAT "Application of selected computer science, communication, and reconnaissance techniques in civilian and military areas".

References

- [1] Wireless Innovation Forum, "Dynamic Spectrum Sharing Annual Report – 2014", Document WINNF-14-P-0001, version V0.2.16 [Online]. Available: https://www.wirelessinnovation.org/assets/work_products/Reports/winnf-14-p-0001-v1.0%20dynamic%20spectrum%20sharing%20annual%20report%202014.pdf
- [2] M. A. McHenry, D. McCloskey, and G. Lane-Roberts, "New York City spectrum occupancy measurements", *Shared Spectrum Company*, 2005 [Online]. Available: http://www.sharespectrum.com/wp-content/uploads/4_NSF_NYC_Report.pdf
- [3] Shared Spectrum Company, "General survey of radio frequency bands – 30 MHz to 3 GHz", 2010 [Online]. Available: https://www.sharespectrum.com/wp-content/uploads/2021/01/2010_0923-General-Band-Survey-30MHz-to-3GHz.pdf
- [4] E. Biglieri, A. J. Goldsmith, L. J. Greenstein, N. B. Mandayam, and H. V. Poor, *Principles of cognitive radio*. Cambridge: Cambridge University Press, 2012 (ISBN: 9781139236850).
- [5] L. E. Doyle, *Essentials of cognitive radio*. Cambridge: Cambridge University Press, 2009 (ISBN: 9780511576577).
- [6] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction, second edition*. Cambridge: The MIT Press, 2018 (ISBN: 9780262039246).
- [7] K.-L. A. Yau, P. Komisarczuk, and P. D. Teal, "Applications of reinforcement learning to cognitive radio networks", in *Proc. IEEE Int. Conf. on Commun. Workshops*, Cape Town, South Africa, 2010, pp. 1–6 (DOI: 10.1109/ICCW.2010.5503970).
- [8] N. Morozs, T. Clarke, and D. Grace, "Distributed heuristically accelerated Q-learning for robust cognitive spectrum management in LTE cellular systems", *IEEE Transac. on Mobile Comput.*, vol. 15, no. 4, pp. 817–825, 2016 (DOI: 10.1109/TMC.2015.2442529).
- [9] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems", in *Proc. of the fifteenth national/tenth Conf. on Artif. Intell./Innovat. Applicat. of Artif. Intell.*, Madison, WI, USA, 1998, pp. 746–752 [Online]. Available: <https://www.aaai.org/Papers/AAAI/1998/AAAI98-106.pdf>

- [10] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios", *IEEE Commun. Surveys & Tut.*, vol. 15, no. 3, pp. 1136–1159, 2013 (DOI: 10.1109/SURV.2012.100412.00017).
- [11] K. Malon, P. Skokowski, and J. Łopatka, "Optimization of wireless sensor network deployment for electromagnetic situation monitoring", *Int. J. of Microwave and Wireless Technol.*, vol. 10, no. 7, pp. 746–753, 2018 (DOI: 10.1017/S1759078718000211).
- [12] K. Malon, P. Skokowski, and J. Łopatka, "Optimization of the MANET topology in urban area using redundant relay points", *Int. Conf. on Military Commun. and Informat. Systems (ICMCIS)*, Warsaw, Poland, 2018, pp. 1–4 (DOI: 10.1109/ICMCIS.2018.8398720).
- [13] P. Skokowski, K. Malon, and J. Łopatka, "Properties of centralized cooperative sensing in cognitive radio networks", in *Proc. XI Conf. on Reconnaissance and Electron. Warfare Systems*, J. Łopatka, Eds. *Int. Society for Optics and Photon.*, vol. 10418, pp. 54–62. Ołtarzew, Poland: SPIE, 2017 (DOI: 10.1117/12.2269996).
- [14] P. Skokowski, "Electromagnetic situation awareness building in ad-hoc networks with cognitive nodes", *Military University of Technology*, Warsaw, Poland, 2021 (in Polish, in print).
- [15] K. Sithamparanathan and A. Giorgetti, *Cognitive Radio Techniques*. Artech House, 2012 (ISBN: 9781608072040).
- [16] ITU-R Report SM.2256-1, "Spectrum occupancy measurements and evaluation", *Int. Telecommun. Union*, 2016 [Online]. Available: <https://www.itu.int/pub/R-REP-SM.2256>



Krzysztof Malon received his M.Sc. and Ph.D. degrees from the Military University of Technology in 2011 and 2019, respectively. Since 2016 he has been working at the Institute of Communications Systems of MUT. His main research interests are wireless communications, cognitive radio, dynamic spectrum access, and radio

spectrum monitoring. Krzysztof Malon has participated in many national and international research projects for the European Defence Agency and the National Centre for Research and Development. Recently, he is also involved in the NATO working group related to the 5G technologies application to NATO operations.

 <https://orcid.org/0000-0001-9257-1166>

E-mail: krzysztof.malon@wat.edu.pl
 Institute of Communications Systems
 Faculty of Electronics
 Military University of Technology
 Warsaw, Poland

Developing RF Power Sensor Calibration Station in Direct Comparison Transfer System using Vector Network Analyzer

Jarosław Szatkowski

National Institute of Telecommunications, Warsaw, Poland

<https://doi.org/10.26636/jtit.2021.155021>

Abstract—Calibration of RF power sensors is crucial issue in RF power measurements. Many calibration laboratories use the direct comparison transfer system with a signal generator and a power splitter. Increasing performance of modern vector network analyzers makes it possible to perform a power sensor calibration with acceptable uncertainties. The main advantage when using a VNA is a simple measurement setup with a wide frequency range (up to 50 GHz, limited only by the VNA and the standard power sensor), where all of required components, i.e. signal generator, a directional coupler and a reference power indicator are built in the VNA technology. This paper reports performing a VNA-based RF power sensors calibration for 10 MHz – 18 GHz band, carried out in the Laboratory of Electric, Electronic and Optoelectronic Metrology at the National Institute of Telecommunications in Warsaw, Poland. In order to validate the proposed solution two of power sensors were calibrated at a reference laboratory. The validation consisted of two steps. At first, one of those characterized power sensors was calibrated at our laboratory in direct comparison transfer system. Finally, the results obtained from the VNA-based system were compared with the previously obtained ones.

Keywords—direct comparison transfer, microwave power measurements, power sensor calibration, measurement uncertainty, VNA.

1. Introduction

Measurement of RF signal power is one of the most important metrology issues in microwaves. Power sensors together with suitable power meters are used in test stand. To achieve best accuracy of the measurement, a standard power sensor is calibrated by one of the national laboratory, for instance NIST, METAS, PTB, etc. For power sensor calibration the direct comparison transfer method is widely used [1], [2]. The calibration system consists of a power splitter or a directional coupler where input port is connected to signal generator, and have a reference power sensor attached to one of the two output ports as shown in the Fig. 1. The second output port is connected to the stan-

dard power sensor and to the power sensor being calibrated (device under test, DUT) interchangeably. The calibration performance depends on the effective source match of the splitter or the directional coupler, and on the accuracy of the used standard power sensor.

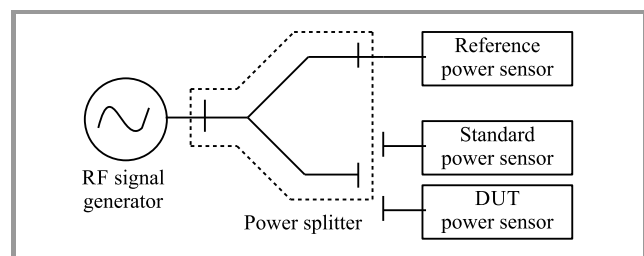


Fig. 1. Power sensor calibration system in generator – splitter connection for direct comparison transfer.

The use of vector network analyzer (VNA) for calibration power sensors was presented in [3], [4]. The principle of operation of the VNA-based system is same as for generator-splitter method, hence it is the direct comparison transfer system with identical sources of measurement uncertainty. However, the quality of VNA output signal is generally much worse than that of the signal obtained from RF generators, hence it can significantly affect the uncertainty budget. A comparison of these two systems was presented in [4]. The main benefit of using the VNA versus the generator-splitter system is its great simplicity. Figure 2 shows the block diagram of VNA port 1.

The standard power sensor and the unit to be calibrated are alternately connected to the VNA. The signal generator, the directional coupler, and the reference power sensor are circuits built into VNA, thus the setup of the system is very simple.

This paper presents the VNA-based power sensor calibration system for frequency range 10 MHz–18 GHz. This is a preliminary work proving the concept of VNA usage for this purpose. The aim is to set up the measurement station covering the VNA full range, up to 50 GHz.

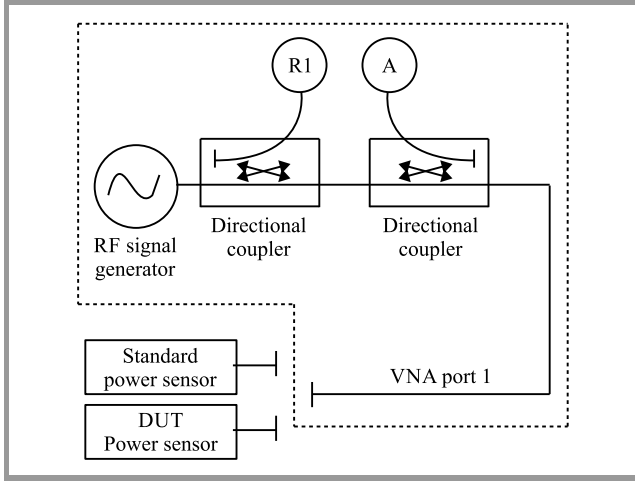


Fig. 2. Block diagram of VNA port 1.

In order to validate the developed system two power sensors were calibrated in a reference laboratory. One of them was then used as the standard power sensor, whereas the second one was used as the DUT. The results retrieved from the reference lab and the data obtained with the generator-splitter method and with the VNA system were compared.

2. Direct Comparison Transfer Overview

The measurement system consists of a RF signal generator, a power splitter (or directional coupler), and a reference power sensor connected to one of the output ports of the splitter. The standard power sensor and DUT are alternately connected to the second port of the splitter (see Fig. 1). Using the reference power sensor and the power splitter instead of connecting the sensors directly to the output of the generator improves the effective source match. The effective source match Γ_{e2} depends only on the scattering parameters of the power splitter itself and is generally much better and stable than the source match of the generator [5]. Importantly, it is quite easy to measure using VNA, either by determining the scattering parameters or by Jureshek method [6]. In contrast, source match of signal generators is quite difficult to determine:

$$\Gamma_{e2} = S_{33} - \frac{S_{31}S_{23}}{S_{21}}. \quad (1)$$

The DUT calibration factor K_D is calculating using [4]:

$$K_D = K_S \frac{P_D}{P_{MD}} \cdot \frac{P_{MS}}{P_S} \cdot \left(\frac{|1 - \Gamma_{e2}\Gamma_D|}{|1 - \Gamma_{e2}\Gamma_S|} \right)^2, \quad (2)$$

where P_S, P_{MS}, P_D, P_{MD} are power levels measured by the standard power sensor, by the reference power sensor with the standard sensor connected, by the DUT power sensor, and by the reference power sensor with the DUT sensor connected, respectively. Γ_S and Γ_D are the reflection coefficients of the standard and of the DUT. The B type uncertainty of the measured calibration factor K_D mainly consists

of the uncertainty of the transferred calibration factor K_S and the uncertainties of the reflection coefficients in the mismatch factor in Eq. 2. It is worth noting, that the calibration factor K_D does not depend on the calibration factor and the reflection coefficient of the reference power sensor, thus this power sensor does not affect the B type calibration uncertainty. The A type uncertainty mainly includes power readings instabilities of each power sensor used, as well as the connection repeatability. The uncertainty budget for a calibration of a power sensor with the direct comparison transfer method is described in [1].

3. VNA in Power Sensor Calibration

A vector network analyzer is used for a measurement of a reflection of a calibrated power sensor. It is a very important parameter specified by all manufacturers of power sensors. In addition, reflection coefficients of standard and DUT power sensors are required to calculate the calibration factor K_D by Eq. (2) and are taking into account when calculating measurement uncertainties.

This article shows the usage of a VNA in a direct comparison transfer system to determine the calibration factor K_D . Furthermore, VNA can be used to calibrate the linearity factor, thus it makes it possible to perform a complete power sensor calibration in a simple way [3].

The VNA integrates all required components for measurement of the calibration factor with the direct comparison transfer system. When looking at Fig. 2 and comparing it with Fig. 1 one can clearly deduce that the directional coupler with R1 receiver works the same way as the power splitter with the reference power sensor. The signal generator is built into VNA. Thus the formula for the calibration factor of DUT sensor is similar to the Eq. (2):

$$K_D = K_S \frac{P_D}{|RI_D|^2} \cdot \frac{|RI_S|^2}{P_S} \cdot \left(\frac{|1 - \Gamma_{es}\Gamma_D|}{|1 - \Gamma_{es}\Gamma_S|} \right)^2, \quad (3)$$

where RI_D and RI_S are the measured complex incident voltage waves when the standard power sensor and the DUT are connected to the VNA port, respectively, and Γ_{es} is the source match of the port used. Γ_{es} is relatively easy to determine as it is one of the error terms retrieved from reflection calibration of the VNA port [7]. The similarity of Eqs. (2) and (3) reflects in the uncertainty analysis, where uncertainty factors similar to those in the generator-splitter method are taken into account [7], [8].

4. Measurements

4.1. Measurement Setups

The DUT and standard power sensors were respectively a diode power sensor and a thermal power sensor manufactured by Keysight Technologies. Both power sensors had been calibrated in a reference laboratory at frequencies ranging from 10 MHz to 18 GHz. Figure 3 presents the

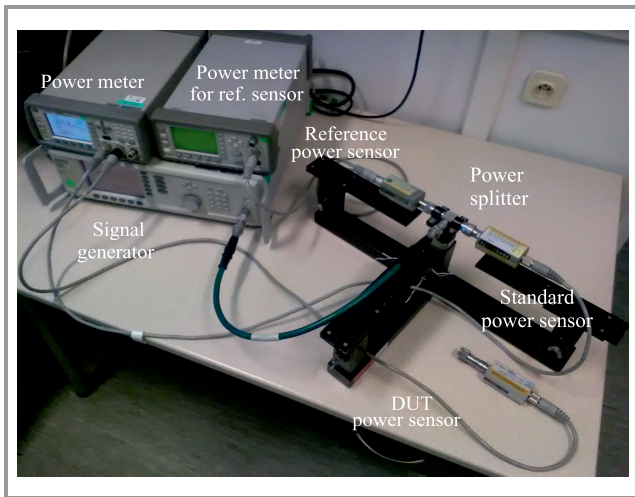


Fig. 3. Measurement setup with a signal generator, a power splitter and with a reference power sensor for power sensor calibration in direct comparison transfer system.

calibration system setup with a signal generator, a power splitter and a reference power sensor. The VNA-based system with port 1 involved is shown in Fig. 4. Both circuits consisted of a standard and a DUT power sensors, as well as a power meter for power readings from the sensors. The generator-splitter setup in addition to a signal generator and a power splitter has a reference power sensor with its corresponding power meter. It is evident from the photos that using the system with a VNA requires less equipment, although is more expensive. However, as mentioned earlier, the VNA is necessary for power sensor calibration and most likely it is already present in the calibration laboratories.

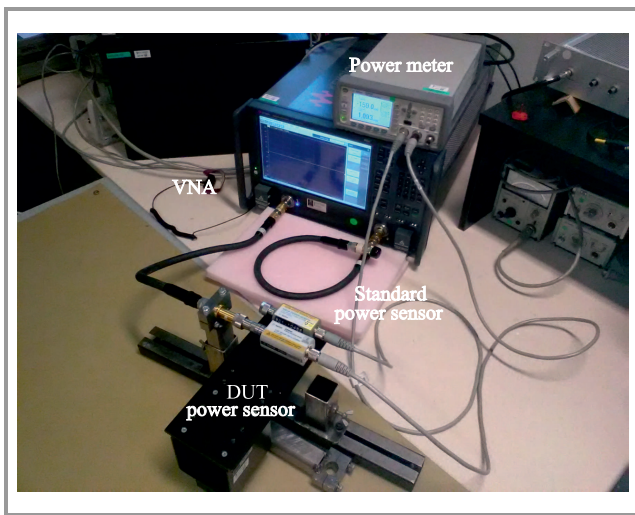


Fig. 4. VNA-based measurement system for power sensor calibration in direct comparison transfer system.

4.2. Source Match Evaluation

First, the source reflections at the splitter output and at port 1 of the VNA with cable were evaluated. With this purpose

an electronic calibration unit was used. The source match characterization of the splitter was performed using the Juroshek method [6]. The source match of the VNA system was retrieved as an error term after oneport reflection calibration [9]. Figure 5 shows the results of both source reflection measurements. The equivalent source match of the power splitter is better than the source match of the VNA in the whole frequency range of interest. For the frequencies up to 2 GHz a source reflection of the splitter is extremely low.

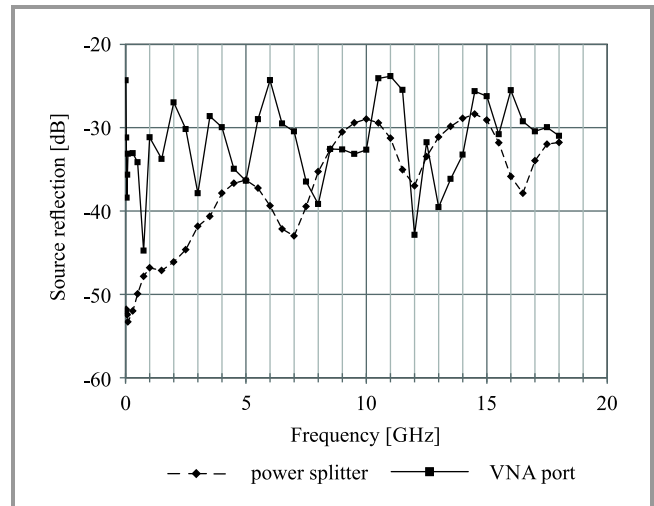


Fig. 5. Source reflections of the power splitter and VNA port 1 with cable.

4.3. Calibration Results

The results of power sensor calibration using a generator-splitter setup, using a VNA-based system, and those obtained from the reference lab are depicted in Fig. 6. Corresponding data for both systems are presented in Table 1. Results from a generator-splitter system seem a bit more consistent with results from the reference laboratory. Discrepancies are especially noticeable around 10 GHz. In order to evaluate the performance of the VNA system a figure of merit En was calculated [10]:

$$En = \left| \frac{K_{D,A} - K_{D,B}}{\sqrt{U_A^2 + U_B^2}} \right|, \quad (4)$$

where $K_{D,A}$, $K_{D,B}$ are the calibration factor results from the VNA method and from the generator-splitter method, respectively. U_A and U_B are corresponding expanded uncertainties. Uncertainties for both methods were calculated according to ISO GUM [11], to internal laboratory procedures, and to uncertainty budgets presented in [1], [3], [8]. A graph of the En parameter is shown in Fig. 7, and the corresponding data is presented in Table 1. According to [10] the comparison result is accepted when $En < 1$. In this case a maximum value of 0.79 is obtained for 10 MHz, which means that the obtained variances are acceptable.

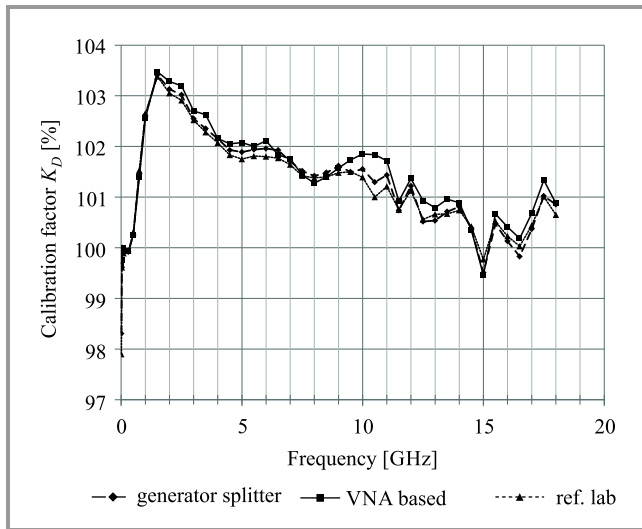


Fig. 6. Calibration factors of DUT power sensor obtained from generator-splitter system, VNA system, and from reference lab.

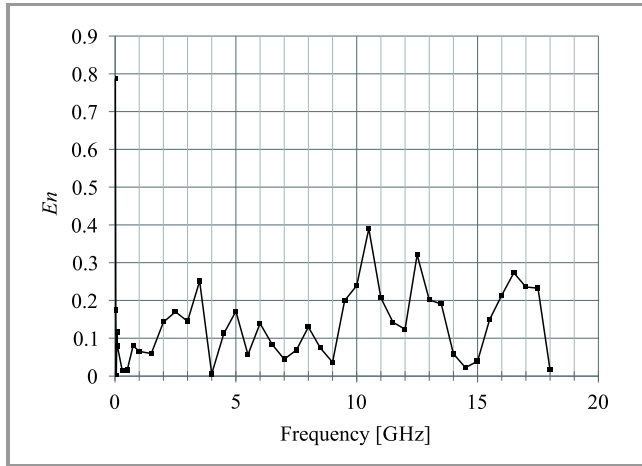


Fig. 7. E_n parameter for VNA method and generator-splitter method.

Generally quite good consistency is observable between the VNA-based method and the generator-splitter method. Second highest value of E_n is 0.39 for 10.5 GHz.

5. Conclusions

This article presented the development of a RF power sensor calibration setup which makes use of a VNA at frequencies ranging from 10 MHz to 18 GHz, in the Laboratory of Electric, Electronic and Optoelectronic Metrology at the National Institute of Telecommunications in Warsaw, Poland. As the preliminary work for full-range measurement capability to calibrate the power sensors up to 50 GHz, it proves the concept, that VNA can successfully be used to determine the calibration factor of power sensors in the direct comparison transfer system instead of a well-known generator-splitter method. The comparison between the two methods showed a good mutual consistency.

Table 1
Power sensor calibration results

Freq. [GHz]	$K_{D,A}$ [%]	U_A [%]	$K_{D,B}$ [%]	U_B [%]	E_n
0.01	99.69	1.50	98.31	0.92	0.79
0.03	99.76	0.61	99.61	0.57	0.17
0.05	100.00	0.38	100.00	0.37	0.00
0.075	100.00	0.63	99.90	0.62	0.12
0.1	99.99	0.63	99.92	0.62	0.08
0.3	99.93	0.71	99.92	0.70	0.01
0.5	100.25	0.71	100.27	0.71	0.02
0.75	101.39	0.72	101.47	0.72	0.08
1.0	102.55	0.74	102.62	0.72	0.07
1.5	103.47	0.75	103.41	0.74	0.06
2.0	103.28	0.76	103.13	0.73	0.14
2.5	103.20	0.76	103.02	0.73	0.17
3.0	102.70	0.74	102.55	0.73	0.14
3.5	102.63	0.81	102.35	0.73	0.25
4.0	102.16	0.80	102.15	0.73	0.01
4.5	102.05	0.79	101.93	0.77	0.11
5.0	102.08	0.78	101.89	0.76	0.17
5.5	102.00	0.80	101.94	0.76	0.06
6.0	102.11	0.82	101.96	0.75	0.14
6.5	101.84	0.77	101.93	0.76	0.08
7.0	101.75	0.78	101.71	0.75	0.04
7.5	101.44	0.78	101.51	0.78	0.07
8.0	101.27	0.78	101.42	0.79	0.13
8.5	101.40	0.83	101.48	0.83	0.07
9.0	101.57	0.84	101.62	0.84	0.04
9.5	101.74	0.85	101.49	0.87	0.20
10.0	101.85	0.86	101.55	0.90	0.24
10.5	101.84	1.07	101.29	0.90	0.39
11.0	101.72	1.06	101.43	0.87	0.21
11.5	100.92	0.99	100.74	0.84	0.14
12.0	101.37	0.82	101.22	0.84	0.12
12.5	100.93	0.91	100.52	0.90	0.32
13.0	100.79	0.86	100.54	0.93	0.20
13.5	100.96	0.87	100.71	0.96	0.19
14.0	100.89	0.89	100.81	0.98	0.06
14.5	100.35	1.05	100.38	0.97	0.02
15.0	99.46	0.92	99.51	0.87	0.04
15.5	100.67	0.93	100.47	0.93	0.15
16.0	100.41	0.94	100.13	0.92	0.21
16.5	100.18	0.92	99.83	0.92	0.27
17.0	100.68	0.94	100.37	0.94	0.24
17.5	101.34	0.97	101.02	0.96	0.23
18.0	100.88	0.99	100.86	0.98	0.02

References

- [1] M. P. Weidman, "Direct comparison transfer of microwave power sensor calibrations", NIST Technical Note 1379, Washington: 1996 [Online]. Available: <https://www.nist.gov/document-14769>

- [2] Y. Shan, Y. S. Meng, and Z. Lin, "Generic model and case studies of microwave power sensor calibration using direct comparison transfer", *IEEE Transac. on Instrumentat. and Measurement*, vol. 62, no. 6, pp. 1834–1839, 2013 (DOI: 10.1109/TIM.2012.2225961).
- [3] K. Wong, "Complete power sensor calibration using a VNA", in *Proc. 80th ARFTG Microwave Measurement Conf.*, San Diego, CA, USA, 2012, pp. 1–5 (DOI: 10.1109/ARFTG.2012.6422420).
- [4] W. K. P. Angin, J. Kwon, T. Kang, and N. Kang, "Comparison of RF power sensor calibration using a vector network analyzer and a direct transfer system", in *2016 URSI Asia-Pacific Radio Sci. Conf. (URSI AP-RASC)*, Seoul, Korea (South), 2016, pp. 1754–1756 (DOI: 10.1109/URSIAP-RASC.2016.7601209).
- [5] R. A. Johnson, "Understanding microwave power splitters", *Microwave J.*, pp. 49–56, 1975.
- [6] J. R. Juroshek, "A direct calibration method for measuring equivalent source mismatch", *Microwave J.*, vol. 40, no. 10, pp. 106–118, 1997.
- [7] J. R. Fenton, "Vector-corrected power sensor calibration", in *51st ARFTG Conf. Digest*, Baltimore, MD, USA, 1998 (DOI: 10.1109/ARFTG.1998.327286).
- [8] F. Aldossary, Z. Huneiti, Z. Hunaiti, and W. Balachandran, "The network analyser (HP8510C) as a transfer instrument for power sensor calibration", in *2008 IEEE Instrument. and Measurement Technol. Conf.*, Victoria, BC, Canada, 2008, pp. 1249–1253 (DOI: 10.1109/IMTC.2008.4547233).
- [9] J. Fitzpatrick, "Error models for systems measurements", *Microwave J.*, pp. 63–66, 1978.
- [10] ISO 13528, "Statistical methods for use in proficiency testing by interlaboratory comparisons", *Int. Organization for Standardization*, Geneva, 2005 [Online]. Available: <https://www.iso.org/standard/35664.html>
- [11] "Evaluation of measurement data – Guide to the expression of uncertainty in measurement", *Joint Committee for Guides in Metrology*, 2008 [Online]. Available: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6



Jarosław Szatkowski received his B.Sc. and M.Sc. in Electronics from Warsaw University of Technology, Poland, in 2015 and 2017 respectively. Since 2017 he is with Laboratory of Electrical, Electronic and Optoelectronic Metrology in National Institute of Telecommunications where he is involved in calibration of RF metrology

instrumentation. His research interests focus mainly on microwave metrology such as vector network analysis and RF power measurements.

E-mail: j.szatkowski@il-pib.pl

National Institute of Telecommunications

Szachowa 1

04-894 Warsaw, Poland

Network Traffic Classification in an NFV Environment using Supervised ML Algorithms

Gjorgji Ilievski¹ and Pero Latkoski²

¹ Makedonski Telekom AD Skopje, Skopje, RN Macedonia

² Ss. Cyril & Methodius University, Skopje, RN Macedonia

<https://doi.org/10.26636/jtit.2021.153421>

Abstract—We have conducted research on the performance of six supervised machine learning (ML) algorithms used for network traffic classification in a virtual environment driven by network function virtualization (NFV). The performance-related analysis focused on the precision of the classification process, but also in time-intensity (speed) of the supervised ML algorithms. We devised specific traffic taxonomy using commonly used categories, with particular emphasis placed on VoIP and encrypted VoIP protocols serve as a basis of the 5G architecture. NFV is considered to be one of the foundations of 5G development, as the traditional networking components are fully virtualized, in many cases relaying on mixed cloud solutions, both of the premise- and public cloud-based variety. Virtual machines are being replaced by containers and application functions while most of the network traffic is flowing in the east-west direction within the cloud. The analysis performed has shown that in such an environment, the Decision Tree algorithm is best suited, among the six algorithms considered, for performing classification-related tasks, and offers the required speed that will introduce minimal delays in network flows, which is crucial in 5G networks, where packet delay requirements are of great significance. It has proven to be reliable and offered excellent overall performance across multiple network packet classes within a virtualized NFV network architecture. While performing the classification procedure, we were working only with the statistical network flow features, leaving out packet payload, source, destination- and port-related information, thus making the analysis valid not only from the technical, but also from the regulatory point of view.

Keywords—classification, machine learning, network functions virtualization, network traffic.

1. Introduction

Classification of network traffic is always important, as network architectures are changing continuously, especially now, when virtual machines (VM), software defined networking (SDN), private, public, and mixed clouds are commonplace solutions used in the IT world. The current trend favors microservices, containers, application functions and, network functions in network functions virtualization (NFV) environments [1], meaning that network

flows are becoming ever more complex. Currently, the majority of network traffic is moving in the cloud, usually within the same datacenter, in the east-west direction. This traffic never leaves the virtual plane and is often managed by SDN components in the NFV environment, thus obstructing the capture or any other operations over the same traffic. This is important both for cloud operators and for entities using the services provided via public clouds. Operations which are common practice and are considered trivial, such as quality of service (QoS), network security, optimization, application management and monitoring functionalities, are becoming a challenge.

In this paper, we are performing an experimental test to reveal network traffic classification efficiency of several supervised machine learning (ML) algorithms. We have created a unique test environment that resembles real life processes and simulates the east-west traffic on the virtual plane, exchanged between virtual hosts, with NFV established. Efficiency of ML algorithms is explored from the point of view of classification precision, but also from the point of view of computational speed. This is very important when we take into consideration the penetration of 5G, as it is tightly integrated with the cloudification of networking operations. For example, the 5G specification calls for a user plane latency of as little as 1 ms for ultra-reliable low-latency communications (URLLC) [2]. This is why the speed of the ML algorithm is crucial and why the process must be performed in a manner that will minimize the expected latency added by the classification.

The study we have conducted provides a novel scenario that is comparable to emerging architectures with NFV and 5G implemented therein. It involves 6 different supervised ML algorithms: Bayes Net, NaiveBayes, J48, K-Nearest Neighbors (KNN), Decision Tree and AdaBoost, as they are the ones that are widely used in traditional computer networks, are proven to be reliable while simultaneously providing valid classification results, and are easy to implement in practice. We have used Weka [3] as a tool for classification.

The taxonomy used in this paper relies on 6 classes which are chosen based on our experience in traditional networks

and remain in alignment with the network traffic expected within 5G radio, as well as 5G core networks: VoIP, encrypted VoIP, DNS, Management, SSH, HTTP and HTTPS traffic. It is our intention to highlight VoIP and encrypted VoIP classifications which are crucial for ensuring QoS capabilities of 5G networks, thus enabling smart connectivity and providing the ability to steer, secure and break out network traffic.

The NFV architecture is becoming a true 5G enabler, providing the ability to place initial workloads within the network and allowing them grow towards the edge, thus offering the basis needed for the expansion of IoT expected with the growth in 5G penetration.

Numerous previous papers have been devoted to the issue of ML algorithms used for performing packet inspection [4]–[7]. The novel experimental testbed and the method classifying network data based on the statistical parameters of packets and on packet flows only, without relying on source and destination addresses (both MAC and IP addresses), without any examination of the payload and without analysis of the communication ports, are the features that distinguish the approach we have adopted.

The volume of encrypted network traffic is growing fast. Significant numbers of services and applications are using encryption as a primary method of securing information. But this has made traffic classification a challenge. The solution that we propose is applicable in practice without compromising data privacy and integrity. It provides an insight into the performance of supervised ML algorithms and determines which one is best suited for NFV-based environments.

There are also many examples of ML algorithms used for deep packet inspection (DPI) in traditional networks [3], [8], [9]. Unlike the aforementioned works, we focus on virtualization and the NFV environment. In such a scenario, network packets are mostly moving in the east-west direction and are often encrypted, meaning that no classic DPI may be conducted. In the proposed approach, it is not important whether the payload is encrypted or not. Legal requirements related to performing DPI in a cloud environment (especially a public cloud) are satisfied as well, since the data carried within the payload is not compromised. We are using the statistical features of the network packets and the network flows only to create datasets that are later used for training and testing the ML algorithms. During the testing phase, we are evaluating the efficiency of the algorithm from the point of view of its precision, but also from the point of view of its speed. Network traffic is sniffed directly inside an open vSwitch. We are not introducing any additional probes or SDN components to capture the traffic. We take into consideration all network traffic between the specific virtual elements making up the environment, but also traffic that is used in managing that environment (including that originating from controllers). Incoming and outgoing Internet traffic is dealt with as well. Such a scenario is realistic with majority of cloud solutions. In addition to its precision, the speed of an ML algorithm is even more important in many instances. If the time con-

sumed to classify the data is adding significant latency to network traffic, and if it is consuming the resources (CPU time, memory usage) of the cloud, precision of the classification process is not as relevant.

In the remainder of the paper, we will go through the related work on the subject, briefly explained in Section 2. The experimental setup and the dataset creation procedure are explained in Section 3, while the results are analyzed in Section 4. Section 5 is devoted to the conclusion and our plans for future work.

2. Related Work

Many researches focus on DPI-related aspects and scenarios involving SDN components [10]–[12]. Others research security-related aspects of performing DPI [13], [14] by using SDN probes for sniffing network traffic and for processing data. This work may be distinguished by its NFV-based setup and targets to ensure complete isolation of the packet payload. Some authors consider the classification of network traffic in traditional networks [15], [16] without tackling the specifics of virtualization which is a very trendy solution and forms an important aspect of our work. Parsaei *et al.* [17] are using SDN to categorize traffic by application, using different variants of the neural network estimator. They are using data mining techniques based on different ML algorithms and propose a controller that could dynamically allocate bandwidth to network flows thus optimizing resource allocation. They achieve a classification accuracy rate of over 97%. Unlike in the work described herein, they use source and destination IPs, as well as the transport layer port for classification purposes. In [18], QoS in an SDN based network is researched with an emphasis placed on overcoming the limitations of traditional networking architectures. Different flow routing mechanisms are categorized there. In this research, we explore classification as a basic concept from which QoS may benefit significantly.

Paper [4] is a study in which the NFV environment is created to classify different types of TCP traffic using three supervised ML algorithms: NaiveBayes, Bayes Net and J48. Network packets are analyzed individually, meaning that three different datasets are obtained: traditional, virtual and combined, in order to compare the performance of different classification approaches. Only statistical parameters of the packets are used. In our case, we use TCP- and UDP-based traffic and analyze the statistical parameters of packet flows within an NFV environment that closely resembles cloud platforms.

Le *et al.* [19] applied big data, ML algorithms, SDN, and NFV to build a practical and powerful framework for clustering, forecasting, and managing traffic behaviors for a huge number of base stations with different statistical traffic characteristics typical of different types of cellular networks (GSM, 3G, 4G). The framework was intended for developing future 5G self-organizing network (SON) applications. Several traffic forecasting-based applications are

introduced as well. Five ML algorithms are used to classify traffic generated by mobile applications, with QoS implemented to enable bandwidth guarantees. The conclusion is that Decision Tree offers the best overall performance of all the algorithms tested. Our experiment is limited to the transport network layer, with the aim to classify traffic that is mostly exchanged along the east-west route, using ML algorithms, but also to evaluate the time needed to conclude the classification process, as it is crucial for the future 5G environments.

Alshammari *et al.* [5] focused on VoIP traffic within traditional networks. Data is extracted from the existing network environment with a complex topology. The authors evaluate the classification of both encrypted and unencrypted VoIP using three ML algorithms: C5.0, ADA Boost and GP Classifier, and relying on the subset sampling technique. In the experiments, C5.0 showed the best performance and the highest precision rate. Here, a cloud-based environment with NFV is used to rate the individual ML algorithms dealing with various types of network traffic.

In [20], a machine learning-based classification of multi-service Internet traffic is used to evaluate the use of resources (CPU time and system memory). We are complementing this research, as we are evaluating the time needed by the ML algorithms to perform the classification.

Article [21] proposes a network traffic classification method based on a deep learning network structure. The experimental dataset is created from ten types of data, each of which abstracted from a complete TCP bidirectional stream containing 249 network flow attributes. Google's TensorFlow deep learning framework is used in the experimental environment. NaiveBayes and Decision Tree ML algorithms are used to compare the efficiency of classification performed by the deep learning network. Compared to this work, we are targeting different supervised ML algorithms, having in mind that not only classification precision, but also the time needed to perform the classification is important, as any delay added to the network packet's speed may be a source of a functional problem in the environment.

The effect of attaching NFV elements to network traffic, especially in terms of an increase or decrease in the volume of traffic processed, is researched in [22]. The authors develop an algorithm that determines the flow path and then proposed a least-first-greatest-last routing.

Bonfiglio *et al.* [23] are researching traffic specifics of Skype, as the application is based on encrypted VoIP for voice calls. Traffic is explored in real time, by applying two different approaches and using the statistical parameters of the traffic generated traffic by Skype. The approaches are then assessed using the flow correlation technique.

To summarize, our testing setup is similar to that introduced in [4], with additional elements added to the environment, such as virtual machines connected to the Internet and virtual network elements with bridged IP addresses. Both TCP and UDP traffic is generated, with and without encryption. The classification groups and labels are chosen in a manner allowing to classify various types of traffic. Viber and

Skype are used to generate VoIP traffic, whereas scripts are used to open SSH management sessions for different hosts. Furthermore, a novel testbed is proposed in the context of 5G and to accommodate the usage of NFV elements within the virtualized environment, as expected in the real-life setup. Network packets are analyzed directly within the virtual switch, without the use of a probe or an SDN element. Statistical characteristics are extracted from TCP and UDP packet flows and are used to perform further steps of the analysis.

3. Experimental Setup and Dataset Creation

To simulate the east-west traffic within a virtualized NFV-based network, the proposed experimental environment is based on Oracle VirtualBox [24] which is installed on a single physical host with an Ubuntu 18.04 Server. All components are connected with an Open vSwitch (OVS) [25], [26] that ensures network connectivity. The switch is connected to the Internet through the host in a bridge mode. All network packets flow through the OVS switch – this includes east-west traffic packets and north-south traffic packets, both sent to and originating from the Internet. Traffic is captured directly on the OVS using Wireshark and tshark [27].

Mininet [28] is used as a network simulator. Two different installations on two separate virtual machines are used, each with a different network topology having 100 hosts, 20 switches and links between them and to the OVS. The hosts within the simulated networks have private IP addresses and are capable of communicating with each other. GRE tunneling is used to link the two simulated Mininet networks. Some of the hosts within Mininet have NAT-ed IP addresses and are able to communicate with the Internet. The Ryu Controller [29] is used to control the simulated Mininet networks. It is installed and configured on a separate virtual machine.

There are four other virtual machines connected to the OVS which are also used for traffic generation. Skype and Viber are installed thereon to simulate VoIP traffic. When initiated, VoIP needs access to the Internet, but later on peer-to-peer communications may be observed within the OVS, in a fully east-west direction. The script that initiates ssh sessions is enabled on the VMs. We have developed a Python script that automatically starts SSH sessions with the Mininet hosts as well. The SSH sessions were started in time intervals that are following Poisson distribution.

A distributed Internet traffic generator (D-ITG) [30] generates various types of TCP and UDP traffic among the hosts within the Mininet. Different scripts are used to generate traffic at packet level, replicating specific stochastic processes for both inter departure time (IDT) and packet size (PS) random variables.

Figure 1 shows an overview of the experimental setup, showing its components symbolically.

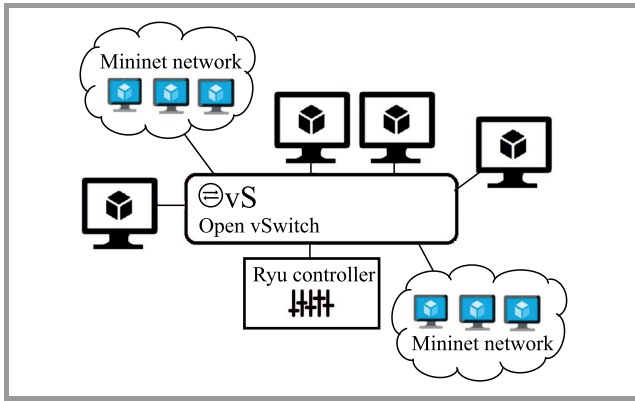


Fig. 1. Experimental environment.

We have performed 50 different experiments to generate various types of traffic (using D-ITG, Skype, Viber, custom scripts) and to analyze it. The experiments were conducted in time intervals varying from 4 to 20 minutes, with VoIP calls lasting from 10 s to 10 minutes, following Poisson distribution. One dataset per experiment was generated. Different D-ITG scripts for different traffic simulations were used in each of the experiments. The scripts used different Mininet hosts and different paths in each attempt. The average number of packets captured was 1.262.375 and the average number of flows was 4090. We have devised a specific classification of traffic, relying on commonly used classes, based on experience from the traditional networks. As it will be shown in the results, precision of the classification process was calculated as an overall figure, but also independently for each of the classes, in order to calculate the macro-average precision level in which the contribution of each class is treated equally (as the number of packets and flows varies for every class).

We used the following labels for the individual classes: DNS – for all traffic used for name resolution, NETMGMT – all traffic used for host and network management, SSH – for the SSH sessions in the environment, WEB – for HTTP and HTTPS traffic, VOIP – for VoIP traffic, SVOIP – for encrypted VoIP. Based on the Wireshark pcap files generated, UDP and TCP packet flows, as well as the classes used for ML training and then for determining and confirming the level of precision, are identified using Argus [31]. Similarly to [5], we define a flow as a bidirectional connection between two hosts. TCP flows are terminated either by flow time-out or by connection tear-down, whereas UDP flows are ended by flow time-out only. When observing flows within the OVS, one could notice that most of the traffic is of the east-west variety, is taking place inside the virtual layout and between the hosts, but flows from the management generated by the hypervisor and the Ryu controller could be detected as well. Because our focus was on the NFV-based environment, some of the flow features, such as the source and destination IP, MAC address, as well as the communication port that can vary inside the virtual environment, were not taken into consideration.

To train and to test the supervised ML algorithms, we have used Weka [3], [32]. 2/3 of each dataset were used for training, while 1/3 was used for testing each of the algorithms. As not all the attributes contribute to the classification equally, the AttributeSelectedClassifier with Ranker as an attribute ranking algorithm was used. InfoGainAttributeEval was used as an evaluator that determines the gain of information that the attributes carry. With this approach, we ranked the attributes that are used for the algorithms, with the information gain of every attribute being evaluated thereafter. This approach prevents potential data leakage. Based on experience from traditional networks and thanks to a careful observation of the datasets obtained, we have selected the attributes given in Table 1 as features that characterize the flows. The payload is not used due to the privacy of cloud environments and due to the use of different encryption methods that will make the payload irrelevant for classification purposes. The labels in the transport layer header (e.g. the port numbers) are not used as well, as they may be changed easily. A short explanation of each of the selected attributes is provided inside the table. The following section presents the results of the test involving the supervised ML algorithms and contains their analysis.

Table 1
Flow attributes

Abbreviation	Feature
proto	Transaction protocol
rate	Packets per second
srate	Source packets per second
drate	Destination packets per second
sintpkt	Source interpacket arrival time
dintpkt	Destination interpacket arrival time
sjit	Source jitter
djit	Destination jitter
mdoffset	Mean of the data offset Values of the packets in the flow
smeansz	Mean of the flow Packet size transmitted by the source
dmeansz	Mean of the flow packet Size transmitted by the destination
smaxsz	Max packet size for source
dmaxsz	Max packet size for destination
sminsz	Min packet size for source
dminsz	Min packet size for destination

4. Results and Analysis

We have conducted 50 experiments, creating 50 datasets. All the ML algorithms were tested on each dataset. The performance of each algorithm was defined as a combination of its precision and the time needed to perform the classification. Since time consumption is correlated to the

performance of the machine on which the analysis is conducted, all classification tasks were performed on the same machine, with all processes active thereon that may influence performance observed carefully. A mean value of 50 results was derived for all target metrics.

True positive (TP), false positive (FP), true negative (TN) and false negative (FN) rates are defined as:

- TP is the number of instances that are correctly identified as belonging to a specific class,
- FP is the number of instances that are not correctly identified as belonging to a specific class,
- TN is the number of instances that are correctly identified as not belonging to a specific class,
- FN is number of instances that are not correctly identified as not belonging to a specific class.

The overall precision of the algorithms is calculated as the proportion between TP instances and all instances in the dataset [32]:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Table 2 shows the average precision of the algorithms in all 50 experiments with the statistical standard deviation across the experiments, as a weighted average value.

Table 2
Algorithm precision

No.	ML algorithm	Precision
1	AdaBoost	0.7440±0.0292
2	BayesNet	0.9672±0.0189
3	J48	0.9906±0.0027
4	KNN	0.9172±0.0438
5	NaiveBayes	0.8634±0.0170
6	Decision Tree	0.9914±0.0033

It can be seen that the Decision Tree algorithm has the best overall precision. It is followed by J48 and BayesNet. On the other hand, the AdaBoost algorithm has the worst overall performance with the lowest precision of 74.4%. In order to perform a deeper analysis of the precision level, micro average precision was calculated – an indicator that aggregates the contribution of all classes and calculates the average metric, as given by Eq. 2. The results are presented in Table 3.

$Precision_{MIC} =$

$$\frac{TP_1 + TP_2 + \dots + TP_N}{TP_1 + FP_1 + TP_2 + FP_2 + \dots + TP_N + FP_N} \quad (2)$$

Not all classes have same or similar number of packets and flows, and the data distribution is skewed. As the class distribution is unequal, the datasets are imbalanced. To avoid the problem of data balancing and to come to valid conclusions, we are calculating macro average precision, recall, the F1-score.

Table 3

Micro average precision of algorithms

No.	ML algorithm	Micro average precision
1	AdaBoost	0.8450±0.0176
2	BayesNet	0.9954±0.0027
3	J48	0.9984±0.0006
4	KNN	0.9856±0.0073
5	NaiveBayes	0.9752±0.0027
6	Decision Tree	0.9984±0.0010

Macro average precision is the average of measure of each class. This means that every class will weigh the same in the macro average precision. Equation 3 is used to calculate macro average precision (Precision_{MAC}), where Pr₁, Pr₂, etc. denote the precision of the algorithm in relation to the individual classes.

$$Precision_{MAC} = \frac{Pr_1 + Pr_2 + \dots + Pr_N}{Count(Pr)} \quad (3)$$

The results are shown in Table 4, where the statistical standard deviation is calculated for the precision between classes.

Table 4

Macro average precision of algorithms

No.	ML algorithm	Macro average precision
1	AdaBoost	0.20335±0.3064
2	BayesNet	0.88990±0.1489
3	J48	0.98240±0.0148
4	KNN	0.82735±0.2202
5	NaiveBayes	0.78915±0.2048
6	Decision Tree	0.98480±0.0107

It becomes clear that the algorithms are not performing in the same manner with regard to all the classes. The Decision Tree algorithm has the highest macro average precision rate and the lowest standard deviation between classes, meaning that it classifies all classes similarly. J48 is very close to Decision Tree, with the precision rate of over 98%. On the other end of the scale, the AdaBoost algorithm shows a very low macro average precision rate with a high standard deviation, meaning that it performs poorly with regard to different classes. The K-Nearest Neighbor algorithm is underperforming as well, with its macro average precision rate equaling 82% only. After comparing these results with the standard weighted precision shown in Table 2, one may see that the algorithms have the same order, but the macro precision rate of the lower-end algorithms is worse, leading to the conclusion that AdaBoost and KNN offer different precision levels for different classes.

In order to evaluate the impact of the false negative classified instances, Recall is used as a model metric. It is the proportion between true positive instances and total actual instances:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Recall was used to calculate the F1-score of the ML algorithms tested in our experiments. It is a metric that balances the precision level and the recall, so that false negative instances are taken into consideration. F1-score is calculated as a harmonic mean of the precision and the recall:

$$F1\text{-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Table 5 shows the F1-score values calculated for our experiments. The Decision Tree ML algorithm has the best F1-score, followed by J48, BayesNet, KNN, NaiveBayes and AdaBoost. The last algorithm has the F1-score of 23.2% only, with very high standard deviation.

Table 5
F1-score

No.	ML algorithm	F1-score
1	AdaBoost	0.231575±0.3356
2	BayesNet	0.913425±0.1055
3	J48	0.975425±0.0212
4	KNN	0.797425±0.2295
5	NaiveBayes	0.782125±0.1510
6	Decision Tree	0.980475±0.0152

The tables are visually represented in Figs. 2 to 5.

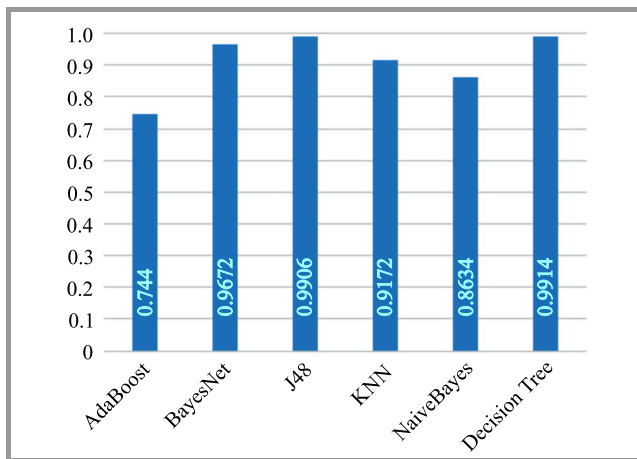


Fig. 2. Algorithm precision.

Precision of the algorithm is only one of the characteristics that determines its actual usability. The time needed to perform the classification is an important aspect as well. If the time needed to complete the classification is too long, the process will add latency to network communications, thus making the benefit of the classification too costly. This is important especially in protocols in which latency may degrade the quality of service, such as VoIP. Furthermore, this is also crucial in 5G scenarios, where latency is one of the major concerns. Consumption of the system's resources (CPU, memory, etc.) is another problem, as it increases if the algorithm operates as a slower pace. The two metrics (precision and time consumption) combined determine the overall performance of the algorithms.

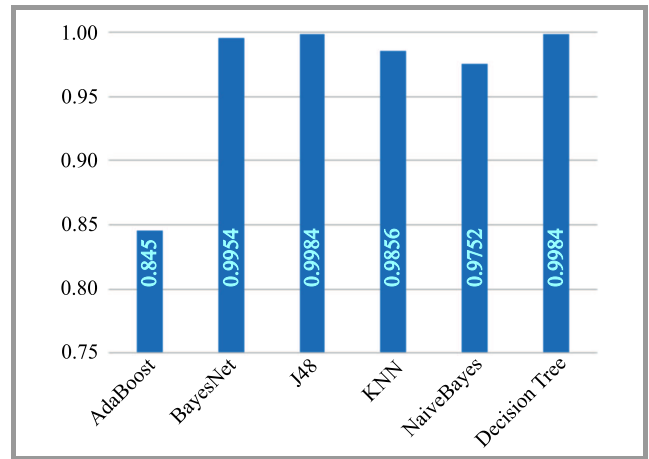


Fig. 3. Micro average precision.

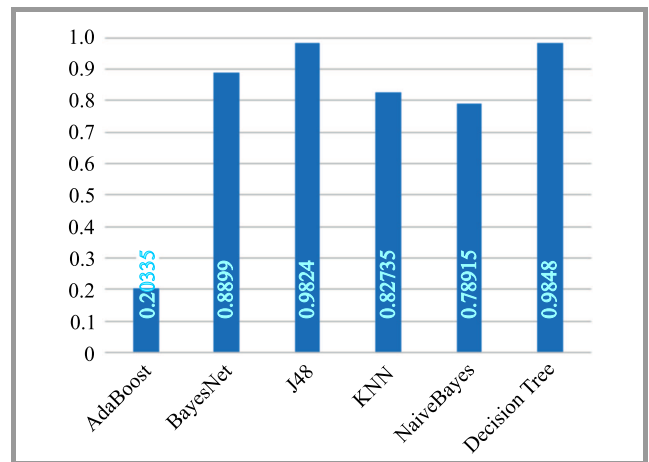


Fig. 4. Macro average precision.

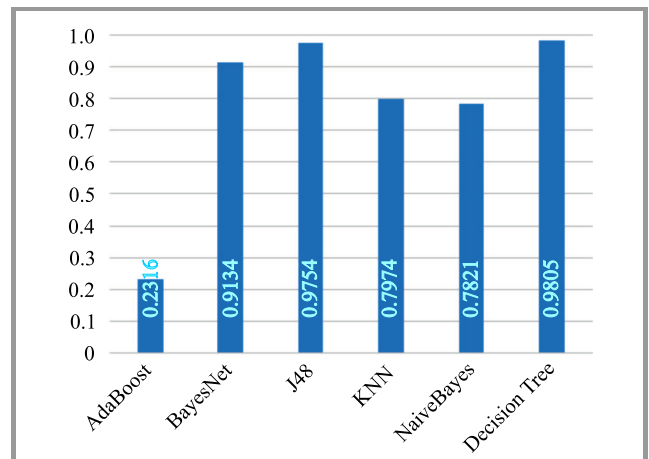


Fig. 5. F1-score.

The time that we have measured is relative to our testbed environment. All experiments are performed in the same environment, with special care taken to isolate all unnecessary processes. The average time consumption value was calculated from 50 experiments.

Table 6 shows the average time needed by the algorithms to perform the classification procedure within the 6 chosen classes.

Table 6
Average time needed for classification

No.	ML algorithm	Average time [s]
1	AdaBoost	0.012
2	BayesNet	0.016
3	J48	0.022
4	KNN	0.272
5	NaiveBayes	0.104
6	Decision Tree	0.016

The results concerning the average time required to perform the classification show that the AdaBoost algorithm is the fastest. Decision Tree and BayesNet algorithms are ranked second and third ex-equo, being 25% slower than AdaBoost. The result of J48 is satisfactory as well. NaiveBayes is almost 9 times slower than AdaBoost and more than 6 times slower than Decision Tree. The KNN algorithm is the slowest. Decision Tree and AdaBoost require only 5.9% of the time needed by KNN to perform the classification.

Figure 6 graphically represents the average time required by the algorithms to perform the classification.

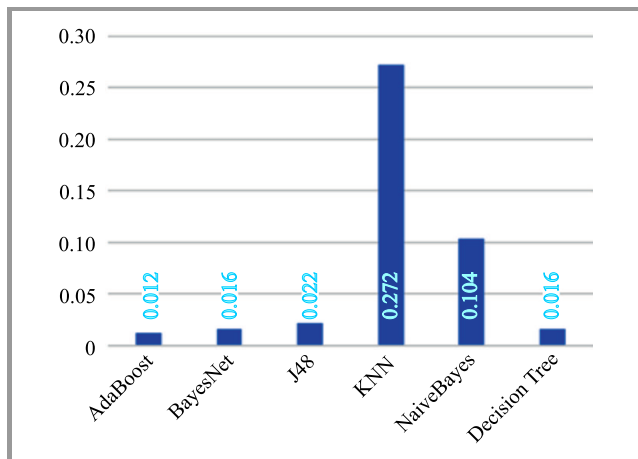


Fig. 6. Time required to perform the classification [s].

To summarize, when we take a look at both the precision and the time needed for classification, the Decision Tree supervised ML algorithm offers the best overall performance. Although AdaBoost is the fastest algorithm, its classification precision is poor and unsteady across different classes, which makes this algorithm unreliable for the scenario in question. J48 also offers a high level of precision that is evenly distributed among the classes, but it is slower than Decision Tree and BayesNet. Nevertheless, its speed similar to that of Decision Tree and BayesNet algorithms, which makes it a valid choice as well. BayesNet offers a high degree of precision, but macro average precision and F1-score values show that the distribution of its precision among the

different classes is not as good as in the case of Decision Tree and J48.

NaiveBayes is in the middle of the scale, both in terms of precision and time. KNN, in turn, offers macro average precision of approximately 83% and F1-score of 80%, but it is by far the slowest algorithm, meaning that it is only useful in situations in which the time needed to perform the classification is of little importance.

5. Conclusion and Future Work

The main idea behind this paper was to present a method for creating datasets based only on the statistical characteristics of network traffic flows, and to test the performance of machine learning algorithms based on the created datasets. All those tasks were performed with the use of an experimental testbed with NFV architecture.

The efficiency of algorithms is examined taking into consideration their precision and the time required to perform the classification. Such an approach is important from the point of view of virtualization point of view, where mixed cloud scenarios are commonplace, but also from the point of view of the growing popularity of 5G, where network latency is crucial.

Our experimental testbed was used to perform multiple experiments and to collect network traffic data from which IP flows were extracted. The statistical features of the flows were used as attributes for the classification procedure. Because such attributes as source and destination IP, MAC addresses and communication ports may vary within a virtualized environment, they are not taken into consideration. Due to encryption and data privacy concerns, the payload of the data packets is also excluded from the datasets and it is not used for classification purposes.

The environment used did not rely on any network probes or SDN elements to collect the data, allowing not to affect the east-west traffic in any manner whatsoever. The traffic was fully intercepted within the virtual layer, where it resides naturally. Such an approach has an impact on resource consumption as well, minimizing additional latency that may be added to network packets by redirecting or by port replication used in the traditional DPI.

The results have shown that the Decision Tree algorithm offers the best overall performance, both from the point of view of classification precision and time consumption. It has proved as a reliable classifier that is performing evenly across different classes. J48 and BayesNet are also performing well, with J48 having slightly better precision and BayesNet being faster. K-Nearest Neighbour and NaiveBayes have an average classification precision of approximately 80%, but they are slow. This applies, in particular, to KNN which is almost 20 times slower than Decision Tree and BayesNet. AdaBoost shows the worst performance with its precision varying considerably among the different classes. The same applies also to its macro average precision and F1-score.

The analysis presented in this paper may be relied upon in practice within multiple systems that are built on top of cloud environments. NFV elements are now an unavoidable part of such infrastructures. The 5G infrastructure relies on these types of systems, and connectivity with such systems is most likely to rely on 5G access technologies. In those examples, QoS, network and application security, data management, system and process monitoring and control all depend on a valid network traffic classification scheme that needs to be precise and fast, without consuming excessive amounts of system resources.

For future work, we are planning to evaluate the impact of the number of classes on the classification results and the time intensity of the supervised ML algorithms, by introducing large numbers of classes and by reducing the classes. Another idea is to expand the experimental testbed to include multiple hosts and distributed switches, and to evaluate a network that is moving across multiple hosts.

References

- [1] M. Chiosi *et al.*, “Network Functions Virtualisation”, *Introductory White Paper*, 2015 [Online]. Available: https://portal.etsi.org/nfv/nfv_white_paper.pdf (accessed on 10.10.2020).
- [2] M. Eiman, “Minimum Technical Performance Requirements for IMT-2020 radio interface(s). Presentation”, 2018 [Online]. Available: https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Documents/S01-1_Requirements%20for%20IMT-2020_Rev.pdf (accessed on 10.10.2020).
- [3] E. Frank, M. A. Hall, and I. H. Witten, *Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition*. San Francisco, CA, USA: Morgan Kaufmann, 2016, pp. 2464–2468 (ISBN: 9780128042915).
- [4] J. Vergara-Reyes, M. C. Martinez-Ordonez, A. Ordonezy, and O. M. C. Rendon, “IP traffic classification in NFV: a benchmarking of supervised machine learning algorithms”, in *IEEE Colombian Conf. on Commun. and Comput.*, Cartagena, Colombia, 2017 (DOI: 10.1109/ColComCon.2017.8088199).
- [5] R. Alshammari and A. Nur Zincir-Heywood, “Identification of VoIP encrypted traffic using a machine learning approach”, *J. of King Saud University – Computer and Informat. Sci. archive*, vol. 27, no. 1, pp. 77–92, 2015 (DOI: 10.1016/j.jksuci.2014.03.013).
- [6] B. Ma, H. Zhang, Y. Guo, Z. Liu, and Y. Zeng, “A Summary of Traffic Identification Method Depended on Machine Learning”, *Sensor Networks and Signal Process. (SNSP) 2018 Int. Conf.*, Xi’an, China, 2018, pp. 469–474 (DOI: 10.1109/SNSP.2018.00094).
- [7] U. Trivedi and M. Patel, “A fully automated deep packet inspection verification system with machine learning”, *IEEE Int. Conf. on Advanced Networks and Telecommun. Systems*, Bangalore, India, 2016 (DOI: 10.1109/ANTS.2016.7947802).
- [8] S. Rezaei and X. Liu, “Deep Learning for Encrypted Traffic Classification: An overview”, *IEEE Commun. Mag.*, vol. 57, no. 5, 2019, pp. 76–81 (DOI: 10.1109/MCOM.2019.1800819).
- [9] M. Shafiq *et al.*, “Network traffic classification techniques and comparative analysis using machine learning algorithms”, in *Proc. 2nd IEEE Int. Conf. on Computer and Commun. (ICCC)*, Chengdu, China, 2016, pp. 2451–2455 (DOI: 10.1109/CompComm.2016.7925139).
- [10] U. Huang, P. Li, and S. Gu, “Traffic scheduling for deep packet inspection in software-defined networks”, *Concurrency and Comput.: Practice and Experience*, 2017 (DOI: 10.1002/cpe.3967).
- [11] M. Mousa, A. Bahaa-Eldin, and M. Sobh, “Software Defined Networking concepts and challenges”, *11th Int. Conf. on Computer Engin. & Systems (ICCES)*, Cairo, Egypt, 2016, pp. 79–90 (DOI: 10.1109/ICCES.2016.7821979).
- [12] L. Polčák *et al.*, “A High Level Policies in SDN”, *Int. Conf. on E-Business and Telecommun.*, Colmar, France, 2016, pp. 39–57 (DOI: 10.1007/978-3-319-30222-5_2).
- [13] J. Arevalo Herrera and J. E. Camargo, “A Survey on Machine Learning Applications for Software Defined Network Security”, *Applied Cryptography and Network Security Workshops ACNS*, Bogotá, Colombia, vol. 11605, 2019, pp. 70–93 (DOI: 10.1007/978-3-030-29729-9_4).
- [14] A. Chowdhary *et al.*, “SDFW: SDN-based Stateful Distributed Firewall”, *Project: Secured and Resilient Networking*, 2018 (DOI: 10.13140/RG.2.2.11001.93281).
- [15] S. Choudhury and A. Bhowal, “Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection”, *Int. Conf. on Smart Technol. and Management for Comput., Commun., Controls, Energy and Materials (ICSTM)*, Avadi, India, 2015, pp. 89–95 (DOI: 10.1109/ICSTM.2015.7225395).
- [16] M. Shafiq *et al.*, “WeChat text and picture messages service flow traffic classification using machine learning technique”, *IEEE 18th Int. Conf. on High Performance Comput. and Commun.; IEEE 14th Int. Conf. on Smart City; IEEE 2nd Int. Conf. on Data Sci. and Systems (HPCC/SmartCity/DSS)*, Sydney, NSW, Australia, 2016, pp. 58–62 (DOI: 10.1109/HPCC-SmartCity-DSS.2016.0019).
- [17] M. Reza, M. J. Sobouti, S. Raouf, and R. Javidan, “Network traffic classification using machine learning techniques over software defined networks”, *Int. J. of Adv. Computer Sci. and App.*, 2017 (DOI: 8.10.14569/IJACSA.2017.080729).
- [18] M. Karakus and A. Duresi, “Quality of Service (QoS) in Software Defined Networking (SDN): A survey”, *J. of Network and Computer App.*, 2016 (DOI: 80. 10.1016/j.jnca.2016.12.019).
- [19] L. Le, D. Sinh, B. P. Lin, and L. Tung, “Applying Big Data, Machine Learning, and SDN/NFV to 5G traffic clustering, forecasting, and management”, *4th IEEE Conf. on Network Softwarization and Workshops (NetSoft)*, Montreal, Canada, 2018 (DOI: 10.1109/NETSOFT.2018.8460129).
- [20] S. Zander and G. Armitage, “Practical machine learning based multimedia traffic classification for distributed QOS management”, *2011 IEEE 36th Conf. on Local Computer Networks*, Bonn, Germany, 2011, pp. 399–406 (DOI: 10.1109/LCN.2011.6115322).
- [21] J. H. Shu *et al.*, “Network traffic classification based on deep learning”, *First Int. Conf. on Advanced Algorithms and Control Engin.*, Pingtung, Taiwan, 2018 (DOI: 10.1088/1742-6596/1087/6/062021).
- [22] W. Ma, C. Medina, and D. Pan, “Traffic-aware placement of NFV middleboxes”, *IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, 2015, pp. 1–6 (DOI: 10.1109/GLOCOM.2015.7417851).
- [23] D. Bonfiglio, M. Mellia, M. Meo, D. Rossi, and P. Tofanelli, “Revealing Skype traffic: when randomness plays with you”, *ACM SIGCOMM Computer Commun. Review*, vol. 37, no. 4, pp. 37–48, 2007 (DOI: 10.1145/1282427.1282386).
- [24] Oracle VirtualBox [Online]. Available: <https://www.virtualbox.org> (accessed on 10.09.2020).
- [25] M. V. Bernal, I. Cerrato, F. Risso, and D. Verbeiren, “Transparent optimization of inter-virtual network function communication in open vSwitch”, *5th IEEE Int. Conf. on Cloud Netw. (Cloudnet)*, Pisa, Italy, 2016, pp. 76–82 (DOI: 10.1109/CloudNet.2016.26).
- [26] Linux Foundation, Open vSwitch Project, 2016 [Online]. Available: <http://www.openvswitch.org>
- [27] Wireshark [Online]. Available: <https://www.wireshark.org/> (accessed on 10.09.2020).
- [28] Mininet: An instant virtual network on your laptop (or other PC) [Online]. Available: <http://mininet.org> (accessed on 12.09.2020).
- [29] Ryu Framework [Online]. Available: <http://osrg.github.io/ryu/> (accessed on 10.09.2020).

- [30] A. Botta, A. Dainotti, and A. Pescapè, "A tool for the generation of realistic network workload for emerging networking scenarios", *Computer Networks (Elsevier)*, 2012, vol. 56, no. 15, pp. 3531–3547 (DOI: 10.1016/j.comnet.2012.02.019).
- [31] Argus Quosient [Online]. Available: <https://qosient.com/argus/> (accessed on 10.09.2020).



Gjorgji Ilievski received his M.Sc. in Computer Engineering in the field of Data Mining Technologies in 2012. His research interests include statistical analysis, cloud computing, computer networking, virtualization, LTE and 5G. He works in telecommunications industry at Makedonski Telekom AD, as a senior cyber security

engineer.

 <https://orcid.org/0000-0003-2109-7027>

E-mail: gjorgji.ilievski@telekom.mk

IT Department

Cyber Security Unit

Makedonski Telekom AD Skopje

Kej 13-ti Noemvri 6
Skopje, Macedonia



Pero Latkoski received his M.Sc. and Ph.D. degrees from the Faculty of Electrical Engineering and Information Technologies, Ss Cyril and Methodius University in Skopje, in 2006 and 2010, respectively. Currently, he holds the position of full professor at the same university's Institute of Telecommunications. His research inter-

ests include communication protocol engineering, software defined networking, and information theory.

 <https://orcid.org/0000-0002-8406-4057>

E-mail: pero@feit.ukim.edu.mk

Faculty of Electrical Engineering and Information Technologies

Telecommunications Institute

Ss Cyril & Methodius University

Rugjer Boshkovikj 18

Skopje, Macedonia

A Shared Cybersecurity Awareness Platform

Marek Amanowicz

NASK National Research Institute, Warsaw, Poland

<https://doi.org/10.26636/jtit.2021.154421>

Abstract—Ensuring a good level of cybersecurity of global IT systems requires that specific procedures and cooperation frameworks be adopted for reporting threats and for coordinating the activities undertaken by individual entities. Technical infrastructure enabling safe and reliable online collaboration between all teams responsible for security is an important element of the system as well. With the above taken into consideration, the paper presents a comprehensive distributed solution for continuous monitoring and detection of threats that may affect services that provision is essential to security and broadly understood the state's economic interests. The said solution allows to collect, process and share distributed knowledge on hazard events. The partnership-based model of cooperation between the system's users allows the teams to undertake specific activities at the central level, facilitates global cyber threat awareness, and enhances the process of predicting and assessing cyber risks in order to ensure a near-real-time response. The paper presents an overview of the system's architecture, its main components, features, and threat intelligence tools supporting the safe sharing of information concerning specific events. It also offers a brief overview of the system's deployment and its testing in an operational environment of NASK's Computer Security Incident Response Team (CSIRT) and Security Operation Center (SOC) of essential services operators.

Keywords—*cybersecurity awareness, risk and threat propagation, threat intelligence.*

1. Introduction

An increasing number of ever more sophisticated and complex cyberattacks may be observed. For instance, the Computer Emergency Response Team being a part of the NASK – National Research Institute registered 10,447 such incidents in 2020. The said number was the largest recorded in history and represented the fastest year-on-year increase. Phishing was the most common type of attack accounting for approximately 67.2% of all incidents. The use of malicious software was the second most common type of threats, with its share equaling approximately 7.1%. Such attacks pose a serious threat to information technology (IT) systems supporting services that are of critical importance for the society. They may lead to the disruption in the provision of such services, breaching national security, impacting public

and economic order, violating civil rights and freedoms, as well as endangering human life.

In order to protect IT systems, new functionalities must be implemented to enable early detection of threats, to assess their negative impact and to prevent them. Good level of situational cyberspace awareness needs to be achieved as well in order to collect, in real-time, information on threats and risks identified and on their impact on the behavior of systems, as well as on the related processes and services. Achievement of global cybersecurity of IT systems requires that procedures and cooperation networks be established to facilitate the reporting of incidents and to coordinate the actions undertaken. It is also recommended to create a technical infrastructure allowing a safe and reliable online collaboration of all teams responsible for cybersecurity. Detailed identification of vulnerabilities of information and operational technologies (IT/OT) and of the impact that cyber threat events exert on the related processes and services is important as well.

However, as presented in [1], [2], the achievement of such a goal is difficult due to the considerable level of interdependence of the systems and the fact that they share the same information and communications technology (ICT) resources. The interrelations between individual services are of a diverse and complex nature [3], and yet they need to be determined precisely in order to identify threats in cyberspace and to assess their impact on the level of security. Many works devoted to this issue, such as [4]–[6], confirm the need of modeling the network of interdependent infrastructures in order to identify its critical components and to better understand the scale and scope of potential threats. Such an approach enables early identification of threats and triggers alerts allowing to take preemptive actions in order to mitigate the risk encountered. However, in order to create a network of services that reliably reflects the actual condition of and the interrelations between its components, it is necessary to obtain detailed and verifiable data from service providers. Furthermore, an effective threat response requires close cooperation between IT security analysis and management teams from all interdependent entities.

The procedures of cooperation between all entities responsible for IT security needs to be developed as well, in order to ensure a clear situational awareness picture and the highest level of protection. It is also desirable to establish spe-

cific solutions encouraging cooperation and ensuring that the vital interests of all cooperating parties are protected. These activities should be supported by technical systems enabling efficient acquisition, processing and distribution of information concerning threats and their potential impacts. This paper presents an innovative and scalable solution that allows organizations to collect, process and share threat-related information in order to predict and assess the risk involved, and to share distributed knowledge in order to provide a near-real-time response. The said solution focuses on procedural and technical aspects of service network modeling, as well as on processes related to aggregation of knowledge distributed across multiple databases, assessment of risk, propagation of threats, building a common operational picture and sharing threat-related information. A brief overview of the process of deploying the system and testing it in a real-life environment is given as well. The main contributions of this paper are:

- presentation of a novel collaborative distributed system facilitating online cyber threat awareness;
- presentation of an innovative concept for building a network of interdependent services and for relying on such a network in assessing the threats and the related risks.

The remaining part of the paper is organized as follows. Section 2 gives a brief review of the initiatives aiming to improve the level of IT security. Section 3 describes the system's architecture, its main components, features and properties in terms of flexibility, extensibility and scalability. Selected solutions enabling to assess the impact of potential threats on the provision of services, as well as those enabling to collect, process, and secure information related to hazardous events between all National Platform for Cybersecurity (NPC) users are presented in Section 4. The paper concludes with an overview of the system's deployment and with proposals concerning future work.

2. Related Work

Many international and national initiatives have been undertaken recently to boost the security of IT systems in order to improve reliability and availability of services. The Directive of the European Parliament and of the Council [8] on security of network and information systems (NIS) of 6 July 2016 (hereinafter referred to as the NIS Directive) encourages a number of such initiatives. The NIS Directive imposes, on the Member States, the obligation to implement several legal measures, including by establishing national strategies for the security of networks and information systems, and sets forth requirements and procedures for reporting cybersecurity incidents by service operators and providers. In response to this, the European Telecommunications Standard Institutes (ETSI) [9] and many European

countries established their strategies to implement the requirements of the NIS Directive¹.

For example, the CS-AWARE project [10], launched under the Horizon 2020 program, focuses on creating solutions targeted for local public administration authorities, non-governmental organizations, as well as small- and medium-sized companies. The tools developed enable automatic detection, classification, and visualization of computer incidents in near-real-time, supporting the prevention or mitigation of the effects of such events. The solutions are based on mechanisms for sharing information about actual threats, relying on big data analysis and processing. By leveraging the existing processes of sharing cybersecurity-related information, CS-AWARE enables and improves incident detection and meets the information sharing-related requirements of the NIS Directive.

Another interesting approach to improve an organization's ongoing awareness of the risk posed to its business by cybersecurity attacks has been developed as part of the PROTECTIVE project [11]. The said approach allows to raise the level of situational awareness by enhancing the correlation and prioritization of security alerts, therefore pinpointing the relevance of the organization's assets to its business. Using the context-awareness approach, any organization may identify its key business goals and may define the relationships between such goals, simultaneously determining information and computer assets of critical importance. This data is combined with near-real-time scoring of the assets' vulnerability levels. This helps rank alerts based on their potential damage to the threatened assets and business.

Many new solutions focus on increasing the awareness of cyber threats and on improving the level network and information system security. For example, an interesting approach to the problem of building common cybersecurity awareness by critical infrastructure operators is presented in [12]. By aggregating, analyzing and correlating data obtained from security management systems, a global cybersecurity picture is created allowing also, due to the links between critical infrastructure elements, to anticipate threat propagation-related risks. An inspiring proposal of an IT system for collaborative cyber incident management for the European interconnected critical infrastructure is presented in [7].

The Polish Parliament passed the Act on the National Cybersecurity System (NCS) [13] which specifies the following: organizational framework of the system, tasks and responsibilities of all entities involved, the manner in which supervision and control over the implementation of the Act is exercised, as well as the scope of the cybersecurity strategy. This aims to create a comprehensive solution for boosting protection against threats in Poland and for enabling effective cooperation with other Member States. The Act is building on service operators, digital service

¹more information on status of NIS Directive implementation in EU countries is available at <https://www.digitaleurope.org/resources/nis-implementation-tracker/>

providers and public entities. The NCS is to be managed at the operational level by three Computer Security Incident Response Teams (CSIRT GOV, CSIRT MON, CSIRT NASK), and – at the central level, the Governmental Representative for Cybersecurity and the Board for Cybersecurity. The NCS concept requires its components and the related entities to assume responsibility for several aspects. In particular, essential service operators are required to implement security management tools within their information systems to support the provision of services. They are obliged, inter alia, to perform risk assessment and to manage incidents on an on-going basis, to collect information on cyber threats and vulnerabilities of the information system supporting the provision of a given service, and to report serious incidents, to the appropriate CSIRT, within 24 hours at the latest. CSIRT teams are required to implement a coherent and comprehensive risk management system at the national level, to undertake actions to mitigate cyber threats of cross-sectoral and cross-border character, and to coordinate the handling of the reported incidents. Each CSIRT has a clearly defined scope of responsibilities and a set of entities it supervises. CSIRT tasks include monitoring cyber threats and estimating risks related to the disclosed cyber threats, including by performing dynamic risk assessment at the national level, classifying incidents and coordinating the process of handling such incidents. The CyberSecIdent research program focusing on “Cybersecurity and e-Identity” was launched for the purpose of implementing the NIS Directive. The research project titled “National Platform for Cybersecurity” (NPC) was conducted between 2017 and 2020 within the framework of the program. Its aim was to develop a prototype integrated system used for continuous monitoring, detection of and warning about threats and risks affecting or likely to affect the quality and continuity of services whose deterioration may cause significant damage to the overall security level.

3. NPC System Overview

3.1. System Architecture

The NPC consists of four systems (Fig. 1):

- Edge systems (ES) located within the customers’² infrastructure, serving as the NPC’s portals to the platform resources,
- Operations center system (OCS), i.e. an application system supporting situational cyberspace awareness and constituting a central point for exchanging information on cybersecurity. By default, there is one OCS instance within the platform, but the architecture allows for the existence of more centers that exchange data with each other,
- Management system (MS) which manages both the application and network layers of the NPC,

²Customer is an entity that provides essential and/or digital services and participates in exchanging cybersecurity data over the edge system

- NPC backbone network (BN), i.e. dedicated communication infrastructure that enables secure information exchange between the platform systems (mainly OCS and ES) in wide area networks.

The operations center and the edge systems ensure integration with the user’s systems that, as a rule, are located in the operator’s private networks and may initiate data exchange with the platform systems, such as the malware information sharing platform (MISP), security information and event management (SIEM) system or incident management (IM) system.

The OCS retrieves and aggregates data from external sources assumed to be located in untrusted networks. These include, for instance, network security incident exchange database (n6), national vulnerability database (NVD) or vulnerability database (vulners.com). The OCS initiates communication and retrieves the data, but the source cannot initiate communication with the OCS.

The management system includes a set of tools and services such as:

- managing a public key infrastructure,
- managing the configuration of systems and applications,
- managing the configuration of network devices of the NPC backbone network,
- monitoring the security status of the platform and all system components,
- maintaining a replica of the system directory.

The key functions of the platform are performed by application systems (i.e. OCS and ES). The system architecture developed is universal (Fig. 2) and enables to implement specific OCS and ES solutions.

The application system architecture includes:

- front-end load-balancing layer for HTTP/HTTPS protocols, designed to provide the users and local systems with a simplified interface for highly available applications,
- application layer implemented by a highly available cluster containing domain-specific, interconnected applications, called microservices, responsible for the system’s business logic. The following applications within this cluster operate in special security zones:
 - application gateway ensuring the secure sharing of resources between the NPC systems,
 - data importer for data acquisition from external sources deployed only in the OCS,
- back-end load balancing layer for various protocols ensuring unified interfaces with the services provided for the applications,

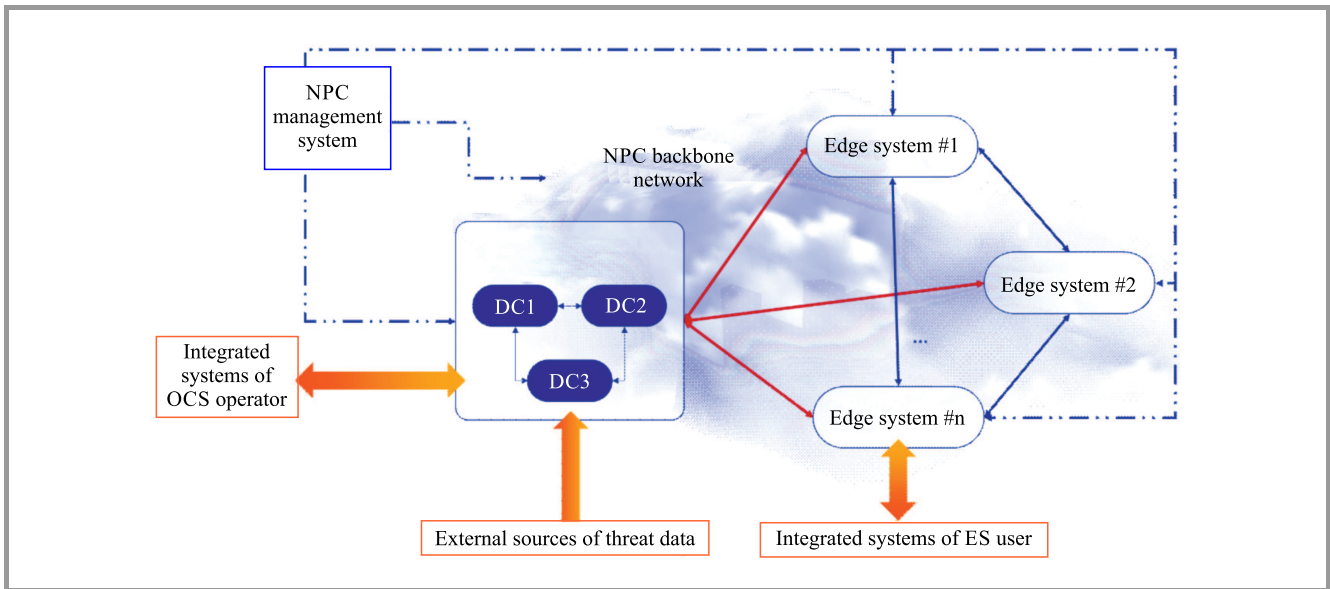


Fig. 1. NPC architecture.

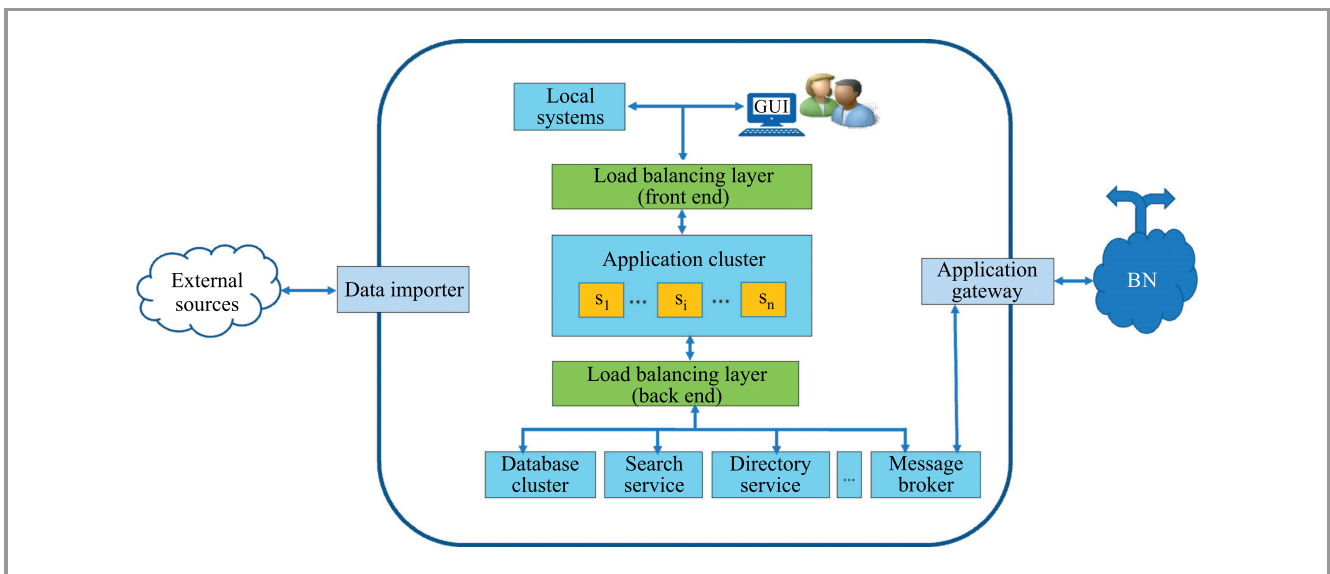


Fig. 2. Application system architecture.

- database back-end layer containing the services provided for the application, such as, for instance:
 - database cluster that stores operational data,
 - directory service, i.e. a database of the NPC users and their permissions,
 - search service enabling full text searches covering the operational data,
 - message broker providing asynchronous message exchange between applications.

The application system may run in a stand-alone or high-availability configuration, and its components, services, and service layers are developed in a high availability configuration, i.e. are distributed over many physical resources. The ES may exchange data without the OCS being present.

In the case of a complete or partial network outage, all application systems are capable of operating properly. Data that have not been sent due to network failure or unavailability of the system are stored locally until the problem is resolved. The NPC applications were deployed on a self-hosted Kubernetes platform that provides scalability and a high level of availability.

3.2. Information Processes

Operation of the NPC system relies on a partner-like collaboration between its users, meaning that service providers are free to decide whether to join the platform and comply with mutually accepted cooperation principles, especially those pertaining to the protection of the data shared.

The security-specific data are exchanged between CSIRT and the related service provider. There is also a possibility of sharing some data between a group of users in compliance with the applicable security requirements. Such an exchange may occur, for instance, when OCS is temporarily unavailable or if the user wants to deliver urgent data to a group of users, such as those belonging to the same business group or sector. It should be stressed that all data exchanged directly between the users have to be submitted to the OCS as well.

The exchange of information is carried out within the functional processes, i.e. when surveying the service providers, handling incident reports, assessing the risk, exchanging information on security events, providing warnings about threats and risks, sharing information on vulnerabilities, and issuing recommendations.

The service providers (ES users) are obliged to provide the following types of data to the OCS:

- detailed information on the services rendered, on the conditions for providing such services and on the potential consequences related to disruptions of their continuity or quality,
- notification of incidents that would exert a significant disruptive effect on the provision of services, including detailed incident descriptions and their potential consequences (i.e. impact on the services rendered),
- outcomes of risk assessment processes associated with the services rendered.

In addition, they may provide the OCS with reports on newly discovered vulnerabilities and the indicators of compromise (IoC), the results of their analysis, information on suspicious data, raw data requiring for detailed studies, and information about the technologies used.

The OCS processes and analyses data collected from external threat sources, as well as those submitted by ES users and shares its own information resources in order to ensure a quick and effective response to existing or potential threats. The OCS performs a significant role in the providing crucial processes that include:

- providing information on the present cybersecurity status of the services, at local and national level,
- managing incident reporting,
- gathering vulnerability data from external sources and sharing the integrated vulnerability database with NPC users,
- modeling the interdependencies between services,
- predicting threats and risks propagation and their impact on cyberspace security,
- analyzing security risks at the national level,
- sharing knowledge supporting technical analysis of threats,

- distributing security warnings,
- providing NPC users with recommendations regarding the desired actions to increase the protection of their information infrastructure.

Data are transferred from the ES in unicast mode, while the OCS may transmit data in unicast or broadcast/multicast mode.

3.3. System Features

The NPC system is based on a universal architecture that relies on the NPC technology stack. The application architecture is based on microservices, ensuring a high degree of system flexibility and allowing the implementation of selected components. It also reduces the need to modify specific services or applications, keeping the system-related costs low. Consequently, the applications used at one place may be easily modified and used in other parts of the system.

The NPC is a scalable and distributed system that may be deployed on a large scale or scaled down to a single rack unit or even less. It is also possible to distribute the system components and functions between multiple physical locations.

The solution ensures low deployment and maintenance costs due to the fact that the ES and the applications may be developed and installed without maintenance downtime and, what's more, implementations are automatically executed in a way imperceptible to the user. It is worth noticing that all applications are managed within the Kubernetes cluster.

Good interoperability of the NPC system is achieved thanks to the availability of all relevant data through:

- documented REST API,
- custom integration with security platforms, such as TAXII, MISP and SIEM, with a potential extension to other platforms as well,
- dedicated API for creating new applications, adding new vulnerability data sources or threat data integration.

The management system enables automated deployment of applications throughout the platform, which is particularly important when adding new entities or upgrading the applications. The system ensures:

- complete control of the entire software supply chain,
- backups and quick data restore for ESs,
- consistency of timescales throughout the NPC,
- monitoring the entire NPC and all its components,
- central analytics of logs from all NPC systems and devices.

It should also be noticed that management functions may be split between the NPC application and the backbone network, for instance, if BN management needs to be performed by a separate entity.

4. Selected Solutions

4.1. Network of Interdependent Services

An expert subsystem supporting decision-making processes and ensuring the safe provision of services by NPC users is an essential part of the platform. It supports the identification of interdependencies between NPC users, their services and the ICT infrastructure used. It also allows to determine the potential impact of incidents (scale, geographic reach, duration), and to obtain the input data required to assess their significance (spread of threats and assessment of their outcomes).

The decision support subsystem is made up of four components, as shown in Fig. 3. All service providers are surveyed before they start the operational use of the system in order to collect the required input data.

Ensuring the consistency of data obtained from the surveys allows to create a network of interdependencies services. The attributes of the network components reflect the criticality (impact on other services) of the individual services and the relationship between them [14]. The process of managing the network of interdependent services allows to conduct several operations, including network upgrades and reconfigurations, depending on the needs of the system analyst.

In order to ensure coherent and reliable security awareness at the national level, a uniform approach to assessing cyber threats by all NPC users is required. A concept of evaluating the risk of unfavorable events by relying on the Markov chain model to calculate an indicator concerning the availability of interdependent services is presented in [15]. Malinowski and Karbowski in [16] adopt a hierarchical approach to risk assessment at the national level, considering cyber threats and vulnerabilities identified by service providers at a local level. The NPC system uses its proprietary risk assessment methodology covering both

the dynamic risk analysis procedure carried out by service providers (the so-called “own risk”) and the static and dynamic risk analysis procedures performed by the OCS [17]. It was assumed that an own risk results from the possibility of violating confidentiality, integrity and availability of the service by using the vulnerabilities of the ICT infrastructure (hardware and software) used to provide it identified by the service provider. The results of the analysis carried out by NPC users are reported to OCS.

The risk assessment performed at the OCS is based on mapping service interdependencies and takes into account threats resulting, inter alia, from the following:

- vulnerabilities identified by service providers in their ICT infrastructure,
- criticality of the services and their interdependencies,
- the extent to which the NPC customer organization ensures the safe rendering of the services,
- reported incidents, IoCs and other security events,
- information on security issues, obtained from various sources, concerning the ICT infrastructure supporting the services reported by NPC users.

Results of the risk assessment procedure are visualized using a network of interdependent services presented in Fig. 4. The node colors correspond to the risk values assigned to the specific services. The width of the lines indicates the strength of specific impacts. More information about a given service and the related risks may be obtained by clicking on the selected node. The panel on the right-hand side of Fig. 4 shows the details of the service chosen (with a blue border), including the risk value and its trend.

By linking the results obtained by OCS, a global cybersecurity awareness picture is created based on a configurable panel with data about the current and predicted state of service in cyberspace. The example presented in Fig. 5 shows the current status of service-related risks, statistical data concerning incidents reported and vulnerabilities identified, the most exposed services, service threshold risk

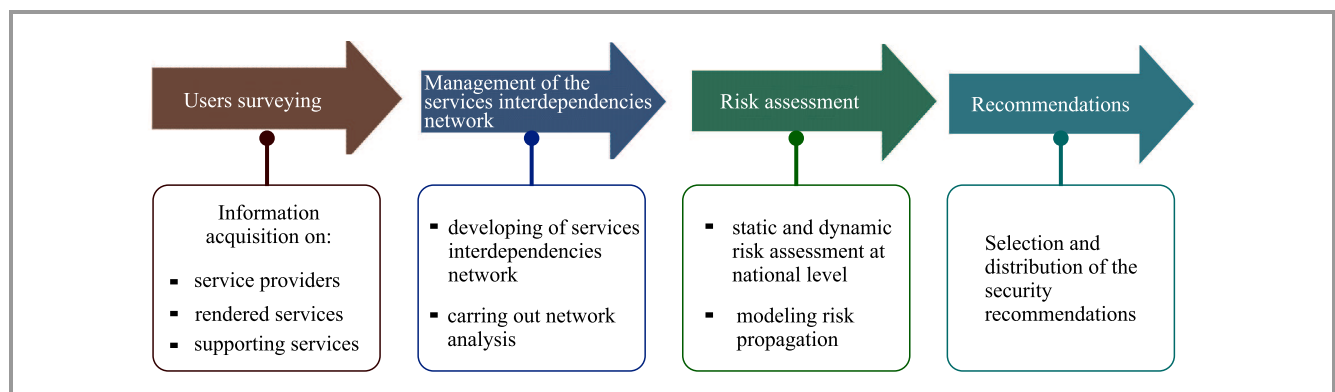


Fig. 3. Components of the decision support subsystem.

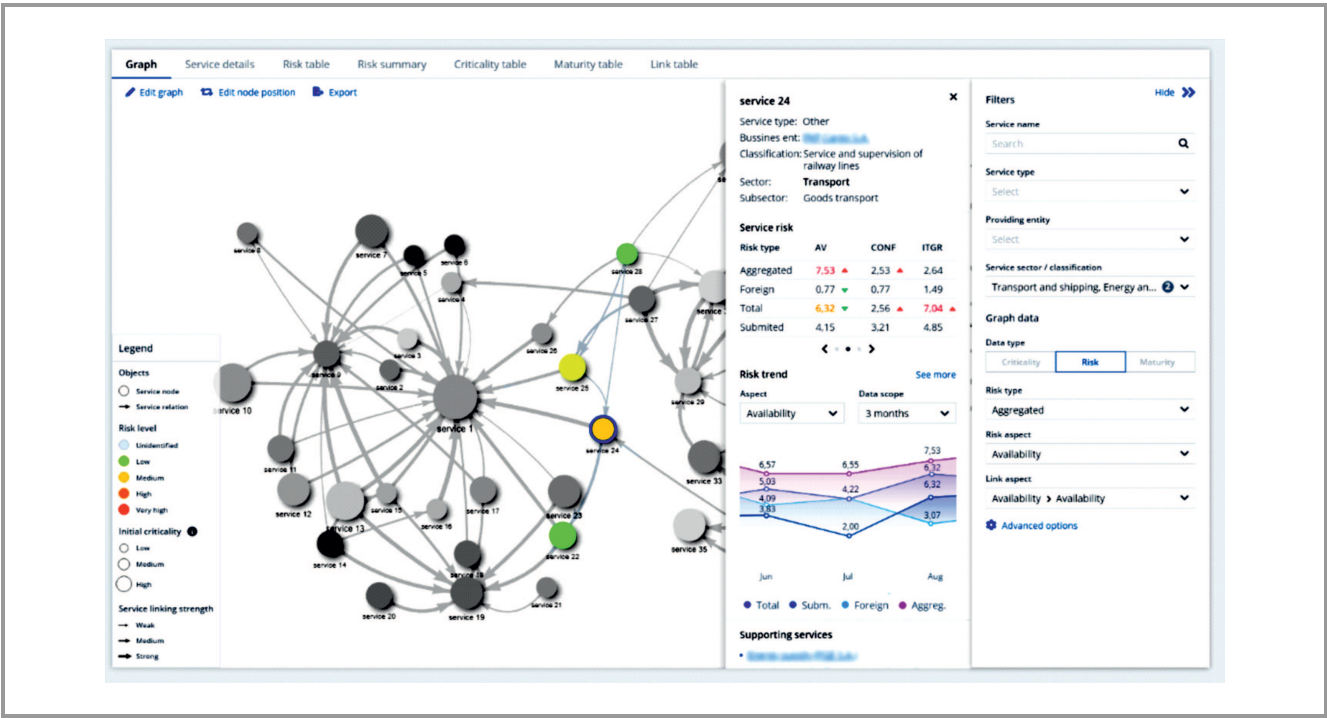


Fig. 4. Results of the risk assessment procedure carried out by OCS in an exemplary network of interdependent services. (see the digital version for color images)

values, and sectoral risks. The supervising analyst is capable of customizing the layout and the content to meet the current needs.

The situational awareness data are shared with NPC users – to the extent and degree of detail resulting from their role, the enabling them to respond quickly and select appropriate measures to eliminate or limit any potential consequences.

The analysis, provided by OCS, contains input data supplied to the rule-based engine that is tasked with selecting appropriate recommendations in order to ensure a high level of security of the services rendered. A dedicated tool is used for the distribution of the recommendations to service providers.

4.2. Threat Intelligence Mechanisms

A set of threat intelligence tools is used to efficiently exchange information on cyber events that enable a coordinated response to the threats that have been identified.

The malware information sharing protocol (MISP) is relied upon to exchange information about network security events and indicators of compromise (IoC). Application services, installed in central and edge systems, perform tasks related to MISP integration, synchronization of the databases, and data distribution within the system. For users who do not have their own MISP instances, a dedicated tool was developed to make this data available. The NPC ensures also integration with the n6 platform designed to collect, process and share information about network events and potential security incident (IoCs). The n6 was created by CERT (Poland) and contains information about sources of the attack, i.e. URL, domain, IP, and name of malicious software as well as other unique information if available.

The application service implemented in the OCS collects and aggregates data on IT/OT systems' vulnerabilities from external public sources and converts these into the format required by the NPC. The aggregated data and source vulnerabilities (i.e. before aggregation) and several related in-

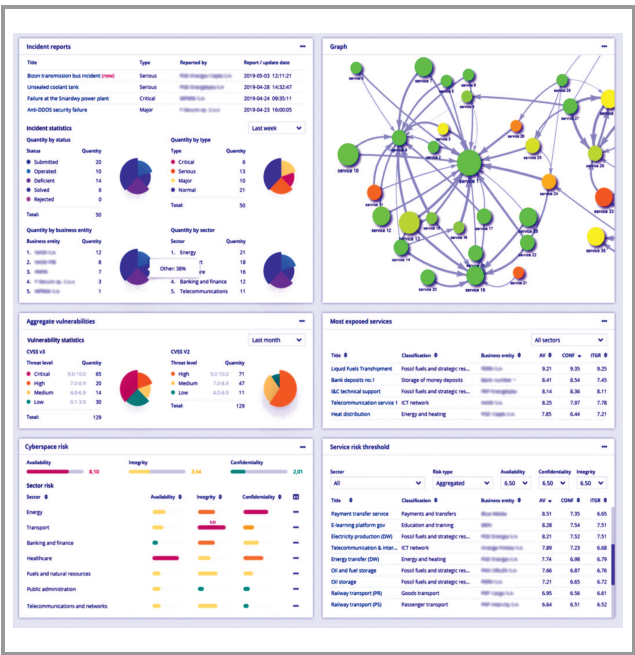


Fig. 5. Example of a situational picture.

formation resources, e.g. technical bulletins, risky products, vulnerability relationships with other objects, are available to all platform users. The tools used for database management ensure the OCS analyst is able to constantly update its contents and allow all users to search the database quickly according to selected criteria. The database enables service providers to identify potential threats and quickly implement the mitigation solutions required.

The NPC users are able to share their knowledge and experiences related to the specific incidents and other security events by using a dedicated application. They can exchange observations and conclusions from their own technical analysis. Such an analysis may also be performed by OCS and ES analysts and may be made available in accordance with the applicable distribution rules. The ES is capable of delivering sensitive data, e.g. malware code, to the OCS. The warning feature is activated when predefined security events occur, enabling the OCS analyst to send a warning message. The operator may also support the recipient's actions taken by adding attachments to the message. Moreover, all threat intelligence tools used have a built-in chat mechanism that supports online communication.

4.3. Security Measures

The NPC incorporates a set of built-in security features for secure sharing of sensitive information and protecting the vital interest of the NPC users, including:

- encrypted end-to-end communications,
- marking sensitive data and configurable anonymization,
- auditing of user actions and extensive logging, assuring non-repudiation and accountability of exchanged data.

The data shared within NPC users are encrypted at the network and application layers of the OSI model. The standard IPsec protocol is used for securing VPN connections between the NPC entities. In addition, the elliptic curves cryptography is used at the application layer for data transferred between the NPC system's components. The system of X.509 certificates is applied for authentication of the system users and signing the shared data, which ensures its credibility.

The NPC security policy assumes that an ES user is not capable of obtaining the names and physical addresses of the other users. The configuration data of the backbone network and a list of NPC users are available only to the management system. Only the identity of the OCS is known, by default, to all NPC users. All data sent from the ES are forwarded to the OCS. Data targeted for other edge systems are addressed using a symbolic recipient name. The complete list of symbolic names is known to the management system only. Unavailability of the data sender relies on changing the value of the selected fields in its header to the constant value anonymous. Sender anonymization is not performed

when messages are exchanged with the OCS. The recipient concealment procedure is used also for "anonymous" data receipt acknowledgement. Full confirmation is made by the OCS only.

The system incorporates a security feature that allows the message sender to hide sensitive data. This type of data is marked by the user, meaning the anonymization feature replaces the selected fragment with a "xxx" of the same length as the original text before sending the message. Anonymization is not performed for messages sent to the OCS. The application system guarantees that the tagged fragment will not be retransferred to the platform users.

Additionally, an audit service is performed to ensure accountability and non-repudiation of user actions and system functionalities. The system acquires and stores information about all events, i.e.:

- time stamp,
- user login data,
- address of the host on which the action was performed,
- name of the acted module,
- action type (e.g. create, update, send),
- subject to which the action relates (e.g. incident, vulnerability),
- optional additional data.

An API for the web application is used for analyzing the collected data, enabling the search function of users' and system actions with the activities filtering, sorting, and correlation finding.

5. System Deployment

The prototype of the NPC was developed in an operational environment of CSIRT NASK with the participation of four service providers from different sectors of the market (financial, transport, energy) and an entity providing cybersecurity services. Three spatially distributed data centers of CSIRT NASK were connected, via the backbone network, with edge systems located within the service providers' IT infrastructure (Fig. 6). The OCS was deployed in a configuration characterized by a high degree of availability, known as dual modular redundancy, where data center number 3 acts as an arbitrator.

A full range of tests was performed to verify the functionality of the system and the results obtained confirmed the system's usability. That enables CSIRT analysts and a number of service providers to perform a trial using a prototype of the system. The scenarios verified included user activities related to the development of a network of services,

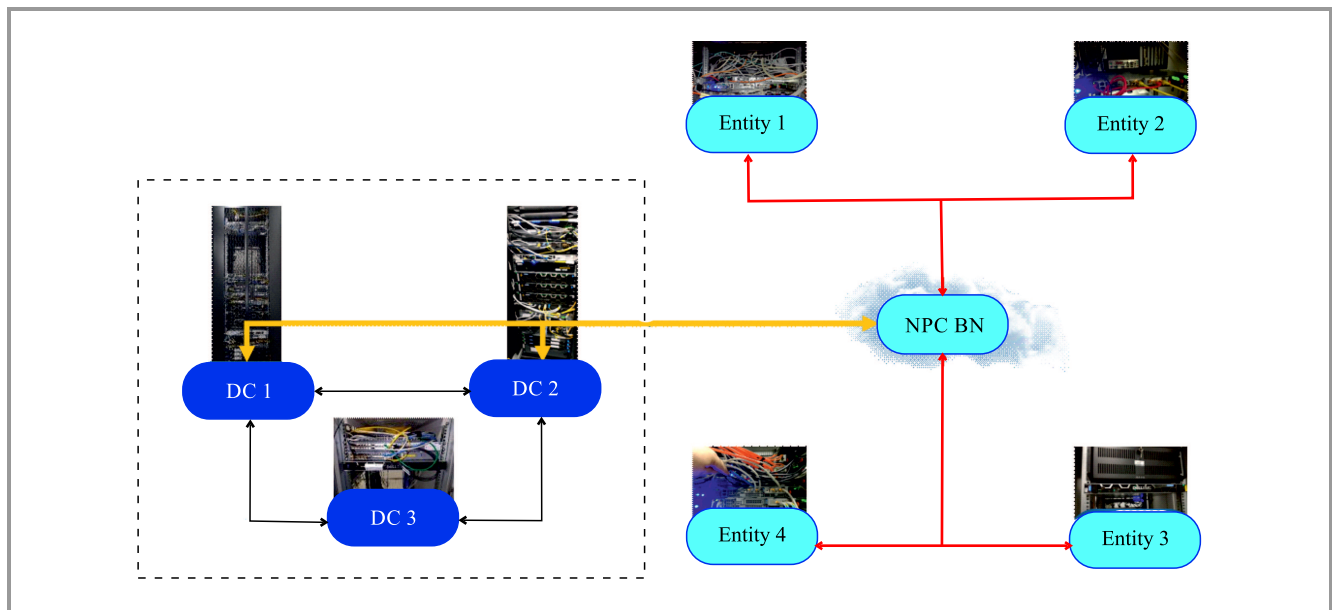


Fig. 6. NPC deployment in an operational environment.

reporting and handling incidents, sharing and using an aggregated vulnerability database, sharing knowledge, using threat intelligence features, assessing the risk and assessing cyberspace security.

All this allowed the users to better understand the functionalities of the NPC system and its operational value. The test results confirmed suitability of the prototype that may serve as a technological foundation for a full-scale implementation of a solution that meets all applicable legal requirements. The system architecture was expanded to incorporate three (instead of one) OCSs and to make the system accessible for all NCS entities.

The lessons learned from the deployment of NPC confirm that the actual level of cyber threat awareness depends on all parties involved in detecting and reacting to cyber threats originating or maliciously installed in their technical infrastructure, as well as on their readiness to share cyber threat-related information. The NPC system presented offers effective features ensuring a high level of trust of the service providers in mutual and/or external relations. It delivers tools for improving the user collaboration, supports secure threat data sharing and allows to develop a shared cybersecurity picture. All these features lead to increasing the level of cyberspace awareness and help react to the actual or potential cyber threats in a more coordinated manner.

Future work needs to be focused on implementing the recommendations formulated based on prototype tests and should lead to developing an operational NPC version for the Polish Cyber Security System. The features of the presented solution rely on the universality and flexibility of the system architecture, support quick and effective implementation of the NPC for use cases (other than NCS) requiring safe sharing of information about threats, creating shared situational awareness and coordinated responses. The pre-

sented system may act as an ICT infrastructure for the Security Incidents Response Teams (SIRTs) or Information Sharing and Analysis Centers (ISACs). In particular, it can be adapted for the safe sharing of information and building a global situational awareness picture for security management in complex and dispersed structure organizations.

6. Acknowledgements

This work was performed under the CYBERSECIDENT/369195/I/NCBR/2017 project supported by the National Centre for Research and Development (CyberSecIdent Program).

The author would like to thank a large group of his highly committed associates from NASK, National Centre for Nuclear Research, National Institute of Telecommunications and Warsaw University of Technology, contributing their extensive expertise to the process of designing and deploying the system in question.

References

- [1] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies", *IEEE Control Syst.*, vol. 21, no. 6, pp. 11–25, 2001 (DOI: 10.1109/37.969131).
- [2] R. Zimmerman, "Decision-making and the vulnerability of interdependent critical infrastructure", in *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, The Hague, Netherlands, vol. 5, 2004, pp. 4059–4063 (DOI: 10.1109/ICSMC.2004.1401166).
- [3] F. Petit and L. P. Lewis, "Incorporating logical dependencies and interdependencies into infrastructure analyses", *George Mason University*, 2016 [Online]. Available: <https://cip.gmu.edu/2016/02/17/incorporating-logical-dependencies-and-interdependencies-into-infrastructure-analyses/>

- [4] A. Nieuwenhuijs, E. Luijff, and M. Klaver, "Modeling dependencies in critical infrastructures", in *Proc. IFIP Int. Federation for Informat. Process.*, 2008, pp. 205–213 (DOI: 10.1007/978-0-387-88523-0_15).
- [5] R. Setola, V. Rosato, E. Kyriakides, and E. Rome, "Managing the complexity of critical infrastructures", vol. 90, *Springer Int. Publishing*, 2016 (DOI: 10.1007/978-3-319-51043-9).
- [6] C.-H. Han, S.-T. Park, and S.-J. Lee, "The enhanced security control model for critical infrastructures with the blocking prioritization process to cyber threats in power system", *Int. J. Crit. Infrastruct. Prot.*, vol. 26, 2019, (DOI: 10.1016/j.ijcip.2019.100312).
- [7] G. Settanni *et al.*, "A collaborative cyber incident management system for European interconnected critical infrastructures", *J. Inf. Secur. Appl.*, vol. 34, pp. 166–182, 2017 (DOI: 10.1016/j.jisa.2016.05.005).
- [8] "Directive (EU) 2016/1148 of the European Parliament and of the Council concerning measures for a high common level of security of network and information systems across the Union" [Online]. Available: <http://data.europa.eu/eli/dir/2016/1148/oj>
- [9] ETSI TR 103 456 v1.1.1, "Implementation of the Network and Information Security (NIS) Directive", 2017 [Online]. Available: https://www.etsi.org/deliver/etsi_tr/103400_103499/103456/01.01.01_60/tr_103456v010101p.pdf
- [10] CS-AWARE Project, *Horizon 2020 Programme* [Online]. Available: <https://cs-aware.eu>
- [11] PROTECTIVE Project, *Horizon 2020 Programme* [Online]. Available: <https://protective-h2020.eu>
- [12] S. Puuska *et al.*, "Nationwide critical infrastructure monitoring using a common operating picture framework", *Int. J. Crit. Infrastruct. Prot.*, vol. 20, pp. 28–47, 2018 (DOI: 10.1016/j.ijcip.2017.11.005).
- [13] "Act on the National Cybersecurity System", *J. of Laws*, item 1560, 2018, [Online]. Available: <https://uodo.gov.pl/en/file/307>
- [14] M. Kamola *et al.*, "Decision support system for identification and security management of essential and digital services", in *Proc. Int. Conf. on Military Commun. and Informat. Systems (ICMCIS)*, Budva, Montenegro, 2019, pp. 1–7 (DOI: 10.1109/ICMCIS.2019.8842769).
- [15] A. Karbowski *et al.*, "Critical infrastructure risk assessment using Markov chain model", *J. Telecommun. Inf. Technol.*, vol. 2, pp. 15–20, 2019 (DOI: 10.26636/jtit.2019.130819).
- [16] K. Malinowski, A. Karbowski, "Hierarchical online risk assessment at national level", in *Proc. Int. Conf. on Military Commun. and Informat. Systems (ICMCIS)*, Budva, Montenegro, 2019, pp. 1–5 (DOI: 10.1109/ICMCIS.2019.8842731).
- [17] M. Janiszewski *et al.*, "A novel approach to national-level cyber risk assessment based on vulnerability management and threat intelligence", *J. Telecommun. Inf. Technol.*, vol. 2, pp. 5–14, 2019 (DOI: 10.26636/jtit.2019.130919).



Marek Amanowicz graduated from the Military University of Technology, Warsaw, Poland, where he held several positions, including that of a faculty dean and Vice Rector for R&D. He was a Deputy Chairman of the Polish National Committee of the International Union of Radio Science. He served as the head national representative to

the Information Systems Technology Panel of the NATO Scientific and Technology Organization. He worked at TAC ONE, an international company in Paris, as a systems V&V manager. In 2017, he joined the NASK – National Research Institute, assuming the position of professor. He is an elected member of the Electronics and Telecommunications Committee of the Polish Academy of Sciences. He has led many national and international research projects focusing on systems engineering, mobile communications, modeling and simulation. He is the author or co-author of more than 200 papers published in scientific journals or at national and international scientific conferences. His current research interests focus on communication systems engineering and information security of complex technical systems.

 <https://orcid.org/0000-0002-9132-5788>

E-mail: marek.amanowicz@nask.pl
NASK – National Research Institute
ul. Kolska 12
Warsaw, Poland

Markov Decision Process based Model for Performance Analysis an Intrusion Detection System in IoT Networks

Gauri Kalnoor and Gowrishankar S

BMS College of Engineering, Bangalore, India

<https://doi.org/10.26636/jit.2021.151221>

Abstract—In this paper, a new reinforcement learning intrusion detection system is developed for IoT networks incorporated with WSNs. A research is carried out and the proposed model RL-IDS plot is shown, where the detection rate is improved. The outcome shows a decrease in false alarm rates and is compared with the current methodologies. Computational analysis is performed, and then the results are compared with the current methodologies, i.e. distributed denial of service (DDoS) attack. The performance of the network is estimated based on security and other metrics.

Keywords—DDoS, intrusion detection, IoT, machine learning, Markov decision process (MDP), Q-learning, NSL-KDD, reinforcement-learning.

1. Introduction

The technology of the Internet of Things (IoT) is relatively new, it connects the Internet to the low hardware resources devices and then susceptible to the various malicious attack, i.e. denial of service (DOS) [1], [2]. The network-based IoT is considered to be one the fastest evolving areas, having 50 billion gadgets connected among them [3], and then vulnerable to security abuse. For example, Mirai is one of the unusual types of a botnet which triggers a large-scale attack like distributed denial-of-service (DDoS) and thus strikes by mistreating some of the IoT devices [4], and even infects the CCTV IP cameras [5].

The safety of IoT is constantly improved [6]. Many frameworks and methods are developed to mitigate most network attacks. The logs with recorded abuse historical data are observed, based on methods using machine learning which can reach a large network – up to millions in a day.

The intrusion detection system (IDS) is an essential component in the security of the network to protect the target network which comprises of irregular actions and threats during interruption of network traffic. Thus, there is a separation of normal activity and anomalous activity in the network. A comprehensive IDS group can be obtained in two classes. Misuse-based IDS is the interrupt that notices

the known strategies. The limit of the primary technique to anticipate new and obscure assaults is restricted. The signature-based IDS is dependent on the irregularity identification and works by making a profile of ordinary conduct of the network, then later recognizing it as any anomalous conduct [3].

In the proposed work, an artificial intelligence (AI) based algorithm has been proposed for developing an IDS for detection of malicious attacks and also monitors the data streams generated from IoT and WSNs [6].

It is an enhanced method of Markov decision process with Q-Network algorithm which gives an optimal best solution in terms of performance of IoT networks. Thus, it is an important and challenging issue to be considered, and decision modeling is applied to obtain the optimal solution. The main contributions of this article are summarized below:

- the RL-based IDS is proposed by exploiting the extended Markov decision process (MDP) algorithm,
- the RL calculation is consolidated on IDS (RL-IDS) with the end goal that the survey for cases like a basic foundation is obtained by unique digital-based hazards for IoT and WSN continuously,
- a Q-network is applied with the end goal that the assessment of Q-work is recognized by conveying IDS into RL. A few tests are performed for the assessment of the execution of the proposed model in the environment considered.

The remaining sections of this paper are presented as follows. Section 2 describes the related work. Section 3 introduces the method for security and reinforcement learning. In Section 4, the system model is formulated and RL-IDS methodology is described. In Section 5, performance is investigated and results are presented. In Section 5, the evaluation carried out for the proposed RL-based IDS scheme is explained and then compared in Section 6 with supervised machine learning schemes. Lastly, the work concludes with the experiments and analysis in Section 7.

2. Related Work

In recent works, many authors have applied standard techniques of machine learning (ML), such as principal component analysis (PCA) and linear discriminant analysis (LDA), as these classification-based algorithms can detect normal records with high precision and identify the abnormal records such that the performance of an IDS can be managed [7]–[11]. In [12], the authors have proposed deep feature embedding to reduce the size or magnitude of data from the network based on IoT in a real-time application by considering the “edge of deep learning”. Likewise, in [13] the preprepared worldview is applied such that the identification and quickness are helped with traditional ML-based calculations.

In [14], the authors have observed that the IoT technology makes possible to connect different smart objects, through the Internet. The authors have formulated a novel QoS management schemes based on power control algorithm. The unexplored R-learning algorithm is used as a docitive paradigm by the authors where the system agents teach other agents to adjust the power levels, thus reducing the complexity in computation and increasing speed in the learning process.

In [15] the optimization has been incorporated into an MDP which can minimize the evaluation metric as long-term average delay. The continuity of state and action space due to the high dimensionality is considered by the author where deep reinforcement learning based dynamic resource management (DDRM) algorithm is proposed. This enables the joint optimization with computing resource and transmission power. The authors have compared the simulated results with conventional URM, RRM and A3C algorithms mainly which reduces the delay in task effectively.

Also, taking as an illustration of the idea-based IDS, Q-learning of reinforcement learning (RL) has been investigated thoroughly by examining and protecting the sensor network that utilizes the dynamic methodology and ideal activities based on the arrangement of states in the respective IoT environment [16]. There are numerous papers on scientific classification, position, and the ML current advancements in data security, i.e. [17], [18]. Structured [19] ML techniques have been applied to location interruption for network information. The exemplary ML models applied to IDS were: support vector machine (SVM), multi-layer perceptron (MLP), k-nearest neighbours (KNN), decision trees (DT), naive Bayes (NB), and random forest.

3. Security in IoT

To meet the ideal security necessities, a complete perspective on network security is required. The accompanying key security properties ought to be viewed when building up a convincing IoT security methodology.

- **confidentiality** – it is a crucial security standard for IoT structures. IoT devices can store and move sensi-

tive information that shouldn't be wrongly found by individuals [21],

- **authentication** – the verification of both communication parties must be completed before performing other procedures,
- **integrity** – the IoT applications need the legitimate constituents to be uniquely altered where the information is moved through the remote correspondence,
- **availability** – the authorized users should be consistently able to access the IoT network,
- **authorization** – this includes granting privileges to clients for an IoT structure [22],

3.1. Reinforcement Learning-based IDS

Beginning by characterizing the idea of RL, and other augmentation of ML dependent on Markov decision process (MDP), first a reward function R is defined providing state s to IDS. It is characterized with five IDS concepts as below.

System state space. The arrangement of states gained by the IDS is $S = s_0$ – ordinary, s_1 – identification, s_2 – no detection, where s_0 demonstrates the typical traffic record in the WSN record, s_1 implies the location of IDS assaults on traffic, and s_2 demonstrates that IDS can't recognize assaults.

Action space. A set of possible actions that the IDS can perform, can be expressed by:

$$A = \{a_0, a_1, a_2, a_3, \dots, a_m\}, \quad (1)$$

where a_k indicates the type of IDS reaction in the k -th attack class and $k = 0, 1, 2, \dots, m, p$, for example, according to Table 1. The shares are sorted according to their risk level: $a_0 < a_1 < a_2 < a_3 < \dots < a_m$.

Table 1
Known attacks and their risk level

Risk	Attack instances
Low	Gues-passwd, Warezcilent, FTP-write
Medium	Satan, Portsweep, Nmap
High	DNS-poisoning, Cross-site-scripting (XSS), ARP-spoofing
Critical	ICMP flood, Land, Smurf, Ping of death, Apache 2

Reward function. The rewarded function is negative when the IDS makes the best move to secure the framework regardless of whether the scheme against the activity is too costly, and positive when the IDS chooses the right activity.

The estimation of the reward is:

$$r_t(s_t, a_t) = \begin{cases} R_p & \text{for } s_t = 0 \text{ and } a_t = a_0 \\ 1 - \mu_j(a_t)R_p & \text{for } s_t = s_0 \text{ and } a_t \in \{a_1, \dots, a_m\} \\ R_p & \text{for } s_t = s_1 \text{ and } a_t = a_k \\ 1 - \lambda_j(a_t)R_p & \text{for } s_t = s_1 \text{ and } a_t \in \{a_0, \dots, a_{k-1}\} \\ R_n & \text{for } s_t = s_1 \text{ and } a_t \in \{a_{k+1}, \dots, a_m\} \\ R_p & \text{for } s_t = s_2 \text{ and } a_t = a_m \\ 1 - \theta_j(a_t)R_p & \text{for } s_t = s_2 \text{ and } a_t \neq a_0 \end{cases}, \quad (2)$$

where $0 < \mu_j(a_j) < 1$, $0 < \lambda_j(a_t) < 1$ and $0 < \theta_j(a_t) < 1$. The r_t refer to the reward s_t is the state of the sensor node, a_t is the action of the sensor at t time.

The reward in each time t is:

$$r_t(S_t = s, a_t = a) = \sum_{s' \in S} P\left(\frac{s}{s'}, a\right) r_t(s', a) \quad (3)$$

State transition probability. The transition probability matrix at time t for $a \in A$ is:

$$\mathbf{P}_a = \begin{bmatrix} \beta_{1,1}^a & \beta_{1,2}^a & \beta_{1,3}^a \\ \beta_{2,1}^a & \beta_{2,2}^a & \beta_{2,3}^a \\ \beta_{3,1}^a & \beta_{3,2}^a & \beta_{3,3}^a \end{bmatrix}. \quad (4)$$

Given by β^a :

$$i, j = p\left(s_t + \frac{1}{s_t}\right) = p\left(\frac{s_t}{s_j}, a\right) \text{ for } i, j = 1, 2, 3.$$

$$\sum_{j=1}^3 \beta_{i,j}^a = 1, i = 1, 2, 3 \text{ and } a \in A. \quad (5)$$

Discount factor. $0 < \gamma < 1$. The IDS arbitrarily choose a_t , and the environment samples the reward $r_t(s_t, a_t)$ according to the state of arrival s . The agent then receives an incentive in the following state s_{t+1} . Besides, π is a specific policy from s_t to s_{t+1} specifying a_t retrieved in each state s_t . Then, the strategy is updated to generate sample paths (s_0, a_0, r_0) , (s_1, a_1, r_1) , $(s_2, a_2, r_2) \dots$. Let us define $\pi = (\pi_1, \pi_2, \dots)$ as the best policy vector. The goal of the data stream is to get π_t , which represents the best pattern based on system status. Therefore, the expected maximum sum of IDS rewards at t , is given by:

$$\pi^* = \arg \max_{a \in A} [r_t(s_t, a_t) + \sum_{s' \in S} P_t(s'|s, a) V_{t-1-t}(s')] . \quad (6)$$

The optimal value function V_{i+1} defines the IDS which can be chosen as the best state. It can be found out from each phase:

$$V_{i+1}(s) = \arg \max_{a \in A} [r_{t-1-t}(s_t, a_t) + \sum_{s' \in S} P_{t-1-t}(s'|s, a) V_i(s')] . \quad (7)$$

Next, the timestamp size is determined, using the concept of Q-learning. In every state, the best action a is chosen and the algorithm Q-learning applied, so that the updates can be performed. The optimal policy π^* is calculated according to the best action. If there are no optimal actions found, then the learning samples $0 < \alpha < 1$ are applied.

$$Q(s_t - a_t) = Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a' \in A} Q(s_{t+1}, a') - Q(s_t, a_t)] . \quad (8)$$

The pair (s, a) is updated to determine the step having the best reward. In each iteration, the prediction of IDS has state value function V_{i+1} and then a Q-table is constructed by using Q-learning, where the lines signify the columns and states s representing the actions a . In each state s_t , the reward r_t is observed corresponding to an action a_t realized by the agent. The action at the next state (s_{t+1}) is also observed in [21], and the approximate value of Q is updated to satisfy the Bellman equation:

$$Q(s_{t+1} - a_{t+1}) = (1 - \alpha)Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a' \in A} Q(s', a')] . \quad (9)$$

4. Proposed Model

The random forests (RF) algorithm is used to classify a large amount of data. Several algorithms like decision trees and merging trees are used during classification to train the sample data available. The final output during classification chooses the most selected class [7].

In this section, the details of the deployment of the Q-learning network-based model are provided aiming to monitor and predict the cyber-attacks in critical infrastructures of sensed big data streams. The discussion is encompassed in the following aspects:

- the attack risks and their different degree,
- the pre-processing details engaged to clean data and filter,
- the strategy of the interaction of IDS model by the agent to secure the attacks,
- the Q-function estimation and its results by considering the best decision.

The architecture of the proposed system is shown in Fig. 1, which presents the sensor data of WSN and the RL-IDS mechanism requested to make a decision.

At pre-processing stage, the network traffic is registered for every type of attack and then invalid and redundant records are removed. Next, the transformation of the record is done based on the type of attack [9]. At the first step, data aggregation obtained by the sensor [20] is performed so that the data volume is reduced.

Next, the Q-network (QN) is applied by using the Q-function for estimation of best action to the attack. It improves the prediction and the estimation of action values effectively among the state's set by applying the non-linear function: $Q(s_t, a_t; \theta) \approx Q(s_{t+1}, a_{t+1})$.

The θ represents neuron weights to be changed by the end of each iterative step i . The implementation of Q-network is further improved by:

- utilizing a step forward for the present state s to get predictive Q values,
- applying the replay (like historical IDS for the interactive process) into data let $Ht =$

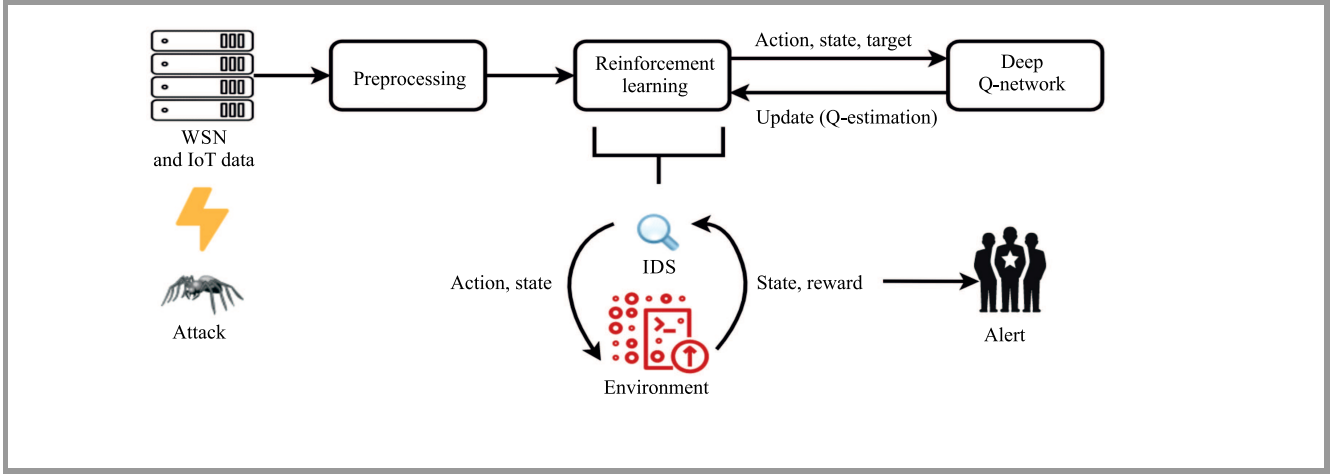


Fig. 1. Proposed method of improvement the IoT and WSN based RL-IDS.

$\{h(1), h(2), \dots, h(t)\}$ within an over-time t as $f_t = (s_t, a_t, r_t, s_{t+1})$,

- updating the Q-network based on the data from training (r, s, a, s) over the target Q-value with optimization of the loss-function during an iterative step noted as:

$$L_i(\theta_i) = E\{[x_i - Q(s, a, \theta_i)]^2\}, \quad (10)$$

$$x_i = r_t + \gamma \arg \max_{a'} Q(s', a', \theta_{i-1}), \quad (11)$$

- applying back-propagation with loss function's gradient, the weights are updated corresponding to the θ parameters as:

$$\nabla_{\theta_i} L_i(\theta_i) = E\{[x_1 - Q(s, a, \theta_i)] \nabla_{\theta_i} Q(s, a, \theta_i)\}. \quad (12)$$

4.1. Model Description

In the proposed scheme, the problem for QoS control is tackled based on the approach of R-learning algorithm. The main aim of every QoS scheduler is maximization the amount of data transmitted with low power consumption. For this fundamental trade-off, the function U is defined to analyze the ratio of throughput to power. Thus, the function for QoS scheduler at i -th position U_i is:

$$U_i(B_j^i, B_{-i}) = \frac{TS_i(B)}{B_j^i}, \text{ s.t., } B_j^i \in B_i, B = \prod_{i \in N} B_i | B_i \in [B_1^i, B_m^i], \quad (13)$$

where B_{-i} is the transmit power vector without B_i , and $TS_i(B)$ is the throughput scheduler.

In wireless communication, the signal to interference noise ratio (SINR) in the given effective range γ_i is measured while computing the throughput at i -th scheduler TS_i and can be expressed using:

$$TS_i(B) = W \cdot \log_2 \left(1 + \frac{\gamma_i(A)}{\Omega} \right), \quad (14)$$

where W is referred to as bandwidth of the channel assigned in through IoT network, Ω ($\Omega \geq 1$) is the gap between capacity and the uncoded M-ary quadrature amplitude modulation (M-QAM).

Algorithm 1: The IoT-WSN-based RL-IDS used for training and testing

Data: sensor data dataset Y

Input: Initialize action, state, environment, parameter θ , targeted Q-network
Initialize reply-memory H space

Output: return vector $Q(s_t, a_t, \theta)$

```

while  $|\widehat{Q}_{i+1} - \widehat{Q}_i| < \sigma$  do
    for  $X = 1, 2, 3, \dots, N$  do
         $s_0$  = starting state
        for  $t = 0, 2, 3, \dots, T - 1$  do
            Select an action (random)  $a_t$  with
            a random-probability  $p$  based on  $\in$ 
            strategy as:
             $a_t = \arg \max_a Q(s, a_k, \theta)$ 
            - Apply  $a_t$  and the reward observed by the
            IDS- $r_t$  and the next state observe chosen
            reward  $r_t$  and store the tuple
             $(s_t, a_t, r_t, s_{t+1})$  in  $H$ 
            - Arbitrary batch selection with this
            selected feature  $(s_t, a_t, r_t, s_{t+1})$  from  $H$ 
            if  $s_{t+1}$  terminal state then
                 $\mu l = rl$ 
            end
            else
                 $\mu l = rl + \delta \arg \max_{a'} Q(s', a', \theta)$ 
            end
            Gradient calculation of the loss function
            based on Eq. (11)
        end
    end
end
    
```

Table 2
Dataset used for evaluation

Category	Port	Attack	Tools	Size [bytes]
Information collect		Scanning of service OS fingerprinting	Nmap, hping3, xprobe2 Nmap	1.4 MB 358 KB
Denial of service	UDP, TCP HTTP	Distributed DoS	hping3 golden-eye hping3	19.5 MB 18.8 MB 19.7 KB
	TCP, HTTP UDP	DoS	hping3 hping3 golden-eye	11.2 MB 21.7 MB 29.7 KB
Information theft		Key-logging data theft	Metasploit	1369
			Metasploit	118

The environment was made by consolidating traffic and Table 2 shows a the used datasets and software tools.

5. Evaluation Criteria

The validation of proposed algorithm is researched by two measures:

- **Accuracy** – this metric is measured as the degree of closeness between the actual and the predicted value,
- **Precision** – this is a metric that describes the accuracy level obtained from the mentioned information and the outcomes anticipated by the executed model. Consequently, accuracy is the proportion of true positive forecasts contrasted with general aftereffects of positive expectation.

Table 3 shows the boundaries or limits used for CNN and MLP algorithms.

Table 3
Parameters of algorithms used for testing

Algorithm	Batch size	Function (activation)	Optimizer	Epochs
Convolution neural network (CNN)	32,64,128	Softmax, ReLu	Adam	10, 30, 50
Multilayer perceptron (MLP)	32,64,128	Softmax, ReLu	Adam	10, 30, 50
Markov decision process (MDP)	32,64,128	Softmax, ReLu	Adam	10, 30, 50

A major drawback of any IoT sensor network is that these devices work in remote networks and have to be sustained on their battery life. Hence the average energy consumed by the device plays a vital role which depends on its performance as shown in Fig. 2, the node shows that the MDP algorithm provides a less amount of energy consumption when compared with CNN and MLP algorithm. MDP provides significant results as false detection is reduced even when the number of nodes is increased as shown in Fig. 3. As the number of nodes increases the false detection is getting reduced as compared with MLP and CNN.

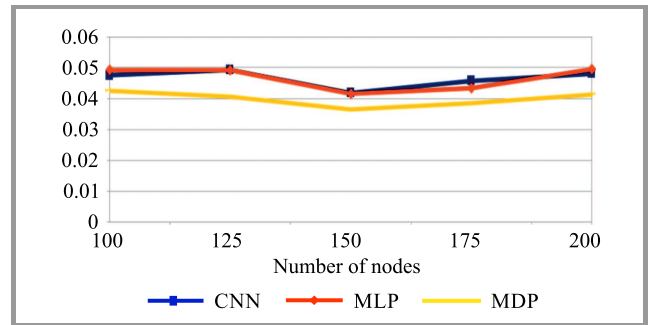


Fig. 2. Average energy consumption by number of nodes.

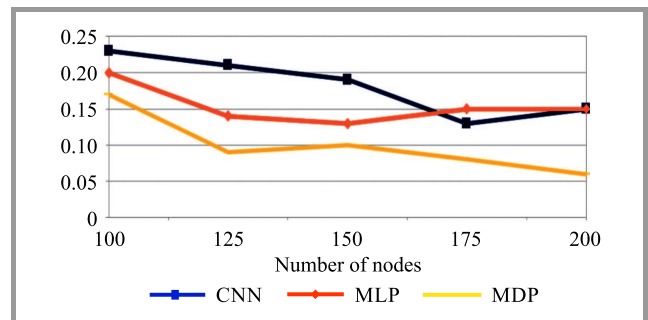


Fig. 3. False alarm rate.

The system of IoT mainly in a wireless system depends on the success rate of message delivery even when the number of nodes are increased and have a successful delivery rate which is provided in Fig. 4. In this plot all algorithm with the proposed algorithm, the throughput is given and can be observed that the MDP performance is good for throughput when nodes are more.

A comparison figure of the detection rate of IoT systems is shown in Fig. 5 which depicts that the detection rate at the receiver node in MDP is better when compared with CNN and MLP.

Figure 6 presents normalized overhead for several nodes in the IoT network when compared with all other algorithms with the reinforced algorithm MDP, it provides better performance for normalized overhead when compared with MLP and CNN. Parameters from Table 4 were used in this plot.

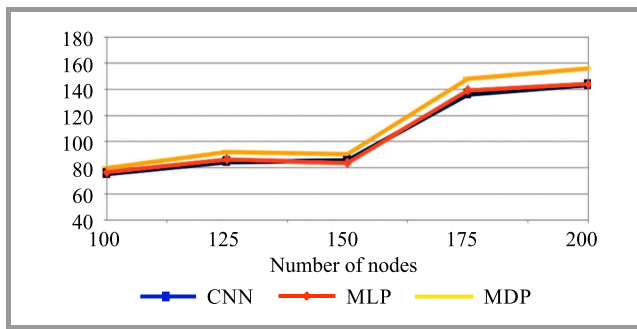


Fig. 4. Throughput rate of change.

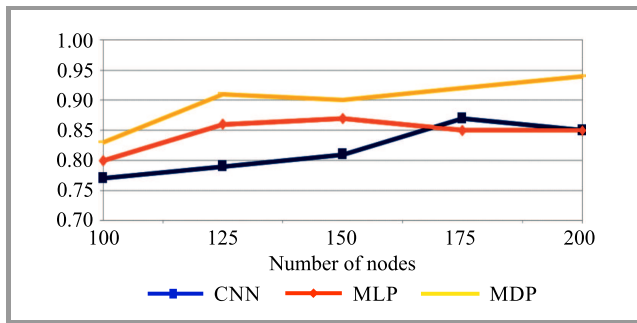


Fig. 5. Detection rate.

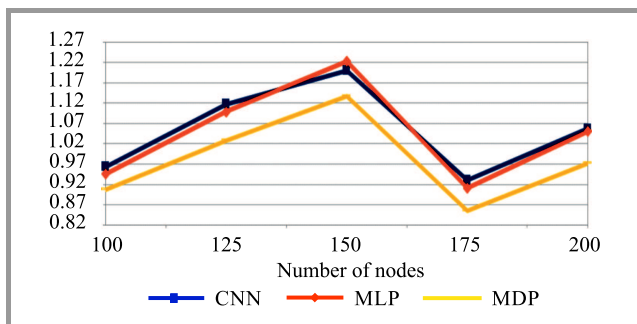


Fig. 6. Normalized overhead.

Table 4
Evaluation metrics (detection rate of attacks)

Algorithm	Metrics				
	DDoS attack	DoS attack	Reconnaissance	Normal (AUC)	Theft (AUC)
MDP	0.99	0.99	0.97	0.99	0.95
CNN	0.98	0.97	0.98	0.98	0.99
MLP	0.55	0.49	0.96	0.97	0.97

Table 5 represents the classification and comparison results based on the feature selection and the AUC precision metrics.

In Table 6, the mean accuracy is expanded as the number of study ages for the MLP classifier. For CNN, there was a decrease as the quantity (in terms of numbers) of epochs increased from 10 to 50.

Table 7 shows the same accuracy evaluation for size of 64. For this situation, the accuracy (batch size 64) diminished

Table 5
Comparison analysis

Algorithm	AUC	Precision	Sensitivity
MDP	0.99	99.80%	98.55%
CNN	0.92	96.75%	97.00%
MLP	0.89	95.05%	93.02%

Table 6
The accuracy evaluation for batch size 32

Algorithm	Epoch	Mean Accuracy	Elapsed time
MDP	10	93.22%	60 min 12 s
CNN	10	91.75%	58 min 39 s
MLP	10	54.07%	39 min 09 s
MDP	30	91.03%	165 min 25 s
CNN	30	89.72%	158 min 30 s
MLP	30	63.95%	124 min 33 s
MDP	50	90.00%	230 min 21 s
CNN	50	89.30%	229 min 22 s
MLP	50	63.00%	186 min 47 s

with the expansion epochs for the classifier (MLP). Data decreasing a bit while the number of epochs is increased from 10 to 50 in CNN.

Table 7
Accuracy for batch size 64

Algorithm	Epoch	Mean accuracy	Elapsed time
MDP	10	92.00%	18 min 40 s
CNN	10	91.15%	20 min 57 s
MLP	10	76.92%	26 min 56 s
MDP	30	92.30%	62 min 17 s
CNN	30	91.02%	64 min 18 s
MLP	30	54.04%	64 min 19 s
MDP	50	92.30%	114 min 60 s
CNN	50	90.64%	112 min 55 s
MLP	50	53.89%	102 min 20 s

Table 8 shows the outcome for block size of 128. The normal exactness seems to increment along with the expanding number of the experiment of epochs for MLP-based classifier. For the CNN, a slight diminishing was observed as the number of epochs rises from 10 to 30. In all cases the larger batch size the shorter application lifetime.

6. Conclusion

In the proposed work, the reinforcement learning in a network is examined. The valuation of the RL-IDS model is incorporated and compared with different ML and DL algorithms such as CNN and LP. The RL calculation gave the best outcome and precision and AUC leads in multiclass

Table 8
Mean accuracy for batch size 128

Algorithm	Epoch	Mean accuracy	Elapsed time
MDP	10	92.50%	12 min 12 s
CNN	10	90.87%	11 min 33 s
MLP	10	54.10%	10 min 16 s
MDP	30	93.00%	40 min 50 s
CNN	30	90.76%	45 min 44 s
MLP	30	54.43%	27 min 58 s
MDP	50	92.03%	55 min 27 s
CNN	50	91.27%	54 min 27 s
MLP	50	79.01%	46 min 18 s

characterization. With epoch increase a slight reduction in precision is observed, while in the 128-batch preliminaries, there was an increase in accuracy. A double change in MLP could make the estimation cycle 1.4 to 2.6 s faster, while CNN could make the figuring cycle 1.8 to 2.4 s shorter. Later on, the models with various calculations are likely created and different calculations for AI or profound learning are joined. Moreover, this calculation ought to be actualized in NIDS so it very well may be utilized progressively to alleviate attacks.

References

- [1] M. Roopak, G. Y. Tian, and J. Chambers, "Deep learning models for cyber security in IoT networks", in *Proc. IEEE 9th Annual Comput. and Commun. Workshop and Conf. (CCWC)*, Las Vegas, NV, USA, 2019, pp. 452–457 (DOI: 10.1109/CCWC.2019.8666588).
- [2] X. Yuan, C. Li, and X. Li, "DeepDefense: identifying DDoS attack via deep learning", in *Proc. of the 2017 IEEE Int. Conf. on Smart Comput. (SMARTCOMP)*, Hong Kong, China, 2017, pp. 1–8 (DOI: 10.1109/SMARTCOMP.2017.7946998).
- [3] D. Evans, "The Internet of Things: how the next evolution of the Internet is changing everything", *Cisco Internet Business Solutions Group (IBSG)*, 2011 [Online]. Available: http://www.cisco.com/c/dam/en-us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
- [4] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets", *Computer*, vol. 50, no. 7, 2017, pp. 80–84 (DOI: 10.1109/MC.2017.201).
- [5] P. Radanliev *et al.*, "Future developments in cyber risk assessment for the Internet of Things", *Computers in Industry*, vol. 102, pp. 14–22, 2018 (DOI: 10.1016/j.compind.2018.08.002).
- [6] E. Bertino and N. Islam, "Botnets and Internet of Things security", *Computer*, vol. 50, no. 2, pp. 76–79, 2017 (DOI: 10.1109/MC.2017.62).
- [7] M. A. Al-Garadi, A. Mohamed, A. Al-Ali, X. Du, and M. Guizani, "A Survey of machine and deep learning methods for Internet of Things (IoT) security", *IEEE Commun. Surveys & Tutorials*, vol. 22, no. 3, pp. 1646–1685, 2020 (DOI: 10.1109/COMST.2020.2988293).
- [8] A. Okwori, "Intrusion detection in Internet of Things (IoT)", *Int. J. of Advanced Res. in Computer Sci.*, vol. 9, pp. 504–509, 2018 (DOI: 10.26483/ijarcs.v9i1.5429).
- [9] Y. Meidan, "ProfilIoT: a machine learning approach for IoT device identification based on network traffic analysis", in *Proc. of the Symp. on Applied Comput. – SAC '17*, Marrakech, Morocco, 2017, pp. 506–509 (DOI: 10.1145/3019612.3019878).
- [10] E. Anthi, L. Williams, M. Slowinska, G. Theodorakopoulos, and P. Burnap, "A Supervised intrusion detection system for smart home IoT devices", *IEEE Internet of Things J.*, vol. 6, no. 5, 2019, pp. 9042–9053 (DOI: 10.1109/JIOT.2019.2926365).
- [11] A. Azmoodeh, A. Dehghantanha, and K.-K. R. Choo, "Robust malware detection for Internet of (battlefield) things devices using deep eigenspace learning", *IEEE Trans. Sustain. Comput.*, vol. 4, no. 1, 2019, pp. 88–95 (DOI: 10.1109/TSUSC.2018.2809665).
- [12] S. Hajiheidari, K. Wakil, M. Badri, and N. J. Navimipour, "Intrusion detection systems in the Internet of Things: A comprehensive investigation", *Comput. Netw.*, vol. 160, pp. 165–191, 2019 (DOI: 10.1016/j.comnet.2019.05.014).
- [13] R. Nicolescu *et al.*, "Mapping the values of IoT", *J. Inf. Technol.*, vol. 33, pp. 345–360, 2019 (DOI: 10.1057/s41265-018-0054-1).
- [14] S. Sheng *et al.*, "Deep reinforcement learning-based task scheduling in IoT edge computing", *Sensors (Basel)*, vol. 21, no. 1666, 2021 (DOI: 10.3390/s21051666).
- [15] Y. Chen *et al.*, "Deep reinforcement learning based dynamic resource management for mobile edge computing in industrial Internet of Things", *IEEE Transac. on Industrial Informat.*, vol. 17, no. 7, pp. 4925–4934, 2021 (DOI: 10.1109/TII.2020.3028963).
- [16] M. Elrawy, A. Awad, and H. Hamed, "Intrusion detection systems for IoT-based smart environments: a survey", *J. Cloud Comput.*, vol. 7, no. 21, 2018 (DOI: 10.1186/s13677-018-0123-6).
- [17] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools", *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2013 (DOI: 10.1109/SURV.2013.052213.00046).
- [18] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in IoT security: current solutions and future challenges", *arXiv [Online]*. Available: <https://arxiv.org/pdf/1904.05735.pdf>
- [19] K. A. P. da Costa, J. P. Papa, C. de Oliveira-Lisboa, R. Munoz, and V. H. C. de Albuquerque, "Internet of Things: a survey on machine learning-based intrusion detection approaches", *Computer Networks*, vol. 151, pp. 147–157, 2019 (DOI: 10.1016/j.comnet.2019.01.023).
- [20] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau, "Autoencoder-based network anomaly detection", in *2018 Wireless Telecommun. Symp. (WTS)*, Phoenix, AZ, USA, 2018, pp. 1–5 (DOI: 10.1109/WTS.2018.8363930).
- [21] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, "Toward a lightweight intrusion detection system for the Internet of Things", *IEEE Access*, vol. 7, pp. 42450–42471, 2019 (DOI: 10.1109/ACCESS.2019.2907965).
- [22] M. Abomhara and G. M. Koien, "Cyber security and the Internet of Things: vulnerabilities, threats, intruders and attacks", *J. of Cyber Secur. and Mobil.*, vol. 4, no. 1, pp. 65–88, 2015 (DOI: 10.13052/jcsm2245-1439.4).



Gauri Kalnoor received her B.E. and M.Tech. from the department of Computer Science and Engineering, Viswesaraya Technological University, Belgavi in 2008 and 2010, respectively. She has worked in Central University of Karnataka as an Assistant Professor and in Wipro Technologies as a Project Engineer. She is a re-

search scholar in B.M.S.C.E. Research Centre and her research area is Internet of Things. She is interested in coding and analysis of machine learning techniques.

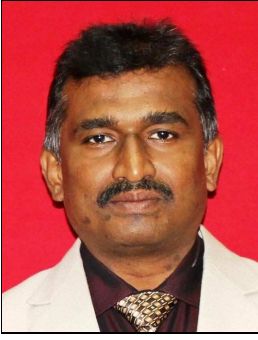
 <https://orcid.org/0000-0001-9970-4697>

E-mail: kalnoorgauri@gmail.com

B.M.S.C.E

Basavangudi

Bangalore, India



Gowrishankar S is a senior Professor at Computer Science & Engineering department at BMS College of Engineering, Bangalore. He served as a Head of the department of CS&E and IS&E of BMSCE. He is actively associated with the Research Collaborative Sabbatical program with University of Alabama, Huntsville (UAH), USA

and he is a visiting professor for UAH. Having an Academic and Research experience of 20 years, he authored more than 80 research publications in reputed international journals and conferences. His research interests include performance evaluation, wireless network and deep learning.

 <https://orcid.org/0000-0002-8119-8711>

E-mail: Gowrishankar.cse@bmsce.ac.in

B.M.S.C.E

Basavangudi

Bangalore, India

Linear and Planar Array Pattern Nulling via Compressed Sensing

Jafar Ramadhan Mohammed¹, Raad H. Thaher², and Ahmed Jameel Abdulqader^{1,2},

¹ College of Electronic Engineering, Ninevah University, Mosul, Iraq

² College of Engineering, Mustansiriyah University, Baghdad, Iraq

<https://doi.org/10.26636/jtit.2021.152921>

Abstract—An optimization method based on compressed sensing is proposed for uniformly excited linear or planar antenna arrays to perturb excitation of the minimum number of array elements in such a way that the required number of nulls is obtained. First, the spares theory is relied upon to formulate the problem and then the convex optimization approach is adopted to find the optimum solution. The optimization process is further developed by using iterative re-weighted l_1 -norm minimization, helping select the least number of the sparse elements and impose the required constraints on the array radiation pattern. Furthermore, the nulls generated are wide enough to cancel a whole specific sidelobe. Simulation results demonstrate the effectiveness of the proposed method and the required nulls are placed with a minimum number of perturbed elements. Thus, in practical implementations of the proposed method, a highly limited number of attenuators and phase shifters is required compared to other, conventional methods.

Keywords—compressed sensing, convex optimization, iterative re-weighted l_1 - norm minimization, linear and planar arrays.

1. Introduction

One of the challenges in current and future wireless communication systems is the presence of interfering signals that may originate either from pre-specified and known or from unknown directions. In such cases, performance of the system may be significantly degraded. This problem becomes more significant in such applications as satellites [1], fifth-generation wireless communications [2] and modern radars [3], as the system of this type are usually expected to operate in environments characterized by severe interference and a very crowded spectrum. One of the simplest and most powerful techniques for eliminating these interfering signals is to point the nulls of the array radiation pattern in the direction of the unwanted interfering signals.

Null placement may be achieved by accurately controlling such array design variables such as element excitation weights and element spacing [4], [5]. Conventionally, all the weights and/or positions of the array elements were perturbed to place the required nulls. Thus, the final phased array systems were usually complex, slow in their convergences, and expensive [6]. Many researchers

have investigated the complexity of such fully perturbed antenna arrays, coming up with some solutions. Some of them suggest simple deterministic approaches, such as iterative Fourier transform method [7] and the edge-element method [8], [9] to identify those elements that need to be perturbed for achieving the required null placement. Other scientists, meanwhile, used numerical optimization algorithms, such as the genetic algorithm [4], [10], particle swarm optimization [11], simulated annealing [12], evolutionary algorithms [13], adaptive algorithms [14], cuckoo search optimization [15], invasive weeds optimization [16], and grey wolf optimization [17], to optimize the excitations of the perturbed elements. None of the aforementioned techniques offers a clear path towards selecting the minimum required number of perturbed elements needed in order to place the required number of nulls. In fact, they always assumed that the number of the perturbed elements should be higher than the total number of the required nulls in order to insure an accurate pattern nulling capability. Thus, the number of the perturbed elements was excessive and the solutions were usually not optimal.

Other methods include the use of clustered arrays in which the main arrays were divided into clusters that may consist of either regular or irregular clustered elements [18]–[20]. Furthermore, paper [21] suggested a partially thinned array approach that was applied to side elements only, thus creating a relatively low complexity null placement method. In addition, the structure of a conventional adaptive sidelobe canceller system used in spaced radars has been greatly simplified by using different auxiliary configurations [22] in order to create another solution to this important issue.

In light of the above discussions, there is a great need for a new optimized method that is capable of perturbing only the exactly required number of elements in order to place the required number of nulls in an efficient manner. In a bid to solve the problem, compressed sensing was suggested in [23], [24] in order to significantly reduce complexity of array feeding networks. In [25], Bayesian compressed sensing was suggested to find the best match between the sparse array and the reference patterns. Generally, several sparse recovery algorithms exist, such as Yalli [26], smoothed l_0 -norm [27], orthogonal matching pursuit [28], and iterative hard threshold [29] that may be used to solve

the complexity problem and achieve the desired patterns. Some of these algorithms, like Yalli and smoothed l_0 -norm, usually do not accurately recover the sparse solutions. On the other hand, the iterative reweighted l_1 -norm [30] and the two-steps l_0 [31] methods were used to efficiently determine the minimum number of perturbed elements and to achieve the desired constraints.

In this paper, radiation patterns of linear and planar arrays are optimized by means of the compressed sensing approach, making sure that the required nulls are placed under a minimum number of perturbed elements. First, the sparse recovery array is built, and then it is implemented with convex optimization applied in order to find the optimum solution. The sparsity of the solution is enhanced through the use of the iterative reweighted l_1 -norm algorithm [32]. That approach allowed the required nulls to be placed efficiently with precisely the needed number of perturbed elements.

2. Principles of the Method

For simplicity, consider a linear array of N isotropic elements in which the array pattern may be expressed by:

$$AF_{uniform}(\theta) = \sum_{n=1}^N x_{on} e^{jkd_n u}, \quad (1)$$

where θ is the observation angle around the array axis, $k = \frac{2\pi}{\lambda}$ is the wave number, λ is the wavelength, $u = \sin \theta$, d_n is the position of elements along the x axis which is represented by $d_n = (n - \frac{N+1}{2})d$, and d is the element spacing. Further, $x_{on} = a_{on} e^{-jkd_n u_o}$, where a_{on} is the array amplitude, $u_o = \sin \theta_o$, and θ_o is the steering angle of the main beam. After substitution, Eq. (1) can be rewritten as:

$$AF_{uniform}(\theta) = \sum_{n=1}^N a_{on} e^{jkd_n(u-u_o)}. \quad (2)$$

To place a number of wide nulls equal to Q , where $q = 1, 2, \dots, Q$, we need to perturb the element weights as follows:

$$X_n = x_{on} + x_n, \quad (3)$$

where x_n is the weight of the sparse elements. The array factor at the null directions is equal to zero, $AF(\theta_q) = 0$. Then Eq. (2) can be modified accordingly:

$$AF(\theta_q) = \sum_{n=1}^N X_n e^{jkd_n u_q} = \sum_{n=1}^N (x_{on} + x_n) e^{jkd_n u_q}, \quad (4)$$

which can be rewritten as:

$$AF_{uniform}(\theta_q) = -\sum_{n=1}^N x_n e^{jkd_n u_q}. \quad (5)$$

This is a set of linear equations that can be written as $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A} = \sum_{n=1}^N e^{jkd_n u_q}$, $\mathbf{x} = x_n$, and $\mathbf{b} = -AF_{uniform}(\theta_q)$. Note that vector $\mathbf{x}$ with size $N \times 1$ is the sparse weight that needs to be found and it contains both zero and non-zero values. Vector $\mathbf{b}$ with size $Q \times 1$ is the magnitude of the uniform array pattern at null directions with the opposite

phase and, finally, $\mathbf{A}$ is the matrix with size $Q \times N$. These three parameters can be written as:

$$\mathbf{x} = [x_1, x_2, \dots, x_N]^T, \quad (6)$$

$$\mathbf{b} = [-AF_{uniform}(\theta_1), -AF_{uniform}(\theta_2), \dots, -AF_{uniform}(\theta_Q)]^T, \quad (7)$$

$$\mathbf{A} = \begin{bmatrix} e^{-jkd_1 u_1} & \dots & e^{-jkd_N u_1} \\ \vdots & \ddots & \vdots \\ e^{-jkd_1 u_Q} & \dots & e^{-jkd_N u_Q} \end{bmatrix}. \quad (8)$$

Suppose that the number of array elements is larger than the number of the required nulls, $Q < N$, which is the practical scenario, especially for large arrays that consist of hundreds of elements. In such a case, the system will have infinite solutions. To find the minimum number of the perturbed elements, Eq. (8) can be solved using the l_0 norm minimization approach. However, the problem becomes non-convex and cannot be solved by means of convex optimization. Thus, it is first converted to the convex type by using the l_1 norm that can be solved by:

$$\text{minimize } \|\mathbf{x}\|_1 \text{ subject to } \mathbf{Ax} - \mathbf{b} \leq \varepsilon. \quad (9)$$

This equation can be solved by using the convex optimization approach, with its implementation explained in [33]. To enhance the sparsity of the solutions and hence reduce the number of perturbed elements, Eq. (9) can be iteratively minimized as:

$$\text{minimize } \|\beta(\mathbf{x}^{i-1})\mathbf{x}^i\|_1 \text{ subject to } \mathbf{Ax} - \mathbf{b} \leq \varepsilon, \quad (10)$$

where $\beta(\mathbf{x}^{i-1}) = \frac{1}{\mathbf{x}^{i-1} + \delta}$, δ is a small positive number that is used to provide stability to array weights, such that the minimization process is assured by estimating non-zero values of the sparse elements and is not affected by the zero values in the next new iterations. The non-zero values of Eq. (10) represent the sparse elements that give exactly the needed number of the perturbed elements. The corresponding array pattern of these perturbed elements can be obtained. Then, the overall array pattern can be computed by:

$$AF_{proposed} = AF_{uniform} - AF_{sparse \text{ elements}}. \quad (11)$$

3. Simulation Results

This section demonstrates the performance of the proposed method while generating the required nulls under the minimum number of perturbed elements. In all scenarios, the number of iterations, i , and δ are chosen to be 15 and 10^{-6} respectively.

3.1. Scenario 1 – Linear Array

A uniform linear array of $N = 20$ elements with inter-element spacing equal to $d = 0.5\lambda$ is considered in this scenario. Several cases of null placements are investigated to illustrate the effectiveness of the proposed method. In the first case, a single wide null centered at 49° is placed.

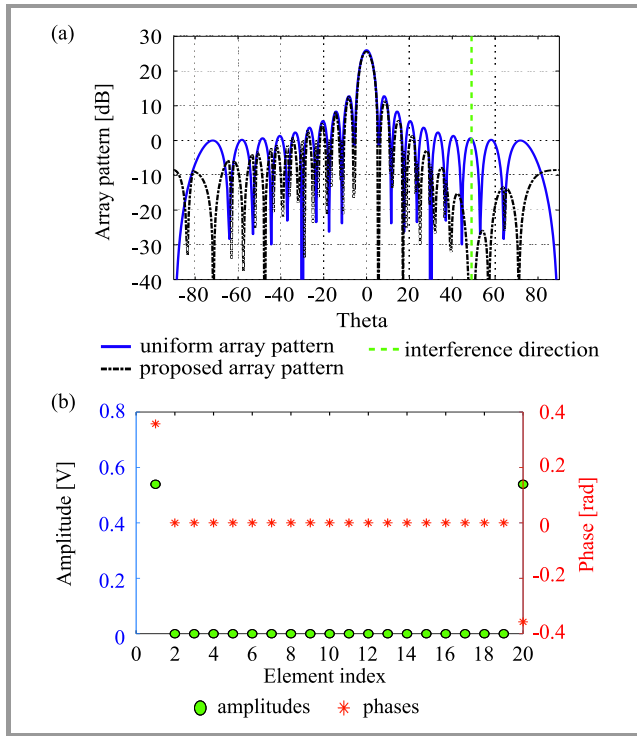


Fig. 1. Radiation patterns of the tested arrays (a) and the corresponding perturbed elements of the proposed array for $N = 20$ and a single wide null at 49° (b). (See digital edition to find the color version).

To obtain such a wide null, two adjacent sharp nulls have to be imposed. For example, to place a wide null centered at 49° two adjacent sharp nulls are placed at 48.95° and 49.05° , respectively. Each sharp null needs one perturbed element and, thus, each wide null will need at least two perturbed elements. Further, to achieve such a wide null, it is required that the patterns of the sparse array and the uniform array according to Eq. (11) are exactly coincident at the null direction of 49° . Figure 1 shows the results of the proposed array using the compressed sensing approach. The uniform array pattern is also shown for comparison. It can be seen that the required wide null has been successfully placed with only two sparse (non-zero) elements. The perturbed complex weights of these two sparse elements selected randomly by the algorithm are $0.5176 + j0.2652$ and $0.5176 - j0.2652$, with indices 1 and 20, respectively. In the second case, multiple wide nulls are generated. Accordingly, the number of the randomly perturbed elements is expected to increase. However, this increment represents the actual need of the algorithm to perform the required null placement. To highlight this important point, relationship between the required number of nulls and the minimum

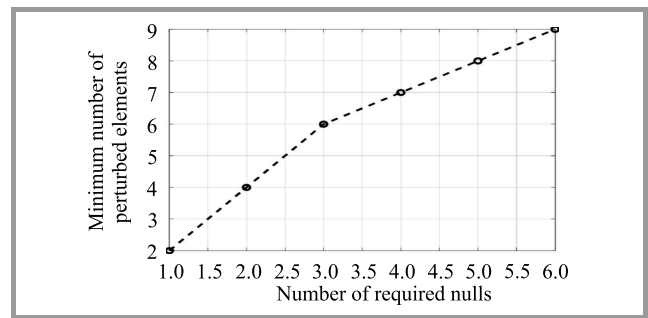


Fig. 2. The minimum number of perturbed elements versus the required number of wide nulls.

Table 1
Design parameters of two and three wide nulls

Two wide nulls			
A	x		b
4×20 matrix	Indices of sparse elements	Complex values of sparse elements	4×1 matrix
	1	$0.5208 + j0.3099$	$-0.1535 - j0.2222$
	4	$-0.2612 + j0.0487$	$0.8294 + j0.1324$
	17	$-0.2612 - j0.0487$	$0.7756 - j0.5410$
	20	$0.5208 - j0.3099$	$0.3115 - j0.5702$
Three wide nulls			
A	x		b
6×20 matrix	Indices of sparse elements	Complex values of sparse elements	6×1 matrix
	1	$0.2905 + j0.1139$	$1.4983 - j0.8935$
	5	$-0.4651 - j0.2487$	$0.9893 - j1.6371$
	7	$0.1439 + j0.3133$	$0.0153 + j0.0281$
	14	$0.1439 - j0.3133$	$0.3423 + j0.2908$
	16	$-0.4651 + j0.2487$	$0.6141 - j0.3653$
	20	$0.2905 - j0.1139$	$0.3006 - j0.3322$

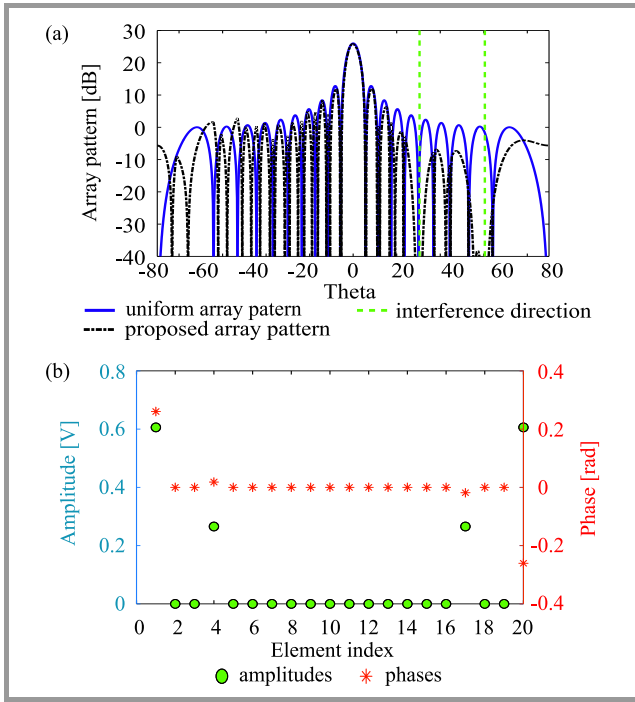


Fig. 3. Radiation patterns of the tested arrays (a) and the corresponding perturbed elements of the proposed array for $N = 20$ and two wide nulls at 30° and 60° (b).

number perturbed elements are plotted, as shown in Fig. 2. It can be observed that the two wide nulls need at least four perturbed elements. Table 1 shows the complex weights of the sparse elements that are required for generating two and three wide nulls.

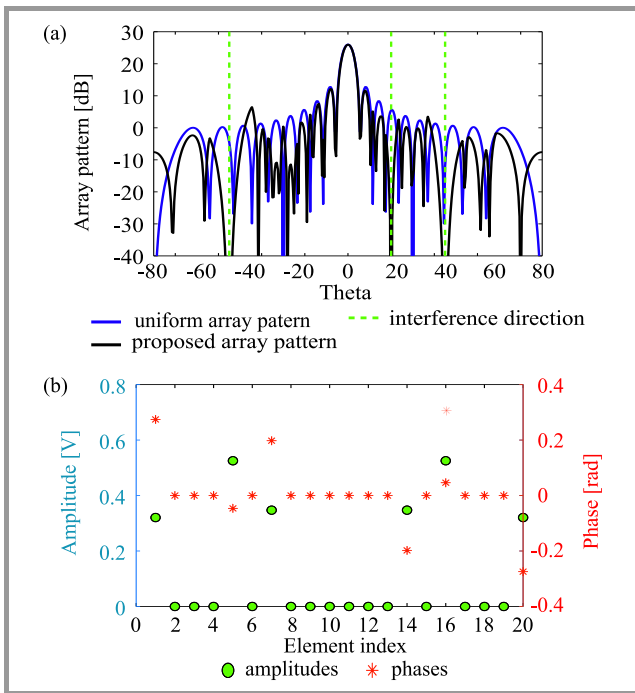


Fig. 4. Radiation patterns of the tested arrays (a) and the corresponding perturbed elements of the proposed array for $N = 20$ and three wide nulls at 20° , 45° , and -55° .

Figures 3 and 4 show the results of the tested arrays of two and three wide nulls, respectively. These results fully confirm the effectiveness of the proposed array in placing the required number of wide nulls. Moreover, the directivity of the proposed array is found to be affected only slightly, as long as the number of perturbed elements is lower than the total number of array elements.

3.2. Scenario 2 – Planar Array

In this scenario, a square planar array with 20×20 elements and $\lambda/2$ inter-element spacing along the x and y axes was considered. In the first use case, the center of the required wide null was chosen to be at $v = -0.5$ and no nulls at u plane were presented, with $v = \sin(\theta) \sin(\phi)$ and $u = \sin(\theta) \cos(\phi)$. Figure 5a shows the three-dimensional pattern of the proposed array obtained with the use of the compressed sensing approach, characterized by the minimum number of perturbed elements, while Fig. 5b shows the results of the original uniform planar array shown for comparison purposes.

In the other use case, two wide nulls centered at $v = -0.5$ and $u = -0.7$ are considered. Figure 6 shows the results of this case, with the two required nulls placed successfully.

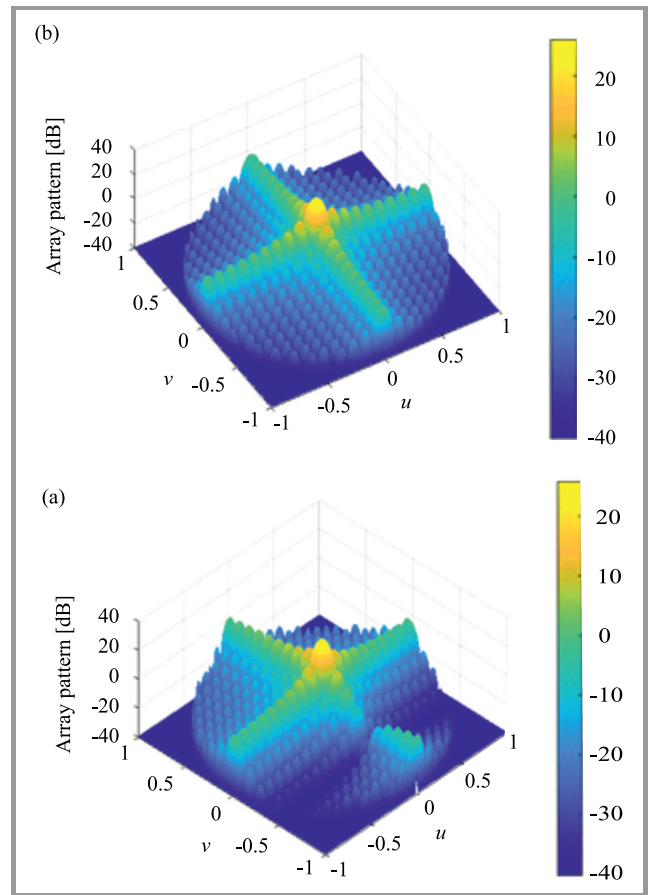


Fig. 5. Radiation patterns of the proposed planar array (a) and the uniform array pattern (b) for 20×20 and a single wide null centered at $v = -0.5$.

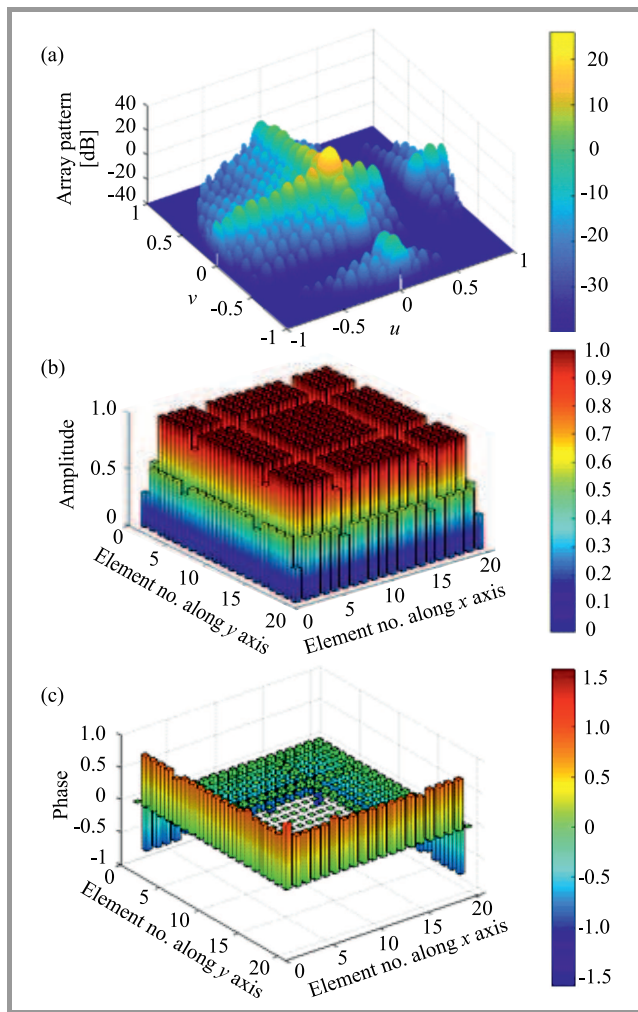


Fig. 6. Radiation patterns of the proposed planar array (a), the corresponding perturbed amplitudes (b), and phases (c).

4. Conclusions

An efficient and simple optimization method based on the sparse theory and on the compressed sensing approach was presented for synthesizing linear and planar array patterns with the minimum number of perturbed elements. The simplicity of the proposed method means that the number of the RF components, such as attenuators and phase shifters, is reduced. The proposed array is capable of placing the required wide nulls in undesired directions. For each single wide null there is a need for at least two perturbed elements. Convex programming has been applied to implement and find sparse elements needed to perform the desired null placements. The results show a significant reduction in the complexity of the array feeding network.

References

- [1] M. Fakharzadeh, S. H. Jamali, P. Mousavi, and S. Safavi-Naeini, "Fast beamforming for mobile satellite receiver phased arrays: theory and experiment", *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1645–1654, 2009 (DOI: 10.1109/TAP.2009.2019911).
- [2] Y. Dong, Z. Chen, P. Fan, and K. B. Letaief, "Mobility-aware uplink interference model for 5G heterogeneous networks", *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2231–2244, 2016 (DOI: 10.1109/TWC.2015.2500566).
- [3] Z. Geng, H. Deng, and B. Himed, "Adaptive radar beamforming for interference mitigation in radar-wireless spectrum sharing", *IEEE Signal Process. Letters*, vol. 22, no. 4, pp. 484–488, 2015 (DOI: 10.1109/LSP.2014.2363585).
- [4] J. R. Mohammed, "Rectangular grid antennas with various boundary square-rings array", *Progress In Electromagnet. Res. Letters*, vol. 96, pp. 27–36, 2021 (DOI: 10.2528/PIERL20112402).
- [5] J. R. Mohammed, "Obtaining wide steered nulls in linear array patterns by controlling the locations of two edge elements", *AEU Int. J. of Electron. and Commun.*, vol. 101, pp. 145–151, 2019 (DOI: 10.1016/j.aeue.2019.02.004).
- [6] C. A. Balanis, *Antenna Theory: Analysis and Design, Third Edition*, Hoboken, New Jersey: John Wiley & Sons, 2005, pp. 385–424 (ISBN: 9780471667827).
- [7] W. P. M. N. Keizer, "Fast low sidelobe synthesis for large planar array antennas utilizing successive fast Fourier transforms of the array factor", *IEEE Trans. Antennas Propag.*, vol. 55, no. 3, pp. 715–722, 2007 (DOI: 10.1109/TAP.2007.891511).
- [8] J. R. Mohammed and K. H. Sayidmarie, "Null steering method by controlling two elements", *IET Microwaves, Antennas Propag.*, vol. 8, no. 15, pp. 1348–1355, 2014 (DOI: 10.1049/iet-map.2014.0213).
- [9] J. R. Mohammed and K. H. Sayidmarie, "Synthesizing asymmetric sidelobe pattern with steered nulling in non-uniformly excited linear arrays by controlling edge elements", *Int. J. of Antennas and Propag.*, vol. 2017, 2017 (DOI: 10.1155/2017/9293031).
- [10] M. Dawoud and M. Nuruzzaman, "Null steering in rectangular planar arrays by amplitude control using genetic algorithms", *Int. J. Electron.*, 2000, vol. 87, no. 12, pp. 1473–1484, 2010 (DOI: 10.1080/00207210050192498).
- [11] M. A. Mangoud and H. M. Elragal, "Antenna arrays pattern synthesis and wide null control using enhanced particle swarm optimization", *Prog. in Electromag. Res. B*, vol. 17, pp. 1–14, 2009 (DOI: 10.2528/PIERB09070205).
- [12] A. Trastoy and F. Ares, "Placing quasi-nulls in planar and conformal arrays", *Electromagnetics*, vol. 19, no. 4, pp. 373–383, 1999 (DOI: 10.1080/02726349908908654).
- [13] A. Slowik and H. Kwasnicka, "Evolutionary algorithms and their applications to engineering problems", *Neural. Comput. Appl.*, vol. 32, no. 16, pp. 12363–12379, 2020 (DOI: 10.1007/s00521-020-04832-8).
- [14] J. R. Mohammed, "A new simple adaptive noise cancellation scheme based on ALE and NLMS filter", in *Proc. 5th Annual Int. Conf. on Commun. Network and Service Res. (CNSR)*, Fredericton, New Brunswick, Canada, 2007, pp. 245–254 (DOI: 10.1109/CNSR.2007.4).
- [15] U. Singh and M. Rattan, "Design of linear and circular antenna arrays using Cuckoo optimization algorithm", *Prog. in Electromag. Res. C*, vol. 46, pp. 1–11, 2014 (DOI: 10.2528/PIERC13110902).
- [16] S. K. Mahto and A. Choubey, "A novel hybrid IWO/WDO algorithm for nulling pattern synthesis of uniformly spaced linear and non-uniform circular array antenna", *AEU Int. J. of Electron. and Commun.*, vol. 70, no. 6, pp. 750–756, 2016 (DOI: 10.1016/j.aeue.2016.02.013).
- [17] P. Saxena and A. Kothari, "Optimal pattern synthesis of linear antenna array using Grey Wolf optimization algorithm", *Int. J. Antennas Propag.*, vol. 2016, pp. 1–11, 2016 (DOI: 10.1155/2016/1205970).
- [18] A. J. Abdulkader, J. R. Mohammed, and R. H. Thaher, "Phase-only nulling with limited number of controllable side elements", *Prog. in Electromag. Res. C*, vol. 99, pp. 167–178, 2020 (DOI: 10.2528/PIERC20010203).
- [19] J. R. Mohammed, A. J. Abdulqader, and R. Hamdan, "Antenna pattern optimization via clustered arrays", *Prog. in Electromag. Res. M*, vol. 95, pp. 177–187, 2020 (DOI: 10.2528/PIERM20042307).

- [20] J. R. Mohammed, A. J. Abdulqader, and R. Hamdan, "Array pattern recovery under amplitude excitation errors using clustered elements", *Prog. In Electromag. Res. M*, vol. 98, pp. 183–192, 2020 (DOI: 10.2528/PIERM20101906).
- [21] J. R. Mohammed, "Thinning a subset of selected elements for null steering using binary genetic algorithm", *Prog. in Electromag. Res. M*, vol. 67, pp. 147–155, 2018 (DOI: 10.2528/PIERM18021604).
- [22] J. R. Mohammed and K. H. Sayidmarie, "Performance evaluation of the adaptive sidelobe canceller with various auxiliary configurations", *AEU Int. J. of Electron. and Commun.*, vol. 80, pp. 179–185, 2017 (DOI: 10.1016/j.aeue.2017.06.039).
- [23] A. Massa, P. Rocca, and G. Oliveri, "Compressive sensing in electromagnetics – a review", *IEEE Antennas Propag. Mag.*, vol. 57, no. 1, pp. 224–238, 2015 (DOI: 10.1109/MAP.2015.2397092).
- [24] J. R. Mohammed, "High-resolution direction-of-arrival estimation method based on sparse arrays with minimum number of elements", *J. of Telecommun. and Informat. Technol.*, vol. 1, pp. 8–14, 2021 (DOI: 10.26636/jtit.2021.143720).
- [25] F. Viani, G. Oliveri, and A. Massa, "Compressive sensing pattern matching techniques for synthesizing planar sparse arrays", *IEEE Trans. Antennas and Propag.*, vol. 6, no. 9, pp. 4577–4587, 2013 (DOI: 10.1109/TAP.2013.2267195).
- [26] J. Yang and Y. Zhang, "Alternating direction algorithms for 11-problems in compressive sensing", *SIAM J. on Scien. Comput.*, vol. 33, no. 1, pp. 250–278, 2011 (DOI: 10.1137/09077761).
- [27] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for over complete sparse decomposition based on smoothed norm", *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, 2009 (DOI: 10.1109/TSP.2008.2007606).
- [28] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007 (DOI: 10.1109/TIT.2007.909108).
- [29] F. Marvasti *et al.*, "A unified approach to sparse signal processing", *EURASIP J. Adv. Signal Process.*, vol. 2012, p. 44, 2012 (DOI: 10.1186/1687-6180-2012-44).
- [30] M. A. Abdelhay, N. O. Korany, and S. E. El-Khamy, "Synthesis of uniformly weighted sparse concentric ring arrays based on off-grid compressive sensing framework", *IEEE Antenn. and Wireless Prop. Letters*, vol. 20, no. 4, pp. 448–452, 2021 (DOI: 10.1109/LAWP.2021.3052174).
- [31] M. Khosravi, M. Fakharzadeh, and M. H. Bastani, "Large array null steering using compressed sensing", *IEEE Sig. Process. Lett.*, vol. 23, no. 8, pp. 1032–1036, 2016 (DOI: 10.1109/LSP.2016.2580587).
- [32] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization", *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008 (DOI: 10.1007/s00041-008-9045-x).
- [33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.2", 2020 [Online]. Available: <http://cvxr.com/cvx>



Jafar Ramadhan Mohammed received the B.Sc. and M.Sc. degrees in Electronics and Communication Engineering in 1998, and 2001, respectively, and the Ph.D. degree in Digital Communication Engineering from Panjab University, India in 2009. He was a Visiting Lecturer in the Faculty of Elec-

tronics and Computer Engineering at the Malaysia Technical University Melaka (UTeM), Melaka, Malaysia in 2011 and Autonoma University of Madrid, Spain in 2013. He is currently a Professor and Vice Chancellor for Scientific Affairs at Ninevah University. His main research interests are in the area of digital signal processing and its applications, antenna, and adaptive arrays.

 <https://orcid.org/0000-0002-8278-6013>

E-mail: jafar.mohammed@uoninevah.edu.iq

College of Electronics Engineering

Ninevah University

Mosul, Iraq



Raad Hamdan Thaher received the M.Sc. degree in Electronics and Communication Engineering in 1981, and the Ph.D. degree in Communication Engineering from Faculty Polytechnic University Bucharest Romania in 1997. He works as a professor in Mustansiriya University, College of Engineering, Electrical Engineering Department, Baghdad, Iraq. His specialization is in the electronic communication engineering and his research interests are in the field of communication systems, electronics, microwaves, antennas, and communication networks.

E-mail: raadthaher55@gmail.com

Department of Electrical Engineering

Al-Mustansiriya University

Baghdad, Iraq



Ahmed Jameel Abdulqader received the B.Sc. and M.Sc. degrees in Electronics and Communication Engineering from the University of Mosul, Iraq, in 2009, and 2013, respectively. He is currently a Ph.D. student at Mustansiriya University, Baghdad, Iraq. He is Lecturer at Ninevah University. His main research interests are

in the area of design and analysis of antenna arrays, array pattern optimization, mobile communication systems, and computer networks.

E-mail: ahmed.abdulqader@uoninevah.edu.iq

College of Electronics Engineering

Ninevah University

Mosul, Iraq

High Temperature Effects in Fused Silica Optical Fibers

Krzysztof Borzycki, Marek Jaworski, and Tomasz Kossek

National Institute of Telecommunications, Warsaw, Poland

<https://doi.org/10.26636/jtit.2021.153521>

Abstract—Fire-resistant fiber optic cables used in safety and monitoring systems playing an essential role in fire fighting and building evacuation procedures are required to temporarily maintain optical continuity when exposed to fire. However, the use of fused silica fiber at temperatures between 800°C and 1000°C is associated with two highly undesirable phenomena. Thermal radiation (incandescence) of optical fibers, with its intensity and spectral distribution being proportional to additional attenuation observed in the fiber's hydroxyl absorption bands ("water peaks") is one of them. The other consists in penetration of thermal radiation from the surroundings into the fiber, due to defects in glass, causing light scattering and resulting in fiber brittleness. Thermal radiation is a source of interference in fiber attenuation measurements performed during fire tests and affects normal operation of fiber optic data links in the event of a fire. In this article, results of laboratory tests performed on a telecom single mode and multimode fibers subjected to temperatures of up to 1000°C are presented.

Keywords—fire-resistant fiber optic cable, fire test, fused silica optical fiber, incandescence spectrum, thermal deterioration, thermal radiation.

1. Introduction

Optical fibers used in fire-resistant cables should be capable of remaining operational, over a specific period of time, during a fire incident. While fused silica fibers are capable of withstanding temperatures of up to 1000°C while retaining their optical continuity, their properties degrade. In particular, time-dependent incandescence (thermal radiation) of the fiber itself occurs, and thermal radiation from the hot surroundings is coupled into the fiber. Both these phenomena are a cause of optical interference affecting transmission of data over hot fibers, and their testing. The performance of an optical fiber in a cable affected by fire depends on the following:

- properties of the fiber itself, i.e. attenuation, incandescence, light scattering, and coupling by defects,
- protection against fire and mechanical damage offered by heat-resistant components of the cable.

This paper focuses primarily on the first of the abovementioned items and describes the variations in optical properties of standard, telecom type fused silica optical fibers

at high temperatures in the event of a fire (or during a fire test). In addition to peak temperature, duration of exposure is important as well, as the performance of a heated fiber degrades slowly due to the following:

- migration of dopants (F_2 , GeO_2), which distorts the fiber's refractive profile,
- appearance of defects that are caused by the fact that fused silica turns into cristobalite and causes the glass to crack.

According to the analysis presented in [1], the migration of the most common GeO_2 dopant was of no importance as a fused silica fiber was heated to 1000°C for up to 100 h. The formation of microscopic defects in fused silica fibers has been reported in [2] and was also observed during the experiments performed. Defects, i.e. cracks in cladding, are capable of introducing infrared radiation from the hot surroundings into the fiber's core. This radiation is routed towards the fiber's end, thus creating interference affecting the operation of the receiver in a data link or of the optical power meter during a fire test. A hot optical fiber no longer is a "dark" interference-free transmission medium.

The investigations presented in this paper were triggered by technical problems encountered in measuring the attenuation of optical fibers during a fire test. Intense thermal emissions originating from a multimode fiber at or above 900°C prevented loss measurements from being performed with the use of a low power signal source and an optical power meter. This issue was previously unknown in the literature, as fire-resistant cables had predominantly the form of power and control cables with copper conductors. Contemporary standards providing for the monitoring of attenuation of optical fibers [3], [4] do not provide for any methods for the elimination of interference resulting from thermal emissions taking place in the tested fiber as well. Section 2 of this paper offers a short review of fire-retardant and fire-resistant fiber cables, while fire testing procedures applicable to the latter are presented briefly in Section 3. Section 4 outlines our test campaign, while Sections 5, 6 and 7 present selected test results for single-mode and multimode fibers, including light scattering and fiber deterioration due to thermally-induced defects. Section 8 reviews physical mechanisms and selected characteristics of thermal radiation encountered in hot fibers. Discussion of the

results and plans concerning further work are presented in Section 9, while Section 10 concludes the paper.

2. Indoor Cables on Fire

Fiber optic cables installed in commercial buildings are required to comply with fire safety regulations and shall pass the fire tests required. Two broad groups of such cables capable of meeting different requirements may be distinguished:

- fire- (or flame-) retardant cables,
- fire-resistant cables.

While these terms sound similar, their meanings are quite different.

2.1. Fire-retardant Cables

As far as cables for data communication networks which are not required to remain operational in the event of a fire are concerned, fire safety requirements applicable in the EU (US standards are substantially different) specify the following:

- a) fire propagation along the cable for given installation conditions: vertical shaft or horizontal space under a false ceiling, single cable or cable harness, type of fire source, its thermal power and duration of exposure,
- b) optical density of smoke (opaque smoke hinders evacuation from the danger zone),
- c) concentration of corrosive and toxic compounds: chlorine, fluorine and bromine compounds, hydrocyanic acid (HCN), sulfur dioxide (SO₂), etc. in the smoke.

Cables meeting the abovementioned requirements are classified as halogen-free fire retardant (HFFR) cables. In order to meet the requirements set out in (b) and (c), the use of materials containing halogens (Cl, F, Br) needs to be avoided, especially in jacket or sheath of the cable. Examples of non-compliant fire-retardant materials widely used in cable manufacturing include polyvinyl chloride (PVC), polyvinylidene fluoride (PVDF) and all polymeric materials with bromine compounds added.

These materials have been replaced with halogen-free thermoplastic polymers, typically polyethylene and its copolymers, or ethylene-vinyl acetate (EVA) which are highly flammable in their pure form, but become flame-retardant after addition of 60–70% (weight-wise) of a fine powder (1–3 µm) inorganic filler serving the purpose of a halogen-free flame-retardant and smoke suppressant. Once heated to its decomposition temperature, the filler releases large amounts of water vapors temporarily blocking access of air to the polymer and carrying the heat away. Instead of burning, slow charring takes place. However, once all water evaporates, the hot polymer begins to burn rapidly.

MDH or magnesium dihydrate – Mg(OH)₂, decomposing to water and solid magnesia (MgO) after being heated to over 300°C, is the most common filler of this type. ATH or alumina trihydrate (Al₂O₃·3H₂O) is a less expensive filler, but its relatively low decomposition temperature of 200–220°C makes it incompatible with several polymers requiring higher extrusion temperatures, including medium and high density polyethylene (MDPE, HDPE), polypropylene (PP), polyamides (PA) or poly(butylene terephthalate) (PBT) widely used for manufacturing fiber optic cables. ATH and MDH fillers are completely non-toxic. Other similar materials exist as well, but are rarely used.

Halogen-free fire/flame-retardant jacketing compounds are known as low smoke zero halogen (LSZH) materials.

A review of fire testing procedures applicable to communication and data transmission cables, along with references to applicable standards, is presented in IEC TR 62222 [5]. Detailed requirements, as well as descriptions of the test methods and hardware, are presented in several IEC/EN standards, applying to thin (maximum diameter of 20 mm) and metal-free indoor fiber optic cables [6]–[15]. The remaining standards apply to HFFR power, control and data transmission cables with metallic conductors.

HFFR cables are used to protect humans' life, as well as to ensure the safety of equipment and buildings by limiting or delaying the spread of fire and the amount of smoke and toxic or corrosive substances produced. However, ordinary fire-retardant cables are not required to remain electrically or optically functional, to retain continuity, to ensure stable transmission properties of optical or electrical circuits, to prevent short circuits, etc., when exposed to fire. A system utilizing these cables fails under such circumstances, either due to a disruption in communications or to the loss of power supply to remote hardware, such as video cameras. Cables from this group are out of the scope of this paper.

2.2. Fire-resistant Cables

Substantially different requirements apply to cables – either of the copper or fiber optic variety used in fire detection and protection systems, e.g. in smoke detectors, alarm devices, or surveillance cameras, and for ensuring emergency communications in the event of a fire. These cables must remain functional for a period of time that is necessary to alert security staff, police and fire services, to evacuate the affected area and to organize fire-fighting efforts. Cables of this type are known as fire-resistant (FR).

In addition to the requirements set forth in Subsection 2.1, fire-resistant cables shall:

- d) retain continuity and stability of critical parameters of electrical or optical circuits in the event of a fire, for 15 to 120 minutes, depending on cable fire classification,
- d) survive mild mechanical shocks and periodic sprinkling with water (optionally).

The applicable tests are defined in standards [16]–[21]. While fire test methods and severity, defined primarily by

peak temperature and test duration, vary considerably, the attenuation range permitted for optical fibers within the cable and its test methods are identical, depending on the type of fiber only [22].

Standardized fire test temperatures range from 800°C to over 1000°C. Hence, only optical fibers made of fused silica (SiO_2) are suitable for such applications. In order to retain compatibility with industry-wide splicing techniques and tooling, standard single-mode or multi-mode fibers with 125 μm cladding diameter in polymer protective coatings are used. While fused silica is capable of temporarily withstanding exposure to 1000°C without melting or significant degradation of its properties, polymer coatings fail: they decompose, leaving carbon residues, and then begin to burn at 500–600°C when exposed to oxygen. Special high-temperature fibers, such as fused silica fibers with nickel coating or fibers made of crystalline alumina (Al_2O_3), are capable of surviving heating to 1000°C without damage and are used in high-temperature fiber sensors, but not suitable for communication cables due to their high attenuation and lack of cladding, unsuitability for fusion splicing and brittleness (alumina fibers).

Telecom fibers have primary protective coatings made of UV – curable dual acrylate. Their maximum operating temperature is 85°C (continuous) and 200°C over a period of several days [23], [24]. Fibers with high-temperature acrylate coatings for continuous use in temperatures of approx. 150°C are available [25] as well, but are not used in communication cables.

A fire-resistant fiber optic cable is a single-use product. Its structure is destroyed during a fire due to:

- burning or charring of its jacket or sheath,
- melting or disintegration of glass or basalt-based strength members,
- decomposition (carbonization) of fiber coatings, buffer tubes, filling compounds, etc.

Temporary fire resistance is achieved by wrapping the cable core and/or buffer tubes in refractory tapes made of synthetic fluorophlogopite mica with a silicone binder. Fluorophlogopite has the melting point of 1387°C, is non-flammable and serves as a barrier to fire, while simultaneously preventing rapid passage of gases. Two examples of fire-resistant cables are shown in Figs. 1 and 2.

High degree of fire- and high temperature resistance is exhibited by mineral insulated (MI) electric cables, where copper or nickel-clad copper wires are surrounded by insulation made of compressed magnesium oxide powder, and are tightly encased in a metallic sheath made of copper, stainless steel or heat-resistant alloy, such as Inconel 825. Basic versions of MI cables contain no flammable materials and some of their special variants may resist a 1000°C fire for 3 hours or more, retaining circuit integrity. However, MI cables are expensive and are characterized by poor bending tolerance. Their mineral insulation easily absorbs moisture,

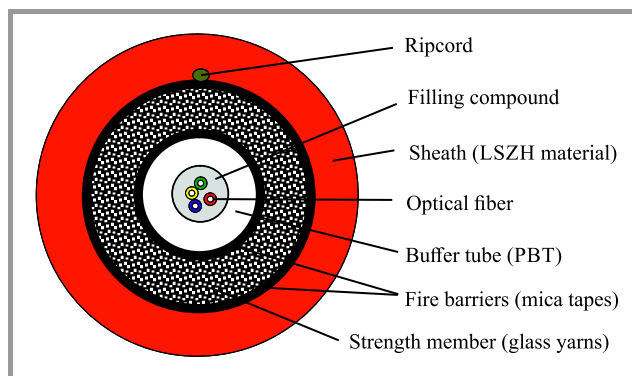


Fig. 1. Dielectric flexible fire resistant cable with a central loose tube – Technokabel FOC-2-SLT-HFFR 4G50, outer diameter 7.8 mm [26].

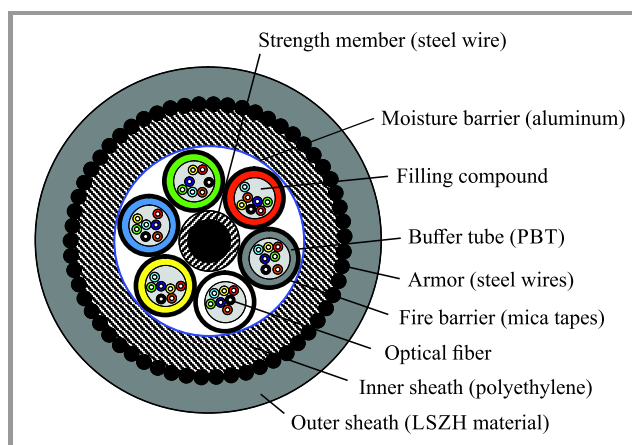


Fig. 2. Armored fire-resistant cable with stranded buffer tubes – FiberTek FTSF-FLTFMAPSZ (FR), outer diameter 19.8 mm [27].

and cable termination is rather labor-intensive. Additionally, the design and, in particular, the technology relied upon to manufacture MI cables (involving high pressure extrusion and drawing) are not suitable for fragile optical fibers.

2.3. Optical Fibers in Indoor Cables

Cables for indoor networks incorporate standardized telecom fibers:

- graded-index multimode with 50/125 μm diameter, type OM2, OM3, OM4 or OM5 [28];
- single-mode designated as ITU-T G.652.B/D [29] and G.657.A1/2 [30]. The equivalent IEC designations are B-652 and B-657.A1/A2 [31]. ISO designations of such fibers are OS1, OS1a, or OS2 [32]. However, the technical specifications are not identical.

Multimode fibers of OM2 or OM3 type are most widely employed in emergency and security systems with relatively low bit rates. More advanced (and expensive) OM4 and OM5 multimode fibers are used in high-speed data links,

particularly in data centers, which normally do not need to operate during a fire.

The cheapest single mode fibers (requiring, however, costly active devices) are used typically in the “vertical” sections of structural cabling carrying traffic between floors, and in links to public networks. Several of them are required to continue operating in the event of a fire, carrying emergency traffic.

3. Testing of Fire-resistant Cables

The examples of fire tests presented in Subsections 3.1 and 3.2 are of relatively low and high severity, respectively. The test defined in the German DIN 4102-12 standard is a “system test” covering both the cable itself and the installation hardware.

3.1. EN 50200 Fire Test

EN 50200 [19] simulates the effects of fire using a short length of a straight cable attached to a wall made of lightweight, refractory material. The cable subjected to the test is bent upwards at both ends of the hot zone, in line with the minimum static bending radius declared in the product’s specification sheet. The entire length of the cable subjected to fire is 0.8–1.0 m. A constant temperature of 842°C is generated by a 0.5 m long gas burner fed by a mixture of propane and air. Additionally, the vertical support holding the cable is subjected to a mechanical shock produced, every 5 minutes, by a steel rod hammer, in order to verify its mechanical integrity. The strength members in most flexible fire-resistant cables made of S-glass or E-glass yarns survive this test without melting, protecting the optical fibers inside.

The fire resistance class depends on how long the attenuation of fibers in the cable remains within the limits prescribed in EN 50582 [22]. The EN 50575 standard [20] provides for the following fire resistance classes:

- PH 15: over 15 min,
- PH 30: over 30 min,
- PH 60: over 60 min,
- PH 90: over 90 min,
- PH 120: over 120 min.

An optional test involves heating the cable for 30 min while simultaneously sprinkling it with cold water for the last 15 min, in order to simulate the operation of sprinklers in a building.

3.2. DIN 4102-12 Fire Test

DIN 4102-12 simulates the operation of cables passing through a room affected by a large fire. The cables undergoing the test (their number often exceeds 100) are placed

on steel racks below the ceiling or are fastened to the ceiling with clips, and are then bent at the entry to the test chamber made of refractory cinder blocks, in a manner typical of a real-world installation. All cable entry holes are sealed with fire resistant mortar. The minimum length of the test chamber is 3 m, and the cable may pass through the chamber many times to increase its length exposed to fire.

The interior of the test chamber is heated by multiple propane burners located in the lower part of the room, causing the temperature to rise in the manner shown in Fig. 3 and to exceed, 1000°C after 90 min. This is the “standard” or “cellulosic” temperature curve for a fire where burning cellulose-based materials (wood, plywood, paper, cardboard, cotton, etc.) are the heat source, with the formula defined in the ISO 834-1 standard [33]:

$$T = T_0 + 345 \log(8t + 1), \quad (1)$$

where T is temperature [°C], T_0 – initial temperature [°C], and t – time elapsed from the start of test [min]. The ISO curve is defined for the period of up to 180 min, but cable tests are usually shorter.

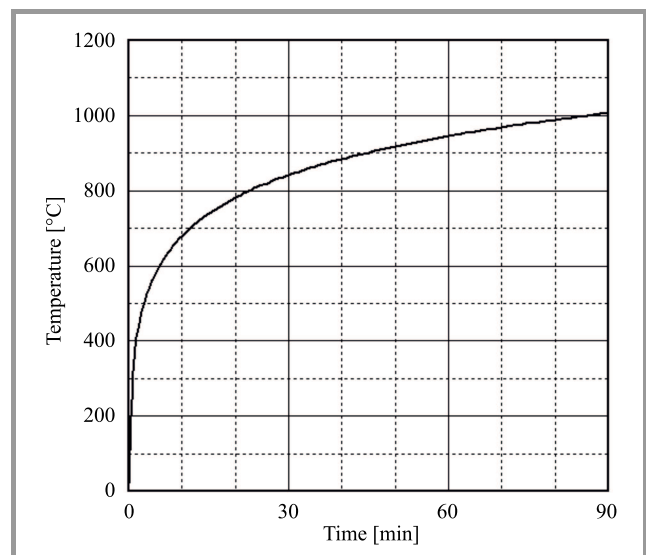


Fig. 3. Temperature vs. time during a 90-minute DIN 4102-12 fire test [21], [33].

The peak temperature is high enough to melt E-glass yarns in cables, so whether the cable will be able to retain its shape and integrity depends on the fire barrier and steel armor used (if present). Strength members made of basalt yarns do not melt, but become brittle and prone to disintegrating.

The fire resistance classification depends on length of time during which the increase of attenuation of optical fibers in the cable remains within EN 50582 limits. Three classes are distinguished: E30, E60 and E90, corresponding to the minimum cable survival time expressed in minutes.

DIN 4102-12 provides for a test covering a combination of cables and their supporting hardware. The latter determines



Fig. 4. Interior of a test chamber 25 min after commencing the DIN 4102-12 test. Temperature equals approximately 820°C. The sagging of cable trays is already considerable. The photograph was taken at the FIRES test lab in Batizovce, Slovakia, in 2019. Photo courtesy of Krzysztof Borzycki.

the sagging of the cable during the test (Fig. 4), affecting the overall results.

3.3. Test Requirements for Optical Fibers

The fire test pass criterion defined in EN 50582 [22] is based on the maximum increase of attenuation of optical fibers in the cable under test:

- for single-mode fibers: ≤ 1 dB/m at 1550 nm wavelength,
- for multimode fibers: ≤ 2 dB/m at 1300 nm wavelength.

Such limits assume that the length of the cable affected by fire is short, up to 5 m, as the 3–15 dB reserve in the attenuation budget of a fiber optic link is usually close to this range.

4. High Temperature Testing of Optical Fibers

The samples of single mode and 50/125 μm graded index multimode fibers (OM2) in a 250 μm primary protective coating were heated in a purpose-built electric tube oven made by a Warsaw-based Termtech company. The oven has a 1 m long straight tube made of alumina-rich ceramics, with a 15 mm hole. The fiber under test was protected against scratching and contamination by a 5/7 mm fused silica tube.

The tube was open at the ends enabling the products of the decomposition process and then of the burning of the remaining carbon (soot) to vent. This means that the conditions inside a cable are not reconstructed faithfully, as the mica fire barrier and molten glass yarns block oxygen

access, thus preventing the carbon residues from burning. However, the setup (Fig. 5) was helpful in investigating incandescence and thermal deterioration of optical fibers.

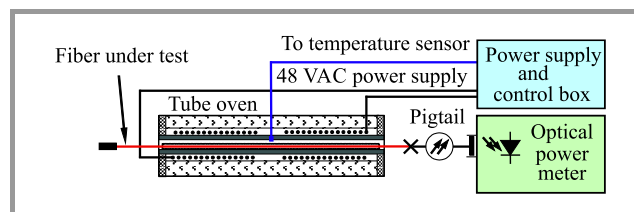


Fig. 5. Setup for measuring thermal emissions of the fiber. To measure the spectra, an optical spectrum analyzer was used instead of a power meter, and for attenuation measurements, a second pigtail and a source with the Fabry-Perot laser were added on the left-hand side instead of light-tight termination.

Test program provided for the measurements of the following:

- power and spectra of fiber thermal emissions (incandescence) vs. temperature,
- fiber attenuation vs. temperature (tests run separately due to the different setups required),
- fiber degradation after heating to 1000°C for up to 4 h.

In the course of the experiments, the fiber was first heated to 400°C, and then to 1000°C in 50°C increments. One step included an 8–10 minute heating phase and a phase in which the temperature was kept constant for at least 5 min. The total duration was 15 min.

The experiments involving loose tubes with fibers placed in sealed protective tubes are planned in the future, in order to simulate conditions inside the cable during a fire.

Optical power was measured using the Agilent HP8153A optical multimeter with the HP81532A plug-in module equipped with an InGaAs 800–1700 nm photodetector, calibrated for the wavelength of 1300 nm.

Spectra measurements take 15–25 min to complete due to weak signals and were performed at 800, 900 and 1000°C with Yokogawa AQ-6315B spectrum analyzer in the 700–1700 nm range. Incandescence of “OH-free” single mode fibers was too weak to acquire spectrum data even at 1000°C. Tests on some fiber samples were repeated in order to measure thermal emissions at 400–700°C, when the results are not affected by the decomposition and burning of fiber coating and by changes in emission power and spectrum after a longer (1–4 h) heating phase, usually up to 1000°C.

Repetition of the test did not reflect the conditions prevalent inside a fire-resistant cable, but provided some useful data for other high temperature applications in which telecom fibers are used, e.g. fiber sensors.

The following uncertainties affecting power and attenuation measurements were identified:

- Emission power (Figs. 6, 8, and 13): 0.5 dB for power values higher than -80 dBm. At lower values, the uncertainty increases due to the ingress of stray light into connectors (darkroom conditions were required for most experiments) and due to a periodic variation of power caused by on/off cycling of the oven's heater and by the temperature of the fiber under test, reaching approximately 1.5 dB at -95 dBm, despite averaging;
- Fiber loss (Figs. 7 and 20): 0.02 dB and 0.02 dB/m for 1 m samples.

Additionally, an attempt to investigate mechanisms of fiber deterioration after exposure to high temperatures, specifically distribution of light scattering defects and their appearance, was made as well – see Subsection 6.2.

5. Test Results of OH-free Single Mode Fibers

Two examples of cable fibers were tested:

- OFS AllWave Plus (G.652.D standard) [34] produced in 2019,
- OFS AllWave FLEX Plus (G.657.A2) [35] produced in 2018.

5.1. Thermal Radiation

The results obtained for both fibers are shown in Fig. 6. Radiation was too weak to acquire spectra.

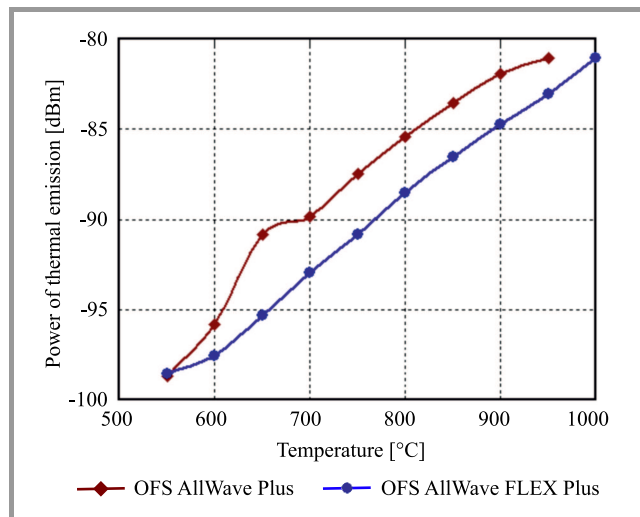


Fig. 6. Power of thermal emission vs. temperature for OH-free single mode fibers. Values below -90 dBm are affected by considerable uncertainty for reasons indicated in Section 4.

The OFS AllWave Plus fiber failed when the temperature exceeded 960°C . The emitted power increased 1000 times in approx. 2 min, and was slowly rising afterwards, as long

as the fiber was kept at 1000°C . The radiation spectrum indicated the coupling of radiation emitted by the hot ceramic tube of the oven. The sample lost optical continuity, but did not break. The OFS AllWave Plus fiber did not fail during the same test performed on 3 different samples. A microscope inspection revealed that some of the fiber failures were caused by contact with pieces of mineral wool fibers from the thermal insulation of the furnace.

The temporary increase of emission power at 600 and 650°C observed in the AllWave Plus fiber was a result of strong incandescence of black carbon soot left after decomposition of the coating materials. Some of this radiation was coupled into the fiber's core. This carbon was burned out at temperatures exceeding 700°C . A negligible increase in power while testing the AllWave FLEX Plus "bending insensitive" fiber may be explained by its strong light guidance and weak coupling of external radiation.

5.2. Change of Attenuation

The results shown in Fig. 7 are for the OFS AllWave FLEX Plus fiber. The sample was heated twice. The first test involved measurements at 1304 nm and 1000°C for 40 min, while the other at 1552.8 nm.

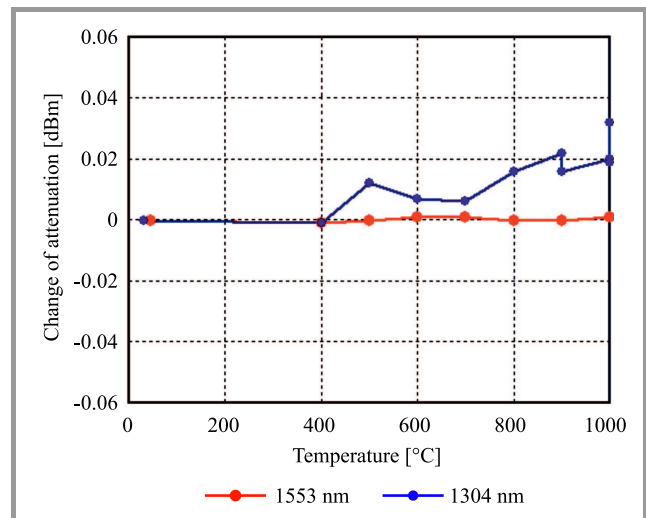


Fig. 7. Change of attenuation with temperature for OFS AllWave FLEX Plus fiber.

During the first test, the fiber coating first shrank because of decomposition and carbonization at approx. $400\text{--}500^{\circ}\text{C}$, which most likely caused the fiber to bend in a wave-like fashion. The fiber undergoing the test was fixed using small weights attached thereto 15 cm away from the furnace, on both sides, meaning that some extra length could be dragged into the tube. Next, the carbon soot burned without residue at $600\text{--}700^{\circ}\text{C}$, but the excessive length of the fiber could not be pushed out and it remained bent. This increased attenuation, which kept rising slowly with time at 1000°C , potentially due to the softening of the fiber and a gradual increase in its curvature. The extrapolated rise

of attenuation after 120 min was 0.06 dB/m. This corresponds to approximately 0.20 dB/m at 1550 nm, taking into account the typical wavelength dependence of losses caused by bending.

When the fiber was heated again, changes in attenuation at 1553 nm were below the measurement's uncertainty threshold, estimated at 0.02 dB/m, as there was no coating to exert forces on the glass fiber. The fiber attenuation changes that were predicted and measured were much lower than the 1 dB/m acceptance limit set forth in EN 50582 [22], and most of this attenuation budget is available to cover the attenuation rise caused by fiber macrobending and/or crush when the cable is deformed during the fire.

During the tests performed in [1] on the Corning SMF-28 single-mode fiber, the attenuation of bare fiber heated to 1100°C remained stable (≤ 0.5 dB/m) for 48 h, but at 1150°C it began to rise after 12 h. The report [1] does not specify the test wavelength.

6. Test Results for Old Single Mode Fiber

Here, the results obtained for the Siecor SMF-1528 fiber (G.652.A) manufactured in 1993 are presented. This was an equivalent of the popular Corning SMF-28 fiber, made in Germany under license by Siecor GmbH.

6.1. Thermal Radiation

The same sample was tested twice, showing some signs of decomposition and burning of the fiber coating and removal of OH ions from the core of the fiber (the so-called drying effect).

During the second part of the test, heating was continued up to 1000°C for 120 min, with amplitude of the main

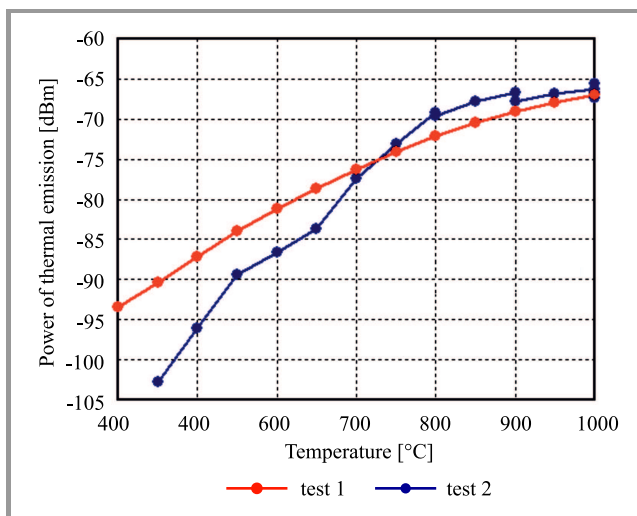


Fig. 8. Power of thermal emission vs. temperature for the Siecor SMF-1528 fiber. The values below -90 dBm are affected by considerable uncertainty, as described in Section 4.

water peak at 1383 nm decreased by approx. 50%. The first test also produced relatively few microscopic defects in the fiber, made evident by weak, distributed scattering of light injected by a 650 nm (red) laser, and by a considerable increase in thermal radiation power at temperatures up to 700°C, presumably due to the coupling of radiation emitted into the fiber core by carbonized coating. The escape of water at higher temperatures reduced the aforementioned value again.

The power data obtained is presented in Fig. 8. The visible discontinuities found at 800 and 900°C correspond to spectral measurements, both lasting 25–30 min. At the end of the first test, the fiber was kept at 1000°C for 100 min. During this time its thermal radiation decreased by 40% – see Figs. 11 and 12.

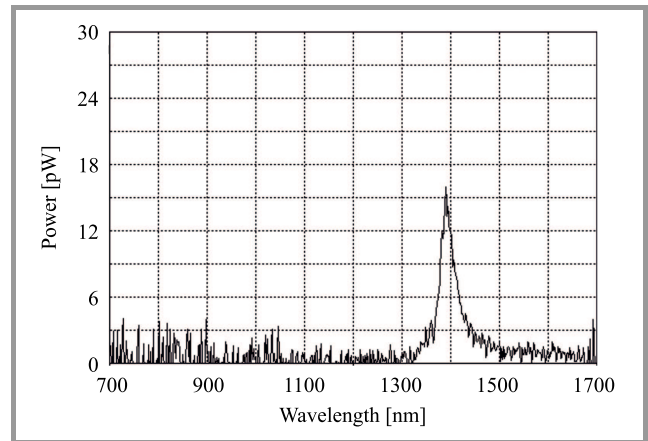


Fig. 9. Spectrum of thermal emission acquired for the Siecor SMF-1528 fiber at 800°C. The vertical axis is optical power in each spectrum slice of width equal to the analyzer's resolution bandwidth (10 nm).

The radiation spectra presented in Figs. 9–12 were acquired during test 1. The 1246, 1383, 1393, and 1407 nm OH bands listed in Table 1 are visible despite the 10 nm resolution, and their relative amplitudes are in agreement. In the last spectrum shown in Fig. 12, a continuous compo-

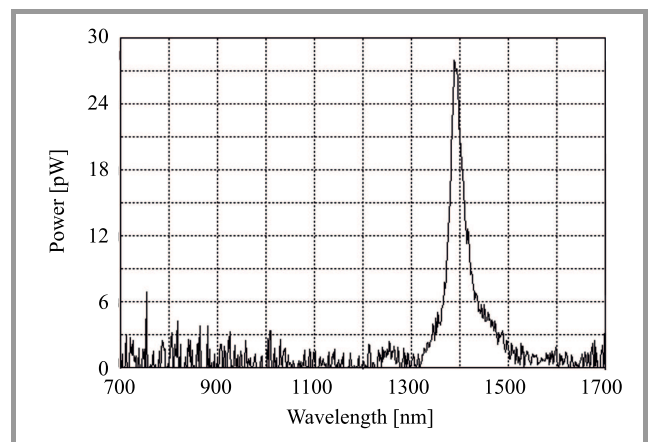


Fig. 10. Spectrum of thermal emission for the Siecor SMF-1528 fiber at 900°C.

nent specific to the incandescence of ceramics appears. It was coupled to the fiber's core by the rising number of defects in the glass – see Subsection 6.2.

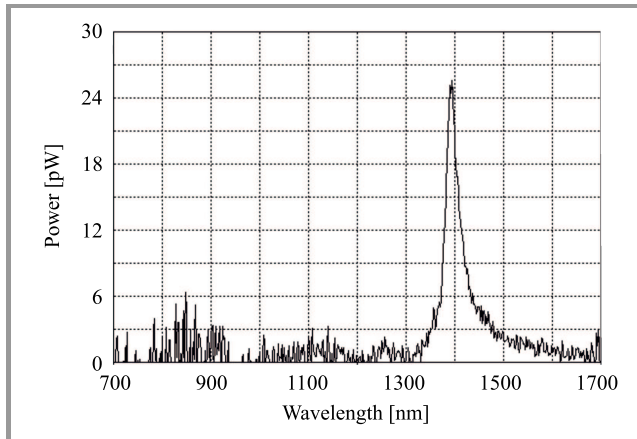


Fig. 11. Spectrum of thermal emission for the Siecor SMF-1528 fiber at 1000°C.

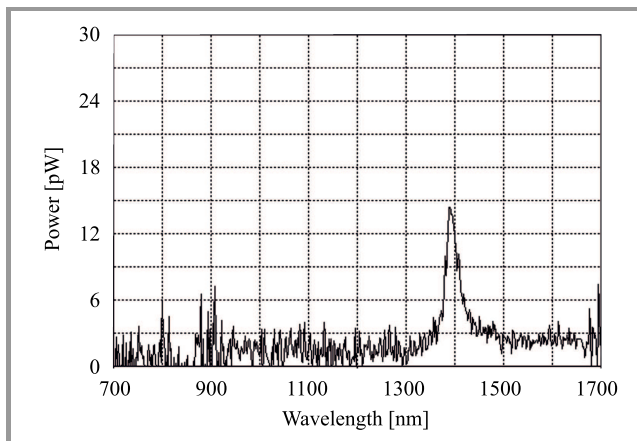


Fig. 12. Spectrum of thermal emission for the Siecor SMF-1528 fiber after 100 minutes at 1000°C.

6.2. Light Scattering and Defects

Uneven and randomly distributed scattering of red light from a 650 nm laser was observed in fiber samples heated, over the period of several hours, to 900–1000°C, but never in adjacent sections of the same fiber not exposed to high temperatures.

An example of longitudinal distribution of light scattering intensity in a heat-degraded Siecor SMF-1528 fiber is shown in Fig. 13. This was the sample tested in Subsection 6.1, but part of it was damaged during handling. Interestingly, there is no full correlation between intensity of lateral scattering and backscattering.

Observations made with the use of an optical microscope showed that the locations scattering visible light launched into the fiber sample turned out to be small spots and long wavy lines formed on the fiber's surface, from which cracks

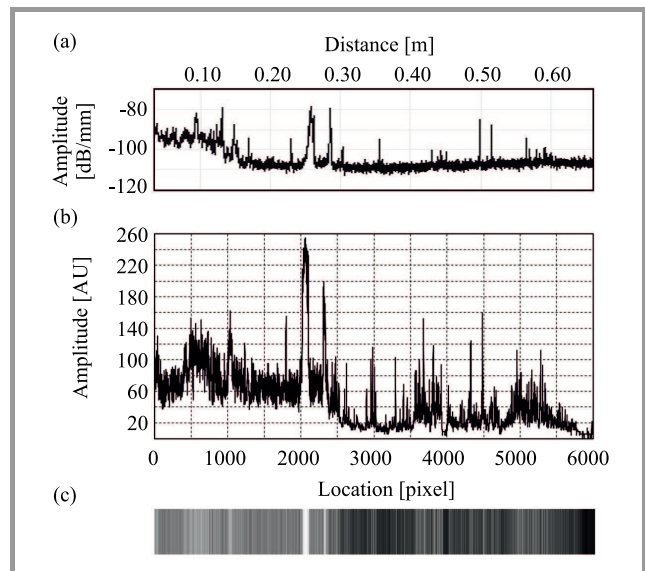


Fig. 13. Intensity of 650 nm light scattered laterally along a 62 cm section of the Siecor SMF-1528 single-mode fiber heated in free air to 900–1000°C for 4.5 hours and observed at room temperature. The image was obtained by averaging all pixels in each column in a photo of the fiber (c) and corresponding distribution of luminosity (b). The top graph is an OTDR trace acquired at 1550 nm with the OTDR – Luna Technologies OBR 4600 millimeter-resolution reflectometer.

usually extended into the cladding at different angles relative to the surface (Fig. 4). After cooling to room temperature, the fiber was very brittle and could not be cleaned without breaking. No defects were observed in the core. The physical degradation of a fused silica fiber subjected to high temperatures may be described in the following way. At temperatures above 850–900°C, fused (glassy) silica is slowly converted into its crystalline form known as cristoballite, which has higher specific gravity of 2.35 vs. 2.20, and a refractive index of 1.485 vs. 1.458 at 589.3 nm (n_D) wavelength. The crystallization of fused silica is intensified by the presence of water vapor in the gas mixture in the glass fiber's outline [1]. Water vapor is a product of thermal decomposition of oxygen-containing polymers, including poly(butylene terephthalate) [36], the most popular material of which loose tubes in fiber optic cables are made, and acrylates [37], of which primary coatings of optical fibers are made. It is also a product of burning of virtually all polymers (e.g. jacket of cables) and wood. Rose and Bruno [38] presented a short description of this process and listed the literature focusing on the subject.

The crystallization causes a localized strain after the loss of approximately 6% of the material's volume. This strain can initiate cracks penetrating the fiber cladding. A crack in the vicinity of or penetrating the core will disturb light guidance by deflecting some light away, thus resulting in scattering. Minuscule grains of cristoballite embedded in fused silica will cause omnidirectional scattering of both guided light – when inclusions are in the core, and external light – when inclusions are in the cladding. Defects

of the first type produce backscattering and loss of guided light from the fiber core, increasing its attenuation. Defects of the other type redirect some of the external thermal radiation to the fiber's acceptance cone.

The size of crystal grains observed on the surface of moderately heat-degraded fibers, after heating to 1000°C for several hours, as reported in literature [1], [38]–[40] usually equals 3–30 μm , which is in agreement with our observations (Fig. 14). Cracks in fiber cladding after heating to 850°C for more than 8 hours, or to a higher temperature, were reported by Rose [39], who concluded that a strong rise in attenuation of the fiber kept at 1100°C is explained by the formation of microcrystals – not only on the surface, but also inside the fiber core. The fiber tested had the same design and dimensions as SMF-1528 (type of fiber was not indicated). Shikama *et al.* [40] observed “spots” on the fiber surface only, which were approximately $3\times$ larger than those showed in Fig. 14, but no cracks after heating to 1000°C for 15 hours. The fiber tested had a cladding made of fluorine-doped fused silica instead of pure silica used in SMF-1528.

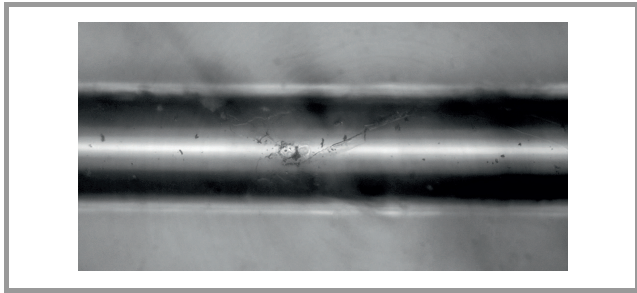


Fig. 14. A light-scattering location in fiber from Fig. 13, at OTDR distance about 45 cm, seen under microscope. The right-hand part of defect is a long crack extending into fiber cladding.

The entry of broadband thermal radiation from the hot surroundings into the fiber has been reported back in the days, together with an observation that fiber deterioration responsible for this phenomenon is enhanced by defects produced by gamma radiation from a ^{60}Co source [2], [40]. Both papers did not elaborate on any optical coupling mechanisms.

7. Test Results for 50/125 μm OM2 Multimode Fibers

7.1. Thermal Emission vs. Temperature

The fiber under test was OFS 50 μm graded-index OM2 [41], manufactured in 2019. The sample was tested four times, with the maximum temperature during test 1 set at 900°C to avoid a rapid fiber failure observed during several other experiments. The results are similar to those obtained for the old single mode fiber, except for stronger radiation.

Figure 15 shows power measurement data, while the spectra acquired are presented in Figs. 16–19. After an initial

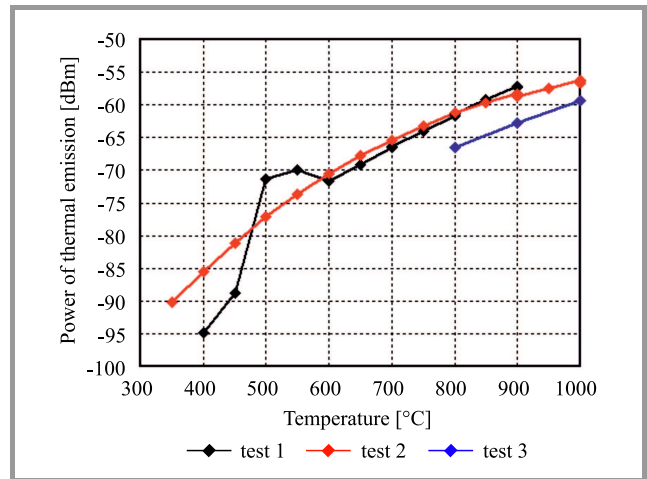


Fig. 15. Power of thermal emission vs. temperature for the OFS MM50 fiber. Values below -90 dBm are affected by measurement uncertainty.

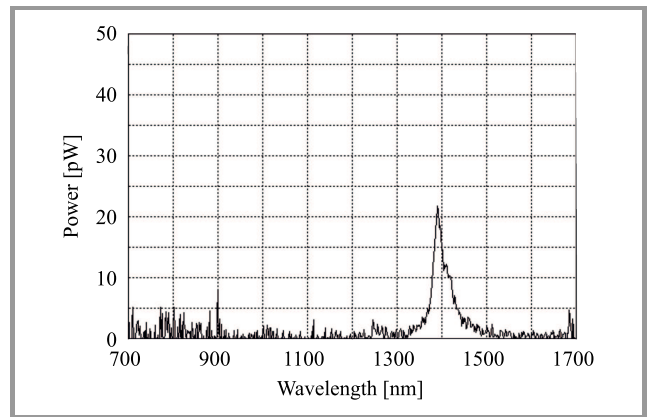


Fig. 16. Spectrum of thermal emission for the OFS MM50 fiber at 800°C in test 1.

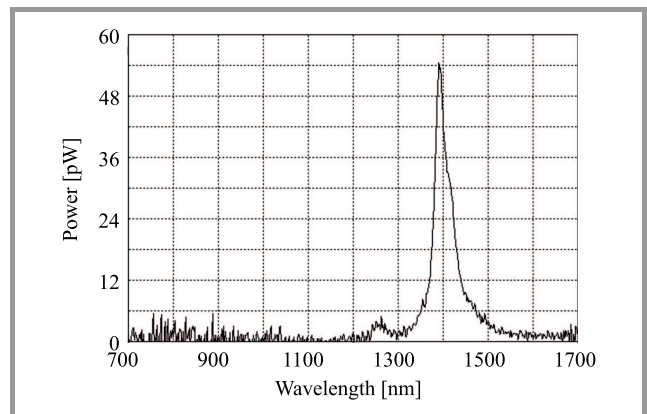


Fig. 17. Spectrum of thermal emission for the OFS MM50 fiber at 900°C in test 1.

rise in emission intensity (Figs. 16 and 17), a further temperature increases to 800–1000°C for 4 hours triggered the water escape and reduced the amplitude of OH emission peaks by approx. 80% (7 dB). As shown in Fig. 12, the last two spectra include a broadband component. This in-

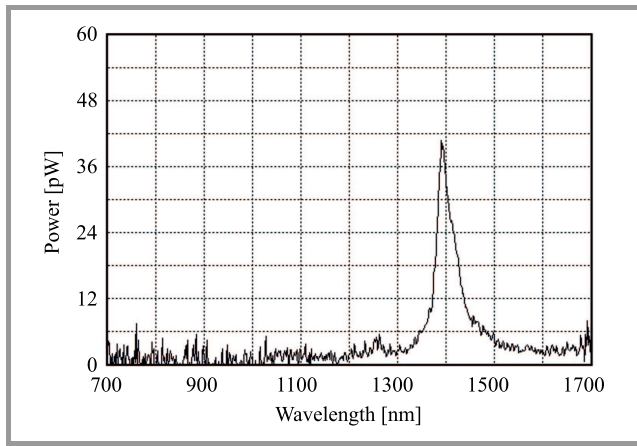


Fig. 18. Spectrum of thermal emission for the OFS MM50 fiber at 1000°C in test 2.

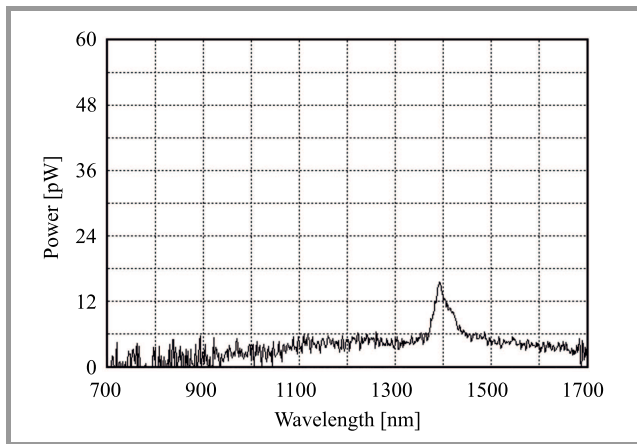


Fig. 19. Spectrum of thermal emission for the OFS MM50 fiber at 1000°C in test 4, after 4 hours of heating to 800–1000°C.

indicates the coupling, into the fiber, of radiation emitted by the glowing ceramic tube of the oven. Interestingly, none of the multimode fibers have showed any detectable emissions at 945 nm (Table 1).

7.2. Change of Attenuation vs. Temperature

The tested fiber was the Fujikura FutureGuide-MM50 [42] manufactured 2019. The sample was heated twice: the first test involved measurements performed at 1304 nm, while the other at 842 nm.

The changes in attenuation observed (Fig. 20) did not exceed 0.025 dB/m. As in the single mode fiber, higher values appeared during the first test, when the fiber coating was carbonized and burned. During this experiment, we experienced random variations of indicated attenuation due to imperfect coupling between the Fabry-Perot laser sources and the fiber tested, and due to reflections in the fiber optic connectors back into the laser.

During tests performed on the 50/125 μm multimode fiber of a similar design [1], the attenuation of fiber heated in the air to 1100°C remained fairly stable (≤ 1 dB/m) for over 300 h, but at 1200°C, it began to rise fast after less than

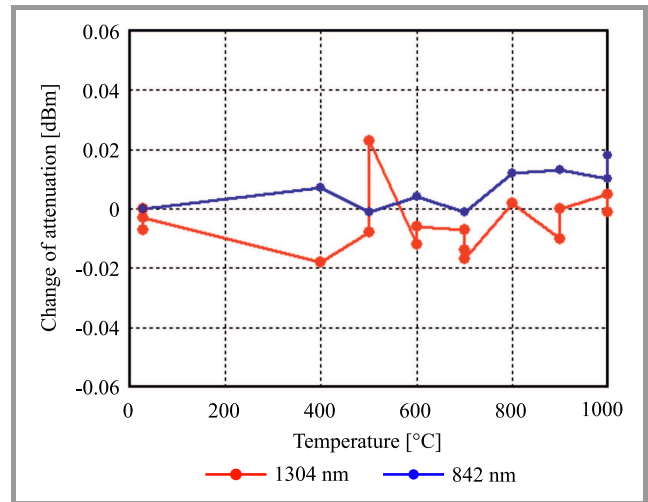


Fig. 20. Attenuation vs. temperature for the Fujikura FutureGuide-MM50 fiber.

20 h and the fiber essentially lost its optical continuity after 40 h. The report does not indicate clearly the wavelength at which the attenuation was monitored – most likely it was 1064 nm.

8. Overview of Spurious Radiation in Hot Silica Fibers

This section describes the characteristics of spurious radiation appearing in telecom type fused silica fibers subjected to a fire. The data presented comes from literature and from the experiments conducted. Other cable families, e.g. specialty fibers with a pure silica core and fluorine-doped cladding, intended for high temperature and radiation environments, may behave differently [1].

8.1. Thermal Radiation Generated in Fiber

Thermal emission of radiation in an optical fiber is proportional to the intensity of radiation absorption at a given wavelength, but not to scattering. The peaks correspond to absorption bands associated with the presence of OH ions in the glass constituting the fiber core, and intensity of emission peaks is proportional to the added attenuation in OH absorption bands, also known as “water peaks”. This phenomenon was reported thanks to research on high temperature fiber sensors [1], [2], but the tests were performed mostly on specialty fibers with a high OH content.

Peak wavelengths with relative absorption (and thus the expected emission) intensities of OH absorption bands between 700 and 1700 nm calculated using the data published in [43] are presented in Table 1. Intensity of OH emission at wavelengths close to 1390 nm rises along with temperature, in accordance with the Arrhenius formula and 1.13 eV activation energy, as reported by Shikama *et al.* [40]. This dependence is useful for temperature sensing with a short piece of a special high-OH fiber.

Table 1
OH absorption bands in fused silica

Peak wavelength [nm]	Relative intensity [dB]
724	−29.1
825	−42.2
878	−28.9
943	−16.1
1139	−29.5
1246	−13.7
1383	0.0
1393	−2.4
1407	−7.0

The increase of attenuation in the three strongest adjacent OH absorption bands: 1383 nm, 1393 nm and 1407 nm in old single-mode fibers conforming to the ITU-T G.652.A standard [29] may reach 2 dB/km at 20°C. In “low water peak” single-mode fibers conforming to G.652.D or G.657.A/B [30] standards, values of up to 0.05 dB/km are typical. All three OH bands are frequently combined as a single “1390 nm water peak” due to low resolution of spectral measurements. The examples of spectra are presented in Subsections 6.1 and 7.1.

Two other mechanisms, namely phonon absorption and UV absorption, noticeably contribute to the increase in attenuation at wavelengths longer than 1750 nm or shorter than 500 nm, respectively. While the first type of radiation lies outside the sensitivity range of common fiber optic receivers and power meters, receivers based on 850 nm silicon photodiodes are somewhat sensitive to short-wave radiation of the second type.

The measurements performed by the authors in the 700–1700 nm range with the Yokogawa AQ-6315B optical spectrum analyzer only detected radiation peaks coinciding with the four strongest OH bands listed in Table 1. The probable explanation is that spectral measurements are characterized by low sensitivity. Attenuation added at the absorption peak had to exceed 0.2 dB/km at room temperature for the corresponding emission from a 1 m long fiber sample at 1000°C to be detectable with the spectrum analyzer available.

After commercial introduction of “OH-free” or “low water peak” dispersion unshifted single mode fiber in 1999, this product was steadily increasing its market share to reach the value of 90% today. However, cheap OH-contaminated fibers are still sold and permitted under the ISO/IEC 11801 standard for structural cabling [32]. The authors expected “OH-free” fibers to exhibit weak incandescence. Pre-2000 single mode fibers contaminated with OH ions behave differently. Multimode fibers widely used in local area networks, data centers and indoor communication systems exhibit the strongest water peaks.

8.2. Thermal Radiation Coupled into Fiber

The highest level of interference is experienced when the fiber is surrounded by a carbonaceous char whose emissivity is close to 1, as for a blackbody radiator. The wavelength corresponding to peak blackbody emission λ_{peak} may be calculated in accordance with the Wien’s displacement law:

$$\lambda_{peak} = \frac{b}{T + 273.16}, \quad (2)$$

where b is the Wien’s displacement constant of $2.898 \times 10^{-3} \text{ m} \cdot \text{K}$, and T is the temperature [°C]. Under fire conditions, the peak wavelengths are 2700–2277 nm. In the spectral range relevant for fiber optic transmissions and testing, i.e. 800–1650 nm, spectral density of blackbody thermal radiation rises with an increase in wavelength.

However, spectra of radiation emitted by char and delivered by the optical fiber to the spectrum analyzer differ from the theoretical curves – see Fig. 21. In particular, the fall of power at wavelengths close to 1390 nm results from high fiber attenuation at three combined OH absorption peaks (Table 1).

Attenuation at 1390 nm, as tested by Rose and Bruno [38] in a single-mode fiber similar to SMF-1528, was tem-

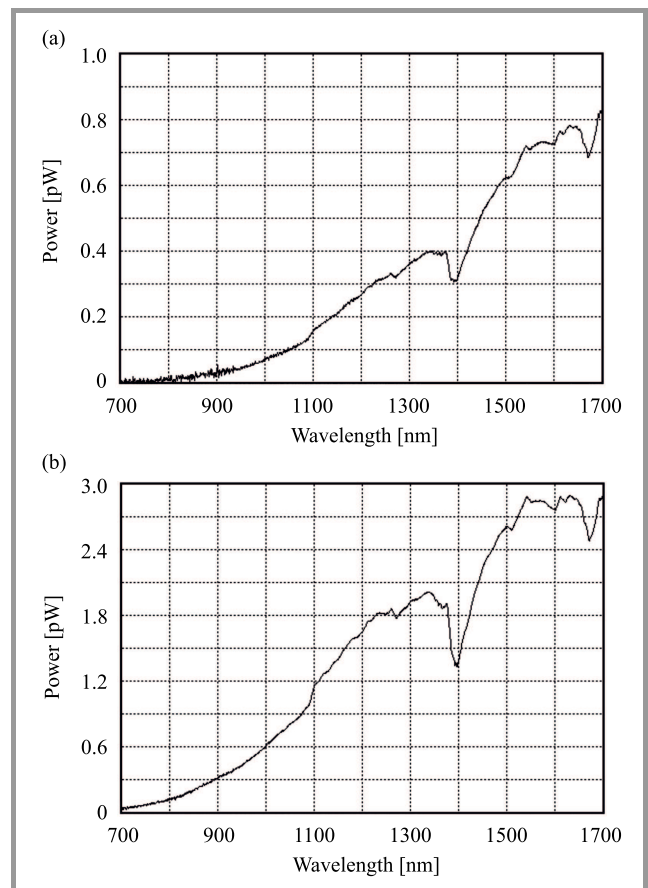


Fig. 21. Spectra of radiation emitted by carbon char left from pyrolysis of acrylate polymer the fiber-tight buffer was made of, at 800°C (a) and 1000°C (b), collected at the end of a 2.5 m long 50/125 μm multimode fiber, of which approx. 0.3 m was placed in the hot zone of the furnace.

perature-dependent, reaching 8.5 dB/m at 950°C. However, the fall to 3.5 dB/m at 1000°C reported there appears to be caused by the “drying” of fiber [39] – see Subsection 8.3. In our experiment, the estimated attenuation of a multi-mode fiber [38] for 1390 nm was 6.3 dB/m at 800°C and 7.6 dB/m at 1000°C.

8.3. Intensity of Thermal Radiation vs. Time

The power of both kinds of thermal radiation described above varies considerably with the duration of heating, but in opposite ways.

The heating of a bare fused silica fiber to 800°C or more decomposes hydroxyl (OH) ions, producing hydrogen which diffuses through the fiber cladding [39], and the OH-related emission peaks fade gradually [1], [2]. This effect reduces optical interference and is beneficial in high temperature applications of optical fibers. In the experiments we conducted, the heating of both single-mode and multi-mode telecom fibers with pure silica cladding to 1000°C for 1 h (the approximate duration of such an exposure during a DIN 4102-12 fire test) decreased the amplitude of emission peaks by approx. 50%. In a depressed-clad single-mode fiber (Lycom SM-DC) with fluorine-doped inner cladding this process was slower, with the half-fading time equaling approximately 2 h. During the tests performed by Honda *et al.* [2] on a pure silica core fiber with 25 µm thick fluorine-doped cladding (compared to approx. 5 µm in the SM-DC fiber), the half-fading time at 1000°C was about 80 h. A chemical reaction between fluorine and hydrogen, slowing down the latter’s diffusion through the layer of fluorine-doped silica glass, is a potential explanation here.

Such thermally induced fading of fiber incandescence has been reported recently [1], [2], [38] for several types of fused silica fibers. Additionally, it was documented that specialty fibers with high OH content emit considerably stronger radiation than low-OH telecom fibers [1].

Crystallization and the related cracking of fused silica at 900–1000°C progress steadily with time, and the power of coupled radiation rises slowly, starting from a very low initial value. This effect was reported i.e. in [2], [38], [40], but without any explanations of the physical phenomena behind the coupling of external radiation to the fiber.

In addition to slow degradation after heating for 1 h (and more), the authors observed several sudden and rapid failures (within 30 seconds) of fiber samples at temperatures over 900°C. The samples began to emit powerful radiation and lost their optical continuity. The damaged sections were short (approx. 2 cm), had milky-white appearance, and scattered away all light launched into the fiber. This type of fiber failure occurred only in some cases, when multiple samples were tested in the same conditions. Localized runaway crystallization reaching the fiber core, potentially initiated by surface contamination, is the presumed cause here.

Consequently, the radiation spectrum evolves during the heating (and exposure to gamma radiation, if present) –

the share of broadband thermal radiation coupled from the outside increases. This is proved by the spectra measured in our experiments (Figs. 11–12, 16–19) and in [2], [38].

8.4. Radiation Intensity vs. Fiber Type

Power of radiation emitted by a unit of the fiber’s length depends on the following:

- attenuation caused by OH content,
- fiber ability to collect radiation emitted omnidirectionally.

Telecom-grade single-mode fibers have a lower core diameter and numerical aperture (NA) than their multimode counterparts (OM2–OM5): 7–9 µm vs. 50 µm and approx. 0.14 vs. 0.20, respectively. Single mode fibers are characterized by a low diameter of the light-guiding core, typically equaling 7–9 µm. The mode field diameter (MFD) listed in fiber data sheets is 20–40% larger due to diffraction effects, and increases with wavelength.

The power of thermal radiation generated in a unit of the fiber’s length is proportional to the core area, while the portion of this emission accepted by the fiber is proportional to the square of NA. Correspondingly, the difference between thermal emission exiting equal lengths of single-mode and multimode fibers with equal OH content is between 1:50 and 1:100 or 17–20 dB. A similar relationship may be assumed for the radiation coupled from the outside.

Values of thermal emission power measured by the authors with the InGaAs detector (sensitivity range of 800–1700 nm) are presented in Table 2. The results apply to a relatively short heating period lasting 1–2 h.

Table 2
Power of thermal radiation from 1 m samples
of fibers at 900°C

Type of fiber	Standard	Power of thermal radiation [dBm]
Single-mode – low water peak	G.652.D, G.657.A	–85 ... –82
Single-mode – with water peak	G.652.A	–71 ... –66
50/125 µm graded index multimode	OM2	–60 ... –52

8.5. Radiation Intensity vs. Fiber Length

radiation launched into the fiber is proportional to its length being heated, the power exiting this section of the fiber is not. This is due to the rise of fiber attenuation in the fire affected zone. Under such conditions, the power of thermal radiation rises initially with length, and later saturates when the power added in each imaginary fiber section just compensates for the losses caused by high attenuation.

For the maximum attenuation values defined in EN 50582 [19]: 1 dB/m (SM fiber) or 2 dB/m (MM fiber),

the saturation length is approx. 4 m and 2 m, respectively. While the high level of thermally induced attenuation in the vicinity of the 1390 nm emission peak may shorten saturation length to much less than 1 m in a multimode fiber, the authors have measured the total power of up to -37 dBm at 950°C during a fire test of a 32 m multimode cable. However, this particular OM2 fiber could be easily thermally damaged during other experiments.

9. Discussion

The influence of high temperatures on the optical properties of telecom fibers made of fused silica has been covered in relatively few research papers with most of these focusing on fibers used for sensing in harsh environments. We failed to find a single publication (other than a few cable fire test reports) dealing with the optical phenomena in fibers inside a fire-resistant cable under fire conditions.

Results of the majority of the experiments conducted, e.g. measurements of spectra and time dependence of thermal radiation, are confirmed by data published by other authors. Alas, most of such data [1], [2], [40] is related to specialty multimode fibers with a core diameter of 50–200 μm , often with a pure silica core and fluorine-doped cladding, and with OH content substantially higher than in telecom fibers. The only test data for single-mode and multimode fibers of the telecom type is presented in [1]. This paper also includes formulas and parameters concerning fiber “drying”, fading of OH thermal emission with time, as well as temperature dependence of the power of emission. However, article [40] a different activation energy: 1.13 eV vs. 0.76–0.86 eV.

The available data on defects caused by fiber heating and on the associated optical effects, such as light scattering and coupling into fiber, is scarce, but is generally in alignment with our findings. Fiber cracking after heating to 1000–1100 $^{\circ}\text{C}$ has been reported recently [1], [39], while Honda *et al.* [2] and Shikama *et al.* [40] detected the coupling of external radiation emitted by a hot ceramic tube of the oven, the intensity of which rose with the duration of exposure to high temperature (up to 100 h at 1000 $^{\circ}\text{C}$) and with the dose of γ radiation received by the fiber during the test. Despite the fact that the type of fiber tested was different: 200/250 μm [2] or 125 μm [40] step-index multimode with a pure silica core and fluorine-doped cladding, the radiation spectra were essentially identical to ours. However, intensity of the main radiation peak at 1390 nm rose steadily with time (no fading), and the authors of paper [2] suggested that γ irradiation damage was the only source of light-coupling defects.

This hypothesis is contradicted by measurements of emission spectra from a fiber heated without γ irradiation, published in [2], [40], where emergence of broadband background radiation is visible - although it was much stronger with γ irradiation.

The literature and the standards fail to address the detrimental effects that optical interference emerging in hot fibers

may have on their use as transmission medium, and during attenuation testing.

The key issues related to fibers used in fire resistant cables are:

- high levels of thermal radiation in multimode fibers, up to -36 dBm from a 33 m long fiber at 1000 $^{\circ}\text{C}$,
- fast deterioration of the fiber, occurring after 1–2 h at 1000 $^{\circ}\text{C}$,
- sudden fiber failures above 900 $^{\circ}\text{C}$.

9.1. Plans Concerning Further Work

The nature of defects responsible for optical and mechanical deterioration of fibers at high temperatures deserves further research that will focus on explaining rapid fiber failures that occurred in our experiments. Cracks initiate a fiber break even with moderate cable bends (see Fig. 4). The cable will fail faster during the fire test and will be given a lower certification rating.

In the future, we plan to test fibers in a typical loose tube, as well as fibers encased in a hermetic metallic tube to prevent ingress of air and simulate real conditions in a cable subjected to fire. A detailed inspection of rapidly damaged samples will be performed as well to explain the failure mechanism and to determine whether it may affect a fiber in a cable.

After documenting, in this paper, the effects of high temperatures on fused silica fibers, we plan to perform a separate study focusing on means capable of remedying this type of interference. Proper selection of wavelength and the photodetector, together with optional use of an optical bandpass filter, may eliminate the problem during transmissions. For attenuation measurements of fibers during fire tests, a cyclic procedure of automatic background calibration and compensation was found to be effective; filtering offers the extra improvement needed. A switch of the wavelength at which attenuation changes of OM1–OM5 multimode fibers are measured from 1300 nm (stipulated in most standards) to 850 nm is strongly recommended in order to reflect all current applications.

10. Conclusions

The experiments performed on standard telecom fibers confirmed that:

- fused silica fibers heated to temperatures typical of fires (800–1000 $^{\circ}\text{C}$) may emit thermal radiation that is strong enough to disrupt attenuation measurements and data transmission;
- intensity of this radiation depends on the diameter of the fiber core and OH content, being the strongest in multimode fibers and negligible in OH-free (“low water peak”) single mode fibers;

- radiation originating from OH-contaminated fibers is concentrated in OH absorption bands, mostly 1383/1393/1407 nm, and can be blocked by band-pass filters;
- prolonged heating of pristine fibers to 800–1000°C results in a gradual escape of water, in a decrease of the fiber's thermal emission (beneficial), in the creation of defects causing light scattering, in coupling of thermal radiation from hot surroundings into the fiber and in its mechanical deterioration by cracking (harmful);
- the change of fiber attenuation resulting from heating alone is below 0.1 dB/m;
- several fiber samples failed rapidly when heated above 900°C, losing their optical continuity due to potential surface contamination.

References

- [1] D. Homa, G. Pickrell, and A. Wang, "Investigation of high temperature silica based fiber optic materials", DOE Award No. DE-FE0027891, *Virginia Polytechnic Institute & State University*, 2018 [Online]. Available: <https://www.osti.gov/servlets/purl/1489125>
- [2] A. Honda, K. Toh, S. Nagata, B. Tsuchiya, and T. Shikama, "Effect of temperature and irradiation on fused silica optical fiber for temperature measurement", *J. of Nuclear Materials*, vol. 367, pp. 1117–1121, 2007 (DOI: 10.1016/j.jnucmat.2007.03.193).
- [3] Standard EN-IEC 60793-1-40, "Optical fibres – Part 1–40: Attenuation measurement methods" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/a4ce5f5b-006b-4ad2-1c1-9dddbb796f22/en-iec-60793-1-40-2019>
- [4] Standard EN 60793-1-46, "Optical fibres – Part 1–46: Measurement methods and test procedures – Monitoring of changes in optical transmittance" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/dd4c92a8-dba8-498c-84a4-412ff1cc3d9a/en-60793-1-46-2002>
- [5] Standard IEC TR 62222, "Fire performance of communication cables installed in buildings" [Online]. Available: <https://standards.iteh.ai/catalog/standards/iec/53462b01-8540-4799-8986-57812ff68c23f/iec-tr-62222-2012>
- [6] Standard EN 60332-1-2, "Tests on electric and optical fibre cables under fire conditions – Part 1–2: Test for vertical flame propagation for a single insulated wire or cable – Procedure for 1 kW pre-mixed flame" [Online]. Available: <http://bit.ly/plnGQV>
- [7] Standard IEC 60332-3-10, "Tests on electric and optical fibre cables under fire conditions – Part 3–10: Test for vertical flame spread of vertically-mounted bunched wires or cables – Apparatus" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/6e572ecb-4b4c-4f6a-8478-27c773b76c8f/en-iec-60332-3-10-2018>
- [8] Standard IEC 60332-3-24, "Tests on electric and optical fibre cables under fire conditions – Part 3–24: Test for vertical flame spread of vertically-mounted bunched wires or cables – Category C" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/0fa90869-905f-46a2-9f58-40c8ea2647c4/en-iec-60332-3-24-2018>
- [9] Standard IEC 60332-3-25, "Tests on electric and optical fibre cables under fire conditions – Part 3–25: Test for vertical flame spread of vertically-mounted bunched wires or cables – Category D" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/ffcbaf8-6378-4298-868f-6267e848e20c/en-iec-60332-3-25-2018>
- [10] Standard EN 50399, "Common test methods for cables under fire conditions. Heat release and smoke production measurement on cables during flame spread test. Test apparatus, procedures, results" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/aec73708-d180-4cf3-b532-fb0850ed0705/pren-50399>
- [11] Standard IEC 61034-1, "Measurement of smoke density of cables burning under defined conditions – Part 1: Test apparatus" [Online]. Available: https://global.ihs.com/doc_detail.cfm?document_name=IEC%2061034%2D1&item_s_key=00134074
- [12] Standard IEC 61034-2, "Measurement of smoke density of cables burning under defined conditions – Part 2: Test procedure and requirements" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/e39b159f-d63f-40a4-835b-df2ec023a31a/en-61034-2-2005-a2-2020>
- [13] Standard IEC 60754-1, "Test on gases evolved during combustion of materials from cables – Part 1: Determination of the halogen acid gas content" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/fe208672-6841-4787-be72-f420fa4a3b5a/en-60754-1-2014>
- [14] Standard IEC 60754-2, "Test on gases evolved during combustion of materials from cables – Part 2: Determination of acidity (by pH measurement) and conductivity" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/6f7e97c5-a648-480f-838a-c0be9ab5fa8e/en-60754-2-2014-a1-2020>
- [15] Standard IEC 60754-3, "Test on gases evolved during combustion of materials from cables – Part 3: Measurement of low level of halogen content by ion chromatography" [Online]. Available: <https://standards.iteh.ai/catalog/standards/iec/2bf8bd77-cea9-44a5-8079-7cbe4f35f8b/iec-60754-3-2018>
- [16] Standard IEC 60331-1, "Tests for electric cables under fire conditions – Circuit integrity – Part 1: Test method for fire with shock at a temperature of at least 830°C for cables of rated voltage up to and including 0,6/1,0 kV and with an overall diameter exceeding 20 mm" [Online]. Available: <https://standards.iteh.ai/catalog/standards/iec/ffa2c2dd-6239-4799-b3a4-1da3312d4c40/iec-60331-1-2018>
- [17] Standard IEC 60331-2, "Tests for electric cables under fire conditions – Circuit integrity – Part 2: Test method for fire with shock at a temperature of at least 830°C for cables of rated voltage up to and including 0,6/1,0 kV and with an overall diameter not exceeding 20 mm" [Online]. Available: <https://standards.iteh.ai/catalog/standards/iec/2740c2d9-7e8d-43ba-aac0-ef2ffe17fd2c/iec-60331-2-2018>
- [18] Standard IEC 60331-3, "Tests for electric cables under fire conditions - Circuit integrity - Part 3: Test method for fire with shock at a temperature of at least 830°C for cables of rated voltage up to and including 0,6/1,0 kV tested in a metal enclosure" [Online]. Available: <https://standards.iteh.ai/catalog/standards/iec/f76a006a-8ae7-4588-adb1-ebd9c419744a/iec-60331-3-2018>
- [19] Standard EN 50200, "Method of test for resistance to fire of unprotected small cables for use in emergency circuits" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/873a9c45-a35b-4ec0-b0d3-ab1fcc792af4/en-50200-2015>
- [20] Standard EN 50575, "Power, control and communication cables. Cables for general applications in construction works subject to reaction to fire requirements" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/b8675c9d-b3b4-4a46-ae9e-7207aca441cb/en-50575-2014>
- [21] Standard DIN 4102-12, "Fire behaviour of building materials and elements – Part 12: Fire resistance of electric cable systems required to maintain circuit integrity – Requirements and testing" [Online]. Available: <https://standards.globalspec.com/std/365477/din-4102-12>
- [22] Standard EN 50582, "Procedure to assess the circuit integrity of optical fibres in a cable under resistance to fire testing" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/7403ddfd-5ea6-4eb0-83bd-feb5fff0e3f2/en-50582-2016>
- [23] Arpita Mitra, Inna Kouzmina, and Maritza Lopez, "Thermal stability of the CPC fiber coating system", *Corning White Paper*, vol. 4250, 2010 [Online]. Available: <https://www.corning.com/microsites/coc/oem/documents/specialty-fiber/WP4250-Thermal-Stability-of-the-CPC-Fiber-Coating-System.pdf>
- [24] A. A. Stolov, D. A. Simoff, and J. Li, "Thermal stability of specialty optical fibers", *J. Lightwave Technology*, vol. 26, no. 20, pp. 3443–3451, 2008 (DOI: 10.1109/JLT.2008.925698).
- [25] "Reliability test report for SR15-9/125-ACL (SM Fiber with high temperature resistant acrylate coating)", *Fujikura*, 2013 [Online]. Available: http://www.fujikura.co.jp/eng/products/optical/appliedoptics/02/_icsFiles/afiedfile/2013/02/04/t4d4013.pdf

- [26] Specification KA2001R1, "Technoflame FOC-2-SLT-HFFR E30/PH120 4x50/125 OM2 fibre optic safety cables", Technokabel S.A., 2020.
- [27] FiberTek, "Armoured(SWA) fire resistant LSZH loose tube fiber optic cable – IEC 60331 FTSE-FLTFMAPSZ(FR): Steel CSM, mica wrapped jelly filled tube, aluminium moisture barrier, PE inner sheath, steel wire armoured, LSZH outer sheath", Rev0.0 (Mar17 C-IN1703111), 2017 [Online]. Available: [https://www.vectorinfotech.com/FileStore/Full/_Product92560_FTSE-FLTFMAPSZ\(FR\)_Rev0.0\(Mar17\).pdf](https://www.vectorinfotech.com/FileStore/Full/_Product92560_FTSE-FLTFMAPSZ(FR)_Rev0.0(Mar17).pdf)
- [28] Standard IEC 60793-2-10:2019: Optical fibres – Part 2–10: Product specifications – Sectional specification for category A1 multimode fibres [Online]. Available: <https://standards.iteh.ai/catalog/standards/iec/b85d8886-7454-46ed-b3e5-a2cc5ef12952/iec-60793-2-10-2019>
- [29] Recommendation ITU-T G.652, "Characteristics of a single-mode optical fibre and cable", 2016 [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.652-201611-I!!PDF-E&type=items
- [30] Recommendation ITU-T G.657, "Characteristics of a bending-loss insensitive single-mode optical fibre and cable", 2016 [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.657-201611-I!!PDF-E&type=items
- [31] Standard EN-IEC 60793-2-50, "Optical fibres – Part 2–50: Product specifications – Sectional specification for class B single-mode fibres" [Online]. Available: <https://standards.iteh.ai/catalog/standards/clc/7ddb02c1-80c1-440c-a200-c95c0ca056fd/en-iec-60793-2-50-2019>
- [32] Standard ISO/IEC 11801, "Information technology – Generic cabling for customer premises", 2017 [Online]. Available: <https://www.iso.org/standard/66182.html>
- [33] Standard ISO 834-1:1999, "Fire-resistance tests – Elements of building construction – Part 1: General requirements" [Online]. Available: <https://www.iso.org/standard/2576.html>
- [34] OFS Fitel, "AllWave optical Fiber – Zero Water Peak: The industry's first zero water peak single-mode fiber for reliable full-spectrum performance", 2017 [Online]. Available: <https://www.ofsoptics.com/wp-content/uploads/AllWave-117-web-7.pdf>
- [35] OFS Fitel, "AllWave FLEX+ Fiber – Zero Water Peak: Optimized bend performance and reliable low loss transmission for in-building, central office and cabinet applications", 2016 [Online]. Available: <https://fiber-optic-catalog.ofsoptics.com/documents/pdf/AllWave-FLEX-PLUS-144-web.pdf>
- [36] M. Pellow-Jarman and M. Hetem, "Comparison of the thermal degradation products of poly(butylene terephthalate) and flame retardant poly(butylene terephthalate) formulations using a pyrolysis FTIR cell", *Polymer Degradation and Stability*, vol. 47, no. 3, pp. 413–421, 1995 (DOI: 10.1016/0141-3910(95)00006-2).
- [37] P. K. Johnston, E. Doyle, and R. A. Orzel, "Acrylics thermal decomposition products and toxicity", *J. of the American College of Toxicology*, vol. 7, no. 2, pp. 139–200 1988 (DOI: 10.3109/10915818809014519).
- [38] A. H. Rose and T. J. Bruno, "The observation of OH in annealed optical fiber", *J. Non-Cryst. Solids*, vol. 231, no. 3, pp. 280–285, 1998 (DOI: 10.1016/S0022-3093(98)00676-0).
- [39] A. H. Rose, "Devitrification in Annealed Optical Fiber", *J. of Light-wave Technol.*, vol. 15, no. 5, pp. 808–814, 1997 (DOI: 10.1109/50.580819).
- [40] T. Shikama, K. Toh, S. Nagata, B. Tsuchiya, and Y. Ohno, "Temperature measurement by thermal luminescence of partially replaced core optical fiber", *J. of Nuclear Materials* vol. 386–388, pp. 1023–1026, 2009 (DOI: 10.1016/j.jnucmat.2008.12.204).
- [41] OFS Fitel, "50 μ m graded-index OM2 – bend-insensitive multimode optical fiber", 2018 [Online]. Available: <https://fiber-optic-catalog.ofsoptics.com/documents/pdf/Graded-Index-50-Fibre-Data-Sheet.pdf>
- [42] Fujikura, "Fujikura FutureGuide-MM50 Multi Mode Fiber (ITU-T G.651): Multi mode fiber with 50 μ m of core diameter (ITU-T G.651) short-reach optical transmission for LAN in offices and premises", 2020 [Online]. Available: https://www.fujikura.co.jp/products/optical/opticalfibers/01/2044175_11306.html

- [43] O. Humbach, H. Fabian, U. Grzesik, U. Haken, and W. Heitmann, "Analysis of OH absorption bands in synthetic silica", *J. of Non-Crystalline Solids*, vol. 203, pp. 19–26, 1996 (DOI: 10.1016/0022-3093(96)00329-8).



Krzysztof Borzycki received his M.Sc. in Electrical Engineering from Warsaw University of Technology, Warsaw, Poland in 1982, and Ph.D. degree in Communications Engineering from the National Institute of Telecommunications (NIT), Warsaw, Poland in 2006. He has been NIT since 1982, except for the time spent on developing DWDM solutions at the Ericsson AB R&D Center in Stockholm, Sweden, in 2001–2002. He is currently an Assistant Professor at the NIT Central Chamber for Telecommunication Metrology. His areas of interest include fiber access networks (FTTx), testing and standardization of fiber cables and passive components, monitoring of fiber and copper cable networks, security of optical networks and effects of high temperatures in fused silica fibers. He also works as an instructor and lecturer specializing in fiber optics, fiber testing and splicing. Dr. Borzycki has participated in multiple European research programs, including COST-270, COST-299, COST TD1001 and NEMO, and is a member of Cables and Fiber Optics Work Groups of the Polish National Standardization Committee (PKN).

<https://orcid.org/0000-0001-6066-6590>

E-mail: k.borzycki@il-pib.pl

National Institute of Telecommunications
Szachowa 1
04-894 Warsaw, Poland



Marek Jaworski received his Ph.D. degree from the National Institute of Telecommunications (NIT), Warsaw, Poland, in 2001. He is currently an Assistant Professor at the NIT Central Chamber for Telecommunication Metrology. He has been with NIT since 1982, working on modeling and design of optical fiber transmission systems, measurement methods, and test equipment for optical networks. His current research interests include numerical simulations of telecommunication systems, advanced modulation formats, and nonlinear photonics.

<https://orcid.org/0000-0002-6742-4874>

E-mail: m.jaworski@il-pib.pl

National Institute of Telecommunications
Szachowa 1
04-894 Warsaw, Poland



Tomasz Kossek received his M.Sc. in Optoelectronics and a Ph.D. degree from Warsaw University of Technology, Warsaw, Poland, in 1996 and 2002, respectively. He is currently an Assistant Professor at the National Institute of Telecommunications Poland, Warsaw. His current research interests include optoelectronic

measurements and their calibration, laser physics, and optical communication.

 <https://orcid.org/0000-0001-6670-2871>

E-mail: t.kossek@il-pib.pl

National Institute of Telecommunications

Szachowa 1

04-894 Warsaw, Poland

Speech-Based Vehicle Movement Control Solution

Gurpreet Kaur¹, Mohit Srivastava², and Amod Kumar³

¹ UIET, Panjab University, Chandigarh, India, ² CEC Landran, India, ³ NITTTR, Chandigarh, India

<https://doi.org/10.26636/jiit.2021.149820>

Abstract—The article describes a speech-based robotic prototype designed to aid the movement of elderly or handicapped individuals. Mel frequency cepstral coefficients (MFCC) are used for the extraction of speech features and a deep belief network (DBN) is trained for the recognition of commands. The prototype was tested in a real-world environment and achieved an accuracy rate of 87.4%.

Keywords—*deep belief networks, mel frequency cepstral coefficients, speech recognition.*

1. Introduction

Speech signal processing is an important research field due to the wide variety of applications it may be useful for. It may be used in banking systems, forensics, hospitals, the military, or in day-to-day activities. Regardless of the domain, people desire intelligent and smart systems facilitating their efforts. Many advancements in this field have led to the development of such smart systems as Apple's Siri, Google Assistant, Amazon Alexa, etc. [1], [2]. Voice operated vehicles may also be used to aid the movement of elderly or handicapped individuals. Unfortunately, voice operated wheelchairs are not quite popular yet, because of numerous challenges in speech recognition [3]–[5]. Every single person has their own way of speaking. Their voice may be characterized by a different accent and variations may exist in the use of language, emotions and the environment. All these parameters exert a great impact on speech recognition.

Considerable progress has been made in various domains, such as education or entertainment. However, much less effort has been put into harnessing speed recognition achievements in the process of physical rehabilitation of patients. All existing feature extraction methods have showed a good recognition rate only in a clean environment. Results obtained in noisy environments still require considerable improvement [6]. The recognition rate suffers when the system is used in a real-world scenario. There is a big difference in the operation of the speaker and speech recognition mechanism when it relies on a stored database and on a real-world database. All existing speaker and speech recognition systems perform efficiently when operating based on a stored database. But in real-world applications, their performance is affected adversely because of the wide variety of speak-

ing styles and background noise [7]. In order to deal with all these challenges, artificial intelligence seems to be the best proposal, as it allows to design systems capable of dealing with variations stemming from speaking styles, gender, language and background noise. Trained neural networks are capable of handling all these details. The problem of speech recognition and speaker recognition has been dealt with by many researchers [8]. Technology-related progress observed thus far aims towards implementing numerous speech recognition systems. Speech recognition depends directly upon feature extraction and classification methods [9]. Some authors have already compared the feature extraction methods known [10]. According to the results of their analyses, MFCC is a solution that is best suited for the task at hand. Classification-related approaches may be template-based [11], stochastic-based [12] or artificial neural network-based systems [13]–[15]. In template-based techniques, voice samples were stored in a database and the comparison was made between the database and the uttered word. There were many popular techniques, such as distance calculation with the use of the dynamic time warping algorithm. However, systems of this type were not capable of providing correct results in real-world applications and, hence, their performance degraded in noisy environments. Later, stochastic-based models were popular in speech recognition. These were based on probability models and could handle incomplete and uncertain input data. Hidden Markov models were best for dealing with variable input data. In the 1970s, the US Department of Defense sponsored numerous projects focusing on stochastic models, for instance Dragon. The technique was widely accepted all over the world before the emergence of artificial intelligence. Artificial intelligence is a knowledge-based system [16] that started to gain in importance after implementation of numerous neural network algorithms. Previously, neural networks were based on a few hidden layers. They were not capable of dealing with the variability of input data. Nowadays, however, deep learning networks [17] are effectively used for this particular purpose. There are many learning-based algorithms used by researchers, for instance convolutional neural network (CNN) [18], multilayer perceptron (MLP), probabilistic neural network (PNN) [19]–[24]. However, the performance of all those systems deteriorates in noisy environments. Learning with the use of noisy data helps

improve the system. Many techniques have been implemented by researchers to make the systems easy to use for humans.

2. System Concept

The proposed system is divided into a software and a hardware module. Speech samples are captured with the use of a microphone. Then, all samples are preprocessed in preparation for applying the feature extraction technique. The MFCC feature extraction method is deployed and the extracted features are then used for training a deep neural network (DNN). DNN is emerging field in speech recognition field and its performance exceed that of other methods. In the DNN approach, a deep belief network (DBN) is used. Many learning techniques are available. In this type of work, unsupervised learning contrastive divergence (CD) is used. Matlab is used to implement the application. In the hardware module, a prototype is made that consists of an RF receiver, a microcontroller, a motor driver and a DC gear motor. The command from Matlab software is received using an RF data modem, via the serial communication protocol. The system works in the 2.4 GHz band its rate is adjustable between 9600 and 115200 bps for direct interfacing with the MCU.

2.1. MFCC Features

Speech signals are converted into the frequency domain using FFT. The FFT output contains a lot of data that is not required. In order to calculate the energy level at each frequency, a mel scale analysis is performed using mel filters. Then energy is calculated and after that logarithmic of filter bank energies is taken. This operation is conducted in order to match the features closer to human hearing. Finally, DCT of the log filter bank energies is taken to decorrelate the overlapped values. Only the first 13 coefficients are selected, known as MFCC features. This is because higher feature numbers degrade the recognition-related accuracy of the system [25]. Those features do not carry any speaker and speech-related information. Mathematically, this may be explained using the vocal tract system. The vocal tract articulation equivalent filter is shown in Eq. (1):

$$s(\Omega) = g(\Omega)h(\Omega) . \quad (1)$$

The equivalent logarithm of $s(\Omega)$ is:

$$\log |s(\Omega)| = \log |g(\Omega)| + \log |h(\Omega)| . \quad (2)$$

In this method, the logarithm of the spectrum is taken and then its inverse Fourier transform is found. The cepstrum $C(\partial)$, or cepstral coefficients, is the inverse Fourier transform of $\log |s(\Omega)|$:

$$C(\partial) = F^{-1} \log |s(\Omega)| = F^{-1} \log |g(\Omega)| + F^{-1} \log |h(\Omega)| . \quad (3)$$

It represents the vocal tract's parameters corresponding to the phonemes of sound.

2.2. Deep Neural Network

A deep neural network (DNN) has many hidden layers. The inspiration is taken from the visual cortex (part of brain). Information is processed in the sequence of regions. So, a neural network may be modeled as a multilayer network consisting of low to high level features. Training is the major issue in DNNs, as optimization is a more difficult step. Performance of the system degrades because of under fitting and over fitting. Under fitting is caused by a vanishing gradient problem and over fitting stems from high variance and low bias situations. The unsupervised pre-training approach may be the solution to be used while training DNNs. Training is performed one layer at a time, from the first to the second to the last. Features are fed to the first hidden layer, then the second layer takes the combinations of features from the first layer. This process is repeated until the last layer is reached. Once all layers have been pre-trained, supervised training is performed for the entire network. This stage is known as fine tuning. This means that deep training can be performed in two steps: pre-training the network and fine tuning, using supervised training. Restricted Boltzmann machines (RBMs) are the building blocks of deep networks. It is an unsupervised, undirected graphical model. It is called restricted, because there is no link between the units of each layer, as shown in Fig. 1.

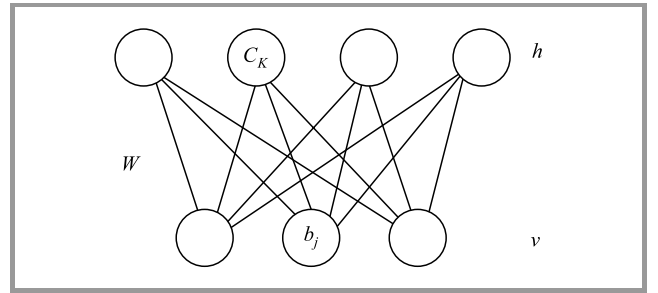


Fig. 1. Undirected graphical RBM model.

The model is of the energy-based variety and its distribution is given by:

$$P(v, h) = \frac{1}{Z} e^{(-E(v, h))} . \quad (4)$$

$E(v, h)$ is the energy function and is defined as:

$$E(v, h) = -b^T v - c^T h - v^T W h , \quad (5)$$

where b, c are vectors and W is the matrix – these are the parameters of the model. The partition function is defined as:

$$Z = \sum_v \sum_h e^{(-E(v, h))} , \quad (6)$$

$$E(v, h) = -b^T v - c^T h - v^T W h \\ = -\sum_k b_k v_k - \sum_j c_j h_j - \sum_j \sum_k c_j h_j W_{jk} . \quad (7)$$

RBM is trained on binary data. Conditional distribution from the joint distribution is:

$$P\left(\frac{h}{v}\right) = \frac{P(h, v)}{P(v)} = \frac{1}{P(v)} \frac{1}{Z} e^{b^T v + c^T h + v^T W h} = \frac{1}{Z'} e^{c^T h + v^T W h} . \quad (8)$$

This expression can be written in a scalar form as:

$$P\left(\frac{h}{v}\right) = \frac{1}{Z'} e^{\sum_{j=1}^n c_j h_j + \sum_{j=1}^n v^T W_{:j} h_j} = (1/Z') \prod_{j=1}^n e^{c_j h_j + v^T W_{:j} h_j} . \quad (9)$$

The distribution over the individual binary h_j is given as:

$$P(h_j = \frac{1}{v}) = \frac{P(h_j = 1, v)}{P(h_j = 0, v) + P(h_j = 1, v)} = \frac{e^{c_j + v^T W_{:j}}}{e^{\{0\}} + e^{c_j + v^T W_{:j}}} = \text{sigmoid}(c_j + v^T W_{:j}) , \quad (10)$$

$$P\left(\frac{h}{v}\right) = \prod_{j=1}^n \text{sigmoid}(c_j + v^T W_{:j}) , \quad (11)$$

$$P\left(\frac{v}{h}\right) = \prod_{i=1}^d \text{sigmoid}(b_i + W_{i:} h) . \quad (12)$$

Parameters of the models are learnt by taking the log likelihood given as:

$$\begin{aligned} l(W, b, c) &= \sum_{t=1}^n \log p(v^{(t)}) = \sum_{t=1}^n \log \sum_h p(v^{(t)}, h) \\ &= \sum_{t=1}^n \log \sum_h e^{-E(v^{(t)}, h)} - n \log Z \\ &= \sum_{t=1}^n \log \sum_h e^{-E(v^{(t)}, h)} - n \log \sum_{v, h} e^{-E(v, h)} . \end{aligned} \quad (13)$$

For maximizing the likelihood, derivative of the log likelihood is taken as:

$$\theta = b, c, W , \quad (14)$$

$$l(\theta) = \sum_{t=1}^n \log \sum_h e^{-E(v^{(t)}, h)} - n \log \sum_{v, h} e^{-E(v, h)} , \quad (15)$$

$$\begin{aligned} \nabla_{\theta}(\theta) &= \nabla_{\theta} \sum_{t=1}^n \log \sum_h e^{-E(v^{(t)}, h)} - n \nabla_{\theta} \log \sum_{v, h} e^{-E(v, h)} \\ &= \sum_{t=1}^n E_{p(h/v^{(t)})} [\nabla_{\theta} (-E(v^{(t)}, h))] - n E_{p(h/v)} [\nabla_{\theta} (-E(v, h))] . \end{aligned} \quad (16)$$

The second half of Eq. (16) i.e. the expectation of h given v , is difficult to learn. Therefore, the contrastive divergence algorithm is used, where the expectation is calculated by a point estimate using Gibb's sampling.

2.3. Deep Belief Network

In a deep belief network (DBN), RBMs are stacked together to form a multilayer network. A graphical model of DBN is shown in Fig. 2. In this model, there are three hidden layers with one input layer. Training is performed with the first layer using the data and then freezing the first layer's parameters. Then, the second layer is trained using the output of the first layer, constituting unsupervised input for the second layer. This process is repeated until the last layer is reached. The following steps are taken for calculating the distribution:

$$P(x) = \sum_{h^{(1)}} p(x, h^{(1)}) , \quad (17)$$

$$P(x, h^{(1)}) = p\left(\frac{x}{h^{(1)}}\right) \sum_{h^{(2)}} p(x, h^{(1)}) , \quad (18)$$

$$P(h^{(1)}, h^{(2)}) = p\left(\frac{h^{(1)}}{h^{(2)}}\right) \sum_{h^{(3)}} p(h^{(2)}, h^{(3)}) . \quad (19)$$

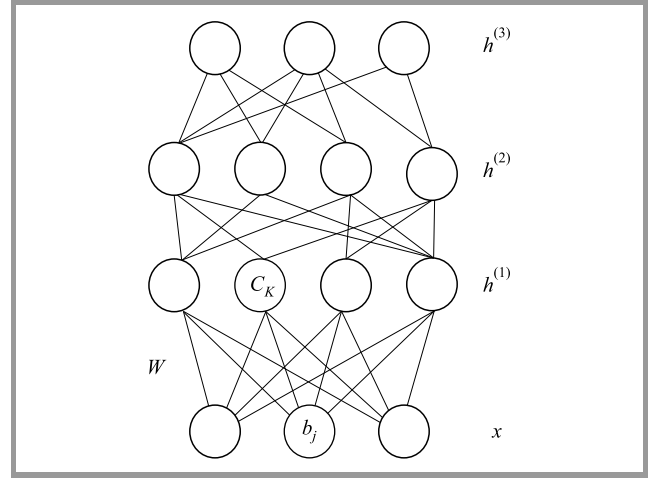


Fig. 2. Graphical model of DBN.

Next, layer-based training is performed for the DBN. Once all the layers have been pre-trained, supervised training is performed for the entire network. This is called fine tuning and the back propagation algorithm can be used to train the network. The output of DBN is, what is recognized who is speaking and what he or she is speaking, and a control signal is generated to validate the proposed technique.

2.4. Hardware Module

The hardware module was based on a 2.4 GHz RF transmitter receiver, a microcontroller (ATMega 8), a motor driver (L293D) and DC gear motors (Fig. 3). The recognized command word is sent from the computer to the hardware unit through RF data modem at 9600 bps to MCU.

The command from Matlab software is received using an RF data modem, using a serial communication protocol. The MCU interprets the commands received. The driver circuit is used to control the DC motors.

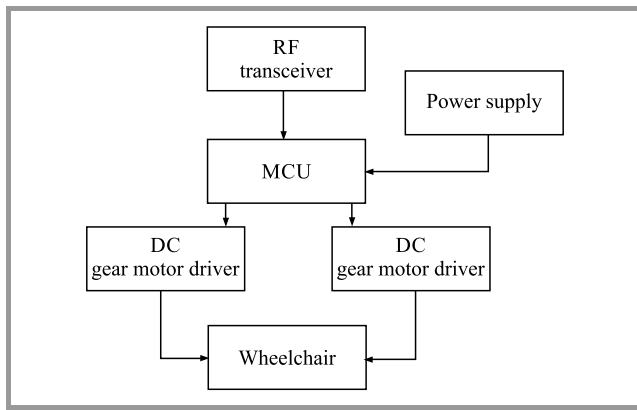


Fig. 3. Hardware of the prototype solution.

3. Results and Discussion

MFCC features are extracted for different words, such as backward, forward, left, right, and stop. Fifty users (25 males and 25 females) took part in the experiment and each of them provided 100 samples of each word. Figure 4 shows one of the examples of the training stage of the speaker-dependent speech recognition with extracted features. Training is performed by the speaker for the word right and the MFCC feature extraction method is used.

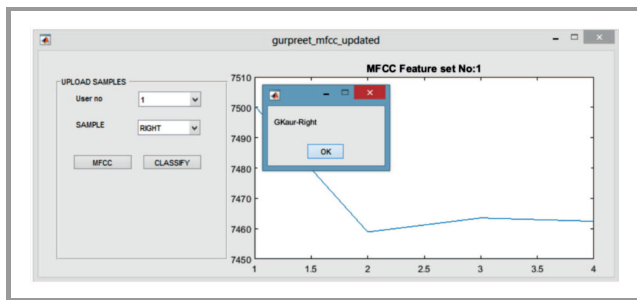


Fig. 4. MFCC extracted features.

The extracted MFCC features are fed to DBN network. In DBN, 13 nodes are taken for input and 13 nodes are taken for output. Two hidden layers h_1 and h_2 are used. Hidden layer 1 consists of 200 nodes and hidden layer 2 consists of 100 nodes. First pre-training is done with the RBM stacks and then fine tuning is performed with supervised learning method. The matrices used to calculate the accuracy are shown in Fig. 5.

In a real-world implementation, the recognized word is sent from the computer to the hardware unit through an RF data modem at 9600 bps to MCU. After this, serial port is read by the program and according to the command received and motors move accordingly. The prototype has been tested in different environments, such as a playground, an office, a cafeteria and also in a hospital as shown in Table 1.

Recognition accuracy was tested at peak hours (10–11 AM), when maximum levels of background noise are present. The average recognition accuracy achieved is 87.4%.

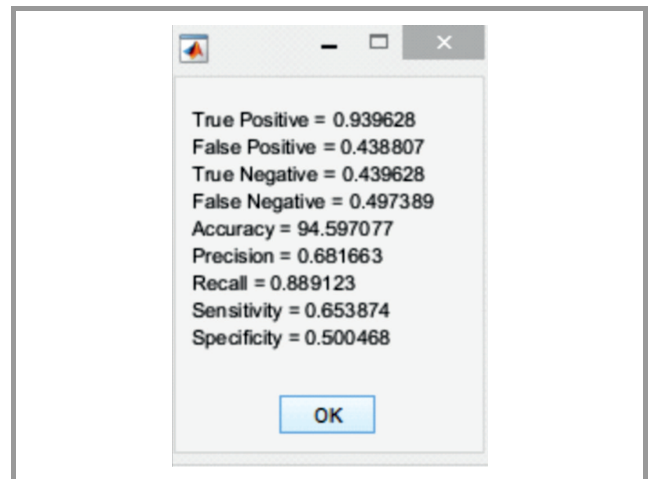


Fig. 5. Measuring matrices values.

Table 1
Accuracy (percentage-wise) in different environments

Command	Environment			
	Playground	Office	Cafeteria	Hospital
Backward	88.93	87.01	86.21	85.21
Forward	88.08	87.09	85.09	85.09
Left	89.17	88.07	86.87	85.17
Right	89.00	88.00	86.80	85.10
Stop	89.03	89.07	86.07	85.07

4. Conclusion

A speech-based system has been developed for controlling the movement of a wheelchair. Speech signals have been processed with the use of the MFCC-based feature extraction method and their classification was performed with the help of a DBN-based neural network technique. Five commands have been used to control the vehicle. The prototype has been tested in different environments, such as a playground, an office, a cafeteria and a hospital. The average word recognition accuracy achieved is 87.4%.

References

- [1] J. H. L. Hansen and T. Hasan, "Speaker recognition by machines and humans: a tutorial review", in *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 74–99, 2015 (DOI: 10.1109/MSP.2015.2462851).
- [2] D. R. Reddy, "Speech recognition by machine: A review", in *Proc. of the IEEE*, vol. 64, no. 4, 1976, pp. 501–531 (DOI: 10.1109/PROC.1976.10158).
- [3] M. Nishimori, T. Saitoh, and R. Konishi, "Voice controlled intelligent wheelchair", in *Proc. of the SICE Annual Conf.*, Takamatsu, Japan, 2007, pp. 336–340 (DOI: 10.1109/SICE.2007.4421003).
- [4] N. Peixoto, H. G. Nik, and H. Charkhkar, "Voice controlled wheelchairs: Fine control by humming", *Computer Methods and Programs in Biomedicine*, vol. 112, pp. 156–165, 2013 (DOI: 10.1016/j.cmpb.2013.06.009).
- [5] V. Partha Saradi and P. Kailasapathi, "Voice-based motion control of a robotic vehicle through visible light communication", *Computers and Electrical Engineer.*, vol. 76, pp. 154–167, 2019 (DOI: 10.1016/j.compeleceng.2019.03.011).

- [6] G. Kaur, M. Srivastava, and A. Kumar, "Integrated speaker and speech recognition for wheel chair movement using artificial intelligence", *Informatica*, vol. 42, pp. 587–594, 2018 (DOI:10.31449/inf.v42i4.2003).
- [7] S. Squartini, E. Principi, R. Rotili, and F. Piazza, "Environmental robust speech and speaker recognition through multi-channel histogram equalization", *Neurocomputing*, vol. 78, no. 1, pp. 111–120, 2012 (DOI:10.1016/j.neucom.2011.05.035).
- [8] S. Furui, "50 Years of progress in speech and speaker recognition", *ECTI Trans. on Computer and Information Technol.*, pp. 64–74, 2012 (DOI:10.37936/ecti-cit.200512.51834).
- [9] G. Kaur, M. Srivastava, and A. Kumar, "Implementation of text dependent speaker verification on Matlab", in *Proc. of 2nd Conf. on Recent Adv. in Engineer. and Comput. Sci. RA ECS*, Chandigarh, India, 2015 (DOI: 10.1109/RA ECS.2015.7453344).
- [10] G. Kaur, M. Srivastava, and A. Kumar, "Analysis of feature extraction methods for speaker dependent speech recognition", *Int. J. of Engineer. and Technol. Innovation*, vol. 7, pp. 78–88, 2017 [Online]. Available: <https://ojs.imeti.org/index.php/IJETI/article/view/382/395>
- [11] S. Narang and D. Gupta, "Speech feature extraction techniques: a review", *Int. J. of Computer Sci. and Mobile Comput.*, vol. 4, no. 3, pp. 107–114, 2015 [Online]. Available: <https://www.ijcsmc.com/docs/papers/March2015/V4I3201545.pdf>
- [12] D. Y. Huang, Z. Zhang, and S. S. Ge, "Speaker state classification based on fusion of asymmetric simple partial least squares (SIMPLS) and support vector machines", *Computer Speech Language*, vol. 28, no. 2, pp. 392–419, 2014 (DOI: 10.1016/j.csl.2013.06.002).
- [13] S. M. Siniscalchi, T. Svendsen, and C. H. Lee, "An artificial neural network approach to automatic speech processing", *Neurocomputing*, vol. 140, pp. 326–338, 2014 (DOI: 10.1016/j.neucom.2014.03.005).
- [14] N. S. Dey, R. Mohanty, and K. L. Chugh, "Speech and speaker recognition system using artificial neural networks and hidden Markov model", in *Proc. IEEE Int. Conf. on Communication System and Network Technology CSNT*, Rajkot, Gujarat, India, 2012, pp. 311–315 (DOI: 10.1109/CSNT.2012.221).
- [15] R. Makhijani and R. Gupta, "Isolated word speech recognition system using Dynamic Time Warping", *Int. J. of Engineering Sciences & Emerging Technologies*, vol. 6, no. 3, pp. 352–367, 2013 [Online]. Available: <https://www.ijeset.com/media/0002/9N13-IJESSET0603130-v6-iss3-352-363.pdf>
- [16] G. Dede and M. H. Sazli, "Speech recognition with artificial neural networks", *Digital Signal Process.: A Review Journal*, vol. 20, no. 3, pp. 763–768, 2010 (DOI: 10.1016/j.dsp.2009.10.004).
- [17] T. Nikoskinen, "From neural network to deep neural network", *Alto University School of Science*, pp. 1–27, 2015 [Online]. Available: <https://sal.aalto.fi/publications/pdf-files/enik15-public.pdf>
- [18] L. Moreno *et al.*, "On the use of deep feed forward neural networks for automatic language identification", *Computer Speech Language*, vol. 40, pp. 46–59, 2016 (DOI: 10.1016/j.csl.2016.03.001).
- [19] T. Alsmadi, H. A. Alissa, E. Trad, and K. Alsmadi, "Artificial intelligence for speech recognition based on neural networks", *J. of Signal and Information Process.*, vol. 6, no. 2, pp. 66–72, 2015 (DOI: 10.4236/jsip.2015.62006).
- [20] V. Mitra *et al.*, "Hybrid convolutional neural networks for articulatory and acoustic information-based speech recognition", *Speech Commun.*, vol. 89, pp. 103–112, 2017 (DOI: 10.1016/j.specom.2017.03.003).
- [21] M. Farahat and R. Halavati, "Noise robust speech recognition using Deep Belief Networks", *Int. J. of Comput. Intelligence and Applicat.*, vol. 15, no. 1, pp. 1–17, 2016 (DOI: 10.1142/S146902681650005X).
- [22] R. Sarikaya, G. E. Hinton, and A. Deoras, "Application of Deep Belief Networks for natural language understanding", *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, pp. 778–784, 2014 (DOI: 10.1109/TASLP.2014.2303296).
- [23] A. R. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic Modeling Using Deep Belief Networks", *IEEE Trans. on Audio, Speech and Language Process.*, vol. 20, no. 1, pp. 14–22, 2011 (DOI: 10.1109/TASL.2011.2109382).
- [24] X. Chen, X. Liu, Y. Wang, M. J. F. Gales, and P. C. Woodland, "Efficient training and evaluation of recurrent neural network language models for automatic speech recognition", *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 24, no. 11, pp. 2146–2157, 2016 (DOI: 10.1109/TASLP.2016.2598304).
- [25] R. Ajgou, S. Sbaa, S. Ghendir, and A. Chemsas, "An efficient approach for MFCC feature extraction for text independent speaker identification system", *Int. J. of Commun.*, vol. 9, pp. 114–122, 2015 [Online]. Available: <http://www.naun.org/main/NAUN/communications/2015/a382006-081.pdf>



Gurpreet Kaur is an Assistant Professor at the Department of Electronics and Communication Engineering at the University Institute of Engineering and Technology, Panjab University, Chandigarh, India. She received her B.Tech. (with Hons) in Electronics and Communication Engineering from Kurukshetra University,

Haryana in 2004, M.E. (with distinction) in Electronics and Communication from the University Institute of Engineering and Technology, Panjab University, Chandigarh in 2007, and Ph.D. in Electronics Engineering from IKG Punjab Technical University, Jalandhar in 2018. Her current research interests focus on speech processing and neural networks.

 <https://orcid.org/0000-0003-3735-0568>

E-mail: regs4gurpreet@yahoo.co.in

UIET

Panjab University
Chandigarh, India



Mohit Srivastava is a Professor at the Department of Electronics and Communication Engineering and R&D Dean at Chandigarh Engineering College, Landran, Mohali, Punjab, India. He received his B.Tech. in Electronics and Communication Engineering from Magadh University, Bodh Gaya, M.Tech. in Digital Electronics and Sys-

tems from K.N.I.T. Sultanpur and Ph.D. in Image Processing & Remote Sensing from Indian Institute of Technology Roorkee in 2000, 2008 and 2013 respectively. His current research interests focus on digital image and speech processing, remote sensing and their applications in land cover mapping, and communication systems.

 <https://orcid.org/0000-0002-4566-4279>

E-mail: mohitsrivastava.78@gmail.com

CEC Landran
India



Amod Kumar received his B.E. (Hons.) in Electrical and Electronics Engineering from Birla Institute of Technology and Science, Pilani (Raj.), M.E. in Electronics from Punjab University, Chandigarh and Ph.D. in Biomedical Signal Processing from IIT Delhi. He has approximately 38 years of experience in research and develop-

ment of different instruments used in process control, environmental monitoring, biomedical engineering and prosthetics. He worked as the Chief Scientist at the Central Scientific Instruments Organization (CSIO), Chandigarh, which is a constituent laboratory of CSIR. Currently he is working at NITTTR, Chandigarh, as Professor at the Electronics Department. His areas of interest focus on digital signal processing, image processing and soft computing.

 <https://orcid.org/0000-0003-1177-3191>

E-mail: csioamod@yahoo.com

NITTTR

Chandigarh, India

Tractography Methods in Preoperative Neurosurgical Planning

Mateusz Koryciński and Konrad A. Ciecierski

Department of Bioinformatics and Machine Recognition,
Research and Academic Computer Network, Warsaw, Poland

<https://doi.org/10.26636/jtit.2021.154521>

Abstract—Knowledge of the location of nerve tracts during the surgical preoperative planning stage and during the surgery itself may help neurosurgeons limit the risk of causing neurological deficits affecting the patient's essential abilities. Development of MRI techniques has helped profoundly with *in vivo* visualization of the brain's anatomy, enabling to obtain images within minutes. Different methodologies are relied upon to identify anatomical or functional details and to determine the movement of water molecules, thus allowing to track nerve fibers. However, precise determination of their location continues to be a labor-intensive task that requires the participation of highly-trained medical experts. With the development of computational methods, machine learning and artificial intelligence, many approaches have been proposed to automate and streamline that process, consequently facilitating image-based diagnostics.

This paper reviews these methods focusing on their potential use in neurosurgery for better planning and intraoperative navigation.

Keywords—artificial intelligence, diffusion tensor imaging, Dijkstra's algorithm, graph traversing, MRI, neural networks, tractography.

1. Introduction

The human body is controlled by the central nervous system, with the brain being a crucial organ of that system. Brain cortex is made of neural cell bodies tasked with processing information and performing cognitive process. The axons of these cells constitute the white matter that is located underneath and connects various areas of the cortex and transmit nerve impulses between them. All those structures are organized into complex neural circuits, serving as a platform for all vital body functions, such as cognitive processes, movement, sight, as well as production and comprehension of speech [1].

Like the rest of the body, the central nervous system is susceptible to numerous illnesses, including Alzheimer's disease, Parkinson's disease or brain cancers, such as glioma. Neurosurgery is the primary treatment in dealing with gliomas [2]. Although neurosurgery is invasive and risky, in many cases it is the only way to extend the patient's life and to improve the quality of their life. Prior to the surgery,

the patient undergoes a series of tests, including magnetic resonance imaging (MRI), with a view of mitigating potential risks. Data collected through this method is used to annotate functional areas of the cortex and to determine the location of nerve pathways connecting them. Although the anatomy of the brain is fairly well known, the precise location of functional areas varies from patient to patient. Furthermore, it has been shown that functional areas may change their location over time. Damage caused to any of the functional areas or nerve pathways may lead to complications and irreversible neurological deficits, making imaging necessary before each surgery (Fig. 1).

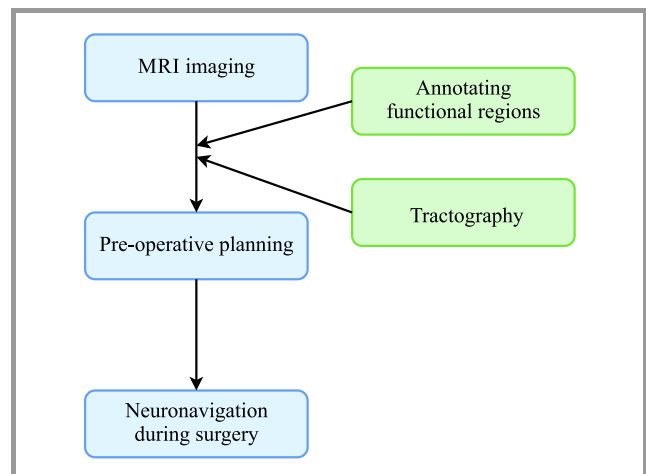


Fig. 1. Outline of the procedure including imaging studies and planning before neurosurgical intervention.

Surgeons may be guided by such knowledge in preoperative planning, deciding on the scope of the intervention and the appropriate entry point. In addition, doctors may use such data during the operation itself, using it as a source of precise information about the layout of the operating field. Tractography is a technique that emerged at the beginning of this century to determine the topology of white matter fibers (nerve tracts) within the brain. Diffusion-weighted magnetic resonance imaging (DW-MRI) [3], [4] uses a specific sequence of pulses and field gradients to produce images where diffusion of water molecules generates the contents. Consequently, it allows to observe,

non-invasively, the process of water diffusion in the living cells. The process of diffusion is anisotropic due to the presence of numerous obstacles in the cells, including organelles and cell membranes [5], [6]. In the axons, it is the myelin sheath that serves as the primary barrier, causing water to travel along the fibers. Hence, tracking water diffusion in the white matter reveals its delicate structure within the brain [7]. The Human Connectome Project [8], a publicly funded multi-institutional initiative, has undertaken the task of constructing a comprehensive map of structural and functional neural connections *in vivo*. An atlas of human brain connectivity is one of the outcomes of the project [9].

Many algorithms have been proposed to produce tractograms [10], [11]. The difficulty in developing the ideal solution lies in numerous problems encountered when analyzing the data, i.e. fiber crossings, false continuities, fiber truncation, as well as when choosing relevant stopping rules or accounting for the presence of edema noise.

This paper provides a review of specific methods that may potentially be used for fiber annotation in neurosurgical applications. We survey primarily algorithms that are suitable for local tractography, as such an approach restricts our analysis to a particular area. In such applications, both deterministic and probabilistic methods may be used. The source of the data is a very important consideration, as it may be a limiting factor affecting clinical applications. Although diffusion tensor imaging (DTI) [12] sequences are usually relied upon, we decided to highlight the advantage of using high-angular resolution DW imaging (HARDI) [13], [14] by focusing on algorithms using this type of data. In the following sections, we describe classical and artificial intelligence methods, then proceeding to the presentation of benchmarking studies. Then, we describe our case study that focuses on tractography in preoperative neurosurgical planning.

2. Mathematical Models

Mathematical models are deemed to be the results of methods used for predicting the orientation of fibers without any support from machine learning algorithms, including from neural networks. Such methods make assumptions based solely on mathematical concepts helping predict local or global tractograms based on diffusion-related data.

Deterministic algorithms focus on envisioning the orientation of a nerve fiber through a single voxel consideration, by expanding the tract from a defined seed point [11]. Seed points are drawn directly from a region of interest (ROI). In neurosurgery, an ROI may be defined as the area of the planned intervention, especially if any part thereof is located on the border between gray and white matter, healthy and tumorous tissue. It may also be defined as a relevant functional region identified with the help of functional magnetic resonance imaging.

Just like their deterministic counterparts, probabilistic models may start from a given seed point. However, they do not follow a single tract. Instead, they keep on building a distribution of probable streams. Such algorithms are computationally more expensive than those relied upon in the deterministic approach. However, they are better suited for tracking in high uncertainty regions (e.g. crossing fibers), especially when noise is present, which is where deterministic approaches are prone to fail [11]. All methods described below are compared in Table 1 in which such factors as diffusion data source, approach to tracking, key underlying concepts and stopping criteria are taken into consideration.

2.1. Linear Forced Vector Differential Equation

One of the first attempts was made in the previous century by Basser in [15]. His methodology generates a dif-

Table 1
Comparison of different classical models for solving tractography tasks

Method	Data source	Approach	Key concepts	Stopping criteria
Basser	DTI	Deterministic	Solving linear forced vector differential equation with Taylor series approximation	Not discussed by the author
TEND	DTI	Deterministic	Incoming vector deflection determining local orientation	Fractional anisotropy below certain threshold; change in the direction exceeding 45°
MRtrix	HARDI	Deterministic, probabilistic	Tracking based on fiber orientation function	FOD peaks below certain threshold; fiber located outside ROI
Complex fiber orientation distribution	HARDI	Deterministic, probabilistic	Utilizes ODF from Q-ball imaging and ODF from a sharpening deconvolution transform	Tracked fiber located outside white matter area
Hough transform voting	DSI, DTI, HARDI	Probabilistic	Selecting the most probable fibers based on Hough transform voting procedure	Tracking is not performed outside specified user-defined area

fusion tensor field from DTI used to obtain fiber trajectory at a given point by solving a linear forced vector differential equation. Nearby points are approximated using the Taylor series expansion. The method uses the eigenvector direction for the highest eigenvalue, assuming that it precisely describes precisely the direction of water diffusion at a given point.

2.2. Tensor Deflection

The TEND algorithm developed by Lazar *et al.* [16] uses the tensor deflection technique to estimate fiber trajectory. In contrast with the method by Basser *et al.* [15], the whole diffusion tensor is used, not only the eigenvector with the highest eigenvalue. Tract reconstruction is performed in a stepwise fashion, where tract direction from the previous step is treated as an incoming vector. This vector is then deflected by the tensor operator towards the major eigenvector direction.

The resulting vector describes the orientation of the fiber at a given position. The curvature of the deflection is limited, resulting in much smoother tract reconstruction. Tracking with TEND is terminated when fractional anisotropy drops below a certain threshold value, or when the change in the direction exceeds 45° .

2.3. MRtrix

MRtrix [17] is a freely available software that combines multiple tools for performing tractography-related tasks. It requires data collected from HARDI imaging. Such data allows to determine fiber orientation by applying constrained spherical deconvolution (CSD) [18] in each voxel to produce a fiber orientation density function (FOD). The resulting FOD holds all information about orientations within a single voxel and is ideal for tracking algorithms. MRtrix relies on two of them a deterministic and a probabilistic one. The deterministic algorithm reconstructs a single fiber along its local orientation. Authors use the Newton-Raphson gradient ascent algorithm to identify the nearest peak in the FOD data iteratively. The peak identification procedure is run once per point, resulting in the most closely aligned direction of the peak, as the current orientation indicates.

In the probabilistic variant, the direction of the next step is provided by a random sampling of the FOD. Sampling is restricted to directions within a certain angle from the current orientation. The selected sample is used to guide towards the next step, as long as the amplitude exceeds a certain threshold. Otherwise, a new sample is generated, and the process is repeated the number of times specified by the user. In comparison to other methods, smoothness and good resolution are achieved by using different data sources and a step size that is smaller than the voxel size. Both algorithms stop tracking when either of the two criteria is met – no satisfying FOD peak can be found (with an amplitude above a certain, user-defined threshold) or when tract propagation reaches the area outside the specified ROI.

2.4. Complex Fiber Orientation Distribution

The method developed by Descoteaux *et al.* [19] uses data from HARDI experiments as well. Their approach uses the sharpening deconvolution function (SDT) to the orientation distribution function (ODF) from Q-ball imaging [20]. The sharpening operation is required due to sparsity of the fiber ODF and the fact that diffusion represented by Q-ball differs from the real direction of the nerve fiber. As it is the case with the MRtrix package, authors propose both deterministic and probabilistic tracking algorithms. The deterministic approach extends classical streamline techniques by considering multiple ODF maxima at each step, where the direction is chosen from 1281 possibilities. The winning direction exceeds all its neighbors' peak values and features an ODF value above 0.5.

The probabilistic algorithm is an interesting extension of the random walk method [21], [22], exploiting the information in multidirectional fiber ODF. Fiber search is performed by particles moving freely from a given seed point, according to the local ODF information. Each voxel is scored based on the number of particles that have reached it. Fiber direction is chosen out of 120 discrete directions, based on the probability derived from the voxel scores, with the step size of 0.5 the voxel size. The procedure is terminated once the particle leaves the white matter. Anisotropy measure map allows the detection of such an event, which is a very precise stopping criterion.

2.5. Hough Transform Voting

Aganji *et al.* [23] propose to use a voting process based on Hough transform to determine the global topology of white matter tracts within the brain or within a selected ROI. Even though global tractography is not optimal in neurosurgical applications, we decided this method was worth mentioning for two reasons. Firstly, even when we consider that the main aim is to obtain the global tractogram, the user can still impose constraints on the area where tracking occurs. Secondly, the method is developed to work with data from various DWI modalities, such as diffusion spectrum imaging (DSI) [24], DTI or HARDI. In the search space mentioned above, i.e. the whole brain in non-restricted tractography, random seed points are drawn. Curves passing through these points are parametrized by the length of the arc, representing potential fiber orientations. Each one is then scored based on the probability of each curve being part of the same fiber as the seed point. Finally, those with the highest scores are selected as the most probable anatomic connections.

The described approach is an exhaustive search, capable of avoiding local minima. There is no clear stopping criterion except of the ROI area constraint and the brain surface in the global variant.

3. Artificial Intelligence

Machine learning (ML) techniques have been successfully applied in many areas of our daily lives, such as spam

Table 2
Comparison of different machine learning approaches used to solve tractography-related tasks

Method	Algorithm	Approach	Key concepts	Stopping criteria
Random forest classification using neighborhood information	Random forest classifier	Probabilistic	Voting process deciding between probable directions or termination	Tracking stops when nonfiber probability exceeds cumulated weighted probabilities for continuation directions
Learn-to-track (1)	Feed-forward neural network	Deterministic	Outputs three-dimensional normalized vector, describing the streamline direction	Not implemented
Learn-to-track (2)	Recurrent neural network	Deterministic	Neural network architecture allows to consider previously seen points to better predict the fiber topology	Not implemented
Bundle-wise deep tracker	Recurrent neural network	Deterministic	Trained to predict tracking direction from a bundle-wise-input data	No information
DeepTract	Recurrent neural network	Deterministic, probabilistic	Estimating local fiber orientation as a discrete probability density function	Track termination probability as one of the predict classes
Neural network regression	Multilayer perceptron	Deterministic	Predicting fiber orientation in a cube of a given size based on diffusion data and existing fiber directions	Learned from the data; tracking stops when white/gray matter boundary is hit
FOD with neural networks	Convolutional neural network, high-resolution network, U-Net	Not applicable	Neural networks trained to regress constrained spherical deconvolution coefficients	Not applicable

filtering, image classification and processing, as well as natural language processing. The same approach is also taken in bioinformatics and medical data, helping design new drugs [25], as well as analyze patient scans [26] or genomes [27]. Not surprisingly, it has also been used to track nerve fibers. The use of ML techniques to predict the location of nerve fibers offers numerous benefits. One of the main advantages is the ability to use raw diffusion data directly, without having to represent the diffusion propagator or tissue microstructure [28]. Furthermore, methods of this type are not limited to relying on particular imaging modalities and are capable of learning from different types of DWI experiments.

Additionally, the learning systems may deal with imperfections, such as noise and distortions, and are suitable for identifying the location of white matter tissue in the scan of the entire brain. All that may be learned directly in the model learning phase. The methods discussed are summarized in Table 2, where the algorithms, approaches, key concepts and stopping criteria are presented.

3.1. Random Forest Classifier

One of the first approaches relying on ML methodology was proposed by Neher *et al.* [28]. This method, similarly to traditional approaches, operates in a step-wise fashion

when extending the current fiber. However, instead of modeling the signal mathematically, a random forest classifier determines local tissue properties directly from raw diffusion data. During the tracking phase, when moving from a given point onwards, an algorithm considers information from voxels in proximity to a given point. The decision about the new orientation is made in a voting process, with the potential directions sampled from a complete sphere or hemisphere in front of the current position. Classifier output provides the probability for each direction, as well as non-fiber probability. Tracking stops when non-fiber probability exceeds the cumulated weighted probabilities of all possible directions. Otherwise, the streamline direction is calculated as the normalized sum of the proposed directions. The voting process makes the fiber extension process less sensitive to noise and local signal ambiguities, as well as to premature termination.

3.2. Learn-to-Track

Neural networks also provide means for analyzing raw diffusion data. Poulin *et al.* proposed two neural network architectures to accomplish this task [29]. Their feedforward neural network (FFNN) returns a three-dimensional vector describing fiber orientation for each point within the diffusion data. The recurrent neural network (RNN)

[30] proposed by the authors takes advantage of the previously seen voxels by remembering features relevant to the entire streamline orientation. The authors note that learning a stopping criterion in the neural network approach is not a trivial task, requiring careful engineering and balancing the loss function. Nevertheless, the method was capable of achieving high spatial coverage of a given test set, while simultaneously controlling the number of false positives.

A subsequent study by the group postulates using RNN in a bundle-wise manner, resulting in improved tracking efficiency, a higher number of valid streamlines and better volume coverage in comparison to all classical algorithms [31].

3.3. DeepTract

Another method using RNNs was developed by Benou *et al.* [32] to address fiber orientation estimation as well as to streamline tractography. Like other methods using ML, it is capable of operating well on various types of raw diffusion data. Unlike the previously described methods, it does not produce merely deterministic predictions of the streamline. It estimates the orientation of local fibers as a discrete probability density function, allowing to randomly sample directions at a given point. The authors address the problem of choosing a new direction as a classification task, predicting the probability for each orientation and the probability of tract termination.

3.4. Neural Network Regression

A method developed by Wegmayr *et al.* proposes a multi-layer perceptron [33] predicting the outgoing direction from an input. The prediction is made in a cube of a given size, being one of the neural network inputs. The second input is a fixed number of incoming vectors, describing the neural fiber entering the central vector of the cube. The network's output is a vector with three values representing the outgoing direction of the fiber. Tracking is done iteratively and begins with random sampling of the ROI area. Prediction of the initial direction is made with the incoming vector set to zeros. The authors claim that the stopping criteria may be learned from the data itself. However, they have implemented a simple rule to stop tracking when the boundary of white and gray matter is reached.

3.5. FOD with Convolutional Neural Networks

Lucena *et al.* propose to use convolutional neural networks (CNNs), broadly used in image classification, to compute more accurate FOD from a single-shell dMRI [34]. The authors trained two three-dimensional CNNs, a three-dimensional high-resolution network (HighResNet) [35] and a three-dimensional U-Net [36] to regress constrained spherical deconvolution coefficients. Although the method does not provide any means of tracking, we decided it was worth mentioning given the produced FODs could be used by other algorithms to track fibers.

4. Benchmarking Studies

It is difficult to assess the accuracy of tractography algorithms, especially if predictions are made *in vivo*, without the possibility to dissect the patient's brain.

In 2015, under an initiative of Klaus H. Maier-Hein, laboratories from across the globe were invited to participate in a tractography competition, where such an assessment was made possible [37]. The methods they relied upon were evaluated based on a specifically crafted data set. It was constructed from multiple whole-brain global tractography maps [38]. A trained radiologist extracted 25 major tracts, comprising association, projection, and commissural fibers. The data prepared for the participants included a structure mimicking clinical-like acquisition of DWI based on a simulated diffusion signal and simulated T1-like contrast. A significant number of methods correctly predicted the topology of most of the fibers under consideration. However, the majority produced, along with correct predictions, many false-positive results, whose number often exceeded that of the correct ones.

A work published by Schilling *et al.* assesses the accuracy of tractography methods using a data set containing a physical phantom and two *ex vivo* brain specimens [39]. Although advances in the tractography methods are significant, the authors remark that anatomical accuracy is still limited. This study confirms previous findings showing a great number of false positives generated by the methods. Most algorithms tested had a low connectivity predictive value and low spatial overlap with the true pathways.

5. Case Study

Currently, we are designing a neurosurgery support system. It will predict functional cortex regions and white matter fibers in oncological patients. Our goal is to provide predictions based on imaging data to facilitate pre-surgery planning. We would like to propose a new tractography method, being a part of that system, combining a variant of the Dijkstra algorithm [40] with a feedforward neural network backbone. The backbone produces information required to traverse the graph representation of the brain's diffusion map, allowing to trace neural fibers.

The general architecture of the underlying neural network is shown in Fig. 2. In the proposed method, the diffusion data is analyzed in small batches of size (5, 5, 5). Thus, the required number of nodes in the hidden layers is significantly reduced. A neural network returns a scalar value for each voxel, describing the probability that a given voxel defines the same neural fiber as the central one. Voxels with values above the specified threshold will constitute the nodes of the graph. Weights for edges between the nodes are computed based on the probability values.

Tracking fibers starts by picking the region of interest (ROI), usually covering the area of the planned surgical intervention. Each voxel within ROI is a tracking seed – the

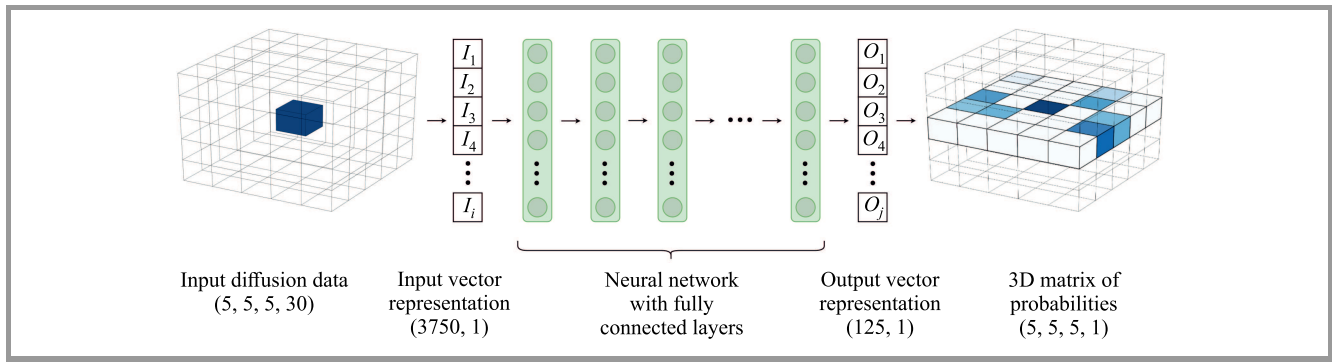


Fig. 2. Architecture of the neural network for predicting probabilities of voxels containing parts of the same fiber.

first node of the graph. In the next step, a three-dimensional matrix of size (5, 5, 5) with the seed located in the center thereof is drawn. A neural network is employed to produce an output 3D matrix with probabilities for all 125 voxels. The central voxel, as the seeding point, has the probability of 1 (Fig. 2). The graph is extended from the initial node to reach the edge of the drawn cube. The voxel at the border, chosen as the graph node, becomes the central voxel of the next cube. This procedure is presented, on two-dimensional slices, in Fig. 3.

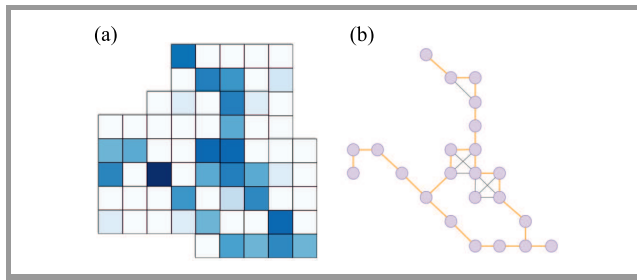


Fig. 3. The method proposed for solving tractography tasks. Part (a) shows path tracking with overlapping cubes obtained from the neural network. Part (b) shows a graph of potential neural paths. Black edges are the possible connections between the nodes. Bold orange edges represent the most probable paths according to the method. (For the color version, please see the digital edition)

In this approach, labeled bands of white matter will quickly cover regions that are remote from the actual ROI. To avoid that, we propose to use two stopping rules. The first is to stop tracking when no satisfying candidates exist near the current node. This principle stems directly from the algorithm used. The additional rule would stop the tracking when a defined Euclidean distance from the seed point is reached.

We are currently collecting data from oncological patients, thanks to long-term cooperation with the Department of Neurosurgery at the Maria Skłodowska-Curie National Research Institute of Oncology. Once this step has been completed, a trained radiology specialist will annotate each diffusion set, producing the expected tractogram. Next, raw diffusion data and tractograms would be used to train the backbone neural network. After the training is completed,

the proposed method would be evaluated on the available benchmark dataset. This will allow to compare its accuracy and the rate of false positive predictions with other state-of-the-art methods.

6. Conclusions

Diffusion-weighted imaging data, including DTI, is not capable of providing satisfactory answers concerning the location of white matter bands within the brain. To provide such answers, algorithms analyzing the data, developed based on the knowledge of brain anatomy and functions, are used. Throughout the years, many algorithms have been proposed to tackle this problem, including both mathematical and artificial intelligence-based approaches. In this paper, we presented a review of the available methods, with an emphasis placed on their potential use in neurosurgery.

The mathematical models are quite simple in their structure. The final solution of the problem is usually a product of a combination of various methods used together. Each step involves different issues, starting from the noise within the data itself, to complex states on the atomic level that cannot be addressed by such methods alone. These models are developed to process specific diffusion-weighted modalities. The methods based on HARDI imaging data tend to produce smoother, higher resolution tractograms. More information is provided for a single voxel, allowing to adopt a step size that is smaller than the voxel itself. The use of probabilistic models may shed more light on the complex topology of human brain connectomics. However, such methods are prone to produce many false positives, as was shown in the benchmarking studies. The use these methods requires expertise in order to adjust their parameters and to correctly interpret the results at each stage of the analysis. Methods based on ML attempt to bypass at least some of the obstacles encountered by the classical approaches. They may be fed directly with data from imaging experiments to evaluate the solution through model parametrization in the learning process. However, these methods are not completely free of any drawbacks. The process of designing and training a successful model takes a lot of time to complete, thus increasing its total cost. Moreover, a good model

requires great amounts of well-annotated data. Preparing such a data set is troublesome and labor-intensive. It requires that numerous imaging experiments be conducted with the use of MRI, and that each of them be annotated by a radiology specialist. The effort put into training the model pays off when the method is used. Well-trained models are capable of producing results quickly and without much expertise of the person using them. Therefore, neurosurgeons could use them in an almost out-of-the-box form in preoperative planning.

We propose an approach that is based on a combination of the Dijkstra algorithm with a simple feed-forward neural network be used to predict the most probable fiber topologies within the brain. Such a method would work by performing an iterative analysis of the diffusion data divided into small batches, to speed up the execution process. Continued training and additional experiments are required with real-world patient data in order to optimize the accuracy of the approach and to compare it with other methods using benchmarking datasets.

Despite the tremendous progress made since the introduction of diffusion-weighted imaging, such methods need to be used cautiously and always have to be supported by specific knowledge concerning anatomy and operation of the brain. Novel architectures of neural networks are proposed on a continuous basis to tackle complex tasks in parallel areas, such as natural language processing and image classification. This means that further development of methodologies used to analyze medical data and facilitate diagnostic imaging should be expected as well.

References

- [1] P. A. Young, P. H. Young, and D. L. Tolbert, *Basic Clinical Neuroscience. Third Edition*. Philadelphia: Lippincott Williams and Wilkins, 2015 (ISBN: 9781451173291).
- [2] R. M. Young, A. Jamshidi, G. Davis, and J. H. Sherman, "Current trends in the surgical management and treatment of adult glioblastoma", *Annals of Translational Medicine*, vol. 3, no. 9, pp. 1–5, 2015 (DOI: 10.3978/j.issn.2305-5839.2015.05.10).
- [3] K. D. Merboldt, W. Hanicke, and J. Frahm, "Self-diffusion NMR imaging using stimulated echoes", *Journal of Magnetic Resonance* (1969), vol. 64, no. 3, pp. 479–486, 1985 (DOI: 10.1016/0022-2364(85)90111-8).
- [4] D. G. Taylor and M. C. Bushell, "The spatial mapping of translational diffusion coefficients by the NMR imaging technique", *Physics in Medicine and Biology*, vol. 30, no. 4, pp. 345–349, 1985 (DOI: 10.1088/0031-9155/30/4/009).
- [5] J. E. Tanner, "Self diffusion of water in frog muscle", *Biophysical Journal*, vol. 28, no. 1, pp. 107–116, 1979 (DOI: 10.1016/S0006-3495(79)85162-0).
- [6] D. F. Scollan, A. Holmes, R. Winslow, and J. Forder, "Histological validation of myocardial microstructure obtained from diffusion tensor magnetic resonance imaging", *American Journal of Physiology – Heart and Circulatory Physiology*, vol. 275, no. 6, pp. 2308–2318, 1998 (DOI: 10.1152/ajpheart.1998.275.6.H2308).
- [7] M. E. Moseley *et al.*, "Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system", *Radiology*, vol. 176, no. 2, pp. 439–445, 1990 (DOI: 10.1148/radiology.176.2.2367658).
- [8] A. W. Toga *et al.*, "Mapping the human connectome", *Neurosurgery*, vol. 71, no. 1, pp. 1–ℓ5, 2012 (DOI: 10.1227/NEU.0b013e318258e9ff).
- [9] R. G. Briggs *et al.*, "A Connectomic Atlas of the Human Cerebrum—Chapter 18: The Connectional Anatomy of Human Brain Networks", *Operative Neurosurgery*, vol. 15, no. 1, pp. 470–480, 2018 (DOI: 10.1093/ons/opy272).
- [10] S. Mori and P. C. M. Van Zijl, "Fiber tracking: Principles and strategies – A technical review", *NMR in Biomedicine*, vol. 15, pp. 468–480, 2002 (DOI: 10.1002/nbm.781).
- [11] B. Jeurissen, M. Descoteaux, S. Mori, and A. Leemans, "Diffusion MRI fiber tractography of the brain", *NMR in Biomedicine*, vol. 32, no. 4, pp. 1–22, 2019 (DOI: 10.1002/nbm.3785).
- [12] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging", *Biophysical Journal*, vol. 66, no. 1, pp. 259–267, 1994 (DOI: 10.1016/S0006-3495(94)80775-1).
- [13] L. R. Frank, "Anisotropy in high angular resolution diffusion-weighted MRI", *Magnetic Resonance in Medicine*, vol. 45, no. 6, pp. 935–939, 2001 (DOI: 10.1002/mrm.1125).
- [14] D. S. Tuch *et al.*, "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity", *Magnetic Resonance in Medicine*, vol. 48, no. 4, pp. 577–582, 2002 (DOI: 10.1002/mrm.10268).
- [15] P. J. Basser, "Fiber-tractography via diffusion tensor MRI (DT-MRI)", in *Proc. of the 6th Annual Meeting ISMRM*, vol. 1, 1998, p. 3 [Online]. Available: <https://nationalbii.com/wp-content/uploads/2020/10/Basser-ISMRM-1998.pdf>
- [16] M. Lazar *et al.*, "White matter tractography using diffusion tensor deflection", *Human Brain Mapping*, vol. 18, no. 4, pp. 306–321, 2003 (DOI: 10.1002/hbm.10102).
- [17] J. D. Tournier, F. Calamante, and A. Connelly, "MRtrix: Diffusion tractography in crossing fiber regions", *Int. J. of Imaging Systems and Technol.*, vol. 22, no. 1, pp. 53–66, 2012 (DOI: 10.1002/ima.22005).
- [18] J. D. Tournier, F. Calamante, and A. Connelly, "Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution", *NeuroImage*, vol. 35, no. 4, pp. 1459–1472, 2007 (DOI: 10.1016/j.neuroimage.2007.02.016).
- [19] M. Descoteaux, R. Deriche, T. R. Knosche, and A. Anwander, "Deterministic and probabilistic tractography based on complex fibre orientation distributions", *IEEE Transac. on Medical Imaging*, vol. 28, no. 2, pp. 269–286, 2009 (DOI: 10.1109/TMI.2008.2004424).
- [20] D. S. Tuch, "Q-ball imaging", *Magnetic Resonance in Medicine*, vol. 52, no. 6, pp. 1358–1372, 2004 (DOI: 10.1002/mrm.20279).
- [21] A. Anwander, M. Tittgemeyer, D. von Cramon, A. Friederici, and T. Knosche, "Connectivity-based parcellation of broca's area", *Cerebral Cortex*, vol. 17, no. 4, pp. 816–825, 2006 (DOI: 10.1093/cercor/bhk034).
- [22] M. A. Koch, D. G. Norris, and M. Hund-Georgiadis, "An investigation of functional and anatomical connectivity using magnetic resonance imaging", *NeuroImage*, vol. 16, no. 1, pp. 241–250, 2002 (DOI: 10.1006/nimg.2001.1052).
- [23] I. Aganj *et al.*, "A Hough transform global probabilistic approach to multiple-subject diffusion MRI tractography", *Medical Image Analysis*, vol. 15, no. 4, pp. 414–425, 2011 (DOI: 10.1016/j.media.2011.01.003).
- [24] V. J. Wedeen, P. Hagmann, W.-Y. I. Tseng, T. G. Reese, and R. M. Weisskoff, "Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging", *Magnetic Resonance in Medicine*, vol. 54, no. 6, pp. 1377–1386, 2005 (DOI: 10.1002/mrm.20642).
- [25] R. Gupta *et al.*, "Artificial intelligence to deep learning: machine intelligence approach for drug discovery", *Molecular Diversity*, 2021 (DOI:10.1007/s11030-021-10217-3).
- [26] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology", *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018 (DOI: 10.1038/s41568-018-0016-5).

- [27] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics", *Nature Reviews Genetics*, vol. 16, no. 6, pp. 321–332, 2015 (DOI: 10.1038/nrg3920).
- [28] P. F. Neher, M. A. Cote, J. Ch. Houde, M. Descoteaux, and K. H. Maier-Hein, "Fiber tractography using machine learning", *NeuroImage*, vol. 158, pp. 417–429, 2017 (DOI: 10.1016/j.neuroimage.2017.07.028).
- [29] P. Poulin *et al.*, "Learn to track: deep learning for tractography", *bioRxiv*, vol. 1, pp. 540–547, 2017 (DOI: 10.1101/146688).
- [30] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997 (DOI: 10.1162/neco.1997.9.8.1735).
- [31] P. Poulin, F. Rheault, E. St-Onge, P.-M. Jodoin, and M. Descoteaux, "Bundle-wise deep tracker: Learning to track bundle-specific streamline paths", in *Proc. of the Int. Society for Magnetic Resonance in medicine ISMRM-ESMRMB*, Paris, France, 2018 [Online]. Available: <https://index.miramir.com/ISMRM2018/PDFfiles/0041.html>
- [32] I. Benou and T. R. Raviv, "DeepTract: A Probabilistic Deep Learning Framework for White Matter Fiber Tractography", D. Shen *et al.* Eds. in *Proc. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 22nd Int. Conf.*, Shenzhen, China, 2019, pp. 626–635 (DOI: 10.1007/978-3-030-32248-9_70).
- [33] V. Wegmayr, G. Giuliarì, S. Holdener, and J. Buhmann, "Data-driven fiber tractography with neural networks", in *Proc. IEEE Int. Symp. on Biomedical Imag. (ISBI)*, Washington, DC, USA, 2018, pp. 1030–1033 (DOI: 10.1109/ISBI.2018.8363747).
- [34] O. Lucena *et al.*, "Using convolution neural networks to learn enhanced fiber orientation distribution models from commercially available diffusion magnetic resonance imaging", *arXiv*, 2020 [Online]. Available: <https://arxiv.org/pdf/2008.05409.pdf>
- [35] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "On the compactness, efficiency, and representation of 3D convolutional networks: Brain parcellation as a pretext task. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)", in *Proc. Int. Conf. on Informat. Process. in Medical Imag.*, Boone, NC, USA, 2017, pp. 348–360 (DOI: 10.1007/978-3-319-59050-9_28).
- [36] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation", 2016 [Online]. Available: <https://arxiv.org/pdf/1606.06650>
- [37] K. H. Maier-Hein *et al.*, "The challenge of mapping the human connectome based on diffusion tractography", *Nature Communications*, vol. 8, no. 1, 2017 (DOI: 10.1038/s41467-017-01285-x).
- [38] M. F. Glasser *et al.*, "The human connectome project's neuroimaging approach", *Nature Neuroscience*, vol. 19, no. 9, pp. 1175–1187, 2016 (DOI: 10.1038/nn.4361).
- [39] K. G. Schilling *et al.*, "Limits to anatomical accuracy of diffusion tractography using modern approaches", *NeuroImage*, vol. 185, pp. 1–11, 2019 (DOI: 10.1016/j.neuroimage.2018.10.029).
- [40] E. W. Dijkstra, "A note on two problems in connection with graphs", *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959 (DOI: 10.1007/BF01386390).



Mateusz Koryciński received his M.Sc. in Bioinformatics from Adam Mickiewicz University Poznań and Poznań University of Technology. He is currently working at the Department of Bioinformatics and Machine Recognition at the NASK research institute (Warsaw, Poland) and is pursuing a Ph.D. at the Doctoral School

of Information and Biomedical Technologies of the Polish Academy of Sciences (Warsaw, Poland). His research interests focus on medical informatics, magnetic resonance imaging and image processing.

 <https://orcid.org/0000-0002-5239-1635>

E-mail: mateusz.korycinski@nask.pl

Department of Bioinformatics and Machine Recognition
Research and Academic Computer Network
ul. Kolska 12

Warsaw, Poland



Konrad A. Ciecierski received his Ph.D. (honors) in computer science from Warsaw University of Technology, Poland, in 2014. He specializes in the application of machine learning in medical science, digital signal processing, natural language processing, and deep learning applications. He is currently an assistant professor and head of

the Bioinformatics and Machine Recognition department at the NASK research institute. He works at the Clinic of Neurosurgery of the Maria Skłodowska-Curie Memorial Oncology Center in Warsaw, where he is a member of a neurosurgical team specializing in the treatment of the Parkinson disease and other movement-related disorders.

 <https://orcid.org/0000-0003-2471-3016>

E-mail: konrad.ciecierski@nask.pl

Department of Bioinformatics and Machine Recognition
Research and Academic Computer Network
ul. Kolska 12

Warsaw, Poland

COVID-19 Pandemic and Internet Traffic in Poland: Evidence from Selected Regional Networks

Michał P. Karpowicz

NASK National Research Institute, Warsaw, Poland

<https://doi.org/10.26636/jtit.2021.154721>

Abstract—The COVID-19 pandemic has forced governments all over the world to impose lockdowns keeping citizens at home in order to limit the virus spread rate. The paper compares weekly traffic samples captured in the selected nodes of the network managed by NASK – National Research Institute during the pre-lockdown period, i.e. between January 27 and February 3, 2020, with those captured between March 30 and April 6, 2020, i.e. after the lockdown was announced. The presented results show changes in network traffic observed during the periods of time in question and illustrate the evolution in the popularity of top network services.

Keywords—COVID-19, Internet traffic.

1. Introduction

The COVID-19 pandemic has reshaped both the economy and our daily routines. Be it education, shopping, sports, traveling, work, health care, entertainment or social interactions – the pandemic has left its mark on all areas of our lives.

To limit the virus spread rate, governments all over the world imposed lockdowns keeping citizens at home. This has resulted in numerous everyday activities being performed online. Consequently, the demand for network services and resources surged. This paper studies network traffic variations observed in the network of the NASK National Research Institute at the beginning of the first lockdown introduced in Poland.

1.1. Related Work

The impact of the COVID-19 pandemic on Internet traffic is a subject of extensive studies. Indeed, reports published so far reveal numerous interesting and similar effects that are correlated with the introduction of lockdowns.

As reported in [1], a sharp surge in traffic was observed worldwide in late March and early April 2020, i.e. after the introduction of the first lockdowns in Europe. The lockdowns led to a surge in the popularity of streaming services, causing a visible change in demand for network resources. On the one hand, that impact was similar to surges caused by planned worldwide events, such as New Year's Eve

celebrations or concerts, or events like natural disasters or flash crowds. On the other hand, numerous differences have been spotted as well. Specific symptoms of ongoing changes in users' habits could be observed. Evolution of e-commerce is visible as well, market models change, remote education is gaining momentum, and interest shifts emerge.

The lockdowns have also left their mark on packet transmission latency. This phenomenon affected many latency-sensitive applications, including online games, video calls, VoIP, and IP geolocation. Inferior service quality was experienced mainly during the evenings. In other words, latency increased due to recreational activities, rather than due to remote working or distance learning [2].

In [3], an academic campus is taken as an example to identify changes in traffic patterns. The study shows that incoming traffic decreased drastically. In contrast, outgoing traffic doubled in order to support online learning and working from home, particularly with the use of streaming platforms, VPNs, and remote desktop services. For a related study of remote learning strategies during the COVID-19 pandemic, see also [4].

Finally, [5] offers a comprehensive study of the lockdown effect observed from the point of view of a Central European ISP, three major European Internet exchange points (IXPs), and one metropolitan educational network in Spain. After the lockdowns were imposed in the second half of March, much more traffic was generated in the mornings and during late evening hours. The study provides details of the changes visible in the transport and application layers. The collected data prove that the distribution of traffic source and destination ports has changed due to lockdowns. Those changes, clearly visible in traffic samples, may be attributed directly to remote working, education, VPN-based communications, and video conferencing. Also, interestingly enough, workday patterns became similar to weekend patterns. That phenomenon needs a further explanation that takes into account the characteristics of local economies and social structures, as the sections below suggest.

This paper presents traffic patterns observed in the NASK – National Research Institute network during the first Polish lockdown.

1.2. Dataset and Ethical Considerations

The dataset used in this study consists of sFlow (RFC 3176) traffic samples collected between January 27 and February 3, 2020, and between March 30 and April 6, 2020, as a part of routine network analysis procedure performed by NASK. It describes the activity of a stable group of (~1,000) commercial customers, government and public web services, and selected academic network users. In the period considered, there were no significant changes to the network infrastructure the study is concerned with. Likewise, the network user base did not change either.

The collected data contain randomly sampled packet headers and do not reveal any packet payload information. Also, special care was taken to anonymize IP addresses in compliance with Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal on the free movement of such data. Therefore, the anonymization process was applied permanently and irreversibly and removed any potential links between IP addresses and both legal entities and natural persons to whom the addresses may relate.

The following procedure was used while working with the dataset. First, the traffic samples collected were pre-processed with the use of a cryptography-based prefix-preserving anonymization algorithm (CryptoPan) to replace the IP addresses observed with new addresses [6]. Next, we replaced the last bits of each new IP address with zeros. As a result, the dataset contained only information regarding the flow of packets between anonymized networks. The activity of any individual IP address is not visible in the dataset. Finally, as presented in the following sections, aggregated statistical data and time series were obtained.

2. Network Traffic Data Study

This section illustrates and comments on the correlation between the changes in network traffic characteristics and the decision to impose the lockdown in Poland in the second half of March 2020.

Firstly, it compares weekly traffic patterns during the pre-lockdown period with weekly traffic patterns during the lockdown. The aggregated traffic is discussed together with its top components, including TCP/UDP and HTTP/HTTPS traffic.

Secondly, it compares the activity of leading source and destination ports before and after the decision to impose the lockdown. This comparison offers an overview of changes attributed to remote working, education, VPN-based communications, and video conferencing. Observations regarding network security are given as well.

To calculate relative changes in traffic patterns, the total number of bytes transferred during the periods studied was compared. The following formula was applied:

$$\Delta = \frac{Y - X}{X} \cdot 100\% , \quad (1)$$

where X denotes pre-lockdown and Y lockdown measurements. The statistics were generated with an improved version of of nfdump software, introducing the correct handling of port 0 traffic.

2.1. Traffic Rate Shifts

Figure 1 illustrates the network traffic rate, measured in bits per second (bps) and packets per second (pps), during the week preceding the lockdown and the week after the lockdown was announced. For the sake of clarity, the presented time series were smoothed with a low-pass zero-phase second order Butterworth filter. As we can see, the overall traffic increased by more than 40%. That number stems from the comparison of the total number of bytes and packets transferred during the periods under analysis.

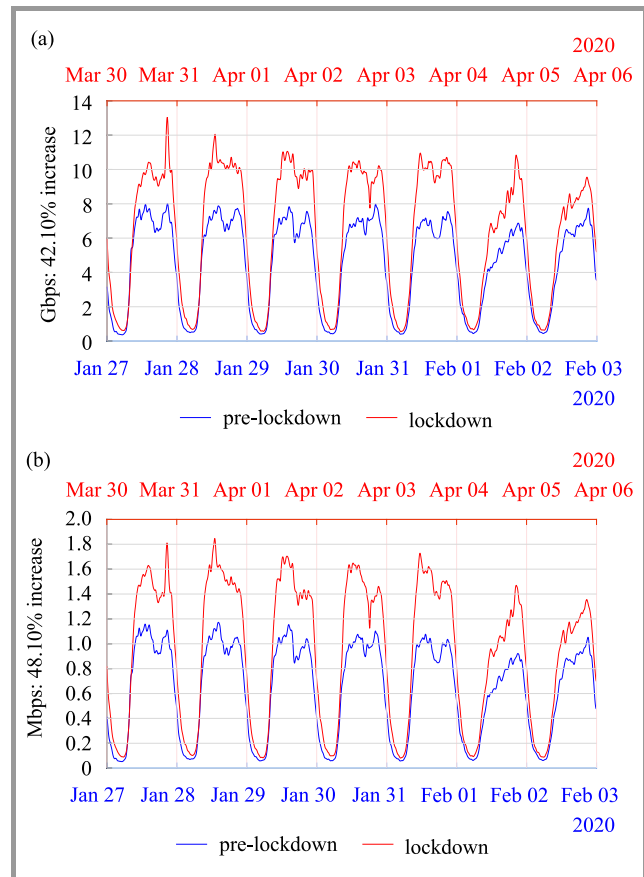


Fig. 1. Aggregated network traffic before and after the lockdown: bits per second (a), packets per second (b). (see the digital edition for color images)

On workdays, traffic increased during the first part of the day, before lunchtime. After lunch, traffic increased again, reaching its peak late in the evening, during the second part of the day. Traffic reached its maximum before lunch. On

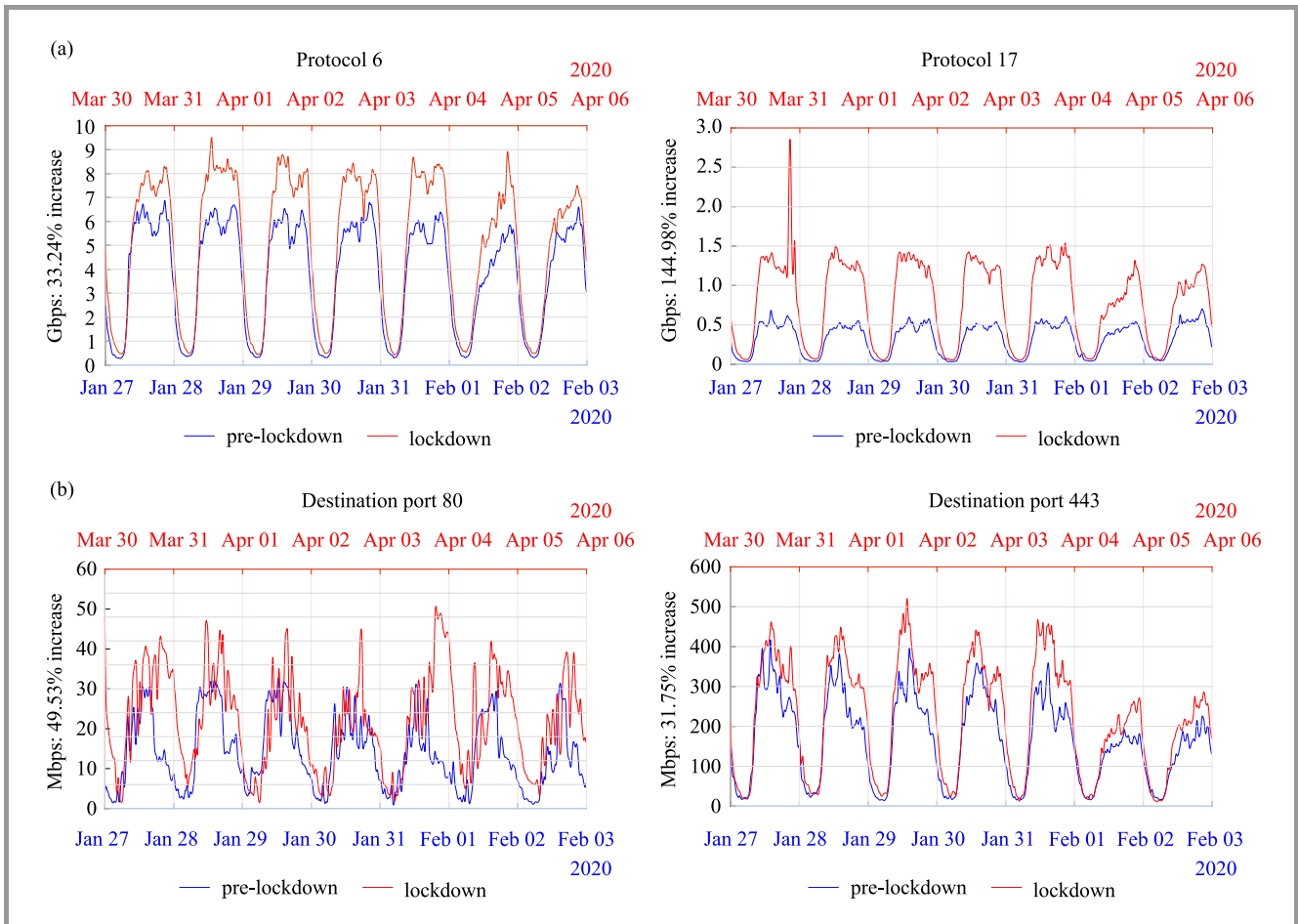


Fig. 2. Traffic shift details: TCP/UDP [Gbps] (a), HTTP/HTTPS [Mbps] (b).

weekends, in contrast, traffic increased gradually to reach its peak in the evening. An increase may also be observed in minimal traffic rates, both on workdays and during weekends.

Figure 2 reveals more details of the traffic shifts. Top figures show TCP (proto 6) and UDP (proto 17) traffic during the period concerned. In the first case, traffic increased by 33%. However, in the second case, that increase reached almost 145%. This may be explained by a rapid increase in UDP-based traffic generated by data streaming and VPN services. HTTP/HTTPS traffic is presented in the middle. The rate of HTTP traffic rose almost by 50%, whereas HTTPS traffic rose by nearly 32%. In the case of HTTP traffic, the peaks moved towards late evening hours, a symptom suggesting user behavior changes.

2.2. Application Activity Shifts

This subsection compares the activity of top source and destination ports before and after the lockdown. A comparison of the top fifteen source ports is presented in Table 1. Top destination ports, in turn, are compared in Table 2. The last column in each table presents activity shifts calculated based on the number of bytes transferred from or to a given port within the period under consideration.

Both before the lockdown and the just after its introduction, ports 443 and 80 remained the most active ones. The dominant role of such application layer protocols as HTTP (port 80) and HTTPS (port 443) may be explained easily. These are the application-layer protocols delivering web page content and transferring data over the Internet. Since the lockdown moved our lives to the online environment, the increase in traffic at ports transmitting HTML documents, images, or videos comes as no surprise. As far as the HTTPS protocol is concerned, source traffic increased by 11.6%, whereas destination traffic increased by 24.1%. In the HTTP protocol, the increase reached 28.8% and 33.8%, respectively.

The remaining part of the list of top ports illustrates critical changes in application activity patterns. The most significant growth in traffic volume was recorded by IPsec, OpenVPN, and new communication services, such as MS Teams or Google Hangouts.

Activation of virtual private networks (VPNs) was the first and straightforward consequence of the lockdown. This explains the increase in IPsec (port 4500) and OpenVPN (port 1194) traffic. In the first instance, that increase reached over 190%, while in the second – over 96% in terms of source traffic. Similarly, since streaming and communica-

Table 1
Top source ports, ordered by bytes

Pre-lockdown				Lockdown				
Source port	Application	TB	Gpackets	Source port	Application	TB	Gpackets	Δ [%]
443	HTTPS	826.5	634.7	443	HTTPS	922.3	729.1	11.6
80	HTTP	326.9	224.0	80	HTTP	421.0	288.3	28.8
53	DNS	10.4	15.8	4500	IPSec	14.5	22.8	190.0
993	IMAPS	8.1	9.4	53	DNS	11.9	16.8	14.4
0		7.0	8.0	0		11.2	14.5	60.0
37777	Video	5.7	4.2	1194	OpenVPN	10.2	12.5	96.2
1194	OpenVPN	5.2	7.0	37777	Video	8.6	6.2	50.9
4500	IPSec	5.0	7.5	5544		6.8	5.2	88.9
6908	Bittorrent	4.5	3.4	993	IMAPS	6.7	7.3	-17.3
6907	Bittorrent	4.0	6.3	8080	HTTP	5.1	3.9	
8080	HTTP	3.7	2.7	8801	Backup	4.9	10.2	
5544		3.6	2.9	8000	Streaming	4.5	3.6	45.2
8000	Streaming	3.1	2.5	6908	Bittorrent	4.5	3.4	0.0
22	SSH	2.6	2.4	10443	SSL/dogtag	3.2	4.4	
995	POP	2.4	1.9	34765		2.8	2.2	

Table 2
Top destination ports, ordered by bytes

Pre-lockdown				Lockdown				
Destination port	Application	TB	Gpackets	Dst port	Application	TB	Gpackets	Δ [%]
443	HTTPS	77.5	323.5	443	HTTPS	96.2	396.8	24.1
80	HTTP	13.6	90.7	80	HTTP	18.2	128.1	33.8
6666	IRC	8.0	5.7	4500	IPSec	13.8	19.7	181.6
0		7.0	8.0	0		11.2	14.5	60.0
1194	Open VPN	5.8	7.5	6666	IRC	9.6	6.9	20.0
4500	IPSec	4.9	7.5	1194	Open VPN	7.9	11.4	36.2
6180		3.7	2.4	38188		4.6	3.4	
1935	RTMP	2.3	4.7	6180		2.9	1.9	-21.6
25	SMTP	2.2	2.3	6901	MSN Messenger	2.8	2.2	33.3
6901	MSN Messenger	2.1	1.7	6902	MSN Messenger	2.8	2.2	
33451	Webex	2.0	1.5	25	SMTP	2.8	2.5	
54536		1.8	2.9	19305	Hangouts	2.1	3.9	
48508		1.8	1.4	1935	RTMP	2.0	3.4	-13.0
48611		1.6	2.6	20	FTP	2.0	1.4	33.3
20	FTP	1.5	1.1	33001	Aspera	1.8	1.2	

tions services supporting online meetings and file sharing became crucial for the remote working model, they significantly boosted traffic rates as well. That is the case for ports 37777, 8000, 6901, 19305, or 20. Additionally, that observation may also explain the decrease in port 993 traffic supporting mailing over SSL. New communication services, such as MS Teams or Google Hangouts, seem to have taken its place.

Interestingly, the data collected reveal also an increase in traffic that is correlated with ports often used by malicious software. This includes port 5544 and port 0. In fact,

during the period from March to May, numerous attacks were registered in the network under observation. Figures 1–2 show a traffic spike caused by a DDoS attack. A campaign of DDoS attacks may also be seen in Fig. 3, illustrating the panel of the FLDX DDoS protection system developed and used in the networks managed by NASK. Detected distributed denial of service (DDoS) attacks can be seen as traffic spikes (each of them was successfully attenuated).

Statistics for port 0 traffic need a more detailed explanation. While being illegal in general cases, port 0 is well known to

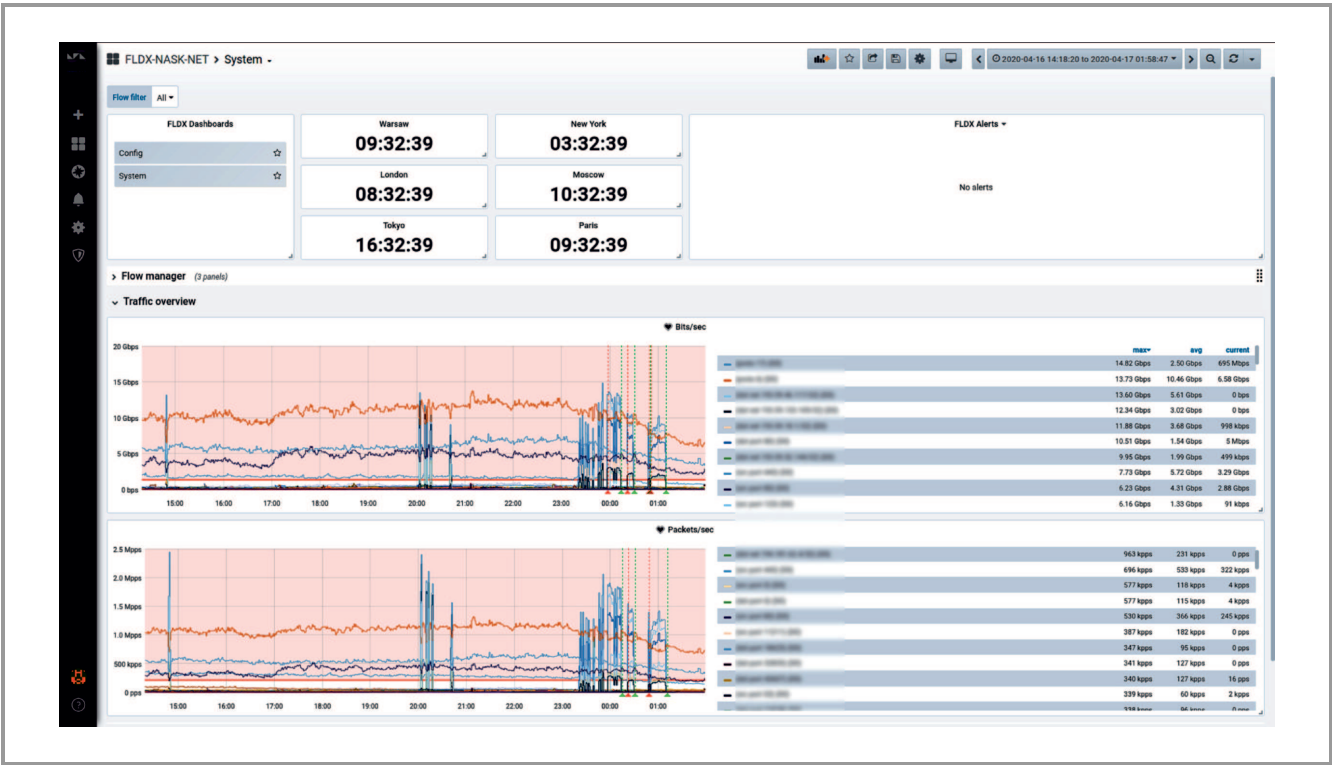


Fig. 3. Example of a series of DDoS attacks (seen as traffic spikes) detected and attenuated by the FLDX DDoS protection system in the second half of April 2020.

have a large traffic share because it aggregates fragmented, malicious, scanning, and wild packets. It is often seen in many network attacks, including DDoS attacks originating from large botnets. However, it is associated with legitimate traffic as well. To be more precise, packets generated by protocols that do not use ports happen to be marked by network devices and packet analyzing software precisely as port 0 traffic. Therefore, much care should be taken when analyzing that particular case. Statistics presented in Tables 1, 2 and 3 compensate for that effect, exposing only traffic with valid source and destination ports. In the

period studied, the overall traffic share of port 0 increased approximately by 60%. For more details, see also [7].

3. Summary

The COVID-19 pandemic has introduced significant changes to our lives. Many of these changes may stay with us for longer or may turn out to be permanent. Remote work and remote education, online medical appointments, online shopping, movie premieres at home, dancing practice in your living room, all these activities have evolved rapidly due to lockdowns. It is so because they help us face the challenges associated with isolation. At the same time, many forms of activity have disappeared as a result of lockdowns. Changes in network traffic reflect changes in our habits. The volume of network traffic has significantly increased, and we tend to be online for longer. Communication platforms, streaming services and VPNs supporting remote work and remote education have gained in importance. A careful observation of web applications and protocols helps us to understand the ongoing changes and pinpoint the potential threats.

Acknowledgements

I express my gratitude to the FLDX system development team, especially to Arkadiusz Piórkowski and Janusz

Table 3
Service port 0 traffic profile: destination ports (top) and protocols ordered by bytes (bottom)

Pre-lockdown		Lockdown	
Destination port	Bytes	Destination port	Bytes
0	7T	0	11.2T
80	60.2M	80	49.7M
443	4.5M	443	29.4M
53	1.9M	6680	2.8M
12001	1.9M	12812	1.5M
Protocol	Bytes	Protocol	Bytes
UDP	6.9T	UDP	11.1T
TCP	61.3T	TCP	110.1G
IPv6	142M	IPv6	109M

Janiszewski, for their continuous and innovative work and support in analyzing data. I am also deeply indebted to Marek Dawidiuk for securing the traffic samples during the difficult period of remote work. Finally, my sincere appreciation goes to Urszula Brochwicz for developing legal recommendations on data anonymization.

References

- [1] T. Boettger, G. Ibrahim, and B. Vallis, "How the Internet reacted to COVID-19: A perspective from Facebook's Edge Network", in *Proc. of the ACM Internet Measurement Conf.*, Virtual Event, USA, 2020, pp. 34–41 (DOI: 10.1145/3419394.3423621).
- [2] M. Candela, V. Luconi, and A. Vecchio, "Impact of the COVID-19 pandemic on the Internet latency: A large-scale study", *Computer Networks*, vol. 182, 2020 (DOI: 10.1016/j.comnet.2020.107495).
- [3] T. Favale, F. Soro, M. Trevisan, I. Drago, and M. Mellia, "Campus traffic and e-Learning during COVID-19 pandemic", *Computer Networks*, vol. 176, 2020 (DOI: 10.1016/j.comnet.2020.107290).
- [4] T. Gonzalez *et al.*, "Influence of COVID-19 confinement on students' performance in higher education", *PloS one*, vol. 15, no. 10, 2020 (DOI: 10.1371/journal.pone.0239490).
- [5] A. Feldmann *et al.*, "Implications of the COVID-19 Pandemic on the Internet Traffic", in *Broadband Coverage in Germany; 15th ITG-Symposium*, pp. 1–5, Online Conf: VDE, 2021 (ISBN: 9783800754748).
- [6] J. Fan, J. Xu, M. H. Ammar, and S. B. Moon, "Prefix-preserving IP address anonymization: measurement-based security evaluation and a new cryptography-based scheme", *Computer Networks*, vol. 46, no. 2, pp. 253–272, 2004 (DOI: 10.1016/j.comnet.2004.03.033).
- [7] A. Maghsoudlou, O. Gasser, and A. Feldmann, "Zeroing in on port 0 traffic in the wild", in *Proc. of the 2021 Passive and Active Measurement Conference (PAM '21)*, Online Conf, 2021, pp. 547–563 (DOI: 10.1007/978-3-030-72582-2_32).



Michał P. Karpowicz, Ph.D. (2010), D.Sc. (2020), is a Professor with the NASK National Research Institute, head of the IT Systems Engineering Department and Assistant Professor with the Institute of Control and Computation Engineering, Warsaw University of Technology. His research interests include control theory, signal processing, and linear algebra. He is an author of more than 30 articles, co-author of one book, designer of two network control systems supporting cybersecurity operations in Poland's nation wide networks.

 <https://orcid.org/0000-0003-1431-3078>

E-mail: michal.karpowicz@nask.pl
 NASK National Research Institute
 ul. Kolska 12
 01-045 Warszawa
 Poland

Information for Authors

Journal of Telecommunications and Information Technology (JTIT) is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review.** Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The authors JTIT LATEX style file is available:

<https://www.itl.waw.pl/en/jtit-for-authors>

Papers published should contain up to 10 printed pages in LATEX authors style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

- [1] Y. Namihiro, Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres, *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.
- [2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.
- [3] S. Demri and E. Orłowska, Informational representability: Abstract models versus concrete models, in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314

Biographies and photographs of authors. A brief professional authors biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Copyright. Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.

For copyright form see:

<https://www.itl.waw.pl/en/jtit-for-authors>

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the worlds most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishings databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(Contents Continued from Front Cover)

Speech-Based Vehicle Movement Control Solution

G. Kaur, M. Srivastava, and A. Kumar

Paper

72

Tractography Methods in Preoperative Neurosurgical Planning

M. Koryciński and K. A. Ciecierski

Paper

78

COVID-19 Pandemic and Internet Traffic in Poland: Evidence from Selected Regional Networks

M. P. Karpowicz

Paper

86



National Institute
of Telecommunications

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
<http://www.nit.eu>