On Pulse Shaping for Generalized Faster than Nyquist Signaling with and without Equalization

Jovan Milojković, Srđan Brkić, *Member, IEEE*, Jelena Ćertić, *Member, IEEE*

***Abstract*—This paper focuses on analyses of generalized Faster than Nyquist (FTN) signaling in the presence of additive white Gaussian noise. A new method for designing pulse shaping filters, that maximize information rate and simultaneously obey constraints related to energy distribution of the pulse autocorrelation function, is proposed. The obtained pulses are coupled with the minimum mean square error (MMSE) equalizer used at the receiving side. In addition, potential for their use without any equalization scheme is also analyzed. Significance of the proposed approach is verified by comparing designed pulses with state-of-the-art FTN schemes, that employ raised cosine pulses, in terms of bit error rate and achievable information rate. We identify cases when the proposed scheme provides the same achievable information rate as the standard FTN system with more than 1.5 dB lower signal-to-noise power (SNR) ratio, without the equalization, and 0.4 dB lower SNR ratio, if MMSE equalization is employed.**

***Index Terms*—Faster than Nyquist signaling, MMSE equalization, pulse shaping**

I. Introduction

In a classical communication system information is transmitted by using orthogonal pulses (with flat frequency spectrum) and ideally no inter-symbol interference is introduced by the transmitter, which is refereed to as the Nyquist signaling approach. Intense research over the past years in the area of channel coding and modulation lead to development of Nyquist transmission systems that operate close to the ultimate Shannon spectral efficiency bounds, and obviously, additional improvement must be followed by the change of Nyquist's transmission paradigm. With the invention of new services, mostly associated with the fifth generation standard for broadband cellular networks (5GNR), the problem of designing spectral efficient transmission system comes again under the spotlight. A technique that is capable to provide a quantum leap in design of spectrally efficient systems is faster then Nyquist (FTN) signalling.

In FTN signaling systems use of orthogonal pulses is abandoned and inter-symbol inference is intentionally introduced. It follows that as the adjacent received symbols are correlated, conventionally used symbol-by-symbol detection becomes inappropriate, and information needs to be extracted by some equalization technique. The foundations of FTN were laid down by Mazo in 1975 [1], who noticed that communicating with symbol rates higher than the Nyquist rate, can provide spectrally efficient transmission, without degradation in Euclidean distances between transmitted symbols. Although the aforementioned insight was revolutionary, it was not fully explored for more than 30 years. Namely, in 2007 Rusek and Anderson [2] created the information-theoretic framework for the analysis of FTN systems and proved that capacity of conventionally used Nyquist systems (for example with raised cosine (RC) pules) can be surpassed with FTN signaling. Furthermore, the same authors showed in [3] that, by increasing the symbol transmission rate, constrained capacity saturates to a fixed value – in other words arbitrary spectral efficiency can be achieved. Their work was refined recently by Ishihara and Sugiura in [4], where it was shown that conventionally used RC pulses employed in precoded FTN systems approach Shannon capacity of the ideal rectangle pulse. For excellent overview on FTN concepts and technologies we direct readers to [5].

In order to keep the advantages of FTN over Nyquist signaling in practical systems, the equalization and channel coding need to be adjusted. The optimal FTN receiver is organized in a form of turbo equalization loop [6], where equalization is performed by employing maximum a posteriori probability (MAP) detector. Complexity of MAP detection grows exponentially with increase of symbol rate, making it infeasible for practical use. On the opposite side, using low complex equalization, for example MMSE (Minimum Mean Square Error), may be insufficient to perform significantly beyond conventional Nyquist systems. This lead to *generalized FTN signaling* approach [7], in which conventional pulses, like RC, are replaced with pulses that are adjusted to a given equalization scheme. Another benefit of custom pulse design is ability to adopt to a given practical requirements, for example peak-to-average power ratio (PAPR) or adjacent channel leakage power (ACLP), which can vary from one communication standard to another.

Over the years, different pulse designs were proposed that can be incorporated into generalized FTN concept. The most prominent approaches include pulses designed: i) to minimize Euclidean distance between different realizations of two random transmitted sequences [8] and bit error rate of uncoded transmission [9], ii) to maximize information rate [2], [10], [11], or iii) to obey predefined frequency domain constraints (and simultaneously limit PAPR) [12]. For example, Rusek and Anderson in [2] proposed an optimization procedure in which pulses with predefined number of taps maximize an upper information rate bound. The procedure was further expanded by Brkic *et al.* in [11] to enable arbitrary

Jovan Milojković, Srđan Brkić and Jelena Ćertić are with the University of Belgrade, School of Electrical Engineering, 11000 Belgrade, Serbia (e-mails: mj205018p@student.etf.bg.ac.rs, srdjan.brkic@etf.rs, jelena.certic@etf.rs).

 ISBN 978-86-7466-930-3

pulse energy distribution in time domain, which can be used to build generalized FTN systems with limited trellis-based equalizer complexity.

In this paper we further extend the optimization procedure presented in [2], [11] in order to make it applicable to FTN systems with MMSE equalizers, or even to FTN transmission systems which do not employ any equalization scheme. Namely, we define additional time domain constraints related to energy of the pulse autocorrelation function. We verify that designed pulses outperform state-of-the-art FTN systems (with RC pulses), for the same symbol rate and ACLP, in terms of bit error rate as well as achievable information rate in additive white Gaussian noise (AWGN) channel.

The rest of the paper is organized as follows. In Section II we briefly describe system model and MMSE equalization scheme. Section III is dedicated to the optimization procedure, while numerical results are given in Section IV. Finally, concluding remarks can be found in Section V.

II. System Model

Consider an output of the baseband equivalent of the transmitter

$$s(t) = \sum_{k=-\infty}^{\infty} a_k h(t - kT), \tag{1}$$

where $a_k \in \{\pm 1\}$ corresponds to the k-th transmitted symbol, T denotes symbol duration and $h(t)$ is the pulse shaping filter. We assume that $h(t)$ is not orthogonal with respect to the transmitting sample rate, i.e, it intentionally introduces ISI and that $h(t)$ can be represented as weighted sum of wider band pulses

$$h(t) = \sum_{l=0}^{L-1} b_l \psi(t - lT), \tag{2}$$

where $\psi(t)$ is a pulse orthogonal to the sampling rate $1/T$, while $\mathbf{b} = (b_0 \ldots, b_{L-1})$ corresponds to a vector of the sampled coefficients, i.e, $b_l = h(lT)$. Note that we restrict the effect of ISI to L consecutive symbols and that energy of the impulse response is considered to be unitary.

The waveform $s(t)$ is transmitted through additive white Gaussian (AWGN) channel, described with energy per symbol to noise power spectral density (E_s/N_0) metric.

At the receiver side in this paper we consider two observation models: i) orthogonal basis model (OBM) [13] and ii) Ungerboeck model [14] without the equalization. According to the OBM, receiving filter is matched to $\psi(t)$ (not $h(t)$), which means that noise at the receiver input is white. In our model we assume that $\psi(t)$ is square root raised cosine pulse with roll-off 0.1. To verify the performance of $h(t)$ on the OBM, we employ MMSE equalizer. Motivated by the recent findings, reported in [15], that FTN signaling can perform satisfactory even without an equalizer implemented in the receiver side, we here consider such detection scheme, for the case of the Ungerboeck observation model. In the Ungerboeck model receiving filter is matched to $h(t)$, which maximizes E_s/N_0 metric; however, received noise sequence becomes correlated.

We next briefly explain MMSE equalization used in the OBM (for more details one could see [16]). Let $\mathbf{r}_n = (r_{n-N_2}, r_{n-N_2+1}, \ldots, r_{n+N_1})^T$ denote a sequence at input of the MMSE equalizer used to estimate transmitted symbol a_n, where T is transposition sign. The parameters N_1 and N_2 specify the length of the noncausal and the causal part of the MMSE filter. Then, finite impulse response (FIR) MMSE filter coefficients $\mathbf{c}_n$ can be obtained as follows

$$\mathbf{c}_n = \mathrm{Cov}(\mathbf{r}_n, \mathbf{r}_n)^{-1} \times \mathrm{Cov}(\mathbf{r}, a_n), \tag{3}$$

where the covariance operator is given by $\mathrm{Cov}(\mathbf{x}, \mathbf{y}) = E(\mathbf{x}\mathbf{y}^H) - E(\mathbf{x})E(\mathbf{y}^H)$, where $E(\cdot)$ denotes mathematical expectation, and H is Hermitian operator. The symbol estimate $\hat{a}_n$ is obtained by $\hat{a}_n = \mathbf{c}_n^H \mathbf{r}_n$. It should be noted that values N_1 and N_2 are dependent on $h(t)$.

III. Optimization of pulse shaping filter

In this section we state the optimization problem for finding pulse shaping filter coefficients $\mathbf{b}$ that maximize the achievable information rate (AIR) and are also adjusted to the equalizing scheme. Fundamentals of the AIR-based optimization can be found in [17], and we first briefly explain key concept of the optimization, and then highlight modifications introduced in order to adjust the optimization procedure to the system model, given in Section II.

The procedure from [17] allows pulse design for arbitrary ACLP value and the filter length L. Let $H(f)$ denotes Fourier transform of $h(t)$. Then we can define, a complement of ACLP, the concentration of pulse energy in W Hz as follows

$$\beta = \frac{\int_{-W}^{W} |H(f)|^2 df}{\int_{-\infty}^{\infty} |H(f)|^2 df}. \tag{4}$$

Without loss of generality, we can consider normalized bandwidth defined as $w = W \times T$, and notice that $w = 0.5$ corresponds to Nyquist signaling while when $w < 0.5$ we are communicating in FTN signaling fashion. The amount of ISI introduced at transmitter side is inversely proportional to w. Alternatively, the energy concentration β can expressed in a more suitable form, as a function of pulse discrete autocorrelation function $\mathbf{g} = (g_{-L+1}, \ldots, g_{L-1})$, as follows [11]

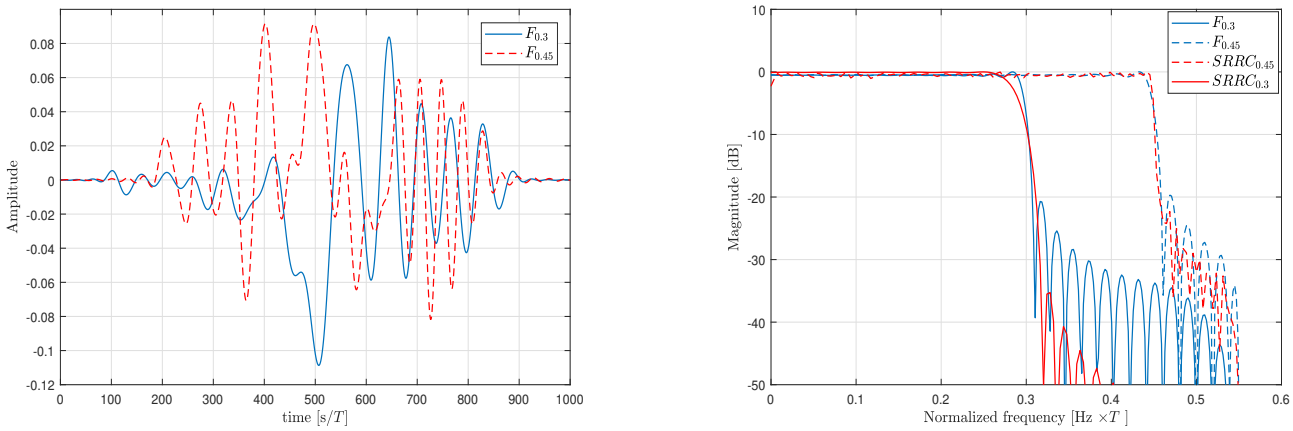
$$\beta(\mathbf{g}, w) = \sum_{l=-L}^{L} 2 g_l w \times \mathrm{sinc}(2\pi w), \tag{5}$$

where $\mathrm{sinc}(x) = \sin(x)/x$, and

$$g_l = \int_{-\infty}^{\infty} h(t) h^{\star}(t - lT) dt. \tag{6}$$

Direct maximization of AIR, for a given discrete modulation set is infeasible as, to the best of our knowledge, a closed form expression does not exist. Instead, it is common to optimize

 ISBN 978-86-7466-930-3



(a) (b)

Fig. 1: Impulse (a) and (b) frequency responses of designed filters for $w = 0.3$ and $w = 0.45$ ($L = 50$, $\beta = 0.999$).

an information rate upper bound $C(\mathbf{b})$, derived assuming Gaussian sample distribution [17]

$$C(\mathbf{b}) = \int_0^{1/2} \log_2 \left[1 + \frac{2E_s|B(f)|^2}{N_0 T} \right] df, \tag{7}$$

where $B(f) = \sum_{k=0}^{L-1} b_k e^{i2\pi kf}$ represents Fourier transform of the vector $\mathbf{b}$.

It was reported in [11] that in large number of cases $C(\mathbf{b})$ monotonically increases with actual AIR obtained through computationally hungry Monte Carlo simulation, which means that optimizing $C(\mathbf{b})$ is meaningful.

By studying typical behavior of pulses optimized by (7), for a fixed β and w constraints we noticed the following: i) relatively small number of filter taps is sufficient to obey strict b constraint (for example $\beta = 0.999$) and ii) significant portions of energy of the pulse autocorrelation are spread across a large number of taps, i.e., the account of introduced ISI is large. Obviously, such pulses cannot be used in systems that do not use equalizers, and are inadequate when equalization is performed with the MMSE equalizer, given its modest ability to suppress ISI.

To resolve the aforementioned issue, we proposed that additional constraint is added into optimization setup that will force the optimization procedure to cluster the majority of the autocorrelation energy to main tap g_0 and potentially $2M$ adjacent taps ($g_{-M},, g_{-1}, g_1, \dots g_M$). Namely, we define a set of thresholds $f\ 0 < f_i < 1$, $0 \leq i \leq M$, forcing the relative energy of central autocorrelation taps to be above the thresholds. However, given frequency-time duality principle, clustering autocorelation energy makes it harder to satisfy frequency domain constraint β. To overcome the problem, the filter lengths L must be increased.

For predefined normalized bandwidth w, with energy concentration β_0 we formally express the optimization problem as follows

$$\begin{aligned} \mathbf{b}_{opt} &= \underset{\mathbf{b}}{\operatorname{argmax}}\ C(\mathbf{b}) \\ s.t.\ \ &\beta(\mathbf{g}, w) = \beta_0, \\ &\frac{g_i^2}{\sum_{\ell=-L+1}^{L-1} g_\ell^2} \geq f,\ 0 \leq i \leq M. \end{aligned} \tag{8}$$

The above optimization problem can be solved similarly as the related problems described in [17] and [11], by sequential quadratic programming (SQP) method.

It should be noted that in OBM system described in Section II, we do not match the receiving filter to transmitting $h(t)$, which means that amount of ISI collected by the receiver is not directly expressed by autoccorelation of $h(t)$. However, we observed strong dependency between ability of MMSE equalizer to suppress ISI and the energy concentration of the pulse autocorrelation function.

To illustrate our optimization procedure, we designed two pulses with $\beta = 0.999$, $L = 50$ and $w = 0.3$ and $w = 0.45$, respectively, denoted by $F_{0.3}$ and $F_{0.45}$, design to obey $f_0 = 0.58$ and $f_0 = 0.85$, respectively. Their impulse and frequency responses are depicted in Fig. 1. For compression we also give frequency responses of square root raised cosine (SRRC) pulses with roll-offs equal to 0.1, designed to meet the same requirements as optimized pulses in terms of length and energy concentrations in frequency domain (denoted by $SRRC_{0.3}$ and $SRRC_{0.45}$).

IV. Numerical results

In this section we provide the performance of designed pulses $F_{0.3}$ and $F_{0.45}$, in terms of bit error rate and achievable information rates, obtained by Monte Carlo simulation (Figs. 2 and 3). We examine pulses behaviour on OBM and Ungerboeck system models, introduced in Section II. Obtained results are compared to SRRC pulses ($SRRC_{0.3}$ and $SRRC_{0.45}$). Given the fact that we only consider binary

 ISBN 978-86-7466-930-3

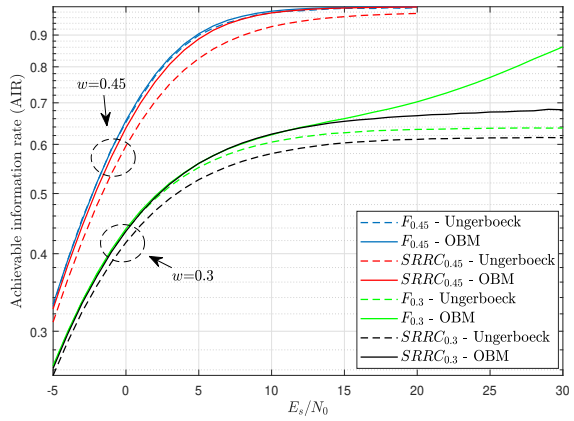


Fig. 2: Achievable information rates of designed pulses.

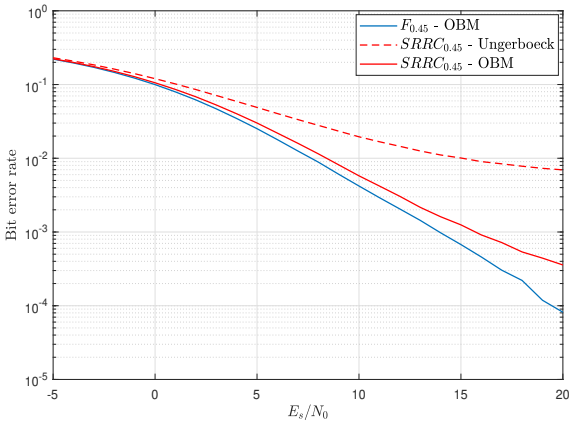


Fig. 3: Bit error rate achievable by $F_{0.45}$ filter.

transmission ($a_k \in \{\pm 1\}$), achievable information rates are calculated numerically as follows

$$AIR = \frac{1}{2} \sum_{a_k=-1,1} \int_{-\infty}^{+\infty} p(z|a_k) \times \log_2 \frac{2p(z|a_k)}{p(z|1) + p(z|-1)} dz, \tag{9}$$

where the conditional probability density function of the log likelihood ratio z of the symbols that corresponds to transmitted a_k, $p(z|a_k)$, is approximated by a histogram.

In Fig. 2 we show that designed pulses outperform SRRC counterparts in terms of AIR on the both system models. For example, $F_{0.45}$ achieves AIR, equal to 0.8 bits/symbols, with 0.4 dB less E_s/N_0 value compared to $SRRC_{0.45}$ on the OBM, while the differences on the Ungerboeck model is approximately 1.8 dB. One can also notice that $SRRC_{0.3}$ cannot achieve information rates above 0.68 bits/symbol in the OBM, while if $F_{0.3}$ is used, higher spectral efficiencies are possible. If we consider uncoded transmission we can, notice that $F_{0.45}$ achieves bit error rate of 10^{-3} with 1.7 dB less E_s/N_0 compared to $SRRC_{0.45}$ on the OBM (Fig. 3).

V. Conclusion

This paper provides novel pulse shaping filters applicable to generalized FTN signaling systems with MMSE equalization and also to systems with no equalizer employed. We show that proposed pulses outperform SRRC pulses of the same structural properties. Our future work will be oriented into examining finite length coded transmission systems that employ proposed pulses.

Acknowledgment

Srđan Brkić acknowledge the support of the Science Fund of the Republic of Serbia, grant No 7750284, Hybrid Integrated Satellite and Terrestrial Access Network - hi-STAR. This work was also supported by the Serbian Ministry of Education, Science and Technological Development.

References

[1] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Tech. J.*, vol. 54, p. 1451–1462, Oct. 1975.
[2] F. Rusek and J. B. Anderson, "Constrained capacities for faster than Nyquist signaling," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, p. 764–775, Feb. 2007.
[3] F. Rusek, D. Kapetanovic, and J. B. Anderson, "The effect of symbol rate on constrained capacity for linear modulation," in *Proc. IEEE Inter, Symp. Inf. Theory 2008*, 6-11 July, 2008, p. 1093–1097.
[4] T. Ishihara and S. Sugiura, "Precoded faster-than-Nyquist signaling with optimal power allocation in frequency-selective channel," in *Proc. IEEE Int. Conf. Commun. Workshop*, June 2021, pp. 1–6.
[5] T. Ishihara, S. Sugiura, and L. Hanzo, "The evolution of faster-than-Nyquist signaling," *IEEE Access*, vol. 9, no. 6, p. 86535–86564, June 2021.
[6] C. Douillard, M. Jezequel, C. Berrou, A. Picart, and P. Didier, "Iterative correction of intersymbol interference: turbo-equalization," *European Trans. Telecomm.*, vol. 6, no. 5, pp. 507–512, June 1995.
[7] J. Zhou, D. Li, and X. Wang, "Generalized faster-than-Nyquist signaling," in *Proc. IEEE Inter, Symp. Inf. Theory 2012*, 1-6 July, 2012, p. 1478–1482.
[8] A. Said and J. B. Anderson, "Design of optimal signals for bandwidth efficient linear coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, p. 701–713, Mar. 1998.
[9] F. Rusek and J. Anderson, "Near bit error rate optimal partial response signaling," in *Proc. Intern. Symp. Information Theory*, Sept. 2005, pp. 538–542.
[10] A. Modenini, F. Rusek, and G. Colavolpe, "Optimal transmit filters for isi channels under channel shortening detection," *IEEE Trans. Commun.*, vol. 61, no. 12, p. 4997–5005, Dec. 2013.
[11] S. Brkic, P. Ivanis, and A. Radosevic, "Faster than Nyquist signaling with limited computational resources," *Physical Communication*, vol. 47, pp. 1–12, June 2021.
[12] T. Delamotte and G. Bauch, "Pulse shaping for satellite systems with time packing: An eigenfilter design," in *Proc. 10th Inter. ITG Conf. on Sys., Commun. and Coding, SCC 2015*, 2-5 Feb. 2015.
[13] J. B. Anderson, A. Prlja, and F. Rusek, "New reduced state space BCJR algorithms for the ISI channel," in *Proc. Inter. Symp. on Inf. Theory*, June 2009, pp. 889–893.
[14] G. Colavolpe and A. Barbieri, "On MAP symbol detection for ISI channels using the ungerboeck observation model," *IEEE Comm. Letters*, vol. 9, no. 8, pp. 720–512, Aug. 2005.
[15] I. Lavrenyuk, A. Ovsyannikova, S. Zavjalov, S. Volvenko, and W. Xue, "Analysis of joint application of optimal FTN signal and 5G error-correction code schemes," in *Proc. 2020 IEEE Inter. Conf. on Electrical Engineering and Photonics (EExPolytech)*, Oct. 2020.
[16] M. Tuchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 673–683, Mar. 2002.
[17] F. Rusek and J. B. Anderson, "Maximal capacity partial response signaling," in *Proc. IEEE 2007 Inter. Conf. Commun.*, 2007, pp. 821–826.

 ISBN 978-86-7466-930-3

Performance simulation for LCR of MIMO Multi-branch SC Diversity System in α-μ fading and α-μ interference channel

Dejan Milić, Suad Suljović, Dejan Rančić, Nenad Petrović, Nenad Milošević

***Abstract* - In order to improve the overall performance of the network in 5G telecommunication networks, Multiple Input and Multi Output Technology (MIMO) is applied. In this paper, the mean number of Level Crossing Rate (LCR) of MIMO systems with L-branch selection combining (SC) receiver is analyzed. During signal transmission, its distortion occurs due to low α-μ fading and α-μ co-channel interference (CCI) effects. Additionally, we applied an accelerated graphics processing unit (GPU) simulation to plan a QoS-efficient 5G mobile network in a smart city. This approach in combination with linear optimization and deep learning significantly optimizes the LCR calculation speed for the observed communication system type, while providing efficient planning - reducing costs, but maximizing performance.**

***Index Terms*—SC combining, α-μ fading, α-μ interference, LCR, QoS, GPU, linear optimization.**

I. Introduction

Due to the unpredictable and dynamic nature of the wireless channel environment, the channel becomes one disruptive element in the transmission chain as it changes the broadcast signal. Therefore, the channel manages the performance of wireless communication systems [1].

Multiple Input Multiple Output (MIMO) is an efficient antenna technology for wireless communication systems where multiple antennas are used on the transmitter and receiver [2]. Antennas at each end of the system are combined to minimize errors during signal transmission, increase data rates and improve channel capacity. This technique allows signals to be transmitted on many different paths at the same time. This mode of signal transmission provides the ability for signals to reach the receiver without fading and co-channel interference. This technique increases the signal-to-noise ratio (SNR) and transmission quality, which creates more stable connections [3].

Fading is a variation of attenuation, ie. signal amplitude fluctuations. During transmission, the signal faces various obstacles such as buildings, trees, etc. present in the signal-causing environment is subject to reflection, diffraction, scattering and shading. This presence of multiple reflectors in the environment of the channel between the transmitting and receiving ends creates more paths for signal passage [4].

The most commonly used signal processing techniques in diversity systems are maximum ratio combination (MRC), equal gain combination (EGC), and selection combination (SC). MRC provides the best improvement in system performance, followed by an equal combination of reinforcements, but it is also the most complicated technique. In order to reduce the complexity of the receiver, this paper considers a simpler combination scheme related to combination selection (SC). The output of the SC receiver is the branch with the highest signal-to-noise ratio. SC has been extended to the case where signals on more than one receiving antenna are combined with the highest current SNR, this scheme is called hybrid maximum ratio selection/combination (HS/MRC). Selection Combination (SC) techniques have been applied in the design of MIMO systems to reduce system complexity and costs. Fifth generation (5G) telecommunications offers a 10-fold increase in spectral efficiency and a 1000-fold increase in system capacity compared to 4G technology [5]. In MIMO systems, there are tens to hundreds of antennas in the receiver and transmitter. Massive MIMO techniques use familiar channel features to deliver superior performance in wireless communications. Channels are modeled to include channel variations in frequency and time.

Describing and modeling channels with fading is of particular importance in mobile communications both for the design of the transceiver system and for performance analysis. During the long period of development of wireless communications, a large number of different models of channels with fading were constructed to describe the statistics of the envelope and the phase of the channel where the signal propagates in several paths, [6]. Examples of such models are Rayleigh's, Rice's, Nakagami-q, Nakagami's, Weibull's, Beckmann's, $\alpha - \mu$ etc. The aim of this paper is to study the statistical properties of first and second order

Dejan Milić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: dejan.milic@elfak.ni.ac.rs)

Suad Suljović is with the Academy of Technical Vocational of Belgrade, Department of Computer Science, Katarine Ambrozić 3, 11000 Belgrade, Serbia, (e-mail: suadsara@gmail.com)

Dejan Rančić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: dejan.rancic@elfak.ni.ac.rs)

Nenad Petrović is with the is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: nenad.petrovic@elfak.ni.ac.rs), (https://orcid.org/0000-0003-2264-7369)

Nenad Milošević with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: nenad.milosevic@elfak.ni.ac.rs)

 ISBN 978-86-7466-930-3

envelopes and phases in these models, with with special reference to the α − μ model.

In mobile communications, the received signal varies in a wide range of values. Therefore, for the design of digital and analog systems, it is necessary to know the statistical characteristics of the signal. There are first and second order statistical characteristics. It is especially necessary to know the statistical characteristics of the second order, such as the average number of axial sections of the LCR and the average duration of AFD fading. These quantities provide additional information that, when combined with other statistics, allows designers to create rational system solutions [7].

Level Crossing Rate (LCR) and Average Duration of Fades (ADF) are important second-order statistical characteristics for describing fading channels. These quantities are useful in the design of mobile radio communication systems and for the analysis of their performance. In digital telecommunications, a sharp drop in the envelope value of the received signal directly leads to a sharp increase in the probability of error [8]. Second-order statistics calculation techniques are particularly applicable to diversity systems that have been shown to be very useful in reducing the impact of fading.

In this paper, we consider a 5G communication system that works over k-ading fading channels and k-μ co-channel interference. One of the most basic statistics of a wireless communication system operating in a fading environment is the mean number of axial cross-sections (LCR). The quality of service (CoS) in wireless communications depends significantly on the LCR as it allows the estimation of the minimum distance between two base stations. We first compute the cumulative distribution function (CDF) for the signal in the described environment, and then we derive a closed-form expression for the LCR. The results of this analysis can be used to design an optimal receiver for a 5G mobile network in smart cities in conditions of small α-μ fading and α -m interference.

II. LCR OF SIGNAL TO INTERFERENCE RATIO AT THE OUTPUT OF THE L-BRANCH SC RECEIVER

In this section, the statistics of the second order 5G wireless communication system with SC receiver with L branches are considered. The received desired signal experiences α-μ fading, while the interference on the co-channel is subjected to α-μ fading. The model of the receiver is shown in Fig. 1.

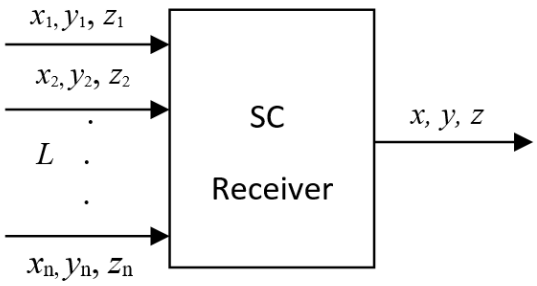


Fig. 1. The model of the SC receiver with *L* branches

Signals $x_1, x_2, \ldots, x_n$ come to the inputs of the SC combiner. The SC receiver with L input branches *(L = 2, 3, ... n)* selects the signal from the antenna with the highest SNR. Also, interference envelopes in the co-channel interference appear as $y_1, y_2, \ldots, y_n$ at each of the L inputs of the SC receiver, while the corresponding output signal is y. The PDF signal envelope on input at SC receiver have the α-μ distribution [9]:

$$p_{x_i}(x_i) = \frac{\alpha\mu^{\mu} x_i^{\alpha\mu-1}}{\Omega_i{}^{\mu}\Gamma(\mu)} e^{-\mu\frac{x_i^{\alpha}}{\Omega_i}} \tag{1}$$

where α is a positive parameter, Ω_i is the mean value of the signal power, μ is the number of clusters and Γ(·) is a Gamma function. In a described channel, the co-channel interference signal follows α-μ distribution:

$$p_{y_i}(y_i) = \frac{\alpha\mu^{\mu} y_i^{\alpha\mu-1}}{s_i{}^{\mu}\Gamma(\mu)} e^{-\mu\frac{y_i^{\alpha}}{s_i}} \tag{2}$$

where *s* is average power of *y*, $y_i \geq 0$. The ratio of the desired signal envelope and interference on the *i*-th input branch of the SC receiver with the L branch can be written as:

$$z_i = \frac{x_i}{y_i}, x_i = z_i y_i \tag{3}$$

SNR for $i = 2.3, \ldots n$ at the output of the SC receiver is:

$$z = \max(z_1, z_2 \ldots z_i) \tag{4}$$

The probability density function (PDF) of the signal z_i is given by [10]:

$$p_{z_i}(z_i) = \int_0^{\infty} \mathrm{d}y_i y_i p_{x_i}(z_i y_i) p_{y_i}(y_i) = \frac{\alpha z_i^{\alpha\mu-1}(\Omega_i s_i)^{\mu}\Gamma(2\mu)}{\Gamma^2(\mu)\left(\Omega_i + s_i z_i^{\alpha}\right)^{2\mu}} \tag{5}$$

Cumulative distribution function (CDF) of z_i is [10]:

$$F_{z_i}(z_i) = \int_0^{z_i} \mathrm{d}t p_{z_i}(t) = \frac{\Gamma(2\mu)}{\Gamma^2(\mu)} B_{\frac{s_i z_i^{\alpha}}{\Omega_i + s_i z_i^{\alpha}}}(\mu, \mu) \tag{6}$$

where $B_z(a, b)$ is the incomplete Beta function, [11; 8.39]. SNR for $i = 2.3, \ldots n$ will be at the output of SC combination:

$$z = \max(z_1, z_2 \ldots z_i) \tag{7}$$

First derivative of z_{ij} can be written:

$$\dot{z}_{ij} = \frac{1}{y_{ij}}\dot{x}_{ij} - \frac{x_{ij}}{y_{ij}^2}\dot{y}_{ij} \tag{8}$$

 ISBN 978-86-7466-930-3

The derivative of the random process α-k-μ is a Gaussian random process, and a linear combination of Gaussian processes is also a Gaussian random process. Therefore, the conditional Gaussian distribution $\dot{z}$ with zero mean and variance applies:

$$\sigma_{\dot{z}_{ij}}^2 = \frac{1}{y_{ij}^2}\sigma_{\dot{x}_{ij}}^2 + \frac{x_{ij}^2}{y_{ij}^4}\sigma_{\dot{y}_{ij}}^2 \tag{9}$$

where respective variances relating to signal and interference are [12]:

$$\sigma_{\dot{x}_{ij}}^2 = \left(\frac{2\pi f_m}{\alpha}\right)^2 \frac{\Omega_i {x_{ij}}^{2-\alpha}}{\mu}, \sigma_{\dot{y}_{ij}}^2 = \left(\frac{2\pi f_m}{\alpha}\right)^2 \frac{s_i {y_{ij}}^{2-\alpha}}{\mu} \tag{10}$$

where f_m denotes the Doppler frequency. After replacing expression (12) with (11), the variance $\dot{z}_{ij}$ becomes:

$$\sigma_{\dot{z}_i}^2 = \frac{1}{y_i^2}\sigma_{\dot{x}_i}^2 + \frac{x_i^2}{y_i^4}\sigma_{\dot{y}_i}^2 = \frac{1}{\mu y_i^{\alpha} z_i^{\alpha-2}}\left(\frac{2\pi f_m}{\alpha}\right)^2 \left(\Omega_i + {z_i}^{\alpha} s_i\right) \tag{11}$$

In equation (13), fm denotes the Doppler frequency. Conditional probability density functions (CPDF) of $\dot{z}_i$ and z_i are [13]:

$$p_{\dot{z}_i}\left(\dot{z}_i \middle| z_i y\right) = \frac{1}{\sqrt{2\pi}\sigma_{\dot{z}_i}} \mathrm{e}^{-\frac{\dot{z}_i^2}{2\sigma_{\dot{z}_i}^2}},$$

$$p_{z_i}\left(z_i \middle| y_i\right) = \left|\frac{\mathrm{d}x_i}{\mathrm{d}z_i}\right| p_{x_i}\left(z_i y_i\right) = y_i p_{x_i}\left(z_i y_i\right) \tag{12}$$

Conditional joint probability density function (CJPDF) of z_i, $\dot{z}_i$ and y_i is [13]:

$$p_{\dot{z}_i z_i y_i}\left(\dot{z}_i z_i y_i\right) = p_{\dot{z}_i}\left(\dot{z}_i \middle| z_i y_i\right) p_{z_i}\left(z_i \middle| y_i\right) p_{y_i}\left(y_i\right) =$$

$$= p_{\dot{z}_i}\left(\dot{z}_i \middle| z_i y_i\right) p_{y_i}\left(y_i\right) y_i p_{x_i}\left(z_i y_i\right) \tag{13}$$

At the output of the SC receiver, the LCR signal is calculated as the mean value of the first derivation of the signal at the output of the SC receiver. The joint probability density function (JPDF) of z_i and $\dot{z}_i$ [13]:

$$p_{\dot{z}_i z_i}\left(\dot{z}_i z_i\right) = \int_0^{\infty} \mathrm{d}y_i \, p_{\dot{z}_i z_i y_i}\left(\dot{z}_i z_i y_i\right) \tag{14}$$

By integrating the first derivative, averaging is obtained. The rate of transition of the LCR level of the random process zi is [10]:

$$N_{z_i}\left(z_i\right) = \int_0^{\infty} \mathrm{d}\dot{z}_i \dot{z}_i p_{\dot{z}_i z_i}\left(\dot{z}_i z_i\right) =$$

$$= \int_0^{\infty} \mathrm{d}\dot{z}_i \dot{z}_i \int_0^{\infty} \mathrm{d}y_i \, p_{\dot{z}_i}\left(\dot{z}_i \middle| z_i y_i\right) p_{y_i}\left(y_i\right) y_i p_{x_i}\left(z_i y_i\right) =$$

$$= \frac{\sqrt{2\pi}\, f_m z_i^{(2\alpha\mu-\alpha)/2} \left(s_i\Omega_i\right)^{(2\mu-1)/2} \Gamma\left(\left(4\mu-1\right)/2\right)}{\Gamma^2\left(\mu\right)\left(s_i {z_i}^{\alpha} + \Omega_i\right)^{2\mu-1}} \tag{15}$$

The LCR envelope of the signal-to-interference ratio at the output of the mD SC with *n* inputs is [14]:

$$N_{x_i|\Omega_i s_i}\left(z_i\right) = L\left(F_{z_{ij}}\left(z_{ij}\right)\right)^{L-1} N_{z_{ij}}\left(z_{ij}\right) = \frac{L\sqrt{2\pi}\, f_m z_i^{\frac{2\alpha\mu-\alpha}{2}}}{\Gamma^{2L}\left(\mu\right)}$$

$$\frac{\left(s_i\Omega_i\right)^{\frac{2\mu-1}{2}} \Gamma^{L-1}\left(2\mu\right)\Gamma\left(\frac{4\mu-1}{2}\right)}{\left(\Omega_i + s_i {z_i}^{\alpha}\right)^{2\mu-1}} \left(B_{\frac{s_i {z_i}^{\alpha}}{\Omega_i + s_i {z_i}^{\alpha}}}\left(\mu,\mu\right)\right)^{L-1} \tag{16}$$

Fig. 2 shows a graphical analysis of the LCR at the SC receiver outputs given, where, $\Omega_1=\Omega_2=\ldots\Omega_n$, and $s_1=s_2=\ldots s_n$; It is assumed that the correlation between the input branches in the SC receiver is minimal.

III. NUMERICAL AND GRAPHICAL RESULTS

Based on expression (16) in Fig. 2 the LCR signal-to-noise ratio is shown in relation to the transition threshold at the SC receiver output from the L branches.

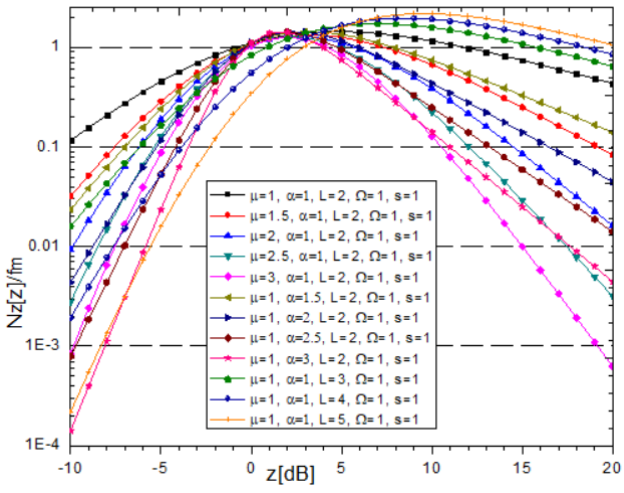


Fig. 2. LCR of system, for different values of parameters μ, α and L.

Fig. 2 shows that for higher values of the signal-to-noise ratio, the increase in the value of the μ parameter LCR decreases. When the parameter α increases then comes to narrowing the LRC function. As the number of L branches at the combinator input increases, the LCR increases by positive values of z [dB], and the system has better performance.

 ISBN 978-86-7466-930-3

IV. PLANNING AND SIMULATION ENVIRONMENT

In this section, we present how the previously derived LCR expression can be leveraged within model-driven network planning environment making use of GPU hardware and multi-objective optimization built upon our previous works [15, 16, 17]. Workflow of the underlying software environment exploiting the synergy of deep leaning and multi-objective optimization is depicted in Fig. 3.

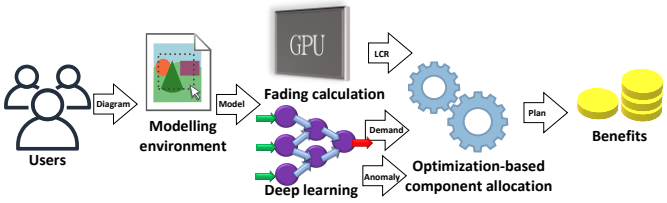


Fig. 3. LCR of system, for different values of parameters μ, α and L.

Within the first step, users construct network diagram inside Eclipse-based environment, according to the structure of underlying network planning metamodel [15, 16]. It considers the following aspects relevant to planning of mobile network inside smart city: operator's base stations, terrain configuration with obstacles, properties of communication channel, adaptive behavior in case of base station anomalies, consumers of telco services (including people equipped with smartphones, autonomous vehicles and smart city infrastructure).

Once modelling is done, the created model is parsed and processed using Ecore [18]. Furthermore, the corresponding parameters are used as input of network-related data processing steps: GPU-enabled LCR determination and deep learning-based predictions. For purpose of fast LCR calculation, we wrote NVIDIA CUDA [19] kernels in C programming language, as described in [17]. In our experiments, it was up to 65.5 times faster than CPU-only program in Mathematica.

On the other side, deep learning module was implemented relying on PyTorch [20] in Python programming language and deployed as Flask service. It covers predictions related to the following factors: number of service users and base station anomalies. For service user count prediction number day, daily temperature and special occasion flag (such as holiday) are considered. On the other side, for base station anomalies ratio prediction the following factors are involved: total downloaded and uploaded data, Quality of Service value, energy consumption and number of users (previously predicted). The layouts of datasets together with achieved prediction performance (Mean Relative Error) are shown in Table I. The problems are treated as regression (real-valued outcome determination), For this purpose, we designed multi-layer perceptron (MLP) with 3 hidden layers, 30 nodes each, making use of Adam optimizer with learning rate 0.01 for training.

TABLE I
OVERVIEW OF PREDICTION MODELS

Model	Input	Approach	Relative error
User number	Day Special Temperature	Regression 3 hidden 30 nodes Adam α=0.01	16%
Anomaly Ratio	Users Download Upload QoS Energy		13%

Finally, in the last step, we make use of our optimization-based model-driven framework for component allocation relying on Pymoo [21], with respect to the model which will be described. The goal of optimization procedure is finding network plan that consists of optimal base station ($bs \in B$) placement for desired smart city locations ($p \in P$), maintaining best possible QoS while keeping lowest costs at the same time. The objective function has following form:

$$minimize \sum_{bs\in BS, l\in L} plan[l,bs](LCR[l,bs]+cost[l,bs]+anomaly[l,bs]) \quad (17)$$

As it can be seen, the sum of LCR (positive impact on performance for lower values), deployment costs and base station anomaly rato is minimized. Here, *plan[l, bs]* denotes decision variable which is 1 in case if base station bs is going to be placed on location l, otherwise 0.

Additionally, we apply a constrint for each location *l* that capacity of base station *bs* (refered to as *maxu[l,bs]*) should be enough to service the predicted number of users (*demandp[l]*) for given location *l*.

$$\sum_{bs\in BS} plan[l,bs]maxu[l,bs] \geq demandp[l], l \in L \quad (18)$$

V. CONCLUSION

In this paper, we have analyzed the MIMO system with a multi-branch SC receiver with L branches when coming to system inputs α - μ fading and CCI with the same distribution. We performed LCR for this model of receiver in the presence of the above-mentioned fading and interference. System performance is more favorable in case when system has lower μ and α parameter, for lower SIR. Analyzing the system, we noticed that the performers are better when the system has a larger number of input branches, because the combiner has the ability to select the branch with the best SIR, leads to better system performance. Finally, for the above fading and CCI system performance is more favorable for lower values of average LCR.

Use of the obtained expression for LCR at the output of a multifaceted SC combinator, we have suggested software environment for simulation. There are several benefits: 1)

 ISBN 978-86-7466-930-3

GPU reduces time for LCR calculation 2) costs are minimized 3) performance is maximized.

ACKNOWLEDGMENT

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Bannour Ahmed, Mohammad Abdul Matin, et al. Coding for MIMOOFDM in future wireless systems. Springer, 2015.
[2] R. S. Kshetrimayum, Fundamentals of MIMO Wireless Communications, Cambridge University Press, 2017
[3] M. K. Simon and M. S. Alouni, Digital Communication Over Fading Channels. 2nd Ed., New Jersey: Wiley-Interscience; 2005.
[4] David Tse and Pramod Viswanath. Fundamentals of wireless communication. Cambridge university press, 2005.
[5] S.K.Yong and J.S.Thompson, "A three dimensional spatial fading correlation model for uniform rectangular arrays", IEEE Antennas and Wireless Propagation, vol. 2, 2003.
[6] M. Patzold Mobile fading channels, John Wiley and Sons, Ltd., England, 2002.
[7] W.C. Jakes, Ed., Microwave Mobile Communications, NJ, Wiley, 1997.
[8] M. Patzold Mobile fading channels, John Wiley and Sons, Ltd., England, 2002.
[9] M. D. Yacoub, "The α−μ distribution: a physical fading model for the Stacy distribution", IEEE Tranactions on Vehicular Technology, vol. 56, no. 1, pp. 27-34, January 2007.
[10] S. Suljović, D. Krstić, D. Bandjur, S. Veljković, and M. Stefanović, "Level crossing rate of macro-diversity system in the presence of fading and co-channel interference", Revue Roumaine des Sciences Techniques, Publisher: Romanian Academy, vol. 64, pp. 63–68, 2019.
[11] I. S. Gradshteyn and I. M. Ryzhik, Tables of Integrals, Series and Products Academic. New York: 1980.
[12] M. Savic, M. Smilic, B. Jaksic, "Analysis of Shannon capacity for SC and MRC diversity system in α-k-μ fading channel", UNIVERSITY THOUGHT, Publication in N. Sciences, Vol.8, No.2, pp. 61-66, 2018.
[13] S. Suljović, D. Milić, S. Panić, Č. Stefanović, and M. Stefanović, "Level crossing rate of macro diversity reception in composite Nakagami-*m* and Gamma fading environment with interference", Digital Signal Processing, Vol. 102, July 2020, 102758.
[14] D. Krstić, S. Suljović, D. Milić, S. Panić, and M. Stefanović, "Outage probability of macro-diversity reception in the presence of Gamma long-term fading, Rayleigh short-term fading and Rician co-channel interference", Annals of Telecommunications, vol. 73, Issue 5–6, pp. 329-339, June 2018. doi: 10.1007/s12243- 017-0593-4.
[15] D. Krstić, N. Petrović, I. Al-Azzoni, "Model-driven approach to fading-aware wireless network planning leveraging multiobjective optimization and deep learning", Mathematical Problems in Engineering", vol. 2022, 4140522, Special Issue: Mathematical Modelling of Data Transmission in Next Generation Wireless Systems, 2022, pp. 1-23, https://doi.org/10.1155/2022/4140522
[16] N. Petrović, S. Koničanin, D. Milić, S. Suljović, and S. Panić, "GPU-enabled framework for modelling, simulation and planning of mobile networks in smart sities", ZINC 2020, pp. 1-6. https://doi.org/10.1109/ZINC50678.2020.9161773
[17] N. Petrović, S. Vasić, D. Milić, S. Suljović, and S. Koničanin: "GPU-supported simulation for ABEP and QoS analysis of a combined macro diversity system in a Gamma-shadowed k-μ fading channel", Facta Universitatis: Electronics and Energetics Vol. 34, No 1, 2021, pp. 89-104. https://doi.org/10.2298/FUEE2101089P
[18] Eclipse Modelling Framework [online], https://www.eclipse.org/modeling/emf/ , last accessed: 22/04/2022.
[19] J. Sanders and E. Kandort, *CUDA by example: an introduction to general-purpose GPU programming*, Addison-Wesley, 2011..
[20] E. Stevens, L. Antiga, and T. Viehmann, *Deep Learning with PyTorch*, *Manning Publications*, Shelter Island, NY, 2020.
[21] I. Al-Azzoni, J. Blank, N. Petrović, "A Model-Driven Approach for Solving the Software Component Allocation Problem", Algorithms 2021; 14(12):354, pp. 1-19, 2021. https://doi.org/10.3390/a14120354

Location Privacy Improvements in Telecommunication Data Management Systems

Milan Simakovic, Zoran Cica, and Dejan Drajic, *Senior Member, IEEE*

***Abstract*—In the era of digital transformation, data are among the most valuable resources. With the development of big data technologies, it is possible to store and process huge amounts of data. Data are possible to collect on every step with high granulation. Such a trend may seriously harm peoples' privacy. Corresponding laws and regulations are declared to protect data privacy. However, even when all the regulations are obeyed, privacy leakage may still happen if the implementation has some flaws. In this paper, we focus on telecommunication data sets and show how user's location information leakage may happen in already privacy-protected data. Moreover, we give a proposition on how this leakage can be prevented while preserving the same data entropy.**

***Index Terms*— data privacy, location tracking, big data.**

I. Introduction

According to [1], more than 66% of the world's population uses the internet at the end of 2021. Dynamic environment, rapid growth, and high competition on the market push companies to further enhance their products and reduce costs. Generated data may contain a huge volume of useful information that will drive such an initiative. And just like that, data becomes one of the main fuels in the industry. This phenomenon is further enhanced with the development of big data technologies [2].

To gather as much information as possible from the data, different ways of data processing are invented. Data correlated from different sources can give new insights that do not exist in separate data sets. Such great potential raises the issue of user privacy. To protect the users' privacy and limit the usage of data, GDPR (General Data Protection Regulation) [3] is defined in the European Union, CCPA (California Consumer Privacy Act) [4] in the USA (United States of America), and PIPL (Personal Information Protection Law) [5] in China. There are also laws that specify data privacy in a particular field, like HIPAA (Health Insurance Portability and Accountability Act) for healthcare in the USA [6]. These laws define data processing by protecting users' privacy and giving individuals the right to control the data collected from them. Although companies comply with all regulations, there are situations that can indirectly harm peoples' privacy. To emphasize this phenomenon, one example of such a scenario is presented in this paper. Namely, we show how the privacy of individuals may be harmed on the already privacy-protected telecommunication data set. Also, a recommendation on how this data set can be further masked to protect users from privacy breach while keeping the same information entropy is presented in the paper.

The remaining of the paper is organized as follows. Related work is discussed in section II. A brief overview of data privacy together with appropriate regulation laws is presented in section III. Section IV presents the main contribution of the paper. It discusses telecommunication data sets, shows the vulnerability of privacy-protected data sets on a sample, and proposes privacy improvement while keeping the same information entropy. Section V concludes the paper.

Milan Simakovic, Zoran Cica and Dejan Drajic are with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia (e-mails: milanrus@hotmail.com, zoran.cica@etf.bg.ac.rs, ddrajic@etf.bg.ac.rs).

Dejan Drajic is with the Innovation Centre of School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia.

II. Related work

Privacy in data technology, especially in big data is a hot topic over the last few years. This section gives a general overview of data privacy challenges, proposed solutions, and frameworks in the literature.

An overview of the privacy-preserving problems in big data stream mining is presented in [7]. Location privacy challenges for mobile applications are discussed in [8]. Location concerns are raised not only to the application provider but to the third parties that are able indirectly to calculate person location from gathered data. In addition, there is a raised concern for services that sell the location data to other parties. Following the people's position and their trajectory is recognized as a serious data privacy concern. To protect trajectory privacy from the data that contain GPS (Global Positioning System) location, a method for data masking that adds noise to the original data based on irregular polygons is proposed in [9]. Due to the rapid development of technology, IoT (Internet of Things) networks are becoming increasingly popular and, thus, becoming significant data generators. Such data may contain privacy-sensitive data. A privacy protection mechanism in industrial IoT based on information tree model is proposed in [10]. Location data is recognized as a huge potential for marketing and advertisement. Considering the number of users, generated amount of data is huge. A method based on big data technologies for location data mining while keeping the data privacy is proposed in [11]. This method uses a clustering algorithm and location entropy to emphasize the most active places.

Companies recognized data privacy as a serious problem and use different methods to solve these challenges. Data

 ISBN 978-86-7466-930-3

privacy models are introduced both for regular [12] and big data systems [13]. Big data models include all the data layers, i.e. collection, storage, and consumption. Data models state for schemas optimized for privacy issues is presented in [13]. In dynamic environments, data is shared both among other teams inside the company and externally. In such a huge data fluctuation, there is a high risk for data privacy issues. Considering that data privacy is protected by law, companies often decide to refrain from using the data which can significantly affect their business efficiency. Model for keeping the big data with a possibility to freely share and explore, and at the same time preserving the data privacy is presented in [14]. Data privacy mechanisms from the k-anonymity, l-diversity, and t-closeness perspectives are discussed in [15]. The advantages and challenges of these mechanisms are analyzed, and a new mechanism based on a combination of these three is proposed. Big data multilayer architecture and utilization of the "differential privacy" approach for sustainable data privacy are discussed in [16].

III. DATA PRIVACY

Data privacy, or information privacy, stands for the ability of the user to control what type of data is collected from him and how these collected data are used. Personal information, depending on the type of application, can be email, location, online search history, preferences, etc. Considering the complexity of modern applications and systems, it is necessary to gather some personal information (e.g. location) to provide the best possible experience to a client. However, applications can often gather more information than they really need and this might bring harm to people's privacy. Collected users' personal data can be used either inside the company to improve internal processes and services or to sell to other companies as a data set. Due to the lack of data security inside the company, data breaches may happen, and personal data can be stolen and used against the clients [17]. Data privacy is often mixed up with data security. While data security is protection from the 3rd party persons to access the data, data privacy is related to data collection methods and regulations that ensure that the user information is not exposed.

To bring closer control of data to the end-user, and restrict the companies in terms of data collection, usage, and surveillance capabilities, many governments around the world have created laws that regulate how data can be used, stored, and protected. Some of the most important regulations are GDPR [3], CCPA [4] and PIPL [5]. GDPR is the data privacy regulation law in European Union. GDPR gives clear instructions on how the data should be collected, transferred, stored, and protected. In addition, it gives users the right to control their personal data, i.e., the "right to be forgotten". This law regulative forces all the companies that collect users' data to have a mechanism to easily wipe all the data for clients without undue delay [3]. CCPA is the data privacy regulation law in the USA. This law stipulates that the user should be aware of what data are collected from him as well as give the company the right to sell his personal data. In addition, CCPA gives a guide to the company on how the law can be implemented [4]. PIPL is the data privacy law in PRC (People's Republic of China). Next to the regulations that prescript GDPR and CCPA, PIPL gives attention to data localization, i.e., that certain categories of personal data must be stored in PRC [5]. Next to the mentioned regulations, there are also many others that are implemented either in other countries or the ones that are industry-focused. For example, HIPAA is the regulation in the USA that governs how personal healthcare data should be handled [6].

To provide data privacy, companies use different techniques for data transformation. Among the most popular techniques are data anonymization and pseudonymization. Anonymization and pseudonymization present the process of transformation of UID (User Identifier) into data sets. Anonymized data stands for one-way encryption meaning that, once the UID is encrypted, there is no theoretical way to re-identify the user from it, neither directly nor indirectly. On the other side, pseudonymization stands for data masking techniques that can be reversed. It means that there is a way to decrypt the data. Pseudonymization is weaker in terms of data protection and should be used carefully. This technique is often used for data that are not related to the UID, but to some attributes. Some of the most popular methods for data pseudonymization are [18]:

- encryption – hiding data by encrypting it,
- shuffling – mixing data inside one column to disassociate attributes from the original user,
- suppression – removing sensitive columns from the dataset,
- redaction – completely removing parts with sensitive data.

Next to these two techniques, there are some other techniques for data privacy protection. Data generalization presents a way to change a value of some column with its range. For example, the value of column age 34 is modified to the range 30-40. Data synthetization presents a method to generate a completely new dataset from the original one using machine learning techniques. A newly generated dataset mimics the properties of original data. Although, these methods are useful in terms of data privacy, they distort information and reduce the data entropy. This is especially evident during the data aggregation.

IV. TELECOMMUNICATION DATA

Telecommunication networks represent a very important factor in the development of humanity. Due to their importance and high competition on the market, telecommunication operators gather data to further optimize services, reduce costs, and improve quality of their networks and services. Data are gathered on all network architecture levels, from physical and core network devices to the application layer. Modern telecommunication network providers typically implement centralized platform based on big data technologies to gather such amount of data.

 ISBN 978-86-7466-930-3

TABLE I
SAMPLE OF SIMPLIFIED MOBILE DATA SET

base station id	calling party	called party	billing	call type	call status	timestamp
bst_drwx_sth	aa9annch2n44	34yv5n9dwqlo	0.2	sms	sent	2022/03/15 08:45:02.000
bst_drwx_sth	i23u6bu546bx	vsf78bt489aa	0.0	call	start	2022/03/15 08:46:16.015
bst_drwx_sth	cw59coj7q33h	64hqu89bu33q	1.4	call	end	2022/03/15 08:46:55.724
…						
bst_mmss_nrt	j4kk9txbryuq	zevynom2w84t	7.2	call	end	2022/03/15 09:33:51.992
bst_mmss_nrt	i23u6bu546bx	quuhe11bcyd4	1.6	call	end	2022/03/15 09:34:08.183
bst_mmss_nrt	2jj3hb56u2bb	oppqnn4bb60u	0.2	sms	sent	2022/03/15 09:34:40.000
…						
bst_lndn_wst	i23u6bu546bx	pp22djqmxirb	1.7	call	end	2022/03/15 12:02:45.501
bst_lndn_wst	curbskoqcgg4	bsgwivg4611b	0.0	call	start	2022/03/15 12:03:10.017

Since telecommunications are always on the top of IT standards, all these systems already implement mechanisms for data privacy that are prescribed in their countries. Although, the implemented mechanisms mask the UIDs, due to the complexity and variety of data, some data privacy breaches may happen. In this section, we show one example of how the data privacy breach in terms of user tracking may happen and propose a masking mechanism that solves the problem for such a scenario while keeping the data entropy at the same level.

Mobile network operators gather data (for example, CDRs (Call Detail Record) and XDRs (Extended Detection and Response)) from base stations. This data contain information about the mobile phone device id, used frequency channel, signal strength, service type, call duration, etc. Such data helps the operators to tune the base stations, maximize the capacity of the cell, and quality of service. An example of simplified data set is shown in Table 1. In the example data set, UID is hashed due to the data privacy regulations. Assuming the anonymization is taken as the hashing mechanism, there is no way to get the UID original value.

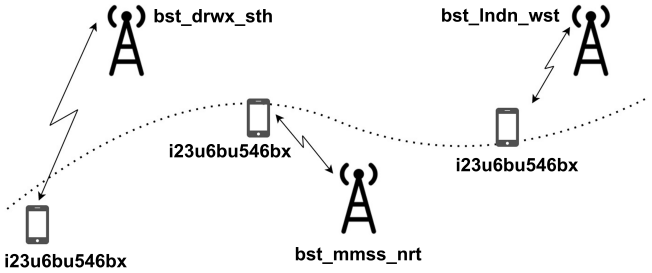


Fig. 1. User location tracking using data from base stations.

Since the same mechanism for data hashing is implemented on the data from every base station, it generates the same hashed UID for a user on every base station. This means that even though the user's UID is not known, it is possible to track user's movement as shown in Fig. 1. From the dataset shown in Table 1, it can be seen how the UID "i23u6bu546bx" is located on different base stations during the day which implies the possibility to track the location of this person. By matching the hashed UID with associated data (like base station ID, call records, etc.) it might be even possible to determine the identity of the user.

Although collected data are used for network improvement, people's privacy can be indirectly harmed in terms of location tracking. Since this information is not relevant to mobile network operators, we propose a new mechanism for data masking. The previous mechanism takes the device identifier (e.g. phone number or MAC address) and creates the hashed id by using some hashing function like shown in (1).

$$hashed\_UID = hash\_function(UID) \quad (1)$$

This mechanism creates the same hashed UID on all base stations. To improve this, we propose to create the hash UID from a combination of base station identifier and UID, as shown in (2). In this way, hashed value of UID will be different at each base station. Thus, information about user movement is removed, but all the necessary information relevant for quality-of-service monitoring are kept.

$$hashed\_UID = hash\_function(base\_station\_ID + UID) \quad (2)$$

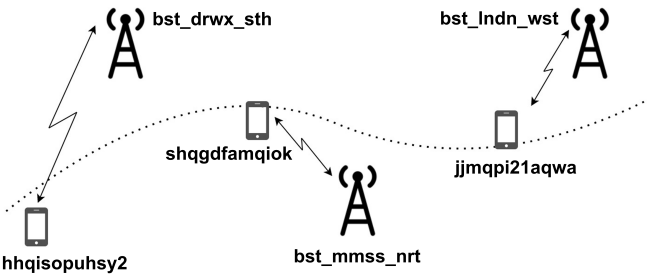


Fig. 2. Avoiding user location tracking.

The data set generated with the proposed hashed function is shown in Table 2. User that corresponds to hashed UID "i23u6bu546bx" in Table 1, has now a different UID on each base station which makes it impossible to correlate the data and track its location. Note that corresponding fields are colored in Tables 1 and 2. Fig. 2 shows the same example as Fig. 1 with a difference that now the UID is hashed with our proposed method. It can be seen in Fig. 2 that hashed UID values for the same user are different at different base stations. Thus, tracking of user movements is not possible anymore. However, if a user is connected to same base station for a

 ISBN 978-86-7466-930-3

TABLE II
SAMPLE OF SIMPLIFIED MOBILE DATA SET WITH PROPOSED HASHING MECHANISM

base station id	calling party	called party	billing	call type	call status	timestamp
bst_drwx_sth	qheuzhnq91nw	34yv5n9dwqlo	0.2	sms	sent	2022/03/15 08:45:02.000
bst_drwx_sth	hhqisopuhsy2	vsf78bt489aa	0.0	call	start	2022/03/15 08:46:16.015
bst_drwx_sth	soek28dh17h3	64hqu89bu33q	1.4	call	end	2022/03/15 08:46:55.724
…						
bst_mmss_nrt	h1ikvgwotggk	zevynom2w84t	7.2	call	end	2022/03/15 09:33:51.992
bst_mmss_nrt	shqgdfamqiok	quuhe11bcyd4	1.6	call	end	2022/03/15 09:34:08.183
bst_mmss_nrt	q11y2ms9h5b6	oppqnn4bb60u	0.2	sms	sent	2022/03/15 09:34:40.000
…						
bst_lndn_wst	jjmqpi21aqwa	pp22djqmxirb	1.7	call	end	2022/03/15 12:02:45.501
bst_lndn_wst	rxihb6ihqwhb	bsgwivg4611b	0.0	call	start	2022/03/15 12:03:10.017

period of time, hashed UID value will be the same, thus, the information about the signal quality for the session and user is preserved.

The same data privacy leak can happen not only in mobile but in other networks as well. For example, HFC (Hybrid Fiber Coaxial) network operators gather data from their cable modems. Considering that nowadays most of the clients are connected to the internet using WiFi (Wireless Fidelity), poor quality of service can be caused either by the poor signal quality on a cable modem or poor WiFi signal. The quality of the WiFi signal depends on many aspects such are the position of the cable modem, the schema of the building, types of walls, etc. Even though HFC operators are not in charge of the quality of WiFi networks, they tend to help clients to improve quality of network either by reconfiguring the WiFi (switching WiFi channel, changing channel width), relocating cable modem to some other place where it will better cover the whole apartment or by adding WiFi extenders.

Information about the cause of the poor signal quality operators find in the data gathered from the cable modems. Modern cable modems have embedded WiFi transmitters. Next to the basic information regarding the quality of the signal, data about the WiFi can be collected as well. Example of data that are gathered is the MAC (Media Access Control) address and WiFi username of the connected device (e.g., from a laptop, tablet, or mobile phone). If the device is connected to several locations that are covered by the same provider (e.g., at home, at café, store, work) the data privacy in terms of movement tracing can be breached. To solve such problem, the same hashing mechanism, (2), we propose for mobile networks can be used. Creating a different hash for UID on different cable modems, would completely remove the possibility to track the user movements.

V. CONCLUSION

Data privacy concerns are raised a few years ago and present one of the most important aspects during the development of data management systems. Depending on country, data management systems implement the appropriate privacy laws. Although the laws are obeyed, data privacy leakage may still happen if implementation is not carefully done. This paper shows an example of location data privacy violations in telecommunication data systems. In addition, we show methodology for how such violations can be solved with domain-specific knowledge. Finally, a proposal for data privacy improvement while keeping the same data entropy is given.

ACKNOWLEDGMENT

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Internet World Stats (site: https://www.internetworldstats.com/stats.htm), Accessed: Mar. 18, 2022.
[2] J. Dean, S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *OSDI - Operating Systems Design and Implementation*, 2004.
[3] Complete Guide to GDPR Compliance (site: https://gdpr.eu/), Accessed: Mar. 18, 2022.
[4] California Consumer Privacy Act(site: https://oag.ca.gov/privacy/ccpa), Accessed: Mar. 18, 2022.
[5] The PRC Personal Information Protection Law (site: https://www.china-briefing.com/news/the-prc-personal-information-protection-law-final-a-full-translation/), Accessed: Mar. 18, 2022.
[6] Health Insurance Portability and Accountability Act of 1996 (site: https://www.cdc.gov/phlp/publications/topic/hipaa.html), Accessed: Mar. 18, 2022.
[7] A. Cuzzocrea, "Privacy-Preserving Big Data Stream Mining: Opportunities, Challenges, Directions," in *proc. of ICDMW 2017*, New Orleans, LA, USA, Nov. 2017.
[8] M. L. Damiani, C. Cuijpers, "Privacy Challenges in Third-Party Location Services," in *proc. of MDM 2013*, Milan, Italy, June 2013.
[9] H. Liu, W. Di, "Application of Differential Privacy in Location Trajectory Big Data," in *proc. of ICITBS 2020*, Vientiane, Laos, Jan. 2020.
[10] C. Yin, J. Xi, R. Sun, J. Wang, "Location Privacy Protection Based on Differential Privacy Strategy for Big Data in Industrial Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3628 - 3636, Aug. 2018.
[11] S. Wang, R. Sinnott, S. Nepal, "Privacy-protected place of activity mining on big location data," in *proc. of Big Data 2017*, Boston, MA, USA, Dec. 2017.
[12] C. Wu, Y. Guo, "Enhanced user data privacy with pay-by-data model," in *proc. of Big Data 2013*, Silicon Valley, CA, USA, Oct. 2013.
[13] X. Feng, "The Optimization of Privacy Data Management Model In Big Data Era," in *proc. of IAEAC 2021*, Chongqing, China, Mar. 2021.

 ISBN 978-86-7466-930-3

[14] Y. Canbay, Y. Vural, S. Sagiroglu, "Privacy Preserving Big Data Publishing," in *proc. of IBIGDELFT 2018*, Ankara, Turkey, Dec. 2018.
[15] R. Mahesh, T. Meyyappan, "Anonymization technique through record elimination to preserve privacy of published data," in *proc. of PRIME 2013*, Salem, India, Feb. 2013.
[16] K. M. Shrivastva, M. A. Rizvi, S. Singh, "Big Data Privacy Based on Differential Privacy a Hope for Big Data," in *proc. of ICRCICN 2014*, Bhopal, India, Nov. 2014.
[17] What is data privacy? (site: https://www.cloudflare.com/learning/privacy/what-is-data-privacy/), Accessed: Mar. 18, 2022.
[18] Which data protection methods do you need to guarantee privacy? (site: https://www.statice.ai/post/data-protection-techniques-need-to-guarantee-privacy#:~:text=Encryption%3A%20hiding%20sensitive%20data%20using,entirety%20of%20a%20column's%20values), Accessed: Mar. 18, 2022.

 ISBN 978-86-7466-930-3

Introducing IoT to Big Data Platform for Network Performance Monitoring

Milan Simakovic, Zoran Cica, and Dejan Drajic, *Senior Member, IEEE*

***Abstract*—Telecommunication operators are collecting large amounts of data for various purposes such as network performance monitoring, network planning, better customer support, etc. Nowadays, big data technologies are commonly used to couple with enormous amounts of data. Thus, telecommunication operators use big data platforms to process and store data collected from their networks. Data are collected from the network and user devices. Since network operators cover most of the residential users, it is possible to use this access to introduce IoT (Internet of Things) and smart city support. By extending the already existing big data platforms to support IoT devices placed at network users' premises, smooth integration of various IoT devices with the smart city concept can be achieved. Such integration would have significant benefits for network operators, users and local communities. In this paper, we propose an extension that introduces IoT support to the existing big data platform used for HFC (Hybrid Fiber-Coaxial) network monitoring. An overview of the most attractive IoT use cases that can benefit from the proposed extension is also presented in the paper.**

***Index Terms*—Big data, IoT, Smart city, Smart home, HFC network.**

I. Introduction

Telecommunication operators typically serve large number of users. In order to successfully perform such task, operators use very complex and heterogeneous telecommunication networks which need to be constantly monitored. For monitoring purposes, operators collect and store data from numerous network devices [1,2]. By processing the collected data, it is possible to achieve optimal network performance and provide high QoS (Quality of Service) to users. Since the amount of collected data is enormous, big data technologies need to be used to efficiently store and process such amount of data [3].

Smart home and smart city are IoT concepts that are becoming very popular nowadays [4,5]. Given the large number of residents and the fact that each user will typically have multiple IoT devices and sensors at home, it is obvious that amount of data collected can be very large. Thus, big data technologies would provide excellent solution for storing and processing such amounts of data [6]. Most of the residential users are covered by some telecommunication operator (HFC, passive optical network (PON), etc.) and the operators use big data platforms to store and process data collected from their networks. Given that fact, these big data platforms can be extended to collect and store data from IoT devices and sensors placed at user premises. In this way, same infrastructure could be reused for IoT purposes. By integrating existing telecom operators big data platform in smart city environment, highly economic and efficient smart city solution can be achieved.

In [7], big data platform for HFC network monitoring is extended with support for efficient failure detection and localization. In this paper, we propose extension of this big data platform to support IoT devices at users' premises. It will be shown, that such extension does not require architecture modifications in the already existing big data platform. Only, adjustments in data collection layer need to be done. Once data are collected and stored, it is up to data consumers (smart city solutions and applications, users, operators) to determine which types of processing of the collected data need to be supported. For example, application for users that graphically shows measured values from their sensors, or notification in case of threshold violation (e.g. smoke detectors detect fire). In this paper, we also present the most attractive use cases that would benefit from such IoT extension of big data platform.

Remainder of the paper is organized as follows. In section II, we give related work overview. Section III presents the existing big data platform for HFC network monitoring, and the introduced extensions to support data collection and storage from IoT devices. Section IV contains a survey of use cases that would benefit from the IoT extension of the big data platform introduced in section III. Finally, section V concludes the paper.

II. Related Work

IoT based smart home solutions represent one of the most important IoT markets since they offer a huge variety of different applications for enhancements of resident's quality of life. It is possible to install different devices for monitoring of various activities. The monitoring can be focused on the resident's activities and on the different parameters in the home. In the elderly healthcare solution, different medical sensors are provided in form of smart bracelets and other wearable devices [8], that allow efficient monitoring of movements and condition of a patient. Different sensors for environmental monitoring could be installed like sensors for temperature, relative humidity, light intensity measurements,

Milan Simakovic, Zoran Cica and Dejan Drajic are with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia (e-mails: milanrus@hotmail.com, zoran.cica@etf.bg.ac.rs, ddrajic@etf.bg.ac.rs).

Dejan Drajic is with the Innovation Centre of School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11120 Belgrade, Serbia.

 ISBN 978-86-7466-930-3

fire/smoke detectors, etc. Based on these measurements, temperature and lighting control can be automatized and appropriate alarms can be raised. Monitoring of energy and water consumption, and water leakage detection can be achieved with appropriate sensors. Also, different appliances in home can be remotely monitored, such as sensors for audio and video surveillance, motion detection that are typically installed to increase home safety [9].

Within the smart cities many urban related problems are aimed to be solved such as air pollution, urban noise monitoring, traffic jams, smart parking, energy consumption, waste management, smart infrastructure, street lighting, assistance to senior citizens, etc [10]. This includes different types of technologies (big data, IoT, WSN (Wireless Sensors Networks) and cloud. Particularly interesting elements for planning Wi-Fi networks in large cities are presented in the paper [11]. Wi-Fi as a transmission technology is preferable in many smart city applications since it provides connectivity to smart phones, computers and many wearable gadgets. Different challenges of Smart City IoT system deployment are addressed in [4], where Security and Privacy, Smart Sensors, Networking and Big Data Analytics are recognized as the most important ones.

IoT concept is one of the most important trends in telecommunications. IoT is integral part of many "smart" solutions like smart city, smart buildings, smart agriculture, smart transport, smart healthcare... [12]. Given the significant amount of data that IoT systems generate, big data is recognized as a "key enabling technology for IoT" [12]. For this reason there are numerous papers that deal with IoT and Big Data combination. Surveys on big data in IoT can be found in [12,13]. A large number of papers on the topic of smart city shows that this concept is one of the most popular ones from the IoT area. Survey on big data in smart city solutions is given in [14]. Industry also embraced IoT concept because digitalization of industrial processes increases efficiency [15]. In [15], Industrial Big Data as a result of IoT adoption in manufacturing is discussed. Big data and IoT in smart farming are discussed in [16]. Although, many IoT solutions are already operating along with big data technologies support, there are still some open challenges. Survey on these open challenges is given in [17].

III. BIG DATA PLATFORM ARCHITECTURE

In this section, first we give a brief description of the existing big data platform architecture for HFC network performance monitoring. Then, we describe extensions necessary to introduce IoT in the existing big data platform.

Fig. 1 shows the big data platform architecture for performance monitoring of HFC network. Big data platform comprises two major parts - big data cluster and data collection layer. Big data cluster uses several big data tools to store and process collected data. The used big data tools include OpenTSDB (Open Time Series Database), Apache HBase, HDFS (Hadoop Distributed File System), Apache Spark, Hadoop YARN (Yet Another Resource Negotiator), and Zookeeper. OpenTSDB is used to verify and process incoming messages from data collection layer, and to store these messages in HBase tables. HBase writes table files to HDFS while HDFS is responsible for storing data on physical storage units. Apache Spark is necessary for data aggregations. Hadoop YARN is used to allocate resources to jobs. YARN also enables different processing frameworks to use common hardware. Zookeeper synchronizes distributed services.

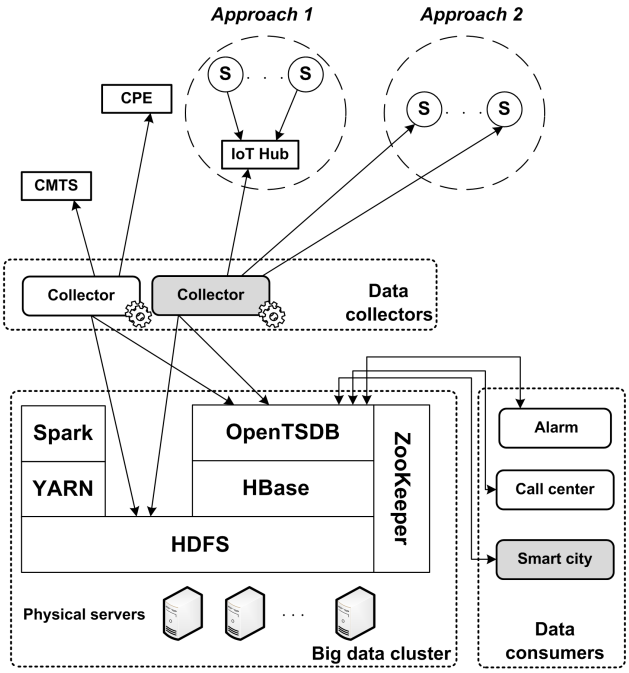


Fig. 1. Big data platform with added support for IoT devices

Data collection layer is responsible for collecting data from CMTS (Cable Modem Termination System) and CPE (Customer Premises Equipment) devices in HFC networks as they represent network devices with a possibility to collect data from them. CPE devices include cable modems and set-top boxes. SNMP (Simple Network Management Protocol) is used for data collection. There are numerous data collectors working in parallel in the data collection layer due to large number of devices in the HFC network from which data need to be collected. Various data (metrics) are collected from CMTS and CPE devices, while the data collection period depends on the data type and importance. Collection period is set in range from 1 minute to 1 hour depending on the importance of collected metrics. Collected data are sent to OpenTSDB via web socket and in parallel data are also written to HDFS.

Fig. 1 also shows the extensions necessary to accommodate IoT support. The updated or added modules are marked with grey coloring. Main update needs to be made in data collection layer. Namely, data collectors need to be aware of new devices and data types that need to be collected. The advantage of the previously described big data platform is its flexibility. Due to large number of different network devices (different vendors, versions, models, etc.), big data platform was developed to efficiently couple with variety challenge

 ISBN 978-86-7466-930-3

that is typical for big data usage [18]. This means that the data collectors are designed to be flexible and adjustable to new devices added to monitored HFC network. Obviously, IoT devices also represent new devices from the data collector point of view. Thus, adding a support for new devices is not a complex task.

The other extension is connection to new data consumers. In the case of IoT introduction that would be IoT data consumers, e.g. smart city environment as shown in Fig. 1. These IoT data consumers can be internal or external depending on the type of integration of the big data platform. For example, the big data platform besides the network monitoring system can serve also as IoT service hub operated by the network operator (internal data consumer). But, the big data platform can provide access to data to external data consumers such as local government smart city platform or external companies like distributors of electricity, water or gas.

At user premises, there are two possible approaches that affect the data collector extension for IoT devices support. One approach is to collect data directly from each IoT device. Second approach is to have a hub for all IoT devices at user premises. In this second case, data collector connects only to a hub which reduces the number of connections necessary to collect all IoT related data from user premises. The other advantage of the hub approach is simplified addition of support for IoT data in data collection layer. Namely, in this case focus is only on the types of IoT metrics not the types of IoT devices. Secondly, this gives more freedom for connecting IoT devices to central hub at home (for example, ZigBee or Bluetooth). Thirdly, if WiFi is used to connect IoT devices to central hub, a central hub might be integrated to cable modem as nowadays these modems typically have WiFi capabilities. Both approaches are illustrated in Fig. 1.

IV. USE CASES

The proposed IoT support extension of the big data platform for HFC network performance monitoring can be applied in many smart monitoring IoT based use cases. IoT and sensor networks represent sources of time series data that can be collected, stored and processed by the proposed extension of big data platform. As discussed in the previous section there are two approaches that can be used to collect data from IoT devices - direct and indirect over IoT hub. Note that these two approaches can coexist if necessary.

Some of the most important and common IoT based smart metering in homes are: indoor air quality monitoring, energy consumption and water consumption monitoring. These metering devices produce periodic and relatively small packet traffic. The most common indoor air quality monitoring parameters are CO, CO_2, PM (Particle matters), VoC (Volatile compounds), temperature and relative humidity. Additionally, some devices support O_2, CH_4, H_2S, NH_3 as well. There are a lot of already available low-cost devices for this purpose [19]. These devices can be used for air quality monitoring and fire detection. Normally, reporting period is set to 5 minutes, but for the alarm purposes it could be set to 1 minute. Expected payload is 0.5 kB per sampling interval. In the case of 1 minute monitoring period, expected data volume per one device would be 720 kB per day, i.e. 21.6 MB per month. If data collectors enrich the collected data with additional information (user id, device model/type,...), the expected data volume can be a bit larger. The required storage space for data collected over one month would be around 5.7 TB under following assumptions: each home has 2 IoT devices in average; data collected from each IoT device requires 30MB per month in average; there are 100000 homes covered by HFC network.

Telecom operators can allow access to the raw data via API (Application Programming Interface). But more attractive for their clients would be aggregation, processing and visualization of their data available via web portal where clients can easily follow the measurements and set alarms and/or notifications, calculate air quality index, etc. In case of different models and versions of IoT devices, operators can perform data aggregations per manufacturer or model to gain insight in the performance of the devices and determine which devices are more preferable and offer them to their clients in future deals. In case of air quality measurement devices, outside units can be mounted as well, for example, on balconies. Data collected from such devices can be aggregated and processed to gain deeper insight in air quality in different parts of the city, or on different height levels in case of multi-storey buildings. This information can be very important for local communities to detect critical zones.

While air quality monitoring requires quite frequent reporting, energy [20] and water [21] consumption can be reported once per day (about 0.4 kB per measurement in both cases, which means 12 kB per month). Energy module collects data about power consumption and other electric parameters, while the water consumption sensor monitors amount of water usage in the home. Measurements can be collected by the platform, where data are stored and processed. The measurements can be presented in form of graphs, and when data reach the specified level of power or water consumption, alerts can be generated and sent to the user. Also, smart city concept assumes interconnection of citizens, local governments and utility companies. Utility companies need to read measurement devices to check the consumption by their clients. If the readings of these devices are integrated with the big data platform, then that data can be passed to utility companies. In this way, multiple benefits can be achieved. Automatized and remote measurement readings, detection of anomalies which can trigger alarm to both utility companies and users. In case of multi-storey buildings, utility measurement devices are typically not part of the homes. In such cases, if the building is covered by HFC network, additional cable modem can be installed that would cover only the utility measurement devices as a part of agreement between the users, operator and utility companies. This actually perfectly represents the idea and aim of smart city concept.

As already explained, presented use cases are not very demanding in terms of storage capacity, so big data platform should be able to easily handle a lot of users. Insight into collected measurements provides to users powerful

 ISBN 978-86-7466-930-3

information how to optimize water and power consumption and how to improve the air quality. Proposed concept facilitates smart home integration and could be expanded for the smart city measurements integration with goal to create pollution and noise maps (devices can be also mounted outside for outdoor air quality monitoring), remote reading of electricity and water consumption, thus, optimizing the work of communal services. On the other hand, the proposed expansion provides an excellent opportunity for the operators to expand their service portfolio and offer new smart services to their users, which would bring a new revenue and increase users' confidence and satisfaction.

The described IoT devices are low-cost and easy for installation and handling. Of course there are a lot of other sensors and devices that could be installed for monitoring of different kinds of parameters, and in this section we presented a few useful examples for the proof of concept purposes.

V. Conclusions

In this paper we propose extension for IoT support of existing HFC network performance platform based on big data technologies. The proposed extension is incremental and not complex, and most importantly does not affect the original purpose (performance monitoring) of the big data platform. The presented use cases that would be enabled by the proposed extension show that significant benefit would be gained by all the parties involved: users, operators, utility companies, local communities. This is exactly in line with the smart city ideas and goals. As a part of our future work, we will focus on tight integration of the proposed extension to smart city solutions with special emphasis on air and noise pollution tracking and creating corresponding noise and pollution maps.

Acknowledgment

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

[1] Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao, R. C. Qiu, "Big Data Analytics in Mobile Cellular Networks," *IEEE Access*, vol. 4, pp. 1985-1996, Mar. 2016.

[2] A. J. Garcia, M. Toril, P. Oliver, S. Luna-Ramirez, R. Garcia, "Big Data Analytics for Automated QoE Management in Mobile Networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 91-97, August 2019.

[3] M. Simaković, Z. Cica, I. Masnikosa, "Big Data Architecture for Mobile Network Operators," in *proc. of TELSIKS 2021*, Nis, Serbia, Oct. 2021.

[4] A. S. Syed, D. Sierra-Sosa, A. Kumar, A. Elmaghraby, "IoT in Smart Cities: A Survey of Technologies, Practices and Challenges," *Smart Cities*, vol. 4, no. 2, pp. 429–475, Mar. 2021.

[5] A. Nag, M.E.E. Alahi, N. Afsarimanesh, S. Prabhu, S.C. Mukhopadhyay, "IoT for smart homes", in book *Sensors in the Age of the Internet of Things: Technologies and applications*, pp. 171-199, 2019.

[6] S. Khare, M. Totaro, "Big Data in IoT," in *proc. of ICCCNT 2019*, Kanpur, India, Jul. 2019.

[7] M. Simakovic, Z. Cica, "Detection and Localization of Failures in Hybrid Fiber–Coaxial Network Using Big Data Platform," *MDPI Electronics*, vol.10, no.23, November 2021.

[8] B.J. Cheng, M.M.A. Jamil, R. Ambar, M.H.A. Wahab, A.A. Ma'radzi, "Elderly Care Monitoring System with IoT Application", in book *Recent Advances in Intelligent Information Systems and Applied Mathematics*, pp. 525-537, 2020.

[9] M.R. Alam, M.B.I. Reaz, M.A.M. Ali, "A Review of Smart Homes—Past, Present, and Future," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1190-1203, Nov. 2012.

[10] J. Temperton, "Bristol is making a Smart City for Actual Humans". Wired, 17 March 2015. http://www.wired.co.uk/news/archive/2015-03/17/bristol-smart-city.

[11] L. Zhang, L. Zhao, Z. Wang, J. Liu, "WiFi Networks in Metropolises: From Access Point and User Perspectives," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 42-48, May 2017

[12] M. Bansal, I. Chana, S. Clarke, "A Survey on IoT Big Data: Current Status, 13 V's Challenges, and Future Directions," *ACM Computing Surveys*, vol.53, no.6, pp. 1-59, Nov. 2021.

[13] E. Ahmed et al., "The role of big data analytics in Internet of Things," *Computer Networks*, vol. 129, no. 2, pp. 459-471, Dec. 2017.

[14] I.A.T. Hashem et al., "The role of big data in smart city," *International Journal of Information Management*, vol. 36, no. 5, pp. 748-758, Oct. 2016.

[15] D. Mourtzis, E. Vlachou, N. Milas, "Industrial Big Data as a Result of IoT Adoption in Manufacturing," *Procedia CIRP*, vol. 55, pp. 290-295, 2016.

[16] N. N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay, A. Martynenko, "IoT, big data and artificial intelligence in agriculture and food industry," *IEEE Internet of Things Journal*, pp. 1-1, May 2020.

[17] S. Kumar, P. Tiwari, M. Zymbler, "Internet of Things is a revolutionary approach for future technology enhancement: a review," *Journal of Big Data*, vol. 6, no. 1, pp. 1-21, Dec. 2019.

[18] M. Simakovic, Z. Cica, "Big Data Applications and Challenges," in *proc. of Infoteh 2016*, Jahorina, BiH, Mar. 2016.

[19] https://www.renkeer.com/product/aqi-sensor/

[20] https://www.acrel.co.id/product/Wireless_energy_meter/IoT_Wireless_energy_meter.html?gclid=Cj0KCQjw5-WRBhCKARIsAAId9FlVSq3lp8SpBQ6TC09F9covmEei2ZZF9_tGw6zH6NRKM8NIb4L8WLMaAj90EALw_wcB

[21] Z.H. Che Soh, M.S. Shafie, M.A. Shafie, S.N. Sulaiman, M.N. Ibrahim, S.A.C. Abdullah, "IoT Water Consumption Monitoring & Alert System," in *proc. of ICELTICs 2018*, Banda Aceh, Indonesia, Sep. 2018.

 ISBN 978-86-7466-930-3

Reliability of Earth-Space Links under Deep Fades with Interleaved Reed-Solomon Codes

Srđan Brkić, *Member, IEEE*, Zoran Čiča, Andreja Radošević, Đorđe Sarač, Predrag Ivaniš, *Senior Member, IEEE*

***Abstract*—This paper contains a reliability analysis of Earth-space links subjected to deep fades, modeled as burst erasure channels. We numerically calculate lower bounds on transmission propagation latency caused by employment of packet erasure codes, when fade duration is represented by random variable with Weibull distribution. Furthermore, we propose coding scheme that involves interleaved short Reed-Solomon (RS) codes to mitigate information loss, caused by long fading events. In order to quickly and accurately evaluate residual packet loss rates of interleaved RS codes, we construct a novel simulator, called segment-based simulator, which is able to predict code performance several orders of magnitude faster than plain Monte Carlo simulation. Finally, we show that for variety of channel parameters and code rates, very short RS codes (even with length 15) can provide near optimal propagation latencies.**

***Keywords*—Fading channels, Packet erasure codes, Reed-Solomon codes, Earth-space Links, Singleton bound**

I. Introduction

SATELLITE communications are experiencing a renaissance over the past years, as their potential in providing broadband transmission is fully recognized and explored. Possibility to transfer information between arbitrary points on the Earth surface via satellites makes this type of communications attractive for emerging services, which require high accessibility, like connected vehicles. Furthermore, inclusion of satellite links into 5G ecosystem is studied in newly published 3GPP standards [1], in an attempt to create powerful hybrid satellite-terrestrial (HST) communication systems [2]. In order to build HST systems different technological challenges need to be resolved, which are partially consequences of reduced accessibility and reliability of Earth-space channels.

Despite widely assumed line-of-site visibility between ground terminals and satellites, Earth-space communications link could be blocked for a variety of reasons. In mobile satellite communications used in urban areas, line-of-site assumption may not hold, while different ground obstacles could, from time to time, disable the communication link, leading to loss of transmitted information. Furthermore, LEO (Low Earth Orbit) satellites are moving along their orbits and each LEO satellite is only shortly visible from a stationary point at Earth surface. To enable uninterrupted transmission, ground terminal periodically executes handover operations and establishes connection with different LEO satellite. Handover operation can not be instantaneously performed, which means that potentially communication channel is for a period of time blocked. Lastly, meteorological conditions, for example heavy rain, could attenuate signal such that its power cannot reach a receiving threshold.

All the above phenomenons can be described via hard-blockage model [3], where on-state (reliable transmission) and off-state (blocked channel with no transmission) are alternatively changed, while state durations are represented as uncorrelated random events. Probability density functions of state durations are chosen to fit empirical data collected over the years, mostly for Ka frequency band [4], [5].

Reliability of Earth-space links could only be improved by forward error correction, as propagation delay prohibits use of retransmission techniques. Given the fact that hard-blockage model corresponds to block erasure channel, using sufficiently long maximum distance separable (MDS) codes, for example Reed-Solomon codes, will be optimal. However, for high throughput communications, during off-states the large amount of symbols are erased, and consequently length of a required MDS code must be high (usually measured in tens of Mb or higher). Decoding complexity of MDS codes increases quadratically with code length, which makes them impractical in Earth-space channels. Instead, suboptimal solutions are of interest, like interleaved RS codes [6] and Raptor codes [7]–[9], whose decoding complexity increases only linearly with code length, or LT (Luby Transform) codes [10], with the worst case $O(n \log n)$ complexity.

Possibility of increasing reliability of Earth-space links with LT codes was analyzed in [11], [12]. The authors in [12] showed that UDP transport protocol equipped with LT codes outperform TCP transfer, when applied in DVB-S systems. Recently, a novel LT-based codes are proposed in [11], specially designed for Earth-space channel. In [3] it was shown that Raptor codes could also be beneficial in blockage mitigation of Earth-space links.

The most powerful Raptor code, called RaptorQ fails to reconstruct K information symbols if it successively receives any $K + O$ codeword symbols with probability close to the failure probability of a random fountain code constructed over Galois field GF(256), which is upper bounded by $1/255 \times 256^{-O}$ [8]. Thus, if a receiver collects only a few additional symbols (for example $O = 4$), the achievable decoding failure becomes negligible compared to any practically required residual target codeword loss rate. It follows that for $K \gg O$ RaptorQ code performs closely to the MDS code, which can recover a transmitted codeword (with zero failure probability) if it collects only K codeword symbols. RaptorQ decoders are based on the inactivation decoding algorithm, which has a serial

S. Brkić, Z. Čiča, P. Ivaniš and Đ. Sarač are with the University of Belgrade, School of Electrical Engineering, 11000 Belgrade, Serbia (e-mails: srdjan.brkic@etf.rs, cicasyl@etf.rs, predrag.ivanis@etf.rs and sarac@etf.rs).

A. Radošević is with the Tannera LLC, 11000 Belgrade, Serbia (e-mail: andreja@tanera.io).

 ISBN 978-86-7466-930-3

structure by nature, and does not represent a very effective solution for hardware implementation. In addition, memory sizes of the largest commercially available programmable chips are not sufficient to store a single Gb long codeword of a RaptorQ code, which may be required in some high data rate Earth-space links. Splitting and decoding a codeword across multiple hardware chips can be identified as the second major drawback to consider long RaptorQ codes for practical implementation.

To resolve drawbacks of long RaptorQ codes, we propose employment of interleaved Reed-Solomon (RS) codes. Multiple independent codecs run in parallel and connected to a single (or distributed across multiple chips) interleaver/deinterleaver, resulting in a design that is free of RaptorQ's drawbacks. However, such erasure protection scheme is applicable if size of the interleaver is close to the length of the optimal MDS code with the same code rate and residual loss rate. It should be noted that the interleaved RS codes are not necessarily MDS codes for arbitrary distributed erasure channel, and therefore, their applicability needs validation. In addition, performance evaluation of inverleaved RS codes by using Monte Carlo simulation is not practical, given the large interleaver sizes.

In this paper, we analyze Earth-space channels subjected to deep fades, which cause blockage of the channel. We assume that probability density of on/off states duration is modeled by Weibull distribution. First, we numerically obtain lower bounds on the propagation latency cased by employment of MDS protection code in considered fading channel, that achieves desired residual loss rate. In addition, we propose a simulation method that enables quick and accurate performance evaluation of interleaved RS codes. With aid of designed simulator we identify interleaved RS codes that perform closely to MDS codes on Earth-space channel. Surprisingly, for majority of code rates and channel parameters extremely short RS codes (of code lengths equal to 15) will provide satisfactory latency.

The rest of the paper is organized as follows. In Section II Earth-space channel model is given. In Section III a methodology for link reliability analysis is presented, while Section IV is dedicated to numerical results. Concluding remarks are given in Section V.

II. Earth-space channel modeling

The fade phenomenon in Earth-space links usually manifests though *fade episodes*, which are defined as time intervals in which signal attenuation due to fading exceeds a *fade threshold*. Similarly, as seen in Fig. 1, *inver-fade intervals* are complementary events, in which signal attenuation is below the fade threshold [13]. It is usually considered that fade episodes have catastrophical effect on transmission through Earth-space channel, i.e., information sent during fade episodes will be erased and will not appear at the receiver side. The information passes the communication channel only in inver-fade intervals. Possible information distortion during inver-fade intervals, due to signal strength variation and additive noise, is overcome by channel coding. If a channel code recovers all information in inver-fade intervals, recovering lost information during fade episodes is done by employing an additional packet erasure

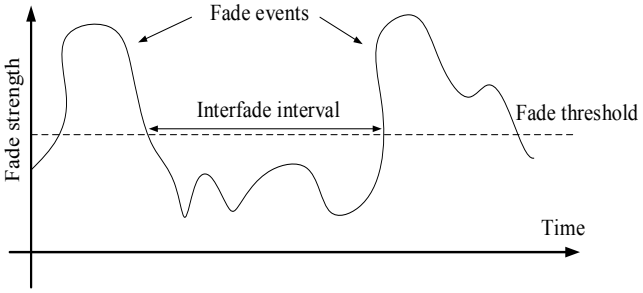


Fig. 1: Illustration of the fading model.

code. Through the paper, we will assume ideal information reconstruction during inver-fade intervals, meaning that probability of residual error after the channel coding is sufficiently low and the information reconstruction in fade episodes is not jeopardized by channel errors.

Under the aforementioned idealization, it is sufficient to describe the fading phenomenon through statistics of duration of fade episodes and inver-fade interval. Modeling such statistics, especially in Ka-band, was heavily investigated during the past years. Long-term measurements have revealed that fade episodes can be classified into two types, short-term episodes (usually less that 1 second) and long-term episodes (usually last significantly longer than 1 second), while each type has different probability density. The cause of long-term episodes is mostly rain, and average duration of long-term episodes is such that the link reliability will not benefit from deployment of packet erasure codes. Thus, will will focus only on short-term episodes, neglecting the effect of long-term fading to link reliability. We simply assume that the long-term fading is treated by some other way, for example through transmission power adaptation.

Specificity of Earth-space channel is its dependency on local climate, meaning that different points on the Earth surface will exhibit different short-term event statistics. Thus, it was proposed that in South Asia, North America and France fade episode duration is modeled with hyperexponential probability density function (PDF), while measurement in Brazil and Vancouver showed that Weibull distribution is a better fit [4]. On the contrary, measurements in Spain revealed that log-normal PDF is the most adequate to model short-term fading episodes [5]. Finally, ITU-R P.1623-1 recommends usage of the power-law distribution [13]. For the purpose of the numerical results, presented in Section IV, we will model fade episode duration as random variable following Weibull distribution. However, simulation mythology presented in Section III is invariant of the used statistical fading model.

Namely, we define the probability that a random fade duration d_f is longer than some D, given that the attenuation A exceeds a threshold a

$$P(d_f > D|a > A) = \exp(-(D/\lambda_f)^{k_f}),$$

where λ_f and k_f, represent shape and scale parameters, respectively. Similarly, we assume that interfade duration d_{if} follows the same probability low, i.e.,

$$P(d_{if} > D|a \leq A) = \exp(-(D/\lambda_{if})^{k_{if}}),$$

 ISBN 978-86-7466-930-3

where λ_{if} and k_{if}, represent shape and scale parameters of the inver-fade distribution, respectively. The channel erasure probability is defined as

$$p = \frac{E[d_f]}{E[d_f] + E[d_{if}]} = \frac{T_f}{T_f + T_{nf}} = \frac{\lambda_f \Gamma(1 + 1/k_f)}{\lambda_f \Gamma(1 + 1/k_f) + \lambda_{if} \Gamma(1 + 1/k_{if})},$$

where $\Gamma(\cdot)$ represents gamma function and $E[\cdot]$ denotes expectation operator.

We further assume that d_f and d_{if} are mutually uncorrelated and that after each fade episode receiver performs acquisition, prolonging the start of an interfade interval by a fixed time period T_{acq}. The acquisition time is included in the reliability analysis presented in the following sections.

III. Link Reliability Analysis

Given the fact that Earth-space links exhibit large propagation delay, reliability of transmission can be increased only by employing codes that can be used to recover erased information. The optimal way to recover information is to use MDS codes, which can produce the lowest residual loss rate, compared to all other codes with the same length and code rate. Alternatively, among all other codes with length N an MDS code achieves desired level of residual loss rate, with the highest possible code rate $r = K/N$, where K denotes information length. Namely, based on the Singleton bound, MSD code can reconstruct a codeword if no more than $N - K$ code symbols are missing. Given the fact that transmission latency is proportional to code length, MDS codes provide bounding latency vs. code rate dependency, for the given fading channel statistics. We only consider the propagation latency which is equal to $L = N/R_s$, where R_s is the transmission bit rate.

Obtaining close form expression for minimal latency, assuming arbitrary fading distribution, is not feasible. Instead, we rely on simulation to provide bounding latencies for desired code rate, which are depicted in Section IV. Given the fact that average fade episode duration T_f is measured in milliseconds, minimal latencies must be higher than T_f. In addition, as it is considered that modern Earth-space links must provide rate measured in Gbps, MDS codes must be at least tens of megabits long. Clearly, only suboptimal solutions are of practical interest, given the fact that decoding complexity of MDS codes grows quadratically with code length. A potential solution must include codes which decoding complexity is significantly lower. Here we investigate possibility of using interleaved MDS codes to provide satisfactory latency vs. code rate trade-offs. The main advantage of using interleaved MDS codes is that decoding complexity grows only linearly with the size of the interleaver (interleaver size represents code length). However, the unanswered question is how close inverleaved codes approach Singleton bound.

Consider transmission of packets with fixed size equal to D_P. Each packet is coded with an Reed-Solomon MDS code (N_c, K_c), with length N_c and code rate $r = K_c/N_c$. It is considered that binary length of the RS code $N_c \times \log_2(N_c) \ll D_P$, meaning that packets are divided and peace-by-peace coded. Before transmission codewords are passed through convolutional interleaver [14], which shuffles I codewords, distancing symbols from a same codeword by at least $I/(R_s * \log_2 N_c)$ seconds, where I represents the interleaver depth. At receiver side the deinterleaver restores the original bit order, while the decoder recovers the missing symbols. The propagation latency is defined as time required to fill the interleaver at the transmitter side and to empty deinterleaver at the receiver side, which based on the principe of convolution interleaving is equivalent to time needed to fill a symbol matrix of size $N_c \times I$, i.e., latency is equal to $L = \log_2(N_c) N_c I/R_s$. Given the fact that fade episodes usually corrupt multiple adjacent codewords, the packet loss rate (PLR) is approximately the same as codeword loss rate.

From the functional point of view the convolutional interleaver is the same as the block interleaver, which stores codewords in columns of $N_c \times I$ symbol matrix. By employing symbol-exact Monte Carlo simulations, it is straightforward to calculate the codeword loss probability of the decoder, i.e., the probability that the number of erased symbols in a any codeword is greater than $N_c - K_c$, for fixed K_c, N_c and I. Symbol-exact simulator alternately generates a large number of fade and non-fade blocks, according to assigned distributions, and memorize positions that correspond to start and end of each fading block. Unrepairable codewords are identified by searching and counting fade overlaps in each block interleaver column. Reliable symbol-exact simulators are computationally hungry since, for the fixed interleaver depth I and code length N_c, they process 10^8 interleaving blocks or higher with sizes measured in Gb (simulation of 10^8 interleaver blocks is needed to reliably estimate the target loss rate of 10^{-6}). Instead, we propose Segment-Based Simulator (SBS), which divides each interleaver row into S long segments, and estimates decoder failure probability assuming that codewords inside a segment are either all correctly reconstructed or all erased. This means that SBS gives an upper bound for the codeword loss rate (CWLR), for an arbitrary chosen RS codes. Inputs to the algorithm are interleaver depth measured in segments $I_{seg} = I/S$ and the number of segments that need to be simulated ($NumSegments$). We formally express SBS using algorithmic notation, as depicted in Algorithm 1.

If segment length is equal $S = 1$ symbol the SBS becomes the symbol-exact simulator. In a limiting case, for sufficiently large interleaver depth, the channel is transformed into symbol erasure channel with loss rate of

$$CWLR = \sum_{i=N_c-K_c+1} \binom{N_c}{i} P^i P^{N_c - i},$$

where P represents the erasure rate of channel that includes acquisition time, i.e,

$$P = \frac{T_f + T_{acq}}{T_f + T_{nf}}.$$

As an illustration, in Fig. 2, we depict dependency of CWLR on the introduced latency, by using (15,11) RS code. With slice modification SBS can be used to determine

 ISBN 978-86-7466-930-3

Algorithm 1: Segment-Based Simulator

Input: I_{seg}, $NumSegments$
Output: $CWLR = errorCount/SegmentCount$

$SegmentCount \leftarrow 0$
$errorCount \leftarrow 0$
while $SegmentCount < I_{seg} \times N_c \times NumSeg$ **do**
  $X = (X_1, X_2, \ldots, X_i, \ldots, X_t), \quad t \gg 1, \quad X_i \sim Weibull(\lambda_{nf}, k_{nf})$
  $Y = (Y_1, Y_2, \ldots, Y_i, \ldots, Y_t), \quad t \gg 1, \quad Y_i \sim Weibull(\lambda_f, k_f)$
  $F = (F_1, F_2, \ldots, F_{2i}, F_{2i+1}, \ldots, F_{2t}), \quad F_j = \begin{cases} \lceil (Y_i + T_{acq}) R_s / S \rceil, & j = 2i \\ \lceil (X_i - T_{acq}) R_s / S \rceil, & j = 2i - 1 \end{cases}$
  $S_i = [a_i, b_i] : \; a_i = 1 + \sum_{j=1}^{2i-1} F_j, \; b_i = a_i + F_{2i} - 1$
  $j \leftarrow 0$
  while $j < \lfloor b_t / (N_c \times I_{seg}) \rfloor$ **do**
    $i \leftarrow 0$
    $j \leftarrow j + 1$
    while $i < I_{seg}$ **do**
      $i \leftarrow i + 1$
      $m \leftarrow i + (j-1) \times N \times I_{seg}$
      $M = \{m, m + I_{seg}, \ldots, m + (N-1) \times I_{seg}\}$
      $E = \{m_q | m_q \in M \wedge m_q \in \{S_1 \cup S_2 \ldots \cup S_n\}\}$
      if $|E| > N_c - K_c$ **then**
        $errorCount \leftarrow errorCount + 1$
  $TotalSegmentCount \leftarrow TotalSegmentCount + I_{seg} \times j$

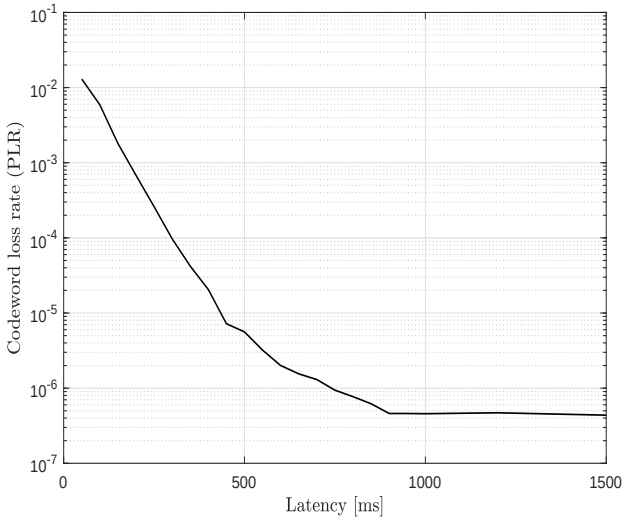


Fig. 2: PLR as a function of latency for (15,11) code ($T_f = 10$ ms, $p = 0.01$ and $k_f = k_{nf} = 1$, $S = 100$ KSymbol).

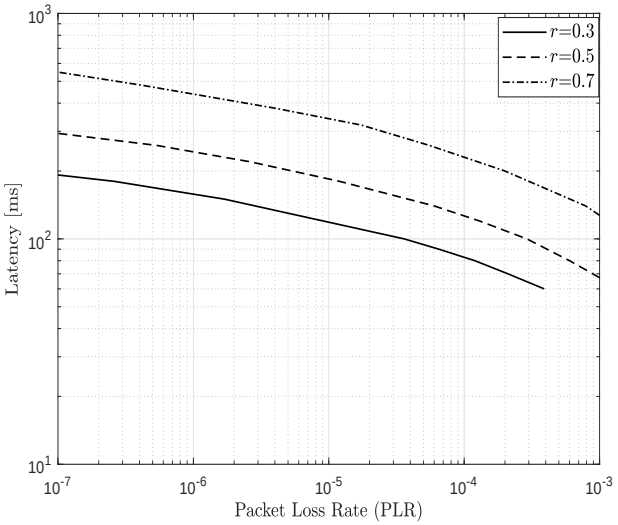


Fig. 3: Latency lower bounds for different PLR values and code rates ($T_f = 10$ ms, $p = 0.01$ and $k_f = k_{nf} = 1$).

a bounding latency for MDS codes also, when code length of a code is equal to size of the interleaver.

Clearly, speed up of SBS simulator, compared to the symbol-exact Monte Carlo simulator is proportional to segment length S. In the following section we use $S = 100$ Ksymbol long segments, for $R_s = 10$ Gbps links to show that in SBS can adequately evaluate code performance. This means that SBS is roughly five orders of magnitude faster than symbol-exact simulator.

IV. Numerical Results

We first, based on the methodology presented in Sections II and III, provide lower propagation latencies required to achieve desired PLR. In all presented results in this section we assume $D_P = 1$ Kb and $R_s = 10$ Gbps. In Fig. 3 required latency dependence of PLR is depicted, for various code rates, assuming $T_f = 10$ ms, $p = 0.01$ and $k_f = k_{nf} = 1$. For example, if $PLR = 10^{-6}$, it follows that it is not possible to construct protection code that introduces latency lower than $L = 250$ ms with code rate $r = 0.5$. If communication service requires lower

 ISBN 978-86-7466-930-3

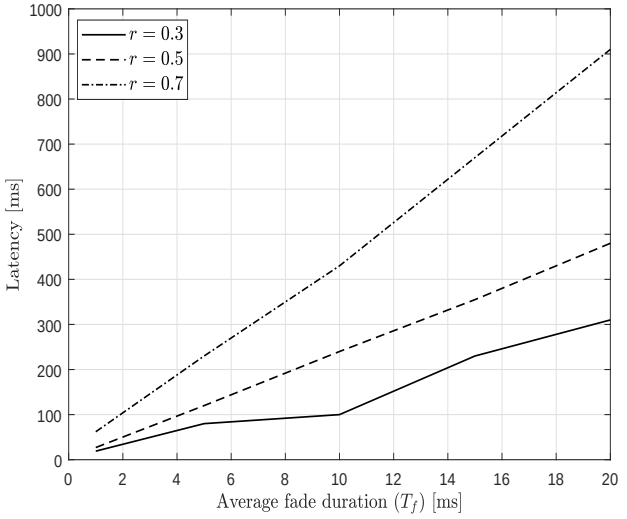


Fig. 4: Latency lower bounds as function of T_f ($PLR = 10^{-6}$, $p = 0.01$ and $k_f = k_{nf} = 1$).

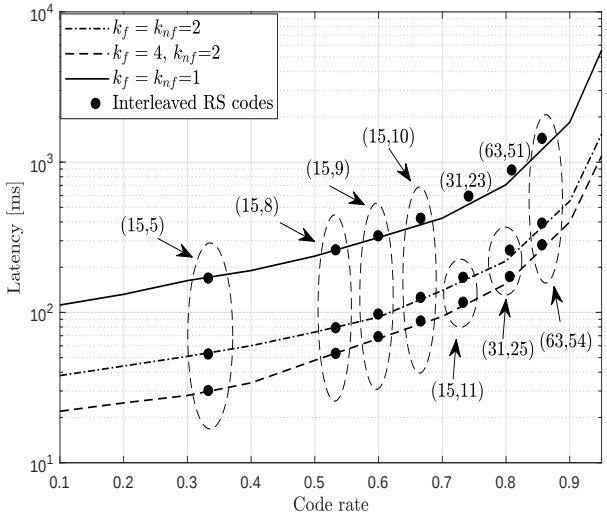


Fig. 5: Latency lower bounds as function of T_f ($PLR = 10^{-6}$, $p = 0.01$ and $k_f = k_{nf} = 1$, $S = 100$ Ksymbol).

latency spatial diversity can be used. For example, voice communications cannot successfully operate with latencies above 50 ms, which means that to operate with code rate $r = 0.5$ and $PLR = 10^{-6}$, five spatially distant links need to be established. It should also be emphasized fundamental trade-off between latency and achievable code rate, i.e., it is not possible to simultaneously increase code rate and reduce bounding latency. In Fig. 4 we depict bounding latencies for different average fade episode durations. We can observe clear linear latency growth, while slope is dependent of a considered code rate.

Finally, in Fig. 5 we investigate latencies achievable with interleaved RS codes, for the fixed $PLR = 10^{-6}$. From implementation perspective RS codes should be as short as possible. We see that it is possible to perform at lower latency bound with code length $N_c = 15$ for low and medium code rates, while code lengths of $N_c = 31$ and $N_c = 63$ are sufficient to operate at low latency bound for higher code rate. As channel condition deteriorate, which is expressed through change of parameters k_f and k_{nf} higher code length are required. For example, for $k_f = k_{nf} = 2$ it is sufficient to use (15,11) RS code to operate at latency bound, while for $k_f = k_{nf} = 1$ at least (31,23) is needed.

V. Conclusion

In this paper we provided a new look to the problem of reliable transmission through Earth-space channels, under deep fade phenomenon. We have shown that short RS coupled with convolutional interleaver ensure near optimal performance in terms of propagation latency, which cannot be avoided if we aim to ensure desired level of residual error loss rate. Interestingly, for large number of code rates it is sufficient to employ short RS codes, of length 15, which represents a low complexity solution.

During further research will use results in presented in this paper, to propose HST system, with increased information capacity, in which handover operations have extremely low blockage probability, which is currently an open research problem.

Acknowledgement

This research was supported by the Science Fund of the Republic of Serbia, grant No 7750284, Hybrid Integrated Satellite and Terrestrial Access Network - hi-STAR and the Serbian Ministry of Science under project TR32028.

References

[1] "Study on new radio (nr) to support non terrestrial networks (release 15)," *3GPP TR 38.811 v15.4.0*, 2020.

[2] M. Jia, X. Gu, Q. Guo, W. Xiang, and N. Zhang, "Broadband hybrid satellite-terrestrial communication systems based on cognitive radio toward 5G," *IEEE Wireless Commun.*, vol. 23, no. 6, p. 96–106, 2016.

[3] M. Blanco, N. Burkhardt, and C. Chen, "Coding strategies for robust link blockage mitigation in satcom," in *Proc. of IEEE Military Commun. Conf. (MILCOM)*, 2013, p. 611–616.

[4] L. E. Braten, C. Amaya, and D. V. Rogers, "Fade durations on earth-space links: Dependence on path and climatic parameters," in *Proc. of the 19th AIAA Inter. Commun. Satellite Systems Conf.*, May 2001, p. 1–7.

[5] J. M. Garcia-Rubia, J. M. Riera, P. G. del Pino, D. P. del Valle, and G. A. Siles, "Fade and interfade duration characteristics in a slant-path Ka-band link," *IEEE Trans. on Antennas and Prop.*, vol. 65, no. 12, pp. 7198–7206, Dec. 2017.

[6] J. Hamkins, "Optimal codes for the burst erasure channel," *IPN Progress Report*, pp. 42–174, Aug. 2008.

[7] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, p. 2551–2567, 2006.

[8] A. Shokrollahi and M. Luby, "Raptor codes," *Foundations and Trends in Commun. and Inform. Theory*, vol. 6, no. 3-4, p. 213–322, 2009.

[9] M. Luby, A. Shokrollahi, M. Watson, T. Stockhammer, and L. Minder, "Raptorq forward error correction scheme for object delivery," *RFC 6330 – Proposed Standard, IETF*, 2007.

[10] M. Luby, "Lt codes," in *Proc. 43rd Annual IEEE Symp. on Foundations of Comp. Science (FOCS '02)*, 2002, p. 271–282.

[11] S. Gu, J. Jiao, Q. Zhang, and X. Gu, "Rateless coding transmission over multi-state dying erasure channel for SATCOM," *EURASIP Journal on Wireless Commun. and Networking*, no. 176, pp. 1–12, 2017.

[12] P. Cataldi, M. Gerla, and F. Zampognaro, "Rateless codes for file transfer over DVB-S," in *Proc. of IEEE Inter. Conf. on Advances in Satellite and Space Commun. (SPACOMM)*, 2009, p. 7–12.

[13] "Predicton method of fade dynamics on Earth-space paths," *Recommendatio ITU-R P.1623-1*, 2003.

[14] L. Lu, K. H. Li, and Y. L. Guan, "Blind identification of convolutional interleaver parameters," in *Proc. of 2009 7th Inter. Conf. on Inform., Commun. and Signal Processing (ICICS)*, 2009.

 ISBN 978-86-7466-930-3

Effect of Phase Noise on Error Probability of MPSK Receiver over TWDP Channel - Simulation Study

Goran T. Djordjevic, *Member, IEEE*, Jarosław Makal, Bata Vasic and Bane Vasic, *Fellow, IEEE*

***Abstract*— The estimation of multilevel phase-shift keying (MPSK) signal phase is adversely affected by the number of factors appearing in transmission channel and related to the receiver. The imperfect reference signal phase estimation causes error probability degradations. The higher the value of frequency, the stronger is the effect of imperfect reference signal phase recovery. The aim of this work is to study the effect of imperfect phase estimation on error probability when signal propagates through a channel where there are two dominant components and a diffuse component, which could be the model for mmWave range. The Two-Wave Diffuse-Power (TWDP) model can accurately characterize this propagation environment. We develop a simulation model to estimate the error floor value and to identify the range of signal power when floor appears. Our simulation results give direct dependence of error rate value on channel parameters, signal power and standard deviation of phase noise. The aim of the paper is not to present the concrete estimator architecture, but to make a further step in studying the effect of a certain phase noise on bit error rate performance.**

***Index Terms*— Error probability, fading channel, phase noise, Monte Carlo simulations.**

I. Introduction

DURING the signal propagation through a wireless telecommunication channel, random changes of the signal parameters occur at the detection point. The signal level is no longer constant, but changes randomly, and the signal phase is a random process. These changes in signal level and phase have a significant impact on the quality of information transmission through that channel. Depending on the weather conditions and the type of the propagation environment including the receiver surrounding, different models have been developed to describe variations in signal intensity [1].

To describe multipath fading, different models of this fading have been proposed, such as Rayleigh, Rice, Nakagami-*m*, Nakagami-*q*, etc. [1]. In the case when two direct waves reach the receiver, but there is also the diffuse component of the emitted EM wave, the so-called Two-Way Diffuse-Power (TWDP) model of fading was proposed in [2]. In that paper in which the TWDP model was initially suggested, an approximation was proposed for the probability density function (PDF) of the signal envelope variations. This model of fading has been constantly the subject of study by many researchers [3]-[9]. In [3], Kim *at all* emphasized some shortcomings of approximation from [2], and they presented exact and approximate formulae for bit error rate in detecting binary phase-shift keying (BPSK) signal transmitted over a TWDP channel for large values of average signal-to-noise ratios (SNR). In [4] and [5], the authors derived some novel expressions for system performance metrics and derived interesting result showing that TWDP fading model has closed form moment generation function (MGF) of received signal envelope. The authors of [6] presented a novel way of parametrization of TWDP channel model. Generally, characterization of TWDP model is in tight relation with propagation environment where there are useful signal, co-channel interference and additive white Gaussian noise [10].

Recently it has been shown experimentally that this model of fading is just appropriate for describing signal propagation in the 60 GHz range [7]-[9].

In all those papers, it is implicitly assumed that the estimation of the phase of the received signal is ideal. However, it is well known that there is a difference between the phase of the incoming signal and the signal phase at the phase estimator output. This difference is a random process that affects signal detection. Due to the existence of this difference, the performance of the system deteriorates. The influence of this phase noise on the error probability when multipath fading is modeled by Nakagami-*m* distribution was considered in [11], while the effect of imperfect reference signal extraction on the error probability over a shadowed multipath fading was analyzed in [12]. In both papers, it was illustrated that error performance of a system can be strongly influenced by imperfect reference signal recovery. With the transition to higher frequency ranges, the influence of the phase noise becomes more pronounced. To the best of our knowledge, the impact of the phase noise on coherent detection in TWDP channel has not been considered so far. However, there is a significant difference between propagation environment considered in [11], [12], where one dominant direct component exist, compared with a TWDP channel model where two dominant waves exist. Our aim in this paper is not to deal with a concrete architecture of phase estimator, but to make a further step in studying the performance of TWDP coherent systems. We suppose the

Goran T. Djordjevic and Bata Vasic are with Faculty of Electronic Engineering, University of Nis, 14 A. Medevedeva, 18000 Nis, Serbia (e-mail: goran@elfak.ni.ac.rs; bata.vasic@elfak.ni.ac.rs)

Jarosław Makal is with Faculty of Electrical Engineering, Białystok University of Technology, Wiejska 45D street, 15-351 Białystok, Poland (e-mail: j.makal@pb.edu.pl).

Bane Vasic is with the Department of Electrical & Computer Engineering, University of Arizona, 1230 E. Speedway Blvd. P.O. Box 210104, Tucson, AZ 85721-0104, USA, (e-mail: vasic@ece.arizona.edu).

 ISBN 978-86-7466-930-3

phase error in the receiver has widely accepted Tikhonov distribution and determine the influence of the phase noise on the bit error rate (BER) performance of the receiver of multilevel phase-shift keying (MPSK) signals transmitted over a TWDP channel. We develop appropriate simulation model and present some Monte Carlo simulations results illustrating the effect of phase noise on the BER performance in detection of MPSK signals. These results obtained under assumption of the presence of phase noise can be compared with the results previously reported in [2]-[6], where the perfect reference signal estimation was considered. The design of phase estimator over a TWDP channel will stay an open problem, but relation between BER, channel parameters and standard deviation of phase noise will be a useful standpoint in design of an estimator.

II. System model

In this Section, we describe briefly the model of the system considered here. We give basic information about modulation/demodulation process, as well as channel characteristics.

At the transmitter, signal is modulated by performing multilevel phase-shift keying. During the duration of one symbol, the modulator output signal has the form $s = Ae^{j\phi_n}$, where

$$\phi_n \in \left[0,\, 2\pi/M\,,\ldots,2\left(M-1\right)\pi/M\right], \tag{1}$$

where M is the number of phase levels.

This signal is transmitted over a wireless channel. During signal transmission, many copies of the transmitted EM wave excite the receiver antenna. The resulting signal envelope consists of specular and diffuse parts. Specular part contains two direct components having constant amplitudes and uniformly distributed phases in the interval from 0 to 2π. The amplitudes of the direct components are denoted by V_1 and V_2, while the phases of these components are denoted by ψ_1 and ψ_2. The scattering component has Rayleigh distribution, i.e., it consist of the in-phase and quadrature components having Gaussian distribution with zero mean value and standard deviation denoted by σ_F. These in-phase and quadrature components are denoted by x_F and y_F, respectively. The resulting received signal envelope is presented as [2]

$$r = V_1 e^{j\psi_1} + V_2 e^{j\psi_2} + x_F + jy_F\,. \tag{2}$$

This model of fading can be described in terms of two parameters denoted by K and Δ. The parameter K denotes the power of the specular components-to-power of the diffuse component. The parameter Δ is related to the ratio of the peak specular components power-to-average specular components power. These two parameters are defined as [2]

$$K = \left(V_1^2 + V_2^2\right)\Big/\left(2{\sigma_F}^2\right), \tag{3}$$

$$\Delta = \frac{\text{Peak Specular Power}}{\text{Average Specular Power}} - 1 = \frac{2V_1V_2}{V_1^2 + V_2^2}\,. \tag{4}$$

The largest the value of the parameter K, the power of the specular components is larger compared with the power of the diffuse component. In other words, fading is shallower. The typical values of parameter K for the terrestrial mobile links are in the range from 0 dB to 15 dB. The values of parameter Δ lie in the range from 0 to 1. When Δ is equal to 1, specular components have the equal amplitudes, while when Δ is equal to zero either specular components amplitude is equal to zero.

It is interesting to note that the current envelope values of the signal propagating through the channel cannot be described using PDF in closed form. One approximation of the PDF was proposed in [2], while a PDF in infinite series form was recently proposed [6]. On the other hand, although there is no exact closed-form solution for PDF, there is a closed-form MGF format for this channel [4], [5]. However, for Monte Carlo simulations, the most important is eq. (2) defining the composite fading signal envelope.

After signal transmission over a TWDP channel, the receiver input signal can be presented as

$$y = re^{j(\phi_n + \psi)} + n\,, \tag{5}$$

where ψ denotes signal phase due to multipath propagation, and n denotes the thermal noise. This thermal noise can be described by Gaussian PDF with zero mean value and standard deviation σ.

In the case of BPSK, the signal in the receiver has the form

$$z = \pm r\cos\left(\varphi\right) + x\ , \tag{6}$$

depending upon "1" or "0" is transmitted. In the case of MPSK, the signal in the upper and down branch of the receiver [13, p. 361] is given, respectively, by

$$z_u = r\cos\left(\phi_n + \varphi\right) + x\,, \quad z_d = r\sin\left(\phi_n + \varphi\right) + y\,. \tag{7}$$

In the previous equations, r denotes the signal envelope defined by (2), ϕ_n denotes the modulated signal phase, φ denotes the phase noise, while x and y are the in-phase and quadrature thermal noise components over a channel. The signal detection is performed based on the value of the angle γ_d that is defined by $\tan\gamma_d = z_d/z_g$.

On the contrast to previous papers considering signal transmission over TWDP channel, in our paper the emphasis is on the imperfect reference signal recovery. Actually, the estimation of the received signal phase is not just perfect and there is a difference between received signal phase and estimated signal phase performed by an estimator. This difference is a random process. The values of this process are denoted by φ. In [14], it was shown that in the case when there is a signal and additive noise at the receiver input, then the phase difference has a Tikhonov distribution. Also, it has been shown that when a useful signal is influenced by fading in such a way that there is one regular component, the Tikhonov distribution is a satisfactory model for the phase error. However, in the case where there is a modulated signal with two regular components (TWDP channel), the phase estimation process is more complicated. The design of the phase estimator itself is beyond the scope of this paper. In the first, rough step, we will consider that the phase estimation is not ideal and that the phase difference has a rough approximation of the Tikhonov distribution. We will establish a connection between the standard deviation of phase noise, channel parameters and signal power. However, the question

 ISBN 978-86-7466-930-3

remains as to how the standard deviation of the phase error is related to the receiver and fading parameters.

The random variable having uniform distribution can be generated according to the algorithm presented in [15, p. 340], and random variable having Gaussian distribution can be generated using Box-Muller method [16, p. 383]. The random values of the signal envelope are generated taking into account eq. (1). The random variables having Tikhonov distribution can be generated by applying the Modified acceptance/rejection method [16, p. 382].

III. Numerical Results

In this Section, we present simulation results and give appropriate comments. The simulation results are obtained by Monte Carlo simulations performed in C++. The maximum number of generated samples for estimating a value of BER is 2×10^9. The methods utilized for generation of samples are described in the previous Section.

Fig. 1 illustrates the dependence of the BER on the average signal-to-noise power ratio (SNR), γ_b. The BER decreases with increasing SNR. However, this decrease in the BER is not uniform in the whole range of SNRs, but is expressed for low and moderate values of SNRs. In the range of high values of the SNR, the BER value does not decrease with increasing SNR. In other words, the BER tends to a constant value called the error floor. The appearance of this floor is a direct consequence of the presence of phase noise, i.e., non-ideal extraction of the reference carrier. The value of the floor depends on the standard deviation of the phase noise. The higher the standard deviation of the phase noise, the lower the value of the floor. This value of the floor cannot be reduced by increasing the signal power, but it can only be influenced by the correct design of the part of the system in which the phase of the received signal is estimated.

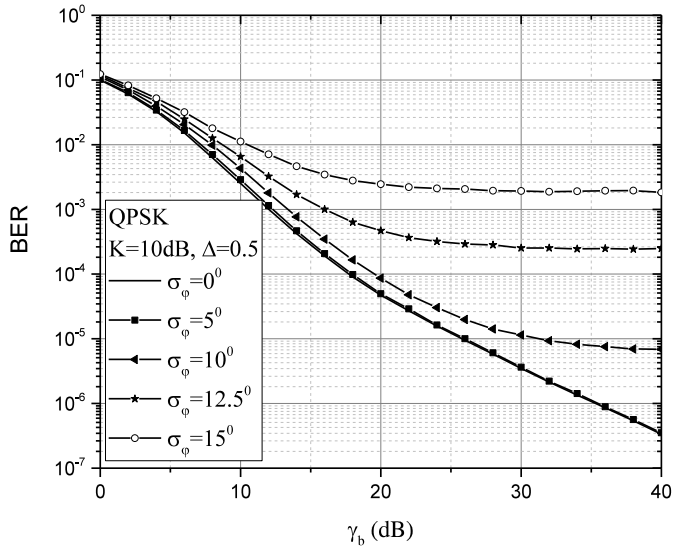


Fig. 1. BER performance for different values of standard deviation of phase noise.

Fig. 2 shows the influence of the standard deviation of the phase noise on the BER values in the detection of BPSK and QPSK signals. Firstly, it is obvious that BER values increase with increasing standard deviation values. Secondly, with the QPSK format, the BER value remains unchanged in the range of standard deviation values up to 8 degrees, while the BPSK format is insensitive to changing the standard deviation up to 16 degrees. In other words, the BPSK format is more resistant to the influence of the phase noise. This result is also logical because the areas of decision-making in the BPSK format are wider, so that the phase noise has less possibility to move the point from one area of decision-making to another.

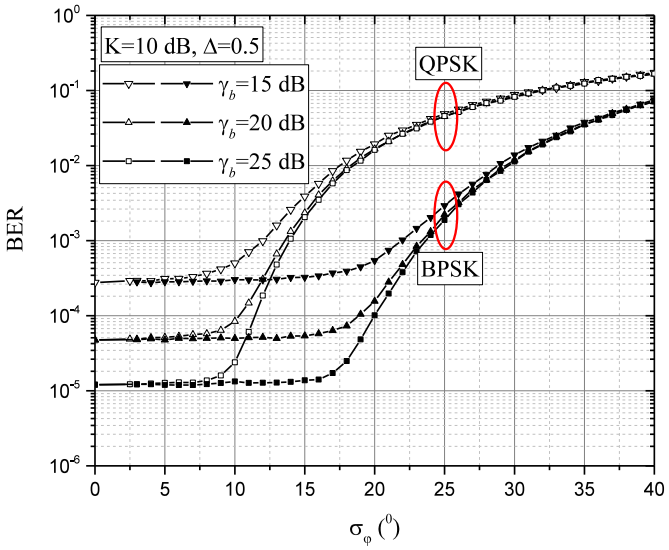


Fig. 2. BER performance for BPSK and QPSK modulation formats.

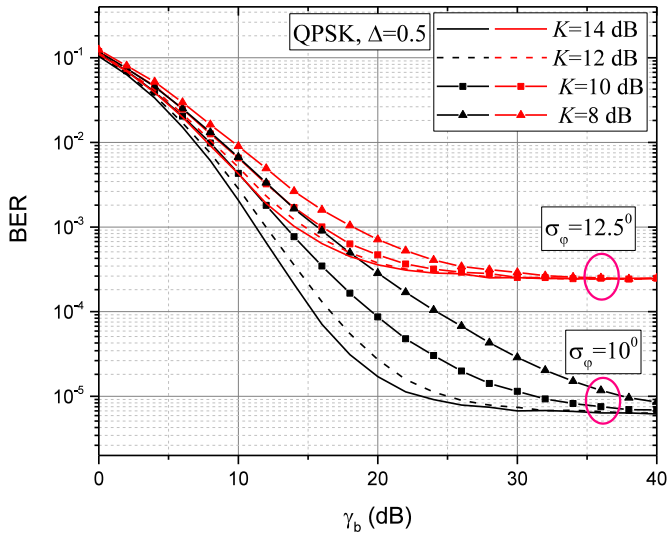


Fig. 3. BER performance for different values of fading parameter *K*.

The influence of the ratio of the average powers of the direct and scattered part of the resulting signal (parameter *K*) is shown in Fig. 3. The influence of the parameter *K* on the BER value is pronounced in the range of moderate values of the SNR and in the case when the standard deviation of the phase noise is smaller. In the region of large values of the SNR, BER tends to a constant value regardless of the value of the parameter *K*. This parameter has no effect on the value of the BER floor. In addition, when the value of the standard deviation is higher, then the influence of the phase noise is

 ISBN 978-86-7466-930-3

more pronounced, so the influence of the parameter *K* is non-dominant.

Similar conclusions as in the previous case can be drawn regarding the influence of the parameter Δ on the BER value. Namely, the parameter Δ also has a significant effect on the BER values in the middle range of the SNR, while there is no effect on the BER values when the signal power is high, i.e., when the BER floor is reached. Also, when the standard deviation of the phase noise is larger, i.e., when the influence of non-ideal extraction is dominant, BER values are less sensitive to changes in parameter Δ compared to the case when the extraction of the reference signal phase is more precise.

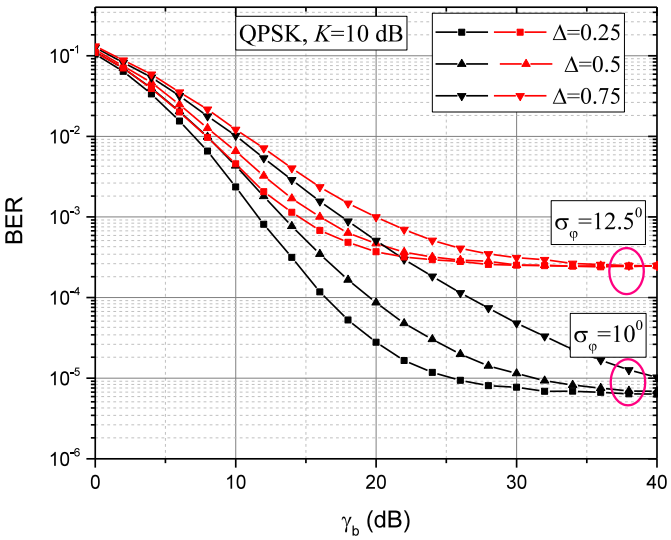


Fig. 4. BER performance for different values of fading parameterΔ.

IV. Conclusion

We have analyzed the influence of non-ideal extraction of the reference carrier on the BER performance of the receiver of MPSK signals propagating through a channel in which variations in signal levels are characterized using the TWDP model. It can be concluded that imperfect estimation of the phase of the incoming signal causes the appearance of BER floor. The value of this BER floor can be reduced by decreasing the value of the standard deviation of the phase noise, i.e., by improving the received signal phase estimation. The influence of the parameters *K* and Δ is pronounced in the range of moderate SNR values. In addition, the results have shown that for typical values of the channel parameters, BPSK modulation format is resistant to phase noise up to the standard deviation value of 16 degrees, while in the same environment QPSK format is resistant to this phenomenon only up to the standard deviation value up to 8 degrees. The design of the phase estimator should be performed so that the predetermined value of the standard deviation of the phase noise is not exceeded.

The results obtained here should be considered as the most optimistic when there is a phase error. The problems of loop locking and phase skipping in the channel with TWDP fading remain open, as well as the connection between the standard deviation of the phase noise and the estimator parameters. Incorporation of some innovative phase noise mitigation techniques, like Viterbi-Viterbi noise estimator under given channel scenario also remains an open problem.

Acknowledgment

This work is partially supported by The Polish National Agency for Academic Exchange (NAWA) under grant No. PPN/ULM/2020/1/00256/DEC/1, and partially supported by the Science Fund of the Republic of Serbia, grant No. 7750284, *Hybrid Integrated Satellite and Terrestrial Access Network - hi-STAR*, as well as by Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

[1] M. K. Simon, M.-S. Alouini, *Digital communications over fading channels*, 2nd ed, Wiley-IEEE Press, 2004.

[2] G. D. Durgin, T. S. Rappaport, D. A. De Wolf, "New analytical models and probability density functions for fading in wireless communications," *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 1005-1015, 2002.

[3] D. Kim, H. Lee, J. Kang, "Comprehensive analysis of the impact of TWDP fading on the achievable error rate performance of BPSK signaling," *IEICE Transactions on Communications*, vol. 101-B, pp. 500-507, 2018.

[4] M. Rao, F. J. Lopez-Martinez, A. Goldsmith, "Statistics and system performance metrics for the two wave diffuse power fading model," 2014 48th Annual Conference on Information Sciences and Systems (CISS), pp. 1-6, Princeton, NJ, USA March 2014.

[5] M. Rao, F. J. Lopez-Martinez, M. –S.Alouini, A. Goldsmith, "MGF approach to the analysis of generalized two-ray fading models," *IEEE Transactions on Wireless Communications*, Vol. 14, no. 5, pp. 2548 – 2561, May 2015.no.

[6] A. Maric, E. Kaljic, P. Njemcevic, "An alternative statistical characterization of TWDP fading model," *Sensors*, vol. 21, no. 22, pp. 1-15, 2021.

[7] T. Mavridis, L. Petrillo, J. Sarrazin, A. Benlarbi-Delai, P. De Doncker, "Near-body shadowing analysis at 60 GHz," *IEEE Transactions on Antennas and Propagation*, vol 63, no. 15, pp. 4505 – 4511, 2015.

[8] D. Kim, H. Lee, J. Kang, "Comments on "Near-body shadowing analysis at 60 GHz," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 6, pp. 3314, 2017.

[9] E. Zochmann, S. Caban, C. F. Mecklenbrauker, S. Pratschner, M. Lerch, S. Schwarts, M. Rupp, "Better than Rician: modelling millimeter wave channels as two-wave with diffuse power," EURASIP Journal on Wireless Communications and Networking, https://doi.org/10.1186/s13638-018-1336-6, pp. 1-17, 2019.

[10] I. M. Kostic, "Envelope probability distribution of the sum of signal, noise and interference", *Telecommunication Forum (TELFOR)*, pp. 301-303, Belgrade, Serbia, 1996.

[11] I. M. Kostic, "Average SEP for M-ary CPSK with noisy phase reference in Nakagami fading and Gaussian noise", *European Transactions on Telecommunications*, vol. 18, no. 2, pp. 109-113, 2007.

[12] J. Anastasov, Z. Marjanovic, D. Milic, G. T. Djordjevic, "Average BER and noisy reference loss of partially coherent PSK demodulation over shadowed multipath fading channel," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7831 – 7835, August 2018.

[13] S. Haykin, *Digital communications systems*, John Wiley & Sons, Inc., NJ, USA, 2014.

[14] W. Weber, "Performance of Phase-Locked Loops in the Presence of Fading Communication Channels", *IEEE Transactions on Communications*, vol. 24, no. 5, pp. 487 – 499, 1976

[15] W. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes, 3rd ed.*, Cambridge University Press, NY, USA, 2007.

[16] M. Jeruchim, P. Balaban, K. S. Shanmugan, *Simulation of communication systems: modeling, methodology and techniques, 2nd ed.*,Springer, 2000.

 ISBN 978-86-7466-930-3

Initial Development of a Program for Drone Micro-Doppler Signature Modelling

Jovan Radivojević, Predrag Petrović, Aleksandar Lebl, Mladen Mileusnić

***Abstract*—Micro-Doppler signature spectrograms obtained by FMCW radars are powerful method for malicious drones detection, identification, localization and classification. Our aim in this investigation has been to replace the base of spectrograms recorded on polygons using high number of available drone types by the spectrograms obtained by the application of originally developed program. Initial program development and verification are described in the paper. It is presented how the calculated spectrograms may be used to determine the important parameters of drones' flight and construction: number of blades in a rotor, rotors' angular rotation rate and blades' length which is the first step in a decision about an applied drone type. The presented results are the starting report on our important development devoted to the improvement of overall public safety.**

***Index Terms*—Malicious drone identification; micro-Doppler signature spectrograms; drone hovering; program for drone spectrograms modelling.**

I. Introduction

DRONES or unmanned aerial vehicles (UAVs) implementation brings great benefits in everyday life by replacing human activities in many areas. Drones may perform some important actions more precisely, more promptly than humans and without people risk exposure. But, on contrary, drones may often be the cause of sudden and unexpected danger for human lives and/or the whole world economy [1], [2].

There are solutions based on several sensor types which may be applied to malicious drones' detection, identification, localization and classification (DILC). The most often applied sensors are radars, cameras (optical and infra-red), audio and radio-frequency (RF) sensors. The benefits and drawbacks of each sensor type application are emphasized in [3], [4]. The applied solutions are usually based on several (even three or four) different sensor types [5], [6]. Among the sensors, radar is especially important due to its relative independence or very low dependence on weather and lighting conditions as fog, rain, smoke or darkness [7]. The present solutions are based on two radar types implementation: Frequency Modulated Constant Wave (FMCW radar whose principles of operation are explained in [8]) and Forward Scatter Radar (FSR whose principles of operation are explained in [9]). Drone spectrograms which are obtained by FMCW and FSR radars are especially important for applied drones identification and classification. They are analyzed in a significant number of existing solutions and [10], [11] are just two examples. In the case of FMCW radar, the obtained spectrograms follow from the Doppler effect of drone parts micro motion to the transmitted radar signal. We speak about drones micro-Doppler signatures, i.e. the obtained graphs are specific for each type of drones [12]. In practical situations the special problem is to distinguish drones from other flying objects with similar dimensions (e.g. birds) whose spectrograms may be similar to drone spectrograms [13].

Drones DILC using spectrograms supposes collecting a great number of practical records on the significant number of different drone types. The records have to be made when drones are at different heights, at different distance from radar (meaning at different elevation angles), when they are hovering or when they are flying, when there are more drones present in the same time (drone swarms) and so on. Besides, the spectrograms appearance depends on some specific drones characteristics as the number of drone's rotors, number of blades on each rotor, the length of blades and the rotors' rotation rate (in rounds per minute – RPM or in rounds per second – RPS). So, it is necessary to have different drone types and to make many spectrograms for each type under different conditions. Due to these problems it is important to develop the program which allows spectrograms calculation and presentation (without practical scenarios recording), especially in the initial phases when DILC criteria have to be defined [7], [11], [14].

IRITEL has a great experience in the development, modernization and implementation of radar systems, development of software for radar systems receivers and simulators of radar operation. The contribution [15] has two-fold relation to the solution presented in this paper: as a realized simulator and as it considers radars with Doppler Effect. Micro-Doppler signatures are formed on the base of the received radar signal and IRITEL has developed both simulators of radar signal receivers [16], [17] and practical solutions of these receivers [18], [19]. The control of radar receivers is the subject of contributions [20], [21]. A good-quality generated signal is also important for FMCW radar operation and IRITEL's solutions in this area have an international verification [22], [23]. IRITEL's complete radar systems in the area of AESA radars [24], [25] and in the area of existing radars modernization [26] - [28] are additional guarantee for future successful practical implementation of solution from this paper.

The main theoretical aspects of FMCW radar operation are explained in the Section II. This explanation includes the method how spectrograms are calculated. The block-scheme of the developed program for spectrograms calculation is

Jovan Radivojević is with IRITEL a.d., Batajnički put 23, 11080 Belgrade, Serbia (e-mail: jovan.radivojevic@iritel.com).

Predrag Petrović is with IRITEL a.d., Batajnički put 23, 11080 Belgrade, Serbia (e-mail: presa@iritel.com).

Aleksandar Lebl is with IRITEL a.d., Batajnički put 23, 11080 Belgrade, Serbia (e-mail: lebl@iritel.com).

Mladen Mileusnić is with IRITEL a.d., Batajnički put 23, 11080 Belgrade, Serbia (e-mail: mladenmi@iritel.com).

 ISBN 978-86-7466-930-3

presented in the Section III. Several calculated spectrograms are presented in the Section IV. The suggested method to determine characteristic parameters of drone flight and drone construction is described in the Section V. At the end, the conclusions are in the Section VI.

II. SPECTROGRAM SIGNAL CALCULATION

The signal generated in FMCW radar is sinusoidal and its frequency is linearly variable as a function of time. The drone spectrograms are obtained on the base of the returned echo signal from the moving rotor blades. The echo signal from all blades forming one rotor may be expressed by [29]:

$$S_{\Sigma}(t) = \sum_{k=0}^{N_b-1} s_{lk}(t) = \\ = L \cdot \exp\left(-j\frac{4\cdot\pi}{\lambda}\cdot\left(R_0 + z_0\cdot\sin\beta\right)\right)\cdot \\ \cdot\sum_{k=0}^{N_b-1}\sin c\left(\Phi_k(t)\right)\cdot\exp\left(-j\cdot\Phi_k(t)\right) \quad (1)$$

where it is $\mathrm{sinc}(\Phi_k(t)) = \sin(\Phi_k(t))/(\Phi_k(t))$ and

$$\Phi_k(t) = \frac{4\cdot\pi}{\lambda}\cdot\frac{L}{2}\cdot\cos\beta\cdot \\ \cdot\cos\left(\Omega\cdot t + \varphi_0 + \frac{2\cdot\pi\cdot k}{N_b}\right)\ \ (k=0,1,2,...N_b\text{-}1). \quad (2)$$

The sign may be + or – before $\Omega\cdot t$, depending on the direction of the rotor rotation for each unique rotor.

The meaning of variables in these two equations may be expressed with the reference to the Fig. 1:

- L – the length of each blade;
- N_b – number of blades in each rotor (two possibilities are presented separately: rotors with two or three blades);
- R_0 - distance between the radar and the drone rotor (approximately the same as between radar and drone);
- z_0 – drone height;
- β – drone elevation angle;
- Ω – rotor angular rotation rate;
- λ – radar signal wavelength;
- φ_0 – rotor starting rotation angle.

The returned signal from all drone rotors is:

$$S_{\Sigma}(t) = \sum_{i=1}^{N_r}\sum_{k=0}^{N_b-1} s_{lk}(t) = \\ = \sum_{i=1}^{N_r} L\cdot\exp\left(-j\frac{4\cdot\pi}{\lambda}\cdot\left(R_{0i} + z_{0i}\cdot\sin\beta_i\right)\right)\cdot \\ \cdot\sum_{k=0}^{N_b-1}\sin c\left(\Phi_{ik}(t)\right)\cdot\exp\left(-j\cdot\Phi_{ik}(t)\right) \quad (3)$$

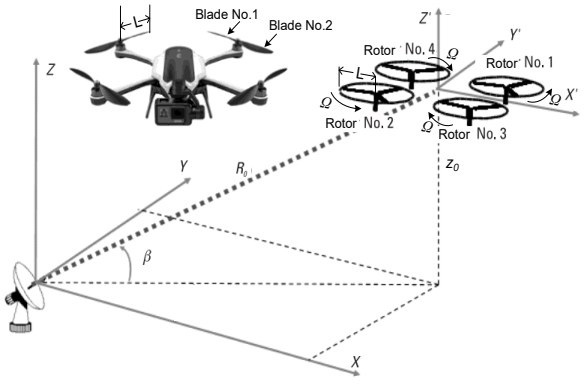


Fig. 1. Parameters included in drone spectrograms calculation

where it is:

$$\Phi_{ik}(t) = \frac{4\cdot\pi}{\lambda}\cdot\frac{L}{2}\cdot\cos\beta_i\cdot \\ \cdot\cos\left(\Omega_i\cdot t + \varphi_{0i} + \frac{2\cdot\pi\cdot k}{N_b}\right)\ \ (k=0,1,2,...N_b\text{-}1) \quad (4)$$

and

- N_r - the number of rotors on the drone.

In general case, all rotors have their own specific angular rotation rate Ω_i. The difference in rates may be even more than 2:1 when drone is flying left or right [30]. In our analysis in this paper we consider that drone is hovering. Thus we suppose that all rotors have the same Ω_i. The range of Ω_i values may be estimated on the base of graphs from [31] which present drone performances when rotors angular rotation rate changes in the range 1400-8600 RPM, or, approximately, 20-150 RPS. The direction of rotors rotation is standardized to allow stable flight. In the case of drone with four rotors (quadcopter) two opposite rotors rotate in the clockwise direction and to other opposite rotors rotate in counter clockwise direction [32]. The length of blades when quadcopters or hexacopters are applied varies in the range between 11.9cm and 38.1cm according to some available literature [11], [13], [33]-[36]. The heights z_{0i} for all rotors are also the same as we may assume that drone is always positioned parallel to the ground. The assumption of all R_{0i} values equality is not quite valid. It would be necessary to precisely involve distances between rotor centres to determine exact values of R_{0i}. Nevertheless, we shall also suppose that there is no difference between R_{0i}s without important loss of generality. The elevation angles β_{0i} for all rotors are also practically the same and they are calculated as the ratio of z_{0i} and R_{0i}. Our investigation has been performed for FMCW radar operating at 24GHz, i.e. λ=12.5mm. The most often applied drones have not more than four rotors (quadcopters), but still exist drones with six (hexacopter) or eight (octocopter) rotors [37].

The standard procedure to calculate spectrogram includes calculation of Short Time Fourier Transform (*STFT*) of the considered signal according to the equation [38]:

$$STFT\left(S_n(m,\omega)\right) = \sum_{n=-\infty}^{\infty} S_n\cdot w_{n-m}\cdot\exp\left(-j\cdot\omega\cdot t_n\right) \quad (5)$$

where w_n is the Hanning window defined by [39]:

$$w_n = \frac{1}{2}\cdot\left(1-\cos\frac{2\cdot\pi\cdot n}{N}\right)\ \ (n=0,1,2...N) \quad (6)$$

III. PROGRAM STRUCTURE SHORT PRESENTATION

The block-diagram of the program for spectrogram calculation is presented in the Fig. 2.

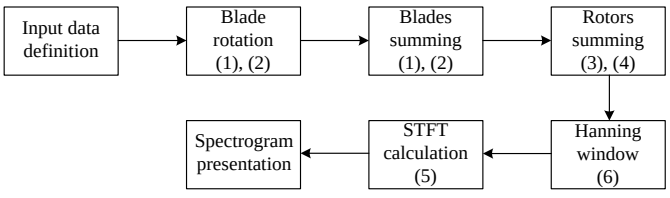


Fig. 2. Block-diagram of the program for spectrogram calculation

The first step in the program flow is to define input data: parameters of drone position towards radar, drone construction and operation, radar signal characteristics and the characteristics of desired spectrogram presentation. The parameters of drone position are its elevation angle (β), its

 ISBN 978-86-7466-930-3

height over ground (z_0) and distance between radar and drone (R_0). The drone construction parameters are the number of drone rotors (N_r), number of blades in each rotor (N_b) and the length of blades (L). Drone operation parameter is its rotors' angular rate (Ω) and direction of rotors rotation (clockwise or counter clockwise). Radar operational frequency (f_r) is the characteristic of radar signal. The characteristic parameters for spectrogram calculation and presentation are time step for spectrogram calculation (t_{step}) and time step for spectrogram display (t_{disp}).

The first step in calculation realization is related to each blade rotation. The calculation procedure follows the equations (1) and (2) in this step. The following step is to sum the echo signals from all single blades forming one rotor. This summing is also the part of equations (1) and (2). The third step in calculation process is to sum all echo signals from drone rotors, according to equation (3) and (4). The final step is related to spectrogram calculation using *STFT* according to the equation (5). Before *STFT* calculation, echo signal is modelled applying Hanning window signal according to the equation (6).

The output result of our calculation is the spectrogram presentation. A new spectrogram is obtained for each selected combination of input data parameters according to their defined values from ranges specified in the previous Section II. This is our initial investigation and this program version is realized in Excel. Some of input parameters could not be selected in the whole specified range according to the data from the Section II. There are three such parameters: the number of rotors is limited to 4 (thus covering the great majority of applied drones), the blade length is limited to 25.4cm (the most drones do not have longer blades according to the presented examples from literature) and the rotor angular rate is limited to about 40 RPS (its usual value is in the range 30-40 RPS [11], [40]).

IV. The Results of Calculation

Figures 3-12 present micro-Doppler signature graphs obtained by the implementation of our program. In all these cases it is considered that a drone has four rotors (N_r=4) and that it is positioned at a distance R_0=100m. The parameters which are varied are: 1) the number of blades constituting each rotor (N_b=3 in the figures 4, 5, 6, 7 and 9, N_b=2 in the remaining figures); 2) the length of blades (L=0.12m in the Fig. 5, L=0.18m in the Fig. 6, L=0.24m in remaining figures); 3) rotor angular rotation rate (Ω=20RPS in the figures 9 and 10, Ω=40RPS in the figures 7 and 8, Ω=30RPS in the remaining figures) and 4) the elevation angle i.e. the drone height over the ground (z_0=77m in the Fig. 11, z_0=94.8m in the Fig. 12, z_0=30m in the remaining figures). The frequency division 0-200Hz on the vertical axis of spectrograms is not the absolute value of Doppler shift for the applied 24GHz radar. It is a consequence of frequency bandwidth compression when STFT is calculated and it approximately corresponds to the compression factor 20.

The legends on the right side of figures 3-12 present the signal level (in dB) at the corresponding figure. The frequency components with the higher level in the range between -20dB and +10dB (which are presented in the brown, red and orange colour) are important for the spectrogram analysis. The other components are of lower or significantly lower level and are not important for consideration. The transition between the part of the graph with the higher frequencies level and the part of the frequencies with the low signal level (lower than -40dB which is presented by turquoise, blue and pink colour) is on all graphs rather sharp. The bandwidth of the area in brown, red and orange colour depends on the value of some parameters for which the graphs are calculated.

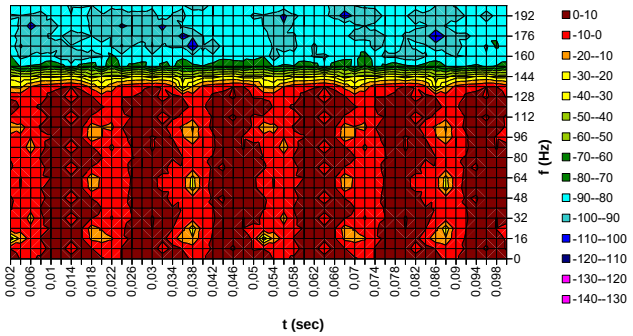


Fig. 3. Micro-Doppler signature of a drone with N_r=4, N_b=2, L=0.24m, R_0=100m, z_0=30m, Ω=30RPS

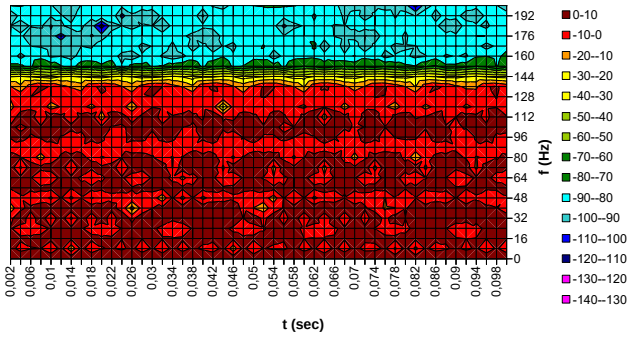


Fig. 4. Micro-Doppler signature of a drone with N_r=4, N_b=3, L=0.24m, R_0=100m, z_0=30m, Ω=30RPS

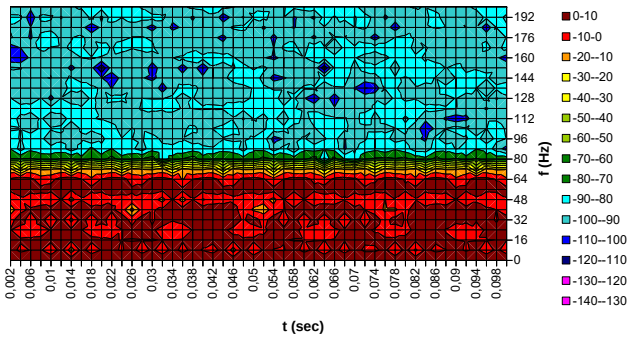


Fig. 5. Micro-Doppler signature of a drone with N_r=4, N_b=3, L=0.12m, R_0=100m, z_0=30m, Ω=30RPS

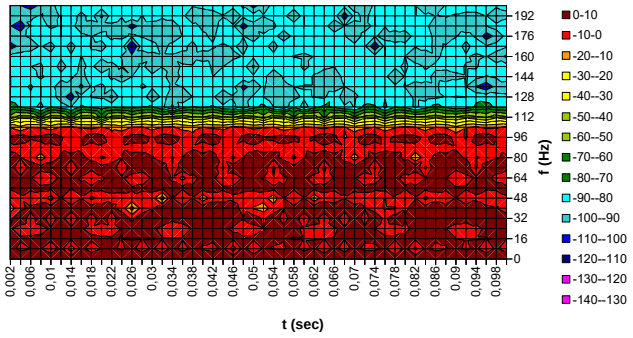


Fig. 6. Micro-Doppler signature of a drone with N_r=4, N_b=3, L=0.18m, R_0=100m, z_0=30m, Ω=30RPS

 ISBN 978-86-7466-930-3

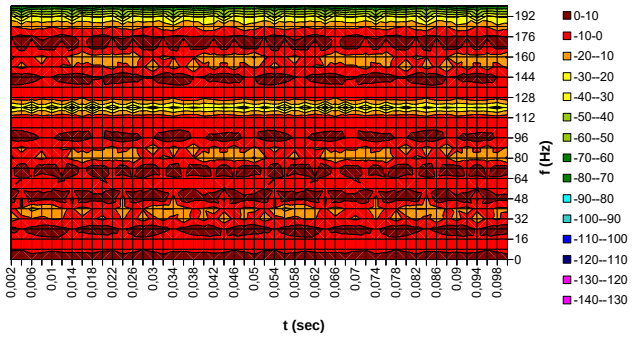


Fig. 7. Micro-Doppler signature of a drone with N_r=4, N_b=3, L=0.24m, R_0=100m, z_0=30m, Ω=40RPS

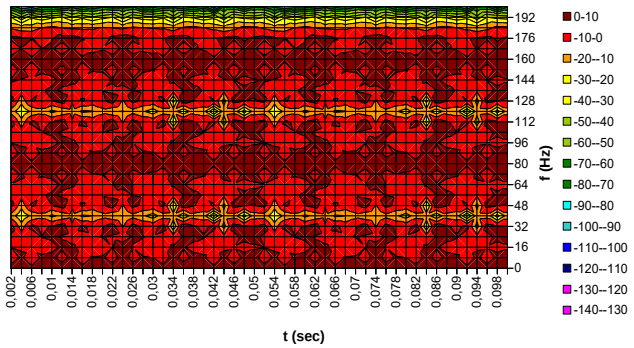


Fig. 8. Micro-Doppler signature of a drone with N_r=4, N_b=2, L=0.24m, R_0=100m, z_0=30m, Ω=40RPS

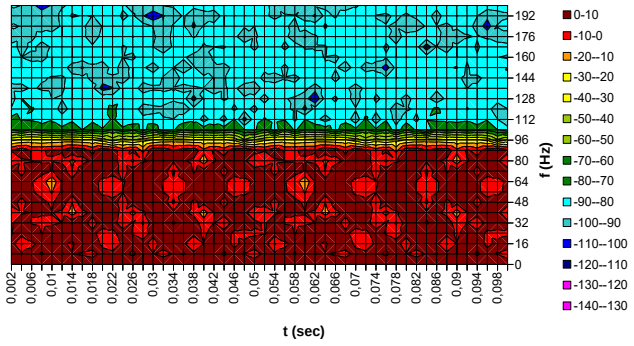


Fig. 9. Micro-Doppler signature of a drone with N_r=4, N_b=3, L=0.24m, R_0=100m, z_0=30m, Ω=20RPS

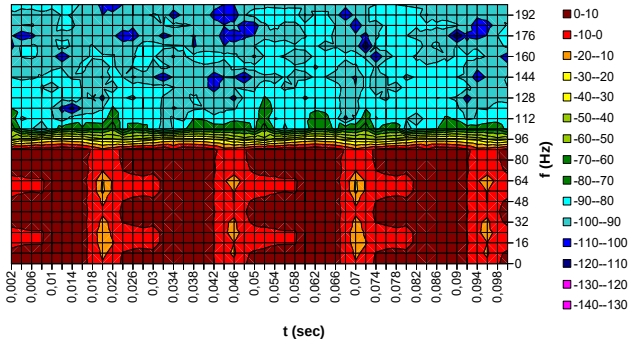


Fig. 10. Micro-Doppler signature of a drone with N_r=4, N_b=2, L=0.24m, R_0=100m, z_0=30m, Ω=20RPS

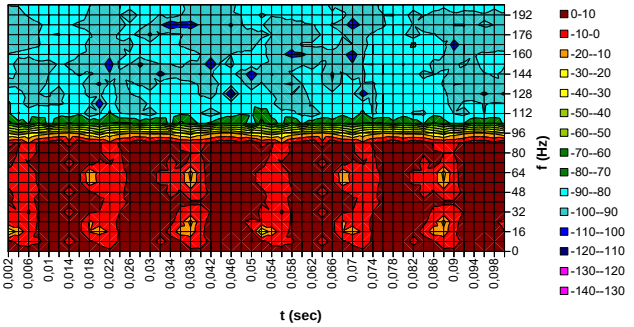


Fig. 11. Micro-Doppler signature of a drone with N_r=4, N_b=2, L=0.24m, R_0=100m, z_0=77m, Ω=30RPS

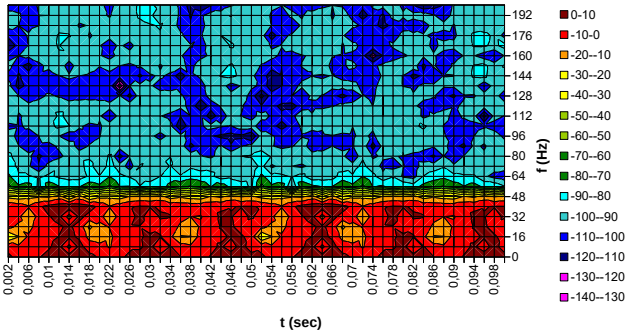


Fig. 12. Micro-Doppler signature of a drone with N_r=4, N_b=2, L=0.24m, R_0=100m, z_0=94.8m, Ω=30RPS

The number of blades in each rotor does not have any influence on the frequency bandwidth of the area with higher signal level. This is approved mutually comparing Fig. 3 (N_b=2, Ω=30RPS) and Fig. 4 (N_b=3, Ω=30RPS), then Fig. 7 (N_b=3, Ω=40RPS) and Fig. 8 (N_b=2, Ω=40RPS) and, finaly, Fig. 9 (N_b=3, Ω=20RPS) and Fig. 10 (N_b=2, Ω=20RPS). When comparing these cases two by two, we see that the number of blades only causes different configuration of brown, red and orange surfaces within the area of higher signal level.

The blades' length has the influence on the frequency bandwidth of the area with higher signal level. Dependence may be considered as linear according to the mutual comparison of the figures 4, 5 and 6. The width is approximately Δ=70Hz when it is L=0.12m (Fig. 5), Δ=104Hz when it is L=0.18m (Fig. 6) and Δ=140Hz when it is L=0.24m (Fig. 4).

The rotor angular rotation rate also has the influence on the frequency bandwidth of the area with higher signal level. Dependence is also linear, as in a case of blades' length. This statement is approved considering the graphs in the figures 9 and 10 (the width is Δ=94Hz at Ω=20RPS), then figures 3 and 4 (the width is Δ=140Hz at Ω=30RPS) and figures 7 and 8 (the width is Δ=186Hz at Ω=40RPS).

The frequency bandwidth of the area with higher signal level is proportional to the cosine value of the elevation angle. This is obvious comparing the graphs in the figures 3, 11 and 12. The drone heights (z_0=30m, z_0=77m and z_0=94.8m) are selected in such way that the ratios $\cos(z_0/R_0)$ in these three cases are 0.954, 0.638 and 0.318 or proportional to 3:2:1. The corresponding widths of the area with higher signal level in the figures 3, 11 and 12 are Δ=140Hz, Δ=96Hz and Δ=48Hz respectively, which is very near to 3:2:1.

 ISBN 978-86-7466-930-3

It is possible to distinguish graphs for the drones which have rotors with two blades (figures 3, 8, 10, 11 and 12) from those which have three blades (the remaining five figures). The characteristics when there are 2 blades usually have clearly separated, periodic parts with the highest signal level between 0dB and +10dB (brown segments). The shape of these parts depends on the rotation starting angle. There are six such parts in the figures 3, 11 and 12, eight in the Fig. 8 and four in the Fig. 10. It means that the number of repeatable parts (N_p) when there are 2 blades may be expressed as

$$N_p = N_r \cdot \Omega \cdot T_p \tag{7}$$

where T_p is the time interval of spectrogram investigation.

When there are 3 blades such clear periodic parts may not be easily islolated.

V. Reading the Parameters of Drone Flight and Construction from Spectrograms

In the Section IV it is investigated how the parameters of drone flight and of the drone construction affect its micro-Doppler signature appearance. Frequency bandwidth of the area with higher signal level depends on 3 parameters: blades' length, rotors' angular rotation rate and drone elevation angle. The spectrogram itself gives two values for the calculation. The first one is the frequency bandwidth of the area with higher signal level and the second one is the number of periodic areas with the highest signal level. So, it is necessary to determine one of the three parameters in some other way, not from spectrogram. The most logical is that this parameter is elevation angle β because it may be also determined by some algorithm implementation on FMCW radar [41], [42] (the concrete algorithm for this function is not studied in this paper).

In order to determine L and Ω from some measured spectrogram let us start from the spectrogram from the Fig. 13, which is calculated for the same parameters as in the Fig. 3 with only the exception that it is z_0=0m, i.e. cos(β)=1. It follows from the graph in the Fig. 13 that it is now Δ=148Hz. In this paper we have concluded that this value of Δ is proportinal to Ω and L. This statement may be expressed mathematically as

$$K_m \cdot L \cdot \Omega = 148 \tag{8}$$

Our goal is to determine the value of multiplication coefficient K_m.

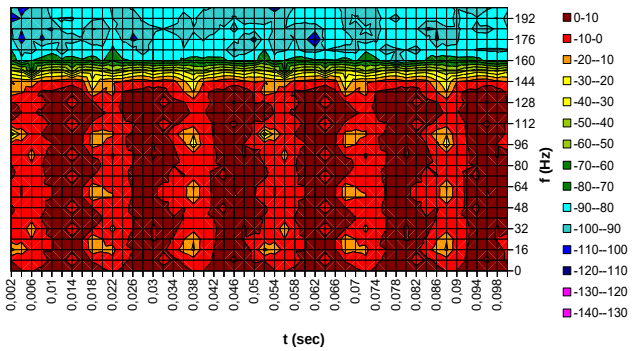


Fig. 13. Micro-Doppler signature of a drone with N_r=4, N_b=2, L=0.24m, R_0=100m, z_0=0m, Ω=30RPS

The graph in the Fig. 13 is obtained as a result of calculation for the condition that it is L=0.24m and Ω=30RPS. If we put now these values in the equation (8), we obtain that K_m=20.55.

Starting from this value of K_m, the value of elevation angle β_m (which is determined by some other algorithm of drone DILC process on FMCW radar) and the value of Δ_m from the analyzed spectrogram, it is possible to calculate the value of the product

$$L_m \cdot \Omega_m = \frac{\Delta_m}{K_m \cdot \cos(\beta_m)} \tag{9}$$

Further, if Ω_m is determined from the appearance of the spectrogram part with higher frequency components, it is possible to calculate the length of blades as

$$L_m = \frac{\Delta_m}{K_m \cdot \Omega_m \cdot \cos(\beta_m)} \tag{10}$$

VI. Conclusions

In this paper we have presented the initial development of a program for calculation and presentation of drones' micro-Doppler spectrograms as well as some results of its implementation. The benefits of such program application are that it is not necessary to have a high number of different drones and to perform significant volume of recording during drones operation in order to form the base of their spectrograms. This program follows the analytical model from [29]. The initial modelling is limited to hovering drones whose rotors have equal angular rotation rate. A number of spectrograms with differently defined input parameters is presented in the paper. The method for determination of drone's flight and construction characteristics is defined on the base of presented spectrograms. The number of blades in each rotor, rotors' angular rotation rate and blades' length, which contribute to specificities of spectrograms, may be concluded on the base of spectrograms appearance.

These initial calculations and presentations are performed in Excel as other, more suitable possibilities, were not available to us. Excel allows to perform good quality presentations of calculation results, as is also demonstrated in our contribution [43]. We plan to perform future investigations in some more powerful surrounding where it would be possible to model drone flying (besides hovering). It is also necessary to model drones with higher number of rotors (hexacopters, octocopters) as well as drone swarms to further improve drones DILC by FMCW radars.

FMCW radars are not the only way to obtain spectrograms as the ones presented in this paper. The similar results may be also obtained by using pulsed radar where the returned pulse delay is the measure of Doppler shift [13]. IRITEL already has experience in the development and improvements of such radars [21], [26]-[28]. In the case of single carrier radar the measure of Doppler shift would be the change of returned signal phase which is more complicate for realization.

References

[1] G. Delauney, "Mystery drone from Ukraine war crashes in Croatia," BBC News, Balkans correspondent, https://www.bbc.com/news/world-europe-60709952.

[2] N. Razzouk, J. Blas, J. Thornhill, "Speed of Saudi Oil Recovery in Focus After Record Supply Loss," Bloomberg, 15. September 2019.,

 ISBN 978-86-7466-930-3

https://www.bloomberg.com/news/articles/2019-09-15/saudis-race-to-restore-oil-output-after-crippling-aramco-attack.
[3] N. Eriksson, "Conceptual study of a future drone detection system Countering a threat posed by a disruptive technology," Master thesis in Product Development, Chalmers University of Technology, Goethenburg, Sweden, 2018.
[4] V. Matić, V. Kosjer, A. Lebl, B. Pavić, J. Radivojević, "Methods for Drone Detection and Jamming," 10th International Conference on Information Society and Technology (ICIST), Kopaonik, March 8-11., 2020., in: Zdravković, M., Konjović, Z., Trajanović, M. (Eds.) ICIST 2020 Proceedings Vol. 1, pp.16-21, 2020.
[5] Advanced protection systems: "Ctrl+sky drone detection and neutralization system," 2017., http://apsystems.tech/wp-content/uploads/2018/01/aps_broszura_web.pdf.
[6] X. Shi, C. Yang, C. Liang, Z. Shi, and J. Chen: "Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges," IEEE Communications Magazine, Vol. 56, Issue 4, April 2018., pp. 68-74., DOI: 10.1109/MCOM.2018.1700430.
[7] F. Fioranelli, O. Krasnov, Y. Cai, A. Yarovoy, J. Yun, D. Anderson, "MSG-SET-183 Specialists' Meeting – Improving the Simulations of Radar Signatures of Small Drone," STO-MP-MSG-SET-183, NATO, S&T organization, pp. 1-1 – 1-12.
[8] V. M. Milovanović, "On Fundamental Operating Principles and Range-Doppler Estimation in Monolithic Frequency-Modulated Continuous-Wave Radar Sensors," *Facta Universitatis, Series: Electronics and Energetics*, Vol. 31, No. 4, pp. 547-570, December 2018, DOI: https://doi.org/10.2298/FUEE1804547M.
[9] A. De Luca, "Forward Scatter Radar: Innovative Configurations and Studies," PhD Thesis, University of Birmingham, February 2018.
[10] C. Zhao, G. Luo, Y. Wang, C. Chen and Z. Wu, "UAV Recognition Based on Micro-Doppler Dynamic Attribute-Guided Augmentation Algorithm," *Remote Sensing*, Vol. 13, No. 6, Article 1205, pp. 1-17., March 2021., DOI: https://doi.org/10.3390/rs13061205.
[11] S. A. Musa, R. A. R.Syamsul Azmir, A. Sali, A. Ismail, N. Emleen, A. Rashid, "Micro-Doppler signature for drone detection using FSR: a theoretical and experimental validation," *The Journal of Engineering*, IET International Radar Conference (IRC2018), 17-19th October 2018., Nanjing, China, pp. 1-6.
[12] M. Passafiume, N. Rojhani, G. Collodi and A. Cidronali, "Modeling Small UAV Micro-Doppler Signature Using Milimeter-Wave FMCW Radar," *Electronics 2021*, Vol. 10, pp. 1-16., https://doi.org/10.3390/electronics10060747.
[13] S. Rahman, D. A. Robertson, "Radar micro-Doppler signatures of drones and birds at K-band and W-band," *Scientific Reports*, Vol. 2018, No. 8, pp. 1-11., November 2018., DOI: 10.1038/s41598-018-35880-9.
[14] A. Lebl, M. Mileusnić, D. Mitić, J. Radivojević, V. Matić, "Verification of Calculation Method for Drone Micro-Doppler Signature Estimation," accepted for publication in *Facta Universitatis, Series: Electronics and Energetics*, ISSN: 0353-3670.
[15] P. Jovanović, M. Mileusnić, B. Pavić, B. Mišković, "DDS based Pulse-Doppler Radar Transmitter Simulator," XIII International Scientific-Professional Symposium INFOTEH Jahorina, March 2014., Vol. 13. Ref. B-II-5, pp. 425-428.
[16] V. Marinković, B. Pavić, A. Toth, "Radar Signal Simulator for New Generation Digital Radar Receiver," INFOTEH Jahorina, Vol. 7. Ref. B-II-18, March 2008., pp. 228-231., in Serbian.
[17] B. Pavić, B. Mišković, V. Marinković-Nedelicki, M. Mileusnić,P. Petrović, "Projekat simulatora impulsnih radarskih signala," tehničko rešenje u kategoriji M85 na projektu tehnološkog razvoja TR32051 pod nazivom "Razvoj i realizacija naredne generacije sistema, uređaja i softvera na bazi softverskog radija za radio i radarske mreže," 2016.
[18] N. Remenski, B. Pavić, M. Mileusnić, P. Petrović, "Practical Realization of Digital Radar Receiver," 49. Conference ETRAN, Budva, June 2005., pp. 105-108., in Serbian.
[19] D. Dramićanin, V. Vlahović, N. Remenski, B. Pavić, P. Petrović, "FPGA Implementation of the Digital Radar Receiver," INFOTEH 2006, March 2006., Vol. 5. Ref. B-II-2, pp. 80-84., in Serbian.
[20] N. Remenski, V. Marinković-Nedelicki, V. Tadić, P. Petrović, "The Control of New Generation Digital Radar Receiver," INFOTEH 2006, March 2006., Vol. 5. Ref. B-II-10, pp. 114-118., in Serbian.
[21] V. Marinković, N. Remenski, V. Tadić, P. Petrović, "The Software for Control of Digital Radar Receiver VHF DP/P-12," 51. Conference ETRAN, Budva, Herceg Novi – Igalo, 2007., in Serbian.
[22] V. Matić, V. Marinković-Nedelicki, V. Tadić, "Comparison of digital signal processing methods for sine wave signal generation," Proceedings of SBT/IEEE International Telecommunications Symposium ITS '98, August 1998., DOI: 10.1109/ITS.1998.713134.
[23] V. Matić, V. Marinković-Nedelicki, "The waveform generator based on digital signal processing," Proceedings of the 2000 Third IEEE International Caracas Conference on Devices, Circuits and Systems 2000, pp. T56/1-T56/6, March 2000., DOI: 10.1109/ICCDCS.2000.869878.
[24] P. Petrović, "Research in Software Defined Radio and AESA Radar Technology, Serbia-Italia/Status and Perspectives of the Scientific and Technological Bilateral Cooperation," 2012., pp. 19-20.
[25] P. Jovanović, M. Mileusnić, P. Petrović, "An Approach to Analysis of AESA Based Radio systems," XII International Scientific-Professional Symposium INFOTEH Jahorina 2013, March 2013., Vol. 12., pp. 372-376.
[26] Land-based air defence radars, Serbia: VHF DR/P-12/18, in the book M. Streetly, *Jane's Radar And Electronic Warfare Systems*, IHS Global Limited, 2011.
[27] B. Pavić, V. Marinković-Nedelicki, M. Mileusnić, N. Remenski, P. Petrović, "Verifikovani modernizovani radar P-12," tehničko rešenje – novi proizvod u kategoriji M81 na projektu tehnološkog razvoja TR32051 pod nazivom "Razvoj i realizacija naredne generacije sistema, uređaja i softvera na bazi softverskog radija za radio i radarske mreže," 2013.
[28] M. Mileusnić, B. Pavić, V. Marinković-Nedelicki, P. Petrović, V. Matić, A. Lebl, "Verifikacija razvoja i realizacije nulte serije nove varijante modernizovanog osmatačko-akvizicijskog radara P-12M," tehničko rešenje u kategoriji M82 na projektu tehnološkog razvoja TR32051 pod nazivom "Razvoj i realizacija naredne generacije sistema, uređaja i softvera na bazi softverskog radija za radio i radarske mreže," 2018.
[29] V. C. Chen, "*The Micro-Doppler Effect in Radar*," Artech House, Second Edition, 2019., ISBN: 978-1-63081-546-2.
[30] C. R. Ferreiro, "Modeling and Analysis of Micro-Doppler Signatures for Radar Target Classification," Thesis for Bachelor in telecommunication Engineering, Telecommunication Engineering School, Universida de Vigo, 2017.
[31] R. W. Deters, S. Kleinke, "Static Testing of Propulsion Elements for Small Multirotor Unmanned Aerial Vehicles," AIAA AVIATION Forum, 35th AIAA Applied Aerodynamics Conference, 2017-3743, 5-9. June 2017., Denver, Colorado, pp. 1-34.
[32] Drone Tech Planet, "How a Quadcopter Works Along With Propellers and Motors," https://www.dronetechplanet.com/how-a-quadcopter-works-along-with-propellers-and-motors/.
[33] "Mavic Mini Propellers," https://store.dji.com/product/mavic-mini-propellers.
[34] "DJI Mavic 3 Low-Noise Propellers," https://store.dji.com/product/dji-mavic-3-low-noise-propellers.
[35] "DJI Propeller Set for Phantom 4 Pro/Pro+ V2.0 (2 Pack)," https://www.bhphotovideo.com/c/product/1407136-REG/dji_cp_pt_00000274_01_propellers_for_phantom_4.html/specs.
[36] "Altair 818 Hornet Propellers," https://altairaerial.com/products/aa818-plus-drone, https://www.amazon.com/Altair-818-Hornet-Propellers/dp/B07819RC9Q.
[37] 911, Rotorcraft, "Types of Drones," https://www.911security.com/learn/airspace-security/drone-fundamentals/types-of-drones-rotorcraft.
[38] M. Ahmadizadeh, "An Introduction to Short-Time Fourier Transform (STFT)," Sharif University of Technology, Department of Civil Engineering, July 2014.
[39] H. A. Gaberson, "A Comprehensive Windows Tutorial,"*Sound and Vibration*, Instrumentation Reference Issue, March 2006., pp. 14-23.
[40] C. Zhao, G. Luo, Y. Wang, C. Chen and Z. Wu, "UAV Recognition Based on Micro-Doppler Dynamic Attribute-Guided Augmentation Algorithm," *Remote Sensing*, Vol. 13, No. 6, Article 1205, pp. 1-17., March 2021., DOI: https://doi.org/10.3390/rs13061205.
[41] P. K. Rai, A. Kumar, M. Z. Ali Khan, J. Soumya, L. Reddy, "Angle and Height Estimation Technique for Aerial Vehicles using mmWave FMCW Radar," 2021 International Conference on COMmunication Systems & NETworkS (COMSNETS), 5-9. January 2021., Bangalore, India, DOI: 10.1109/COMSNETS51098.2021.9352744.
[42] P. K. Rai, H. Idsoe, R. R. Yakkati, A. Kumar, M. Z. Ali Khan, P. K. Yalavarthy, "Localization and Activity Classification of Unmanned Aerial Vehicle Using mmWave FMCW Radars," IEEE Sensors Journal, Vol. 21, No. 14, pp. 16043-16053., July 2021., DOI: 10.1109/JSEN.2021.3075909.
[43] J. Radivojević, B. Pavić, A. Lebl, M. Petrović, "Sweep Jamming with Discrete Subbands – an Advanced Strategy for Malicious Drones Missions Prevention," Scientific Technical Review, Vol. 71, No. 2, March 2022, pp. 46-52., ISSN: 1820-0206, UDK: 355.43:623.624,449.8, COSATI: 03-10, 14-04-01, DOI: 10.5973/str2102046R.

Execution Time Improvement using CPU Parallelization and Non-Uniform High-Resolution Range-Doppler Map Estimation in HFSWR

Dragan Golubović, Nenad Vukmirović, Zoran Lončarević, Marko Marković and Miljko Erić

***Abstract*—High-resolution range-Doppler (RD-HR) map estimation, used for primary signal processing in a High Frequency Surface Wave Radar (HFSWR), is the most computationally demanding step of the vessel detection algorithm. In order to reach real-time processing, which is of great importance in practical implementations of such systems, a very high-speed computation is required. In this paper, we propose non-uniform signal frame selection, to reduce the load with almost no loss in performance, and parallel processing on a CPU to get high-resolution range-Doppler maps in a multi-antenna scenario. The paper contains the description of the proposed algorithm and the performance analysis. The experimental results show a 60- to 130-fold improvement in the execution time of the program for vessel detection.**

***Index Terms*—HFSWR; range-Doppler map; high-resolution methods; multi-core; parallelized algorithms; speedup.**

I. Introduction

The focus of numerous scientific papers is maritime surveillance of vessels at long distances. The reason for that are many illegal crime activities over the horizon, drug trafficking, attacks on petrol platforms and strategic objects, etc. High Frequency Surface Wave Radars (HFSWRs) are widely used for this purpose [1]-[3].

To make these systems have better ship detection accuracy, as well as the ability to detect some ships, which are not visible at all using the currently used primary signal processing algorithms, high-resolution algorithms are used [4].

The downside is that the computational burden of HFSWRs is huge in that case, and high-speed computation is required in order to have real-time processing. It is of great importance in practical implementations.

In recent years, the demand for high performance numerical computing in many radar systems was increased. The focus of many papers is the formulation of parallel algorithms for the MIMO radar based on the Central Processing Unit (CPU)/ Graphical Processor Unit (GPU) architecture [5]-[9]. The architecture must be capable of multitasking, which allows multi-threaded execution of the program for vessel detection. GPUs have an essential role in designing the real-time programs because they are highly parallel, multithreaded processors. Graphic cards are equipped with multi-core GPUs enabling the development of high computationally demanding programs. But the performance improvement of multithreaded programs depends on the algorithms used and their implementations, which is often not a trivial task. This improvement is limited by the fraction of the program source code that can be run on multiple cores simultaneously. It is important to say that high execution efficiency is possible, only if hardware characteristics are appropriate.

The focus of this paper will be a real-time implementation of target detection in HFSWRs by improving the existing implementation of the high-resolution range-Doppler (RD-HR) map estimation, which is the most computationally demanding task of the entire algorithm. The algorithm runs in real-time if its execution time is shorter than the acquisition time of the frames used for the estimation (32.768 s in this particular case).

The existing implementation did not run in real-time, so we solved this problem in two ways. Firstly, we significantly reduced the computational load of the RD-HR map estimation with very little performance loss by using non-uniform sampling across the frames. Secondly, the parallel signal processing on CPU was performed. An advantage of the proposed implementation is that it requires only a general-purpose computer and no expensive dedicated hardware, like graphic cards, or Digital Signal Processing (DSP) cards.

The paper is organized as follows. In Section II, we presented a detailed algorithm description to estimate high-resolution range-Doppler map. In Section III, we explained the numerically efficient method to achieve real-time processing requirement. We discussed some experimental results and made some comparisons in Section IV and

Dragan Golubović is with the University of Belgrade, School of Electrical Engineering, 11120 Belgrade, Serbia and Vlatacom Institute, 11070 Belgrade, Serbia (e-mail: dragan.golubovic@vlatacom.com).

Nenad Vukmirović is with the University of Belgrade, School of Electrical Engineering, 11120 Belgrade, Serbia and the Innovation Center of the School of Electrical Engineering, 11120 Belgrade, Serbia.

Zoran Lončarević is with Vlatacom Institute, 11070 Belgrade, Serbia (e-mail: zoran.loncarevic@vlatacom.com).

Marko Marković is with Vlatacom Institute, 11070 Belgrade, Serbia (e-mail: marko.markovic@vlatacom.com).

Miljko Erić is with Vlatacom Institute, 11070 Belgrade, Serbia and the University of Belgrade, School of Electrical Engineering, 11120 Belgrade, Serbia (e-mail: miljko.eric@vlatacom.com).

 ISBN 978-86-7466-930-3

outlined some conclusions and further research activities in Section V.

II. High-Resolution Range-Doppler Map Estimation

HFSWR to be analyzed in this paper operates in HF spectrum 3-30 MHz and it uses Frequency Modulated Continuous Waves (FMCW). The radar consists of a transmitter (Tx) antenna array, receiver (Rx) antenna array and transceiver hardware. The numerical results in this paper are based on the signals from an *N*-element linear Rx antenna array. Other geometries are also possible. At the receiving channels, acquired complex time samples of IQ branch signals are available for further signal processing. In Fig. 1 the complete high-resolution target detection algorithm in HFSWR was proposed.

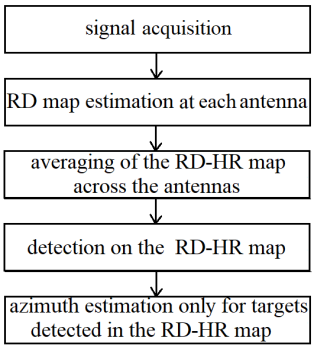


Fig. 1. The overview of the proposed high-resolution algorithm for vessel detection in HFSWR

The proposed algorithm has five steps, but the most computationally demanding task is step two – RD-HR map estimation at each antenna. Because of that, in order to improve program performance in terms of execution time, it is necessary to describe the creation of RD-HR map in detail and find a way to reduce its numerical complexity. So, other steps are not the focus of this section and only RD-HR creation will be explained.

In practical situations, *P* (the number of signal samples in one frame) and *N* are predefined values and *M* can be varied and it represents the number of frames used for the creation of one segment. Based on this segment, RD-HR map is formed. The developed algorithms were tested for the length of the segment of *M*=256, where the successive segments overlap in 128 frames. This ensures that the results are refreshed every 128 frames. The first step in the RD-HR creation process, is the implementation of FFT algorithm in P points for all frames *m*=1, 2, ... , *M* and all antennas *n*=1, 2, ... , *N*. By adding zeros to the vectors with signal samples, better grid resolution can be achieved when applying FFT. For further processing and a better estimate of the covariance matrix later, we need to extend the segment by L-1 additional frames. Fig. 2 shows the forming of the first two segments, with a 128-frame overlap. The same procedure is used for all other segments during signal processing.

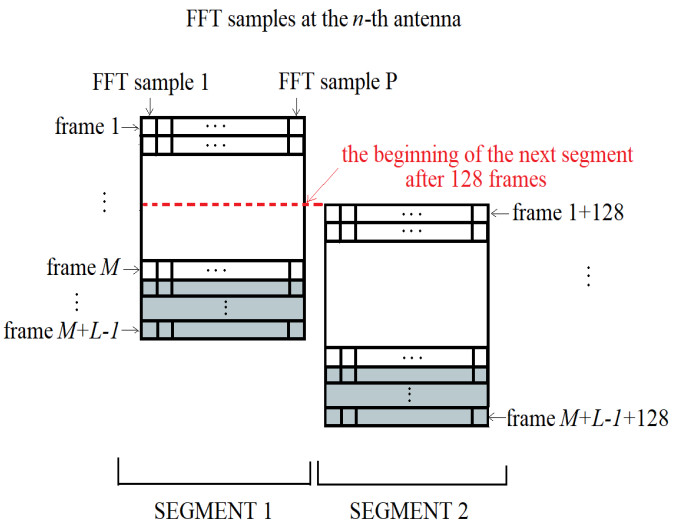


Fig. 2. Segment creation in the proposed algorithm

For each range cell, with index *p*, and each antenna, *n*, we form a matrix $\mathbf{Q}_{p,n} \in C^{M\times L}$ as in Fig. 4 from the *p*-th FFT sample from each frame in Fig. 3. Columns of matrix $\mathbf{Q}_{p,n}$ are vectors $\mathbf{q}_{l,p,n}$, for $1 \le l \le L$, $1 \le p \le P$ and $1 \le n \le N$.

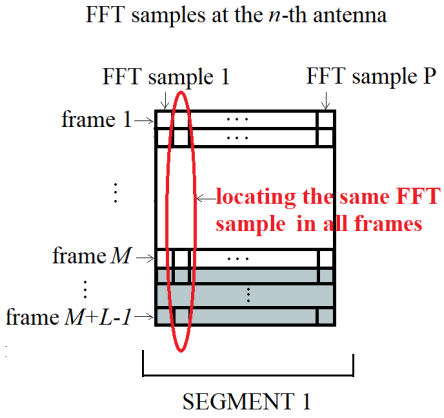


Fig. 3. Locating the same FFT sample in all frames

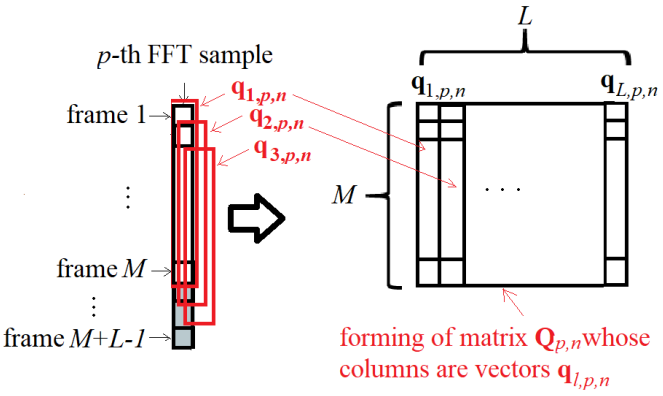


Fig. 4. The creation of the $\mathbf{Q}_{p,n}$ matrix

 ISBN 978-86-7466-930-3

Then the covariance matrices $\mathbf{C}_{p,n} \in \mathrm{C}^{M\times M}$ are formed for n=1, 2, ... , N and p=P−R+1, P−R+2, ... , P, where R represents the maximum projected radar range, as follows:

$$\mathbf{C}_{p,n} = \frac{1}{L}\mathbf{Q}_{p,n}\mathbf{Q}_{p,n}^{\mathrm{H}} \in \mathrm{C}^{M\times M} . \quad (1)$$

The formation of the covariance matrix $\mathbf{C}_{p,n}$ is a key step in the RD-HR map creation. Based on the covariance matrix, the criterion function of high-resolution MUSIC-based algorithm is formulated, as follows:

$$P_{\mathrm{MUS}}^{RD}(\mu,p,n) = \frac{1}{\left\|\mathbf{a}_{\mu}(\mu)^{\mathrm{H}}\,\mathbf{E}_{p,n}\right\|}, \quad (2)$$

where the columns of $\mathbf{E}_{p,n} \in \mathrm{C}^{M\times(M-K)}$ are the M-K eigenvectors from the noise subspace of $\mathbf{C}_{p,n}$, corresponding to the M-K smallest eigenvalues of the covariance matrix $\mathbf{C}_{p,n}$, K is a parameter of the MUSIC-based algorithm, and $\mathbf{a}_{\mu}(\mu) \in \mathrm{C}^{M\times 1}$ is a steering vector formulated in the normalized Doppler domain as

$$\mathbf{a}_{\mu}(\mu) = \left[1, e^{-j\mu}, \dots , e^{-j\mu(M-1)}\right]^{\mathrm{T}} , \quad (3)$$

where the parameter μ denotes the normalized Doppler frequency in radians per frame. The criterion functions are calculated for a set of discrete values of normalized Doppler frequencies and with a grid resolution that is many times better than the Doppler FFT resolution, thus obtaining an RD-HR map.

III. NUMERICALLY EFFICIENT METHOD FOR RD-HR MAP ESTIMATION

As presented in the previous section, the RD-HR map estimation is numerically complex (long execution time). Program for vessel detection of the proposed algorithm must be improved to achieve real-time requirement. Because of that, we propose here a numerically efficient method realized in two steps.

In the first step, we want to make some kind of pattern according to which we select a small subset of the frames from each segment. The reason for that is to reduce numerical complexity. In that case, the dimensionality of the covariance matrix will be smaller ($J \times J$) where $J < M$. The problem of non-uniform sampling method is analogous to the problem formulated in the field of antenna arrays - how to replace a linear uniform antenna array with a non-uniform antenna array with the same aperture and a smaller number of antennas without significant degradation of the antenna array factor. This problem in antenna array theory is known as the problem of minimally redundant linear antenna arrays. Fig. 5 shows the forming of the matrix $\mathbf{Q}^{(l)}{}_{p,n}$ by selecting a subset of J rows of the matrix and according to a chosen mapping ℓ : $\{1, 2, \dots , J\} \rightarrow \{1, 2, \dots , M\}$, $J < M$.

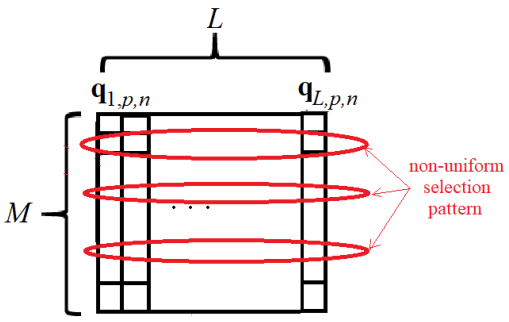


Fig. 5. The creation of the $\mathbf{Q}^{(l)}{}_{p,n}$ matrix using some non-uniform pattern

The same mapping ℓ is used to form the steering vector $\mathbf{a}_{\mu}{}^{(l)}(\mu)$ by non-uniform selection of the elements of the vector $\mathbf{a}_{\mu}(\mu)$.

The criterion function of the high-resolution MUSIC-type algorithm for creating RD-HR map with non-uniform sampling has the same form as the criterion function for the variant with uniform sampling, as follows:

$$P_{\mathrm{MUS}}^{RD\,(l)}(\mu,p,n) = \frac{1}{\left\|\mathbf{a}_{\mu}^{(l)}(\mu)^{\mathrm{H}}\;\mathbf{E}_{p,n}^{(l)}\right\|} . \quad (4)$$

The numerical complexity is significantly reduced, because the eigenvalue decomposition of the covariance matrix is numerically simpler in this case.

Despite the significant reduction in execution time, real-time processing is still not achieved. Therefore, we perform another step in the algorithm optimization.

As can be seen, a similar signal processing is performed on each of the antennas. This leads us to the idea not to form RD-HR maps at all antennas sequentially, one after the other, but simultaneously. Therefore, it is necessary to create a multi-threaded process.

Fig. 6 shows the proposed method for multithreaded RD-HR estimation in order to reduce execution time. A thread is the smallest unit of a processing that can be scheduled by an operating system. The multithreading was realized on CPU cores.

Therefore, the most demanding job, which is the formation of RD maps, was realized through parallel execution on multiple cores, while the rest of the program, which is not computationally demanding, remained to be executed sequentially. The algorithm is also applicable to a system with more antennas, but the performance depends directly on the number of processor cores.

IV. NUMERICAL RESULTS

The results presented in this section are based on the measured radar data. P=1536 and L=64 are predefined values and the developed algorithm was tested for the length of the segment M=256, where the successive segments overlap with 128 frames. Number of antennas directly affect the execution time of the program for vessel detection, and we made some tests with N=16 antennas in the Rx antenna array.

 ISBN 978-86-7466-930-3

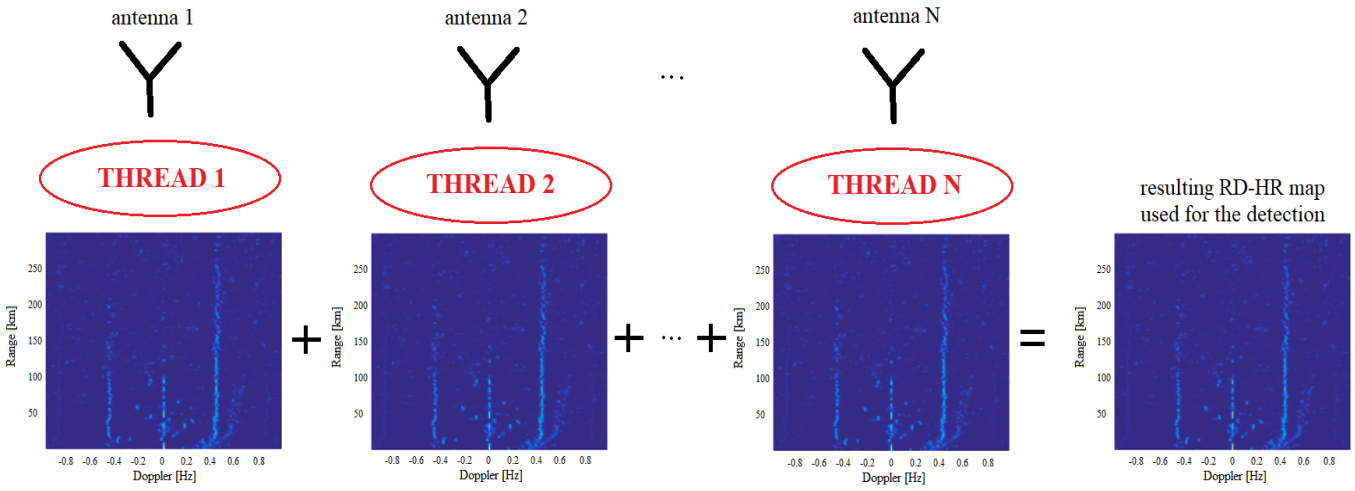


Fig. 6. Multithreaded RD-HR maps creation simultaneously at all antennas

The selected frame duration is 0.256 seconds, and since we want to output the results after every 128 frames, the real-time requirement is 32.768 seconds. Thus, the processing time of one segment must be shorter than 33.28 seconds. We choose also 3 predefined values for the parameter K of the MUSIC-based algorithm: K=10 for ranges up to 120 km, K=20 for ranges from 120 to 200 km, and K=30 for ranges up to 300 km. The number of grid points along the Doppler frequency dimension is 513. These are actually the basic system parameters.

For testing purposes, a program was developed using the programming language C and multithreading is realized too. We first formulate a testing methodology, seen in Fig. 7.

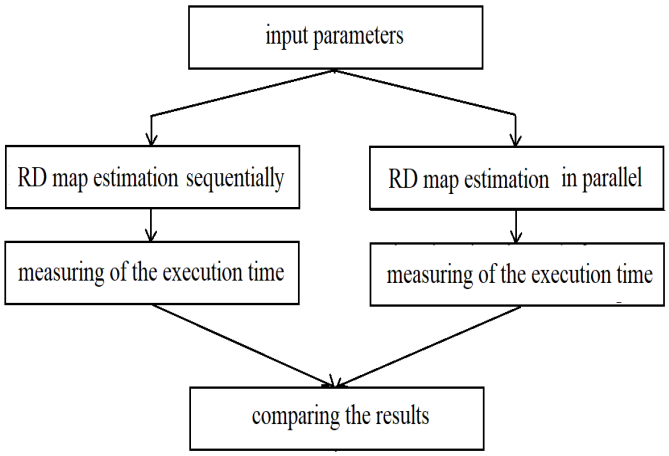


Fig. 7. The overview of the proposed testing methodology

We run the program on a PC with 6-cores CPU (i7), and on a PC with better 8-cores CPU (AMD Ryzen), and for 2 predefined patterns in order to get non-uniform RD-HR maps. The comparison with uniform RD-HR map is also presented. The first pattern is based on prime numbers between 1 and 256 (the last frame number in the segment) and the length of this pattern is J=56, as follows: pattern56={1, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256}.

The length of the second pattern is 88 and it is also based on prime numbers, with some more numbers inserted, in order to reduce the distance between the numbers in the pattern. This leads us to create an RD-HD map that visually resembles a uniform RD HR map, while being numerically simpler: pattern88 ={1, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 35, 37, 41, 45, 48, 50, 53, 55, 59, 61, 63, 67, 71, 73, 76, 77, 83, 85, 89, 92, 97, 101, 103, 105, 107, 109, 111, 113, 115, 118, 121, 125, 127, 129, 131, 134, 136, 137, 139, 142, 145, 147, 149, 151, 155, 157, 160, 163, 167, 171, 173, 179, 181, 184, 187, 191, 193, 197, 199, 202, 205, 209, 211, 215, 219, 223, 227, 229, 233, 236, 239, 241, 245, 247, 251, 254, 256}.

Fig. 8-10 show RD-HR maps for different patterns.

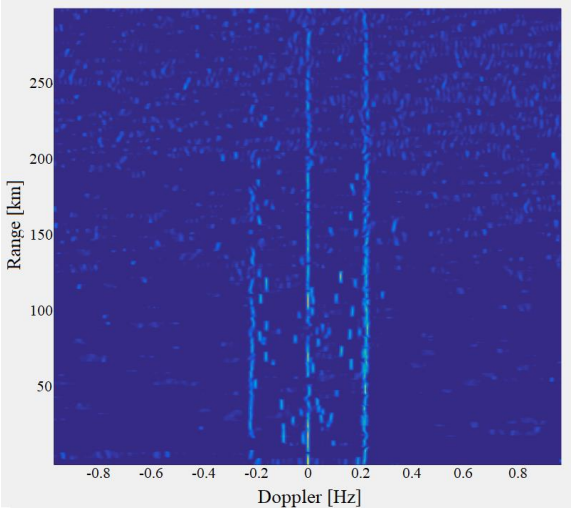


Fig. 8. Uniform obtained RD-HR map of the first segment

 ISBN 978-86-7466-930-3

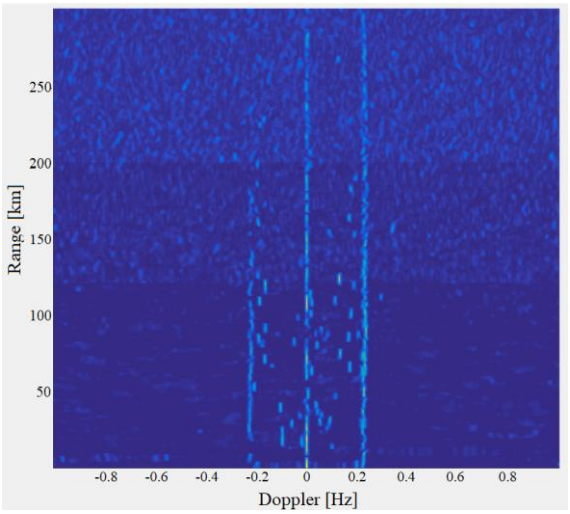


Fig. 9. Non-uniform obtained RD-HR map of the first segment (pattern88)

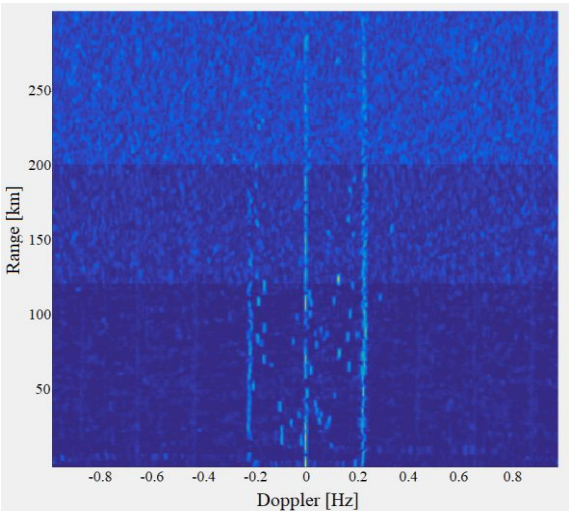


Fig. 10. Non-uniform obtained RD-HR map of the first segment (pattern56)

As can be seen from the figures, non-uniform sampling of frames increases the number of peaks in the RD-HR map compared to the case with uniform sampling, but vessels on the map are clearly visible in all 3 cases.

The dominant tasks in terms of the number of complex multiplications and additions are the eigenvalue decomposition and the calculation of MUSIC criterion function [10]. The approximate number of operations is

$$N_{\text{op}} \sim \frac{13}{3} J^3 RN + N_{\text{d}} J(J+1)RN . \quad (7)$$

In uniform variant $J=M$. Thus, we decrease this number by a factor 44.6 and 15.3 for pattern56 and pattern88, respectively. N_{d} is the length of RD-HR map by Doppler dimension. In this particular case N_{d}=513. Since the numerical complexity is much lower, a non-uniform variant can be used for real-time processing.

Table I shows the execution time of the program (with no-parallelized code) for vessel detection on different computers. Also, the measured execution time to form the RD map is compared to the measured execution time of the rest of the program.

TABLE I
EXECUTION TIME OF THE PROGRAM (NO-PARALLELIZED CODE)

CPU type	Execution time of RD-HR estimation (seconds)	Execution time of the rest of the program (seconds)
Intel CORE i7 1075H	880	3.12
AMD Ryzen 9 5900HX	606	2.64

The results clearly show that the estimation of the RD-HR map is numerically most complex, and the execution time improvement is required in order to reach real-time processing.

Now, we can define the speedup S and the efficiency E by comparing the execution time on one core, T_1, and on n cores, T_n:

$$S = \frac{T_1}{T_n} \quad (5)$$

$$E = \frac{T_1}{nT_n} . \quad (6)$$

Table II shows the comparison between the execution time of the program with and without parallelized code. The obtained results show that the real-time processing is achieved.

TABLE II
EXECUTION TIME OF THE PROGRAM (PARALLELIZED CODE)

CPU type and selected pattern	Execution time of the program without parallelization (seconds)	Execution time of the program with parallelization (seconds)
Intel CORE i7 1075H (uniform)	883.12	132
Intel CORE i7 1075H (pattern88)	91.2	14.21
Intel CORE i7 1075H (pattern56)	32.12	7.54
AMD Ryzen 9 5900HX (uniform)	608.64	90.75
AMD Ryzen 9 5900HX (pattern88)	60.32	9.63
AMD Ryzen 9 5900HX (pattern56)	24.56	4.4

 ISBN 978-86-7466-930-3

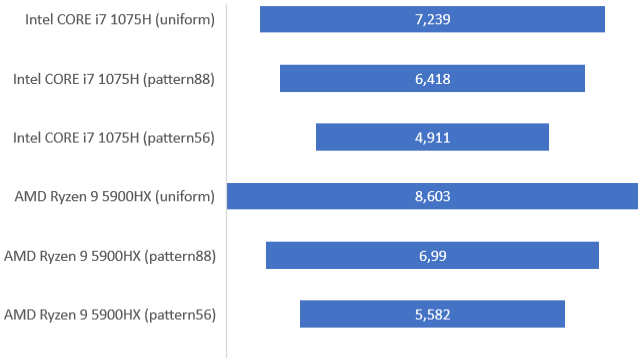


Fig. 11. Speedup of the parallelized algorithm

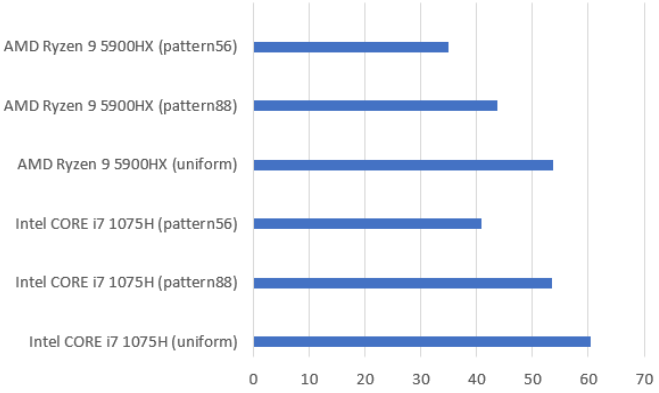


Fig. 12. Efficiency of the parallelized algorithm

Fig. 11 and 12 show speedup and the efficiency of the parallelized algorithm. Logical processors usage in multithread software was shown in Fig. 13.

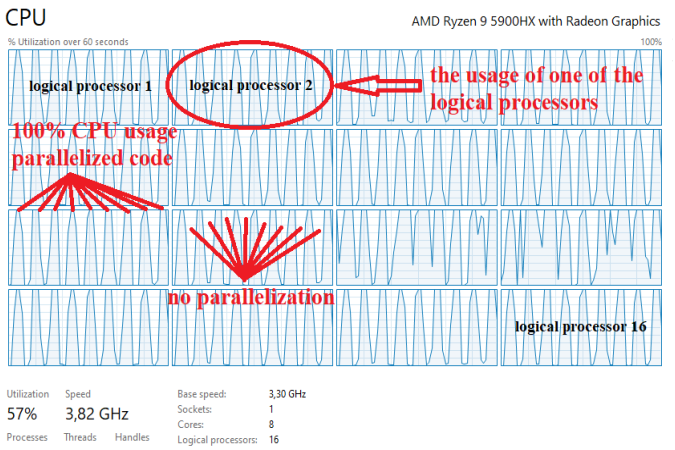


Fig. 13. Logical processors usage in multithread software

V. Conclusion

In this paper, we propose a numerically efficient algorithm that can help many researchers to reach real-time requirement in their programs for vessel detection. The obtained results show that the parallelization of the code is needed, so the parallel signal processing on CPU was performed. Additionally, the proposed method, does not need any specialized hardware, only a general-purpose computer.

Acknowledgment

This research, as a part of P.148 Project, was funded by Vlatacom Institute. The APC was funded by Vlatacom Institute. The research was also supported by the Serbian Ministry of Education, Science and Technological Development.

References

[1] A.M. Ponsford and J. Wang, “A review of high frequency surface wave radar for detection and tracking of ships,” Turk J Elec Eng&Comp Sci, vol. 18, pp. 409–428, 2010.
[2] M. Jankiraman, “FMCW Radar Design,” Kindle ed., Artech House, England, 2018.
[3] K. Gurgel and T. Schlick, “Remarks on Signal Processing in HF Radars Using FMCW Modulation,” Proceedings of the International Radar Symposium IRS 2009, Hamburg, Germany, 2009.
[4] B. Kim, Y. Jin, J. Lee and S. Kim, “High-Efficiency Super-Resolution FMCW Radar Algorithm Based on FFT Estimation,” Sensors 2021, vol. 21, 4018. https://doi.org/10.3390/s21124018 , 2021.
[5] Liu, G.; Yang, W.; Li, P.; Qin, G.; Cai, J.; Wang, Y.; Wang, S.; Yue, N.; Huang, D. “MIMO Radar Parallel Simulation System Based on CPU/GPU Architecture,” Sensors 2022, vol. 22, 396. https://doi.org/10.3390/s22010396
[6] A.C. Sodan, J. Machina, A. Deshmeh, K. Macnaughton, “Parallelism via multithreaded and Multicore CPUs,” IEEE Computer Society, vol. 43, issue: 3, pp. 24-32, Mar. 2010.
[7] D. Dheeraj, B. Nitish, S. Ramesh, “Optimization of Automatic Conversion of Serial C to Parallel OpenMP,” International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover, PES Institute of Technology Bangalore, India, Dec. 2012.
[8] E. Ayguade, N. Copty, A. Duran, J. Hoeflinger, “The Design of OpenMP tasks,” IEEE Transactions on Parallel and Distributed systems, vol. 20, Issue: 3, pp. 404-418, June 2008.
[9] B. Kim, Y. Jin, J. Lee and S. Kim, “Low-Complexity MUSIC-Based Direction-of-Arrival Detection Algorithm for Frequency-Modulated Continuous-Wave Vital Radar”, Sensors 2020, vol. 20, 4295. https://doi.org/10.3390/s20154295
[10] M. Erić, B. Igrić, “Practical Implementation and Performance Estimation of MUSIC Method Implemented on Signal Processor TMS 320c30,” Scientific-Technical Review, vol. LIV, No.1, 2004.

 ISBN 978-86-7466-930-3

Layer 2 Forwarding Using T4P4S: P4 Language and Data Plane Development Kit

Dimitrije Jovanović and Aleksandra Smiljanić, *Member, IEEE*

***Abstract*—P4 is an open-source programming language designed to program protocol-independent packet processors. T4P4S is a P4 compiler that can enable P4 programs to run on network devices using the Data Plane Development Kit (DPDK) framework. This paper covers the test environment for layer 2 forwarding using DPDK. We compare the layer 2 forwarding switch compiled using T4P4S and the appropriate P4 program versus the layer 2 forwarding switch provided with DPDK package.**

***Index Terms*—Programmable data plane, P4 language, DPDK, T4P4S.**

I. Introduction

P4 language (Programming Protocol-Independent Packet Processors) [1] is an open source, domain specific programming language used to specify packet processing of data plane devices. It provides high-level instructions for describing transformations of network data, and enables the end-user to define data plane functionalities, regardless of the network protocol.

Before the introduction of programmable data planes [2], the specialized hardware and low-level programming languages were used to implement data plane logic, as the only way to achieve satisfactory packet processing rates. P4 programmable hardware devices (targets) have so far achieved great performances, but they are still very expensive. Originally, the main goal of P4 is protocol-independence. However, it was not designed to be target-independent.

T4P4S (Translator for P4 Switches) is a P4 compiler able to support a variety of hardware devices [3]. To solve direct target-dependence, the Networking Hardware Abstraction Layer (NetHAL) has been introduced. The compiler translates the P4 program into C language code of a hardware independent core switch, using NetHAL as an interface between hardware and the switch code. The result is a hardware dependent DPDK software switch compatible with a given device. The simplified process of generating a switch program is illustrated in Fig. 1

Introduced software abstraction makes the P4 program target-independent. However, the overall performance is potentially lower. To this end, we conduct a simple layer 2 forwarding experiment (l2fwd), where we first forward packets using DPDK (Data Plane Development Kit) [4]. Then, we implement a P4 program with the same forwarding function by using T4P4S with DPDK. The goal is to verify integration of P4 and DPDK, by showing that the tests T4P4S l2fwd and DPDK l2fwd produce similar results.

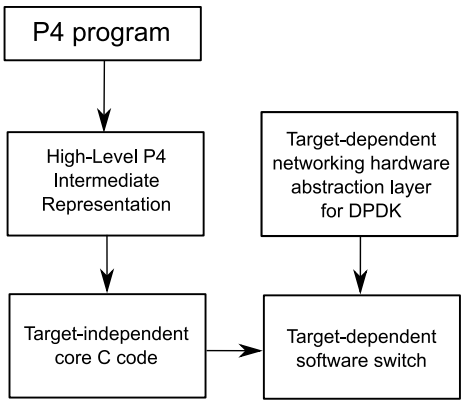


Fig. 1. Generation of the DPDK software switch using T4P4S.

This paper is structured as follows. Sections II, III and IV introduce the P4 language and its components. In section V, we cover an example of P4 program used for layer 2 forwarding. The testing environment is presented in Section VI. Finally, the testing results are shown in Section VII. We conclude the work in Section VIII.

II. Packet Forwarding in P4

While the data plane performs transformations and forwarding of packets, the control plane configures the data plane and supplies it with network data. P4 defines functionalities of the data plane that was traditionally defined by the manufacturer. Inside a P4 switch, the data plane and its set of tables are not fixed in advance but rather defined by a P4 program. Thus, the data plane has no in-built knowledge of the existing network protocols [5].

P4 language adopts an abstract model for the packet processing pipeline which consists of three main stages [6]:

- Parser extracts user-defined packet headers from the received packet. Parsers are written in the form of finite state machines with three predefined states: start, accept and reject.
- Controls perform transformations on extracted headers. They define the control flow, an imperative program which operates on tables called match-action units. These tables consist of keys, matching types and actions that are executed upon matching.
- Deparser assembles the packet from processed headers. It performs the serialization of user-defined packet headers into the packet, and emission of these headers on the proper egress port.

Dimitrije Jovanović is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: dica@etf.bg.ac.rs).

Aleksandra Smiljanić is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: aleksandra@etf.bg.ac.rs).

 ISBN 978-86-7466-930-3

III. P4 Architecture

The P4 architecture of each target consists of its P4 programmable elements and their data plane interfaces [5]. Regardless of the target's architecture type, the core of packet processing is the programmable match-action pipeline, which consists of metadata bus and a sequence of connected match-action units.

Metadata bus allows the communication between different phases of the match-action pipeline. It is considerably wider than common CPU buses, thus, these kinds of architectures can achieve high data processing speeds.

The most general architectural model consists of the ingress and egress match-action pipeline on its ends, and the traffic manager in between. The traffic manager performs packet queueing, replication and scheduling. The pipelines, parser and deparser are P4 programmable, while the traffic manager is fixed function.

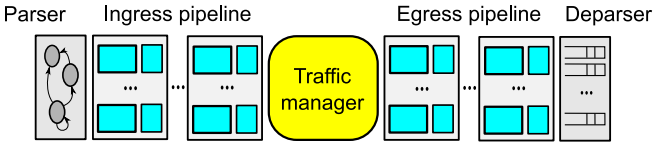


Fig. 2. V1Model P4 architecture.

One of the most common P4 architectures is the V1Model [7], which is illustrated in Fig. 2. Here, the ingress pipeline is placed after the parser and performs checksum verification. Symmetrically, the egress pipeline is placed before the deparser and performs checksum update. For example, T4P4S can use the V1Model architecture, which will also be implemented in this paper for testing.

IV. P4 Programming Components

The core of the P4 language consists of data types, expressions, declarations and statements. The rest of the language is directly used for expressing parsers, match-action units and architectures [5].

P4 is a statically typed language. It provides several built-in base types (void, error, strings, match kind, Boolean, integers) and derived data types (enumeration, header, struct, tuple, extern, parsers, control blocks, packages, etc.).

Certain sorts of expressions in P4 can be executed only on a limited set of data types. The complete grammar production rule for general expressions can be found in the language specification.

Declarations are used for introducing functions, constants, and variables, which are executed at run time. Specially, data types with constructors (extern, control blocks, parsers, packages) use instantiations, which are executed at compilation time.

Statements can appear within parser states, a control block or an action. Some kinds of statements cannot appear in certain blocks. Control blocks support all of the following statements: assignment, empty, block, exit, return, switch.

Control blocks are used to manipulate and transform parsed headers. Data is transformed by match-action units, which are described by tables. One action consists of the code, which is in the P4 program, and the data, which is in table entries populated by the control plane. This is shown in Fig. 3.

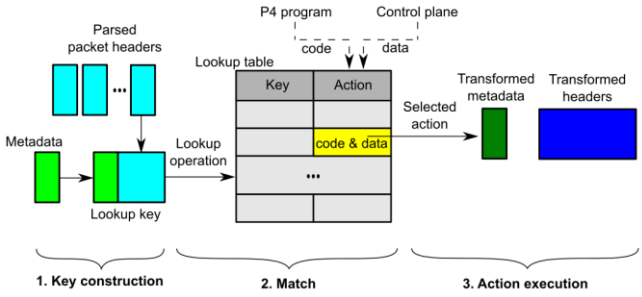


Fig. 3. Match-action unit.

Match-action units perform the following operations:

- Key construction, using packet headers or metadata.
- Match, by searching the key in the lookup table.
- Action execution, by which the input data is transformed.

Control blocks are also used to define deparsing (packet reconstruction), which is an inverse operation of parsing.

V. P4 Layer 2 Forwarding Example Program

A simple configuration of layer 2 forwarding between two servers is illustrated in Fig. 4. A DPDK-compatible traffic generator is used to produce and transmit packets from server 1 (source server) to server 2 (target server). In the first case, server 2 uses a basic DPDK l2fwd example [8] to receive and forward packets. In the second case, we use a P4 program compiled by T4P4S. The same packet processing function is implemented in both cases.

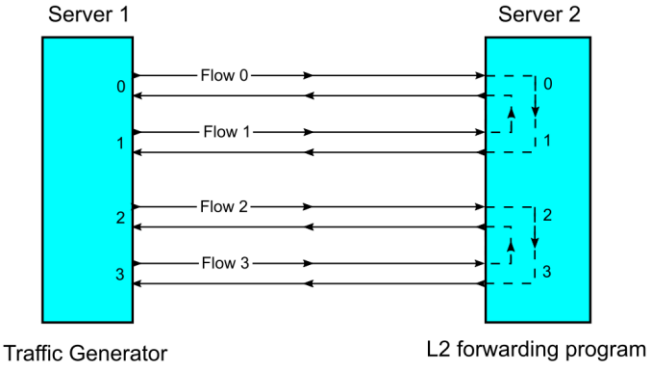


Fig. 4. Test configuration [8].

The P4 program given in Fig. 5 is used to generate our l2fwd switch example. Lines 1-8 define header types to be parsed from the packet. Lines 10-13 define the parser, which in this case extracts the Ethernet header from the ingress packet. Lines 15-58 define the control flow, which consists of tables (smac, dmac), actions (_drop, _nop, mac_learn, forward, broadcast) and controls (ingress, egress).

In this example, exact matching is used for table lookup. If the source MAC address is not found in the smac table, a digest is sent to the control plane to add the source address and the ingress port to both tables (action mac_learn).

If the source MAC address already exists in the table, nothing needs to be done (action _nop). In the dmac table,

 ISBN 978-86-7466-930-3

the exact matching lookup determines the egress port based on the destination MAC adress. If there is a match, the packet is forwarded to that egress port (action forward). Otherwise, if the destination MAC address is not found in the table, the packet is sent to all ports except the ingress port (action bcast).

```
header_type ethernet_t {
   fields {
      dstAddr : 48;
      srcAddr : 48;
      etherType : 16;
   }
}
header ethernet_t ethernet;

parser start {
   extract(ethernet);
   return ingress;
}

action _drop() {
   drop();
}
action _nop() {
}
#define MAC_LEARN_RECEIVER 1024
field_list mac_learn_digest {
   ethernet.srcAddr;
   standard_metadata.ingress_port;
}
action mac_learn() {
   generate_digest(MAC_LEARN_RECEIVER,
                   mac_learn_digest);
}
table smac {
   reads {
      ethernet.srcAddr : exact;
   }
   actions {mac_learn; _nop;}
   size : 512;
}
action forward(port) {
   modify_field(
        standard_metadata.egress_port,
        port);
}
action bcast() {
   modify_field(
        standard_metadata.egress_port,
        100);
}
table dmac {
   reads {
      ethernet.dstAddr : exact;
   }
   actions {forward; bcast;}
   size : 512;
}
control ingress {
   apply(smac);
   apply(dmac);
}
control egress {
}
```

Fig. 5. Layer 2 forwarding P4 program (./examples/p4_14/l2fwd.p4_14) [9]

VI. Test Environment

One of the testing goals is the comparison of running DPDK l2fwd from the CentOS host and the Ubuntu container, to verify the Docker solution and to detect the performance degradations if any. Also, DPDK l2fwd can be compared with T4P4S l2fwd, both on Ubuntu, in terms of different metrics, such as bit rate and packet size. The tests are performed on the target server using 2x2 ports (port 0 forwards to port 1 and port 2 forwards to port 3 and vice versa), as shown in Fig. 4.

The source test server used to generate traffic is Supermicro with 125 GB RAM, 24 cores and four Intel X710 NICs, each with 4 x 10 Gbit/s Ethernet ports. The traffic is generated using the Pktgen traffic generator.

The target test server is Supermicro with 250 GB RAM, 32 cores and five Intel X710 NICs, each with 4 x 10 Gbit/s Ethernet ports. The existing CentOS 7 host operating system contains the working DPDK version 18.11, which is not suitable for T4P4S installation due to incompatibility. To resolve that, we installed the latest version of T4P4S with underlying DPDK 21.11 on the Ubuntu 20.04 Docker container.

The configuration of the CentOS host includes the installation of DPDK required kernel modules uio, uio_pci_generic, vfio_pci and igb_uio (the last one compiled from dpdk-kmods Github repository), as well as the reservation of hugepages that will be used by DPDK. The Ubuntu 20.04 Docker container needs to be started with the option that mounts the host file system /dev/hugepages under /mnt/huge mounting point. Also, the container needs to run in the privileged mode with the host networking option, to allow the full access to the host drivers and the network cards.

In the Ubuntu container, it is required to install several missing packages, because the basic Ubuntu 20.04 Docker image assumes the very minimal configuration without development libraries, tools and compilers.

The detailed instructions on how to build T4P4S are given in the GitHub P4ELTE/t4p4s project [9]. Before testing, it is required to bind network cards to DPDK igb_uio driver, using the dpdk-devbind.py command. The packet generator (GitHub Pktgen-DPDK project) is located on another similar server, where four X710 NICs are directly connected to the selected four X710 NICs on the testing server, as shown in Fig. 4. The test configuration measures internal packet transfer between ports 0-1 and between ports 2-3, and also measures performances of parallel processing of two independent forwarding mechanisms.

TABLE I
DPDK L2FWD AND T4P4S L2FWD COMMAND LINE OPTIONS

./dpdk-l2fwd -c 0xf -n 4 -- -p 0xf
./t4p4s.sh :l2fwd model=v1model coropt prtopt rssopt
grep "^???opt" opts_dpdk.cfg coropt -> ealopts += -c 0xf -n 4 prtopt -> cmdopts += -p 0xf –config "\" (0,0,0),(0,1,1),(0,2,2),(0,3,3),(1,0,0),(1,1,1),(1,2,2),(1,3,3), (2,0,0),(2,1,1),(2,2,2),(2,3,3),(3,0,0),(3,1,1),(3,2,2),(3,3,3)\"" rssopt -> cflags += -DT4P4S_RTE_RSS_HF=0x7ef8

 ISBN 978-86-7466-930-3

DPDK l2fwd is configured using command line options, which consist of Environment Abstraction Layer (EAL) options and application options. T4P4S l2fwd is configured by user-defined command line options, that we need to define in advance in the T4P4S configuration file opts_dpdk.cfg. EAL options are the same, but the application options differ for DPDK l2fwd and T4P4S l2fwd. For T4P4S l2fwd, we also need to specify the P4 architectural model, for example the V1Model.

The command line options presented in Table I will be used in all of the following tests. The number of processor cores is set to 4 (option -c and bit mask 0xf), the number of enabled ports is also 4 (option -p, bit mask 0xf). In the case of DPDK l2fwd, ports 0 and 1 forward packets to each other and ports 2 and 3 forward packets to each other by default. For T4P4S l2fwd, we need to explicitly map each RX port to the specific port, queue and processor core (port, queue, core). Ports that form the same traffic flow must be on the same socket, which is socket 0 in this case. The last option of the T4P4S l2fwd command in Table I is used to provide the proper RSS value, at compilation time. Configuration of the T4P4S l2fwd switch can be verified by looking at the command output presented in Fig. 6. One port handles ingress packets in nb_rxq queues, and egress packets in nb_txq queues. This output is practically the same for both DPDK l2fwd and T4P4S l2fwd.

```
# ./t4p4s.sh :l2fwd model=v1model coropt prtopt rssopt

[RUN CONTROLLER] dpdk_l2fwd_controller (default for l2fwd@std)

[COMPILE  P4-14] ./examples/p4_14/l2fwd.p4_14 @std

[COMPILE SWITCH]
[57/57] Linking target l2fwd.
[RUN SWITCH] ./build/last/build/l2fwd
EAL: Detected CPU lcores: 32
EAL: Detected NUMA nodes: 2
EAL: Detected shared linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No available 1048576 kB hugepages reported
EAL: VFIO support initialized
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:01:00.0 (socket 0)
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:01:00.1 (socket 0)
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:01:00.2 (socket 0)
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:01:00.3 (socket 0)
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:02:00.0 (socket 0)
…      …              …
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:03:00.3 (socket 0)
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:82:00.0 (socket 1)
…      …              …
EAL: Probe PCI driver: net_i40e (8086:1572) device: 0000:84:00.3 (socket 1)
TELEMETRY: No legacy callbacks, legacy socket not created

P4_FWD: entering main loop on lcore 1
P4_FWD: entering main loop on lcore 0
P4_FWD:  -- lcoreid=0 portid=0 rxqueueid=0
P4_FWD:  -- lcoreid=0 portid=1 rxqueueid=0
P4_FWD:  -- lcoreid=0 portid=2 rxqueueid=0
P4_FWD:  -- lcoreid=0 portid=3 rxqueueid=0
P4_FWD: entering main loop on lcore 2
P4_FWD:  -- lcoreid=2 portid=0 rxqueueid=2
P4_FWD:  -- lcoreid=2 portid=1 rxqueueid=2
P4_FWD:  -- lcoreid=2 portid=2 rxqueueid=2
P4_FWD:  -- lcoreid=2 portid=3 rxqueueid=2
P4_FWD:  -- lcoreid=1 portid=0 rxqueueid=1
P4_FWD:  -- lcoreid=1 portid=1 rxqueueid=1
P4_FWD:  -- lcoreid=1 portid=2 rxqueueid=1
P4_FWD:  -- lcoreid=1 portid=3 rxqueueid=1
P4_FWD: entering main loop on lcore 3
P4_FWD:  -- lcoreid=3 portid=0 rxqueueid=3
P4_FWD:  -- lcoreid=3 portid=1 rxqueueid=3
P4_FWD:  -- lcoreid=3 portid=2 rxqueueid=3
P4_FWD:  -- lcoreid=3 portid=3 rxqueueid=3
```

Fig. 6. Initialization of T4P4S l2fwd switch.

VII. Test Results

The relation between the maximal packet rate and bit rate on an Ethernet link is given in (1). Ethernet frame of size L is preceeded by preamble (7 B) and SFD field (1 B), and it is followed by an inter-frame gap (at least 12 B). Therefore, the total size of transmitted data is at least L + 20 B.

$$R_{packet,\max} = \frac{R_{bit}}{L + 20B} \quad (1)$$

There is a hardware upper limit for packet rate when the traffic is generated on all four ports simultaneously. This value depends on the model of used NIC. The maximal TX packet rate on one port for the shortest frames (64 B) is given in the last row of Table II, which equals 10.066 Mp/s. If we had used only two ports in our experiment, the maximal TX packet rate on one port would be 14.88 Mp/s, according to (1).

The results of several tests for layer 2 forwarding, for different frame sizes and TX packet rates, are given in Table II. We initiated the Ethernet traffic with the given parameters on all four ports simultaneously, and measured RX packet rate on each port. Table II gives the achieved RX packet rate for one port. We considered four values for Ethernet frame size, ranging from 1518 B to 64 B (column 1). TX packet rate (column 2) is given as the percentage of the maximal packet rate (column 3), because the rate is set that way in the Pktgen command line. For each frame size, we considered TX packet rate values of 10% and 100% of the maximal packet rate.

The tests produced practically identical results for RX rate comparing DPDK l2fwd on CentOS host and on Ubuntu Docker container. That excludes any performance degradation possibly caused by the Docker container.

TABLE II
Comparing DPDK l2fwd on CentOS host and Ubuntu Docker container: RX packet rate

Frame size [B]	% max packet rate	TX packet rate [Mp/s]	RX packet rate CentOS Host [Mp/s]	RX Packet rate Ubuntu container [Mp/s]
1518	10	0.08128	0.08128	0.08128
512	10	0.24944	0.24944	0.24944
256	10	0.45293	0.45293	0.45293
64	10	1.488	1.488	1.488
1518	100	0.81275	0.81274	0.81274
512	100	2.34966	2.34962	2.3496
256	100	4.52906	4.52896	4.52896
64	100	10.0798	9.40504	9.39915

Pktgen [10] is a software traffic generator that is a part of the DPDK framework. It can configure and display metrics for traffic flows at real time. Pktgen command line options (Fig. 7) are split into Environment Abstraction Layer (EAL) options and application options. The mask of used processor cores and the number of memory channels are the required EAL arguments. We can also specify the PCI devices and

 ISBN 978-86-7466-930-3

the memory allocated from hugepages on specific sockets. In the application options, we can map ports to logical cores. Run time commands are entered in Pktgen prompt while it is running. For each port, we can specify the number of packets to transmit and their size, packet rate (in percentage of the maximal rate), as well as the source and the destination MAC address (or their range).

In our tests, the source MAC address of port 0 is the destination MAC address of port 1 and vice versa. Ports 2 and 3 are paired in the same way. Pktgen displays traffic results for each port at real time, such as the achieved rate and number of transmitted and received packets. The example of Pktgen output for ports 0 and 1 is presented in Fig. 7.

```
# ./app/x86_64-native-linuxapp-gcc/pktgen -c 0xfffff
-w 82:00.0 -w 82:00.1 -w 82:00.2 -w 82:00.3  --file-prefix=pg1
--socket-mem=8192,8192  --  -m  "[6:7].0,  [8:9].1,  [10:11].2,
[17:16].3


| Ports 0-3 of 4   <Main Page>  Copyright (c) <2010-2017>,
  Flags:Port      :   ---------------:0   ---------------:1
Link State        :       <UP-10000-FD>       <UP-10000-FD>
Pkts/s Max/Rx     :       234944/234944       234983/234944
       Max/Tx     :       235008/234944       234988/234962
MBits/s Rx/Tx     :             999/999             999/999
Broadcast         :                   0                   0
Multicast         :                   0                   0
  64 Bytes        :                   0                   0
  65-127          :                   0                   0
  128-255         :                   0                   0
  256-511         :             2583246             2583969
  512-1023        :                   0                   0
  1024-1518       :                   0                   0
Runts/Jumbos      :                 0/0                 0/0
Errors Rx/Tx      :                 0/0                 0/0
Total Rx Pkts     :             2407086             2407680
      Tx Pkts     :             2407680             2407570
      Rx MBs      :               10244               10247
      Tx MBs      :               10247               10246
ARP/ICMP Pkts     :                 0/0                 0/0
                  :
Pattern Type      :               abcd...             abcd...
Tx Count/% Rate   :        Forever /10%        Forever /10%
PktSize/Tx Burst  :          512 /   64          512 /   64
Src/Dest Port     :         1234 / 5678         1234 / 5678
Pkt Type:VLAN ID  :     IPv4 / TCP:0001     IPv4 / TCP:0001
802.1p CoS        :                   0                   0
ToS Value:        :                   0                   0
  - DSCP value    :                   0                   0
  - IPP  value    :                   0                   0
Dst  IP Address   :         192.168.1.1         192.168.0.1
Src  IP Address   :       192.168.0.1/24      192.168.1.1/24
Dst MAC Address   :   ac:1f:6b:2d:1a:c9   ac:1f:6b:2d:1a:c8
Src MAC Address   :   ac:1f:6b:2d:1a:c8   ac:1f:6b:2d:1a:c9
VendID/PCI Addr   :   8086:1572/82:00.0   8086:1572/82:00.1

-- Pktgen Ver: 3.5.2 (DPDK 17.11.4)  Powered by DPDK ------
```

Fig. 7. Pktgen output for T4P4S l2fwd (showing ports 0 and 1 only).

By turning on T4P4S l2fwd debugging, in Fig. 8 one can see how the process of MAC learning and forwarding goes on. We can follow the specified packet by looking at lines that begin with its assigned string. For the packet on the ingress port 3 (lines 0@0), the parser enters the start state and extracts the Ethernet header. The source MAC address ac:1f:6b:2d:1a:cb is not found in the smac table, and therefore that address-port pair is added both to the smac table and the dmac table. The next incoming packet on the ingress port 2 (lines 1@0) has the destination MAC address ac:1f:6b:2d:1a:cb. Since now there is a match in the dmac table, the packet is forwarded to the port 3.

Since we concluded that l2fwd tests behave the same on the CentOS host and the Ubuntu Docker container, we can compare DPDK l2fwd and T4P4S l2fwd, both on the Docker container. We executed the tests with analogous parameters with the parameters for comparing DPDK l2fwd on CentOS host and Ubuntu Docker container. The results are shown in Table III and Table IV.

```
0@0 Handling packet #-01 (port 3, 60B): ac1f 6b2d 1aca ac1f 6b2d
1acb 0800 4500 002e 492e 0000 0406 e749 c0a8 0301 c0a8 0201 04d2
162e 1234 5678 1234 5690 5010 2000 fbe6 0000 7778 797a 3031
0@0  %%%% Parser state start
0@0    :: Parsed header#1 ethernet/14B: .dstAddr/6B=ac1f_6b2d_1aca
.srcAddr/6B=ac1f_6b2d_1acb .etherType/2B=2048=0x0800
0@0  %%%% Packet is accepted, 14B in 1 header, 46B of payload: 4500
002e 492e 0000 0406 e749 c0a8 0301 c0a8 0201 04d2 162e 1234 5678
1234 5690 5010 2000 fbe6 0000 7778 797a 3031
1@0 Handling packet #-01 (port 2, 60B): ac1f 6b2d 1acb ac1f 6b2d
1aca 0800 4500 002e 49b5 0000 0406 e6c2 c0a8 0201 c0a8 0301 04d2
162e 1234 5678 1234 5690 5010 2000 fbe6 0000 7778 797a 3031
1@0  %%%% Parser state start
1@0    :: Parsed header#1 ethernet/14B: .dstAddr/6B=ac1f_6b2d_1acb
.srcAddr/6B=ac1f_6b2d_1aca .etherType/2B=2048=0x0800
1@0  %%%% Packet is accepted, 14B in 1 header, 46B of payload: 4500
002e 49b5 0000 0406 e6c2 c0a8 0201 c0a8 0301 04d2 162e 1234 5678
1234 5690 5010 2000 fbe6 0000 7778 797a 3031
1@0  ++++ Lookup on smac/exact/6B:
ethernet.srcAddr/48b=ac1f_6b2d_1aca -> hit mac_learn
1@0     < Sending digest to port 1024
1@0         : ethernet.srcAddr/48 = ac1f 6b2d 1aca
1@0         : all_metadatas.ingress_port/9 = 0200
0@0  ++++ Lookup on smac/exact/6B:
ethernet.srcAddr/48b=ac1f_6b2d_1acb -> hit mac_learn
0@0     < Sending digest to port 1024
1@0  <<<< Sending digest to port 1024 using extern
extern_Digest_pack for cpd
0@0         : ethernet.srcAddr/48 = ac1f 6b2d 1acb
0@0         : all_metadatas.ingress_port/9 = 0300
0@0  <<<< Sending digest to port 1024 using extern
extern_Digest_pack for cpd
---  ctl> Add dmac/exact: forward($port/9b=2=002) <- ac1f 6b2d 1aca
---  ctl> Add smac/exact: _nop <- ac1f 6b2d 1aca
---  ctl> Add dmac/exact: forward($port/9b=3=003) <- ac1f 6b2d 1acb
---  ctl> Add smac/exact: _nop <- ac1f 6b2d 1acb
1@0  ++++ Lookup on dmac/exact/6B:
ethernet.dstAddr/48b=ac1f_6b2d_1acb -> hit forward($port/9b=3=003)
1@0     = Set all_metadatas.egress_port/9b = 3 = 0x0003
1@0  <<<< Emitting packet #-01 with unchanged structure on port 3:
14B of headers, 46B of payload
```

Fig. 8. Debug output of T4P4S l2fwd program.

TABLE III
COMPARING DPDK L2FWD AND T4P4S L2FWD: RX PACKET RATE

Frame size [B]	% max packet rate	TX packet rate [Mp/s]	RX packet rate DPDK [Mp/s]	RX packet rate T4P4S [Mp/s]
1518	10	0.08128	0.08128	0.08128
512	10	0.24944	0.24944	0.24944
256	10	0.45293	0.45293	0.45293
64	10	1.488	1.488	1.48782
1518	100	0.81275	0.81274	0.81273
512	100	2.34966	2.3496	2.30071
256	100	4.52906	4.52896	2.3005 *
64	100	10.0798	9.39915	2.3109 *

TABLE IV
COMPARING DPDK L2FWD AND T4P4S L2FWD: RX BIT RATE

Frame size [B]	% max bit rate	TX bit rate [Gbit/s]	RX bit rate DPDK [Gbit/s]	RX bit rate T4P4S [Gbit/s]
1518	10	1	1	1
512	10	1	0.999	0.999
256	10	1	0.999	0.999
64	10	1	0.999	0.999
1518	100	10	9.999	9.999
512	100	10	9.999	9.754
256	100	10	9.995	5.059 *
64	100	6.75	6.185	1.535 *

 ISBN 978-86-7466-930-3

In Table III and Table IV, one can observe that the RX rate of T4P4S l2fwd is similar to RX rate of DPDK l2fwd for lower rates and longer packets. But in the case of higher rates and shorter packets, there is a considerable amount of lost packets in the case of T4P4S l2fwd. The values for RX rate for T4P4S l2fwd denoted with (*) in Table III and Table IV represent that case. The encountered issues are caused by a suboptimal software integration of DPDK and P4, which includes a trade-off between flexibility and performance. Target-independence of the T4P4S compiler introduces generalizations related to memory allocation, which are common sources of performance issues. The improvement in that area will be the subject of future work.

VIII. Conclusion

We have demonstrated that P4 features can be readily tested using standard equipment, consisting of two Linux servers with sufficient memory and processor cores, and standard network cards. We have considered two cases: when packet forwarding is done just using DPDK, and when P4 program is executed on the receiver side, by using T4P4S over DPDK. We have shown that, in most cases, there are no considerable differences between RX bit rate measured for DPDK layer 2 forwarding and T4P4S layer 2 forwarding that implements the P4 language program. We have observed a drop in performance of the implemented P4 software for the shortest packets. On the other hand, the P4 language greatly improves the flexibility of packet processing, while T4P4S makes P4 programs portable across multiple targets. In the future, we will work on improving the performance of the integrated DPDK and P4 software.

Acknowledgment

This paper is sponsored in part by the Serbian Ministry of Education, Science and Technological Development.

References

[1] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese and D. Walker, “P4: Programming Protocol-independent Packet Processors,” *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 87–95, July 2014.

[2] R. Bifulco and G. Rétvári, “A Survey on the Programmable Data Plane: Abstractions, Architectures, and Open Problems,” 2018 IEEE 19th International Conference on High Performance Switching and Routing, Bucharest, Romania, June 2018.

[3] P. Vörös, D. Horpácsi, R. Kitlei, D. Leskó, M. Tejfel, and S. Laki, “T4P4S: A Target-independent Compiler for Protocol-independent Packet Processors,” 2018 IEEE 19th International Conference on High Performance Switching and Routing, Bucharest, Romania, June 2018.

[4] H. Bi and Z. Wang, “DPDK-based Improvement of Packet Forwarding,” 3rd Annual International Conference on Information Technology and Applications, Hangzhou, China, July 2016.

[5] P. L. Consortium, “P4 16 Language Specification,” [Online]. Available: https://p4.org/p4-spec/docs/P4-16-v1.2.2.pdf, May 2021.

[6] H. Harkous, M. Jarschel, M. He, R. Pries and W. Kellerer, “Towards Understanding the Performance of P4 Programmable Hardware,” 2019 ACM/IEEE Symposium on Architectures for Networking and Communications Systems, Cambridge, UK, September 2019.

[7] F. Hauser, M. Häberle, D. Merling, S. Lindner, V. Gurevich, F. Zeiger, R. Frank and M. Menth, “A Survey on Data Plane Programming with P4: Fundamentals, Advances, and Applied Research,” [Online]. Available: https://arxiv.org/pdf/2101.10632.pdf, August 2021.

[8] Intel Corporation, “L2 Forwarding Sample Application (in Real and Virtualized Environments),” [Online]. Available: https://doc.dpdk.org/guides/sample_app_ug/l2_forward_real_virtual.html

[9] P4ELTE, “Retargetable compiler for the P4 language,” [Online]. Available: https://github.com/P4ELTE/t4p4s

[10] D. Turull, P. Sjödin, and R. Olsson, “Pktgen: Measuring performance on high speed networks,” *Computer Communications*, vol. 82, pp. 39–48, March 2016.

 ISBN 978-86-7466-930-3

Mogućnost primene beacon tehnologiji za razvoj Covid-19 sistema za praćenje kontakta u visokoškolskim institucijama

Ivana Stefanović, Milutin Nešić i Marko Milivojčević

***Apstrakt*—U ovom radu razmatrana je mogućnost primene *beacon* tehnologije za razvoj rešenja koje omogućava praćenje rizičnih kontakta kao i automatizaciju sprovođenja propisanih mera u realnom vremenu, u cilju sprečavanja zaražavanja i širenja infekcije izazvane virusom Covid-19 u visokoškolskim institucijama. Za realizaciju sistema dovoljno je da studenti i zaposleni kod sebe imaju pametni telefon na kome je instalirana odgovarajuća aplikacija. Sistem se bazira na tome da će student i zaposleni kod kojih je utvrđeno prisustvo virusa Covid-19, putem aplikacije, dobrovoljno obavestiti visokoškolsku ustanovu o pozitivnom testu. Sprovedena je anonimna anketa kako bi se utvrdilo da li su student i zaposleni zainteresovani za implementaciju Covid-19 sistema za praćenje kontakta. Izvršena su i praktična merenja, a dobijeni rezultati potvrđuju da je *beacon* tehnologija odličan izbor za implementaciju sistema.**

***Ključne reči*—Beacon; BLE; Covid-19; pozicioniranje.**

I. Uvod

Globalna pandemija izazvana virusom Covid-19 utiče na sve aspekte ljudskog života, uključujući i obrazovanje. Zavisno od trenutne epidemiološke situacije i u skladu sa preporukama Ministrastva prosvete, nauke i tehnološkog razvoja nastava u visokoškolskim ustanovama odvija se uživo, na daljinu ili kombinovano. Za razliku od samog početka pandemije, kada se celokupna nastava odvijala na daljinu, danas je u većini visokoškolskih ustanova usvojen kombinovani model koji podrazumeva da se deo nastave održava uživo u školama. Većina mera za zaštitu zdravlja studenata i zaposlenih odnosi se na održavanje fizičke distance, nošenje maski, redovno pranje ruku, čišćenje i dezinfekciju školskih prostorija. Uprkos svim merama koje škole sprovede, dolazi do zaražavanja studenata i zaposlenih, što dovodi u rizik druge studente i zaposlene koji su bili u neposrednom kontaktu sa zaraženim.

U okviru ovog rada predstavljen je centralizovani sistem čija je osnovna funkcija identifikacija i obaveštavanje studenata i zaposlenih o neposrednom kontaktu u školi sa osobom kod koje je potvrđeno prisustvo virusa Covid-19. Sistem je baziran na upotrebi *beacon* tehnologije koja pored praćenja kontakta omogućava i automatizaciju sprovođenja skoro svih propisanih mera u realnom vremenu u cilju sprečavanja zaražavanja i širenja infekcije. Pomoću *beacon* tehnologije moguće je u realnom vremenu pratiti kretanje studenata i zaposlenih i obavestiti ih kada je potrebno povećati rastojanje sa sagovornikom. Na ovaj način može se unapred smanjiti broj rizičnih kontakta, a samim tim i broj zaraženih. Takođe, u slučaju velike koncentracije studenata koji se dugo zadržavaju na istom mestu, što se često dešava tokom pauza između nastave, moguće je obavestiti i covid redara. Pomoću *Covid-19* sistema za praćenje kontakta može se vršiti prebrojavanje ljudi u prostorijama i na osnovu toga odrediti kada je potrebno izvršiti provetravanje i čišćenje prostorija. Na osnovu kretanja, broja bliskih kontakata, vremena provedenog u određenim prostorijama moguće je podsetiti studente i zaposlene o pranju ruku.

Beacon tehnologija se bazira na upotrebi BLE (Bluetooth Low Energy) protokola, čime se postiže visoka energetska efikasnost. Srž tehnologije su *beacon* uređaji koji na osnovu snage primljenog signala određuju rastojanje između sebe i drugih uređaja poput pametnih telefona. Glavna prednost *beacon*-a je mogućnost slanja poruka korisnicima u blizini putem aplikacije na pametnom telefonu. Zbog niske cene, jednostavne implementacije, male potrošnje energije i visoke preciznosti, *beacon* tehnologija je postala veoma popularna kod sistema koji uključuju lociranje, navigaciju i praćenje u zatvorenom prostoru. Kompanija Apple je predstavila *beacon* tehnologiju 2013. godine. Već 2016. godine *beacon* tržište procenjeno je na 519.6 miliona dolara, a predviđa se da će do 2026. godine vrednost porasti na 56.6 biliona dolara [1].

Međunarodni aerodrom u Hong Kongu ima preko 17000 *beacon*-a koji se koriste za navigaciju putnika po aerodromu [2]. Macy's robne kuće su 2014. godine instalirale oko 4000 *beacon*-a u preko 850 prodavnica kako bi se povećala prodaja i broj kupaca [3]. Mnogi muzeji širom sveta, poput muzeja umetnosi San Diego, koriste *beacon*-e kako bi poboljšali iskustvo posetioca [4]. Neki univerziteti koriste *beacon* tehnologiju kako bi automatizovali proces evidencije prisustva studenata na predavanjima [5].

Naredne sekcije rada organizovane su na sledeći način. U sekciji II. razmatrane su različite tehnologije koje se koriste za implementaciju lokacijskih servisa u zatvorenom prostoru. Detaljno je objašnjen princip funkcionisanja sistema. Najznačajniji rezultati istraživanja i njihova analiza predstavljeni su u sekciji III. Na kraju rada, prikazane su različite mogućnosti za modifikaciju sistema, čime bi se obezbedila održivost sistema i nakon završetka pandemije.

II. Metod

Implementacija Covid-19 sistema za praćenje kontakta podrazumeva korišćenje proaktivnih lokacijskih servisa u zatvorenom prostoru koji omogućavaju stalno praćenje

Ivana Stefanović – Akademija tehničko-umetničkih stukovnih studija Beograd, Odsek Visoka škola elektrotehnike i računarstva, Starine Novaka 24, 11000 Beograd, Srbija (e-mail: ivanas@ viser.edu.rs).

Milutin Nešić – Akademija tehničko-umetničkih stukovnih studija Beograd, Odsek Visoka škola elektrotehnike i računarstva, Starine Novaka 24, 11000 Beograd, Srbija (e-mail: nesic@ viser.edu.rs).

Marko Milivojčević – Akademija tehničko-umetničkih stukovnih studija Beograd, Odsek Visoka škola elektrotehnike i računarstva, Starine Novaka 24, 11000 Beograd, Srbija (e-mail: markom@ viser.edu.rs).

 ISBN 978-86-7466-930-3

korisnika i iniciraju se automatski. Različiti lokacijski servisi koriste različite tehnike pozicioniranja koje se međusobno razlikuju po pitanju tačnosti, preciznosti, efikasnosti, kašnjenju, ceni i kompleksnosti implementacije.

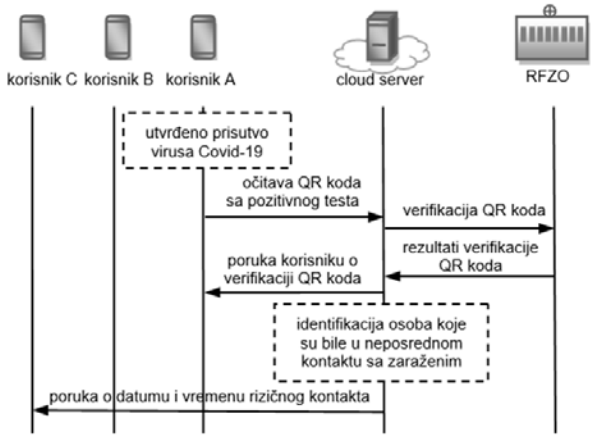
GNSS nude niz pogodnosti kada je u pitanju pozicioniranje na otvorenom, ali u zatvorenom prostoru ne mogu obezbediti zahtevanu tačnost koja je neophodna za implementaciju Covid-19 sistema za praćenje kontakta. Za implementaciju sistema razmatarane su različite tehnike pozicioniranja poput RFID (radio-frequency identification), NFC (Near Field Communication), UWB (Ultra-wideband) tehnologija, kao i pozicioniranje u okviru WLAN mreža koje je bazirano na upotrebi Wi-Fi-a i Bluetooth-a. RFID, NFC i UWB tehnologije podrazumevaju korišćenje različitih tagova i čitača ili senzora za prijem signala sa tagova. Korišćenje bilo koje od navedenih tehnika podrazumevalo bi da svi studenti i zaposleni kod sebe moraju imati tag, dok bi čitači ili senzori bili raspoređeni po školi. Zbog ograničenog dometa, broja tagova koji se istovremeno mogu skenirati, broja čitača i senzora povećava se kompleksnost i cena implementacije sistema. *Beacon* tehnologija je specifična zbog jednostavne integracije i kompatibilnosti između *beacona* i pametnih telefona. Upotrebom *beacon* tehnologije dovoljno je da studenti i zaposleni kod sebe imaju pametni telefon na kome je instalirana odgovarajuća aplikacija. Prema Republičkom zavodu za statistiku, 2020. godine, u Republici Srbiji 94.1% stanovništva koristi mobilni telefon. Za starosnu grupu od 16 do 24 godine, kojoj pripada većina studenata, ovaj procenat iznosi 100%, za mušku populaciju, i 99.6% za žensku populaciju [6].

Wi-Fi i Bluetooth tehnologije su pogodne za implementaciju Covid-19 sistema za praćenje kontakta, jer je kao i u slučaju *beacon* tehnologije dovoljno da studenti i zaposleni kod sebe imaju pametni telefon na kome je aktiviran Wi-Fi, odnosno Bluetooth. Pozicioniranje korisnika primenom Wi-Fi vrši se na osnovu snage primljenog signala sa pristupnih tačaka, a može se koristiti i cirkularna lateracija, gde je pozicija korisnika određena presekom kružnica. Da bi se obezbedila neophodna tačnost prilikom pozicioniranja studenata i zaposlenih u školi potrebno je povećati trenutni broj pristupnih tačaka, dodavanjem novih. Ugradnja novih pristupnih tačaka i povezivanje na postojeću infrastrukturu je znatno komplikovanije i skuplje u poređenju sa ugradnjom *beacon*-a. Takođe, *beacon*-i se lako mogu postaviti na bilo kojoj lokaciji u školi, jer ne zahtevaju mrežno napajanje niti dodatne infrastrukturne radove, za razliku od pristupnih tačaka.

Pozicioniranje pomoću Bluetooth-a nudi niz prednosti u odnosu na Wi-Fi. Jedna od njih je energetska efikasnost koja je postignuta BLE protokolom. Takođe još jedna značajna prednost Bluetooth-a je mogućnost povezivanja pametnog telefona sa drugim pametnim telefonima, a da pri tome nisu neophodne pristupne tačke [7]. Razvijen je veliki broj aplikacija za praćenje Covid-19 kontakta koje koriste Bluetooth tehnologiju, poput COVIDSafe aplikacije koja se koristi u Australiji, eRouska u Češkoj, STOP COVID – Protego u Poljskoj [7]. Sve navedene aplikacije dostupne su na Google PlayStore-u i imaju preko million preuzimanja. Potrebno je naglasiti da se u ovom slučaju zapravo ne vrši pozicioniranje korisnika, nego isključivo dolazi do razmene podataka između korisnika koji su bili u neposrednom kontaktu. Na ovaj način moguće je implementirati Covid-19 sistem za praćenje kontakta, ali se ne mogu implemetirati dodatne funkcionalnosti za automatizaciju sprovođenja propisanih mera u cilju sprečavanja zaražavanja i širenja infekcije.

Na osnovu razmatranih tehnologija koje bi mogle biti implementirane u osmišljeni sistem, *beacon* tehnologija se ističe dobrim odnosom *price/performanse* i jednostavnošću implementacije.

Osnovni funkcija Covid-19 sistema za praćenje kontakta je identifikacija i obaveštavanje osoba koje su bile u neposrednom kontaktu, u školi, sa osobom kod koje je potvđeno prisustvo korona virusa Covid-19. Student ili zaposleni, kod kojeg je utvrđeno prisustvo Covid-19 virusa, putem aplikacije instalirane na svom telefonu, obaveštava školu o pozitivnom SARS-CoV-2 testu, očitavanjem QR koda sa testa. Verifikacija QR koda vrši se na sajtu republičkog fonda za zdravstveno osiguranje. Nakon uspešne verifikacije vrši se identifikacija osoba sa kojima je zaraženi bio u neposrednom kontaktu. Evropski centar za prevenciju i kontrolu bolesti (ECDC) definisao je dva kriterijama na osnovu kojih se vrši identifikacija osoba koje su bile u neposrednom kontaktu sa zaraženom osobom. Prvi kriterijum odnosi se na rastojanje između osoba, a drugi na dužinu trajanja kontakta. Prema ECDC izveštaju [8] i klasifikaciji kontakta, visok rizik od zaražavanja postoji ukoliko je osoba boravila u zatvorenom prostoru sa zaraženom osobom najmanje 15 minuta na udaljenosti manjoj od 2m. Takođe, prema ECDC izveštaju [9], koji se odnosi na školske ustanove, postoji visok rizik od zaražavanja za sve osobe koje su bile u istoj učionici za zaraženom osobom duže od 15 minuta. Sledeći korak je obaveštavanje osoba, za koje je utvrđeno da su bile u neposrednom kontaktu sa zaraženim, putem aplikacije o datumu i vremenu kontakta. Neophodno je naglasiti da prilikom obaveštavanja osoba koje su bile u neposrednom kontaktu sa zaraženom osobom neći biti otkriven identitet zaražene osobe. Osoba koja je bila u kontaktu samo se obaveštava o datumu i vremenu kontakta, zbog eventualne izolacije i praćenja simptoma. Na Slici 1. prikazani su koraci postupka identifikacije i obaveštavanja osoba o neposrednom kontaktu sa zaraženim.



Sl. 1. Koraci postupka identifikacije i obaveštavanja osoba o neposrednom kontaktu sa osobom kod koje je utvrđeno prisustvo Covid-19 virusa.

Nedostatak ovakvog sistema je taj što se bazirana na tome da će osoba kod koje je utvrđeno prisustvo SARS-CoV-2 virusa dobrovoljno obavestiti ustanovu o pozitivnom testu. Kako bi se utvrdilo da li su student i zaposleni

 ISBN 978-86-7466-930-3

zainteresovani za ovakav sistem izvršeno je anonimno anketiranje na Akademiji tehničko-umetničkih strukovnih studija Beograd, odsek Visoka škola elektrotehnike i računarstva. Anketa sadrži 6 pitanja, sa ponuđenim odgovorima "da" i "ne":

1. Da li biste želeli da se u Školi implementira sistem pomoću kojeg se vrši identifikacija studenata i zaposlenih koji su bili u neposrednom kontaktu sa osobom kod koje je potvrđeno prisutvo Covid-19 virusa?
2. Da li biste na svom telefonu instalirali besplatnu aplikaciju za ove potrebe?
3. Da li biste u okviru aplikacije uneli Vaš broj telefona i JMBG?
4. U slučaju da ste bili u neposrednom kontaktu sa zaraženom osobom da li biste želeli da budete obavešteni putem aplikacije o datumu i vremenu kontakta?
5. U slučaju da se Vi zarazite Covid-19 virusom da li biste obavestili Školu o tome, putem aplikacije?
6. U slučaju da se Vi zarazite Covid-19 virusom da li biste želeli da se osobe koje su bile u bliskom kontaktu sa Vama tokom boravka u Školi obaveste putem aplikacije o datumu i vremenu kontakta? Osobe sa kojima ste bili u bliskom kontaktu dobile bi isključivo podatke o datumu i vremenu kontakta, ali ne o osobi sa kojom su bile u kontaktu.

Anketirano je 73 studenata i 17 zaposlenih. Rezultati ankete prikazani su u sekciji III.

III. Glavni rezultati

A. Rezultati ankete

U Tabeli I. prikazani su rezultati ankete za svako pitanje pojedinačno. Iz Tabele I. se vidi da više od 75% anketiranih želi da se u školi implementira sistem za identifikaciju studenata i zaposlenih koji su bili u neposrednom kontaktu sa osobom kod koje je potvrđeno prisustvo virusa Covid-19, i da bi za tu potrebu instalirali besplatnu aplikaciju na svom telefonu. Čak 87.78% anketiranih je spremno da putem aplikacije obavesti školu o pozitivnom SARS-CoV-2 testu. Većina anketiranih, preko 80% želi da bude obavešteno o neposrednom kontaktu, kao i da drugi budu obavešteni. Sa druge strane, 34.44% anketiranih ne želi da unese broj telefona i JMBG u okviru aplikacije. Ovi podaci su neophodni radi verifikacije testa i obaveštavanja korisnika o neposrednom kontaktu.

TABELA I
REZULTATI ANKETE

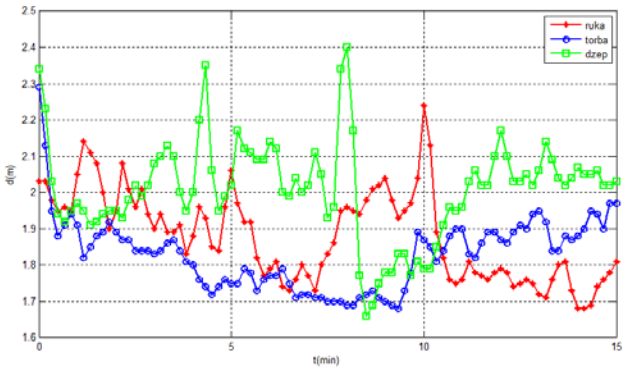
Pitanje	Odgovor DA	Odgovor NE
1.	75.56 %	24.44 %
2.	75.56 %	24.44 %
3.	65.56 %	34.44 %
4.	81.11 %	18.89 %
5.	87.78 %	12.22 %
6.	84.44 %	15.56 %

Prilikom analize odgovora anketiranih izdvojile su se tri grupe odgovora. 52.22% anketiranih je na sva pitanja odgovorilo sa da, dok je svega 12.22% na sva pitanja odgovorilo sa ne. Takođe uočena je i grupa anketiranih, 15.56%, koja je na sva pitanja sem trećeg dala odgovor da.

B. Rezultati merenja

Na osnovu 128-bitnog UUID (*Universally Unique Identifier*) broja *beacon*-a lako se može izvršiti identifikacija osoba koje su boravile u istoj učionici sa zaraženom osobom duže od 15 minuta. Sa druge strane, identifikacija studenata i zaposlenih koji su proveli više od 15 minuta sa zaraženim na rastojanju manjem od 2 m, u nekom drugom prostoru poput hodnika ili kabineta, zahteva veću preciznost.

Za merenje rastojanja i prijemne snage korišćena je Radius Networks aplikacija *Locate*. Telefon marke Huawei p20 pro korišćen je kao *ibeacon*, pri čemu je predajna snaga podešena na -56dBm. Kao prijemnik korišćen je telefon marke Huawei p30 pro. Telefoni su postavljeni na rastojanju od 2 metra, a merenje je vršeno tokom 15 minuta, za tri slučaja. U prvom slučaju telefon koji je korišćen kao prijemnik nalazi se u ruci, u drugom slučaju u torbi, a u trećem u džepu od pantalona. Na Slici 2. dat je grafički prikaz procenjene udaljenosti tokom vremenskog intervala od 15 minuta, pri čemu je na horizontalnoj osi prikazano vreme u minutima, a na vertikalnoj osi procenjeno rastojanje u metrima, za sva tri slučaja, pri čemu su rezultati beleženi svakih 10 sekundi.



Sl. 2. Procena udaljenosti korisnika od *beacon*-a tokom vremenskog intervala od 15 minuta

Najveće izmereno odstupanje procenjenog rastojanja od realnog iznosi 0.4m. Ovo odstupanje dobijeno je u trećem slučaju, kada se telefon nalazio u džepu. Sa Slike 2. se može uočiti da je za slučaj kada se telefon nalazi u torbi procenjeno rastojanje skoro uvek manje od stvarnog rastojanja. Takođe, u slučaju kada se telefon nalazi u ruci, tokom 81% vremena merenja, rastojanje je procenjeno na manje od stvarnog. Sa druge strane, u slučaju kada se telefon nalazi u džepu, procenjeno rastojanje je veće od stvarnog rastojanja tokom 54% vremena merenja. Ovo je najverovatnije posledica statičnosti telefona i njegovog položaja kada je torbi, dok se u ruci i džepu telefon pomera pa samim tim i dijagram zračenja menja svoj pravac. Srednja greška procenjenog rastojanja za sva tri slučaja, dobijena na osnovu 273 merenja iznosi ±0.14m. Na ovaj način postignuta je preciznost od 74.73%. Pri tome najviša preciznost postiže se u slučaju kada se telefon nalazi u torbi, čak 97.8%, dok se najniža preciznost postiže u slučaju kada se telefon nalazi u džepu, samo 39.56%. Preciznost je definisana kao odnos broja merenja kada je detektovano da se korisnik nalazi na rastojanju manjem ili jednakom od 2 m i rastojanju koje je veće od 2 m.

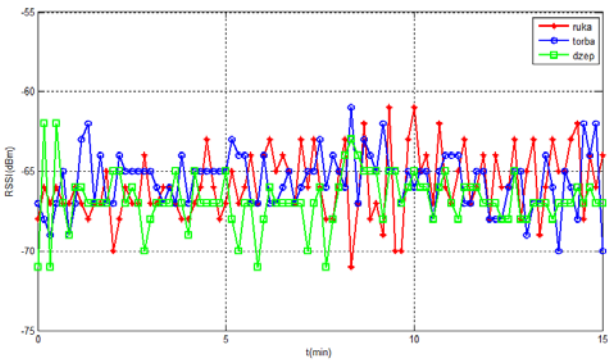
 ISBN 978-86-7466-930-3

Kako je utvrđeno da srednja greška procenjenog rastojanja iznosi ±0.14 m na dobijene rezultate merenja primenjen je novi kriterijum pri čemu se kao kritično rastojanje uzima 2.14 m. Povećanjem rastojanja sa 2 m na 2.14 m postiže se preciznost od 95.97%. U Tabeli II. dat je pregled preciznosti za sva tri slučaja, pojedinačno, kada se kao kritično rastojanje koristi 2 m i 2.14 m. Kao što se vidi iz Tabele II. promena rastojanja sa 2 m na 2.14 m dovodi do povećanja preciznosti u sva tri slučaja, pri čemu se u trećem slučaju beleži najveći porast preciznosti od čak 50.55%.

TABELA II
PRECIZNOST

	kriterijum 2 m	kriterijum 2.14 m
ruka	82.42 %	98.90 %
torba	97.80 %	98.90 %
džep	39.56%	90.11 %

Na Slici 3. dat je grafički prikaz izmerenog RSSI (*Received Signal Strength Indicator*), za sva tri slučaju, tokom vremenskog intervala od 15 minuta. Srednja vrednost RSSI dobijena sumiranjem svih prikupljenih podataka iznosi -66.07 dBm. Potrebno je naglasiti da postoje mala odstupanja srednje vrednosti RSSI za sva tri slučaja pojedinačno. Minimalna izmerena vrednost RSSI je -71 dBm, a maksimalna -61 dBm.



Sl. 3. Izmerene vrednosti RSSI tokom vremenskog interval od 15 minuta

Pomoću jednačine (1) može se odrediti eksponent slabljenja, *n*. $RSSI_{do}$ predstavlja srednju vrednost izmerenog RSSI na rastojanju od 1m, dok $RSSI_d$ predstavlja srednju vrednost izmerenog RSSI na rastojanju *d* [10].

$$n = \frac{RSSI_{d0} - RSSI_d}{10 \cdot \log d} \quad (1)$$

Na osnovu 20 merenja srednja vrednost RSSI na rastojanju od 1m iznosi -58.20 dBm. Zamenom dobijenih vrednosti u jednačinu (1) dobija se eksponent slabljenja *n*=2.61. Dobijeni eksponent slabljenja odgovara eksponentu slabljenja u poslovnim zgradama, isti sprat, pri frekvenciji od 2.4 GHz, čija se vrednost kreće u opsegu od 1.6 do 3.5 [11].

Izvršena su merenja RSSI i na rastojanjima 0.5 m, 1.5 m, 2.5 m i 3 m. Dobijene vrednosti eksponenta slabljenja prikazane su u Tabeli III. Na osnovu dobijenih rezultata srednja vrednost eksponenta slabljenja iznosi 2.32. Pri eksponentu slabljenja od 2.32 i srednjoj vrednosti RSSI od -66.07 dBm procenjuje se da je korisnik udaljen od *beacon*-a 2.18m, odnosno srednja greška procenjenog rastojanja iznosi ±0.18 m. Dakle, dobijeno je odstupanje od svega ±0.04 m u odnosu na rezultate dobijene merenjem i novo uvedeni kriterijum od 2.14m.

TABELA III
EKSPONENT SLABLJENJA NA RAZLIČITIM RASTOJANJIMA

d	0.5 m	1.5 m	2 m	2.5 m	3 m
n	2.01	2.14	2.61	2.15	2.67

IV. ZAKLJUČAK

U ovom radu je pokazano da je *beacon* tehnologija odličan izbor za implementaciju Covid-19 sistema za praćenje kontakta u zatvorenom prostoru jer nudi niz pogodnosti poput visoke preciznosti, niske cene, jednostavne implementacije, održavanja i nadogradnje.

Sistem koji je prikazan u ovom radu namenjen je za visokoškolske institucije ali se može implementirati i u drugim zatvorenim prostorima poput tržnih centara, bankama, bolnicama, domovima zdravlja, firmi i kompanija sa velikim brojem zaposlenih i sl. U slučaju implementacije sistema npr. u tržnim centrima mogu se iskoristiti već postojeće aplikacije, poput aplikacija Tvoj Centar, Ada Loyalty Club i dr.

Nedostatak opisanog sistema je ograničeno trajanje koje zavisi od završetka pandemije, međutim u visokoškolskim institucijama nakon završetka pandemije sistem se lako može prilagoditi i koristiti za druge potrebe, poput navođenja kandidata po školi prilikom predaje dokumenata, polaganja prijemnog ispita i upisa, evidenciju studenata na nastavi i ispitima, obaveštavanje studenata i zaposlenih o rasporedu nastave, lociranje zaposlenih i sl. Takođe, i u drugim institucijama van obrazovanja po završetku pandemije moguće je prilagoditi sistem konkretnim potrebama u cilju unapređenja radnih procesa, povećanja efikasnosti i zadovoljstva klijenata.

LITERATURA

[1] T. Alsop, “Beacons technology market revenue worldwide 2016 and 2026”, Statista, October, 2020.
[2] P. Tedeschi, K. Eun Jeon, J. She, S. Wong, S. Bakiras, R. Di Pietro, “Privacy-Preserving and Sustainable Contact Tracing Using Batteryless Bluetooth Low-Energy Beacons”, *IEEE Security & Privacy*, pp. *2-11*, October, 2021.
[3] L. Chamberlain, “Macy’s To Test Beacon Messages Outside App, Explore Retargeting”, March, 2016. [Online]. Available: https://geomarketing.com/macys-to-test-beacon-messages-outside-app-explore-retargeting
[4] K. Shams, “8 museums successfully using beacons & 8 examples of beacon innovations”, September, 2016. [Online]. Available: https://proxera.net/8-museums-successfully-using-beacons-8-examples-of-beacon-innovations/
[5] S. Perez, “BeHere Lets Teachers Take Attendance Using iBeacon Technology”, March, 2016. [Online]. Available: https://techcrunch.com/2014/03/28/behere-lets-teachers-take-attendance-using-ibeacon/
[6] M. Kovačević, V. Šutić, U. Rajčević, A. Milaković, “Upotreba informaciono-komunikacionih tehnologija u Republici Srbiji”, Republički zavod za statistiku, Beograd, Srbija, 2020. [Online]. Available: https://publikacije.stat.gov.rs/G2020/Pdf/G202016015.pdf
[7] M. Shahroz, F. Ahmada, M. Shahzad Younis, N. Ahmadb, M. N. Kamel Boulos, R. Vinuesa, J. Qadir, “COVID-19 digital contact

 ISBN 978-86-7466-930-3

tracing applications and techniques: A review post initial deployments", *Transportation Engineering 5 (2021): 100072*, May, 2021.
[8] "Contact tracing: public health management of persons, including healthcare workers, who have had contact with COVID-19 cases in the European Union – third update", European Centre for Disease Prevention and Control, Stockholm, Sweden, 18 November 2020. [Online]. Available: https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-contact-tracing-public-health-management-third-update.pdf
[9] "Objectives for COVID-19 testing in school settings – first update", European Centre for Disease Prevention and Control, Stockholm, Sweden, 21 August 2020. [Online]. Available: https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-objectives-school-testing.pdf
[10] J. Röbesaat, P. Zhang, M. Abdelaal, O. Theel, "An Improved BLE Indoor Localization with Kalman-Based Fusion: An Experimental Study", *Sensors 17.5 (2017): 951*, 2017.
[11] O. S. Oguejiofor, A. N. Aniedu, H. C. Ejiofor, G. N. Okechukwu, "Indoor Measurement And Propagation Prediction Of WLAN At 2.4GHz", *International Journal of Engineering Research & Technology*, vol. 2, no. 7, pp. 798-802, July, 2013.

ABSTRACT

This paper discusses the possibility of applying beacon technology to develop solutions that allow monitoring of close contacts and automate the implementation of prescribed measures in real time, in order to prevent spread of infection caused by Covid-19 virus in higher education institutions. For the realization of the system, it is enough that students and employees have a smartphone on which the appropriate application is installed. The system is based on the fact that the student and employees who have been diagnosed with the Covid-19 virus will voluntarily inform the higher education institution about the positive test, through the application. An anonymous survey was conducted to determine whether students and employees are interested in implementing the Covid-19 contact tracking system. Practical measurements were also performed, and the obtained results confirm that beacon technology is an excellent choice for system implementation.

Possibility of applying beacon technologies for the development of Covid-19 contact tracking systems in higher education institutions

Ivana Stefanović, Milutin Nešić i Marko Milivojčević

 ISBN 978-86-7466-930-3

Zaštita prenosa paketskog telefonskog saobraćaja upotrebom tehnologije virtuelnih privatnih mreža

Mićo Živanović, Jovan Bajčetić, Ivan Tot

***Apstrakt*—Istraživanje predstavljeno u ovom radu prikazuje jednu realizaciju zaštite paketskog telefonskog saobraćaja primenom tehnologije virtuelnih privatnih mreža kroz konfiguraciju servera za prenos paketskog telefonskog saobraćaja i zaštićeni prenos uz primenu tehnologije virtuelnih privatnih mreža u tunel modu, primenom odgovarajućeg protokola za zaštitu tajnosti, autentifikaciju, zaštitu integriteta i razmenu kriptografskih ključeva. Izvršeno je snimanje i analiza saobraćaja primenom softvera Wireshark u zaštićenom i nezaštićenom prenosu. Prikazani rezultati omogućavaju lakše razumevanje kompleksnog procesa uspostave tunela upotrebom simulacionog softvera u edukaciji.**

***Ključne reči*—paketski telefonski saobraćaj; virtuelne privatne mreže; zaštićena komunikacija; kriptografski ključ.**

I. Uvod

Stalan razvoj Interneta ima za posledicu da je Internet postao univerzalno sredstvo za komunikaciju. U toku razvoja, postavio se zahtev za bezbednošću prenošenih informacija koji se ogledao u obezbeđenju bezbednosnih servisa: autentifikacije, poverljivosti, neporecivosti i integriteta podataka [1]. Razvijeni su sistemi zaštite u tri ravni: upravljačkoj, kontrolnoj i ravni podataka. Za potrebe ovog rada biće razmotreni mehanizmi zaštite u ravni podataka koji se odnose na informacioni saobraćaj. Ravan podataka se štiti pomoću implementiranja pravila (bezbednosnih polisa) po kojima se informacioni sadržaj prenosi upotrebom mrežnih uređaja.

Jedna od tehnologija koja omogućava zaštitu prenosa podataka u ravni podataka je tehnologija virtuelnih privatnih mreža (VPN – Virtual Private Networks). Navedena tehnologija pruža sledeće mogućnosti umrežavanja:
– Intranet, umrežavanje geografski dislociranih objekata;
– Udaljeni pristup mobilnih korisnika (rad od kuće);
– Ekstranet, ograničeni pristup nekoj mreži iz drugih mreža (pristup poslovnih partnera korporativnom WAN-u) [2].

Za realizaciju virtuelne privatne mreže mogu se koristiti periferni korisnički uređaji (host, ruter ili svič), na lokaciji korisnika (CE – Customer Edge) i periferni mrežni uređaji provajdera (PE – Provider Edge).

Virtuelnu privatnu mrežu čini više udaljenih mreža koje su povezane preko Interneta. Zbog korišćenja zajedničkih resursa na Internetu, komunikacija među korisnicima virtuelne privatne mreže se mora zaštititi. Zaštita virtuelne privatne mreže se ostvaruje pomoću barijera koje implementiraju IPSec protokol u tunel modu [3].

VPN tunel je veza između dva PE rutera ili dva CE uređaja koji predstavljaju krajnje tačke tunela [2].

Prema IETF, IP VPN se mogu klasifikovati u zavisnosti od odgovornosti u pogledu upravljanja na:
– VPN kojima upravlja korisnik (Customer Provisioned VPN, CP VPN);
– VPN kojima upravlja provajder servisa (Provider Provisioned VPN, PP VPN) [2].

Prema lokaciji VPN opreme, PP VNP se mogu podeliti na:
– CE – bazirane, kod kojih su krajnje tačke VPN locirane kod korisnika;
– PE – bazirane, kod kojih su krajnje tačke VPN tunela locirane kod provajdera, na PE ruteru.

U zavisnosti od ponuđenog servisa, PE – bazirane VPN se dele na:
– PE – bazirane L2 VPN (koje pružaju servise OSI sloja 2);
– PE – bazirane L3 VPN (koje pružaju servise OSI sloja 3);
– CE – bazirane IP VPN pružaju samo servise OSI sloja 3.

Istovremeno sa razvojem bezbednosnih servisa razvijaju se arhitekture za pružanje različitih komunikacionih servisa koji koriste Internet protokol (telefonija, video, podaci, multimedijalni servisi). Prenos telefonije preko Interneta razvijao se postupno, pre svega zbog prethodno razvijenih sistema klasičnih javnih telefonskih mreža (PSTN) i digitalnih mreža sa integrisanim servisima (ISDN).

U cilju razvoja telefonije zasnovane na komutaciji paketa (VoIP telefonija), razvijene su grupe protokola za prenos VoIP telefonije i povezivanje VoIP telefonije sa telefonijom koja se prenosi u drugim sistemima prenosa (H.323 i SIP protokol) [4].

Prikaz istraživanja u ovom radu će se sastojati iz opisa načina realizacije zaštite informacije korišćenjem VPN tehnologije, a potom u jednoj realizaciji zaštite prenosa paketskog telefonskog saobraćaja upotrebom tehnologije

Mićo Živanović – Ministarstvo odbrane, Sektor za ljudske resurse, Nemanjina 15, 11000 Beograd, Srbija (e-mail: comiveza@yahoo.com).
Jovan Bajčetić – Vojna Akademija, Univerzitet odbrane u Beogradu, Veljka Lukića Kurjaka 33, 11042 Beograd, Srbija (e-mail: bajce05@gmail.com).
Ivan Tot – Vojna Akademija, Univerzitet odbrane u Beogradu, Veljka Lukića Kurjaka 33, 11042 Beograd, Srbija (e-mail: totivan@gmail.com).

 ISBN 978-86-7466-930-3

virtuelnih privatnih mreža, upotrebom IPSec protokola u tunel modu ka udaljenom korisniku, uz prikaz analize mrežnog saobraćaja korišćenjem programskog alata Wireshark.

II. ZAŠTITA PODATAKA PRIMENOM TEHNOLOGIJE VIRTUELNIH PRIVATNIH MREŽA

Za prenos informacionog sadržaja preko Interneta neophodno je obezbediti zaštitu u prenosu. Čest je slučaj da kompanije koriste Internet kao okosnicu za povezivanje svojih filijala ili klijenata kako bi ostvarili prenos podataka za svoje potrebe. Iz navedenog razloga nameće se potreba za zaštitu prenošenog saobraćaja. U tu svrhu koriste se različite tehnologije, od kojih je jedna - tehnologija virtuelnih privatnih mreža. Za realizaciju virtuelnih privatnih mreža na raspolaganju je više tehnologija, zavisno od toga da li se VPN realizuje kao "oblast – oblast" (site–to–site) ili kao "udaljeni pristup" (remote access). U oba navedena slučaja najčešće se koristi IPSec (IP security) protokol.

IPSec protokol štiti pakete između dva uređaja u mreži [3].

Uređaji kojima se realizuje IPSec su: server, ruteri, korisnički računari ili specijalizovani hardver. IPsec pruža dve vrste zaštite: autentifikaciju i poverljivost.

Mehanizam autentifikacije osigurava da je primljeni paket zaista poslao onaj ko je u zaglavlju paketa naveden kao izvor i da se paket nije promenio tokom prenosa, dok mehanizam poverljivosti omogućava entitetima u komunikaciji da šifruju poruke kako bi sprečili nepozvana lica da dođu do sadržaja poruka [1]. Za šifrovanje podataka se koriste simetrični algoritmi (DES, 3DES, AES), što zahteva pouzdanu razmenu ključeva strana u komunikaciji i za tu svrhu se koriste protokoli za autentifikaciju (neki od protokola iz IETF (IKE - Internet Key Exchange) standarda) [3].

Razvoj bezbednosti u arhitekturi Interneta se odvijao postepeno u čemu je značajno mesto imala Radna grupa za inženjering Interneta (IETF – Internet Engineering Task Force). Sâmo uvođenje standarda je išlo postepeno. Prvi u nizu standard IETF koji se odnosio na bezbednost u arhitekturi Interneta bio je standard RFC 1636 (Request for Comments). Standard se odnosio na osnove bezbednosti Interneta (upotreba firewall – a, servis autentifikacije, privatnost i dr.) [5].

Da bi se adekvatno razumeo način formiranja VPN sesije in a pravi način predstavio prilikom edukacije, biće razmotreno nekoliko najvažnijih dokumenata IETF kojima su definisani režimi rada VPN, mehanizam autentifikacije, razmena kriptografskih ključeva i zaštita poverljivosti.

IPSec koristi dva protokola za bezbednost: AH (Authentication Header) i ESP (Encapsulating Security Payload). Zaglavlje autentifikacije (AH) je definisano specifikacijom RFC 4302 (IP Authentication Header), dok je ESP enkapsulirajuće bezbedno pakovanje (Encapsulating Security Payload) definisan specifikacijom RFC 4303. AH i ESP podržavaju dva režima rada: transportni režim i tunelovanje.

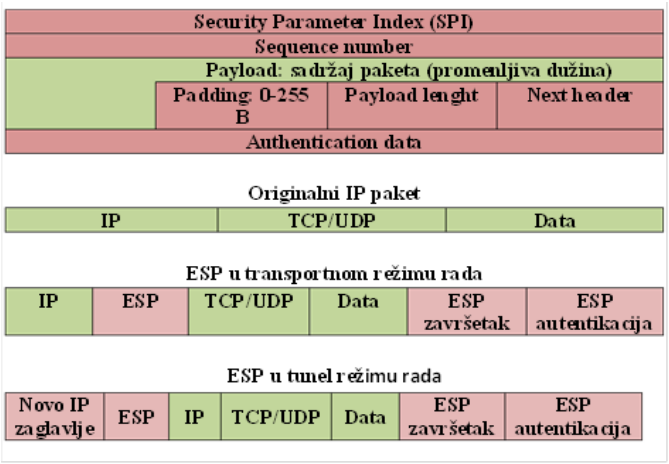
U transportnom režimu AH autentifikuje IP koristan sadržaj i odabrane delove IP zaglavlja, dok ESP šifruje i opciono autentifikuje IP koristan sadržaj. Tunelovanje vrši zaštitu celog IP paketa. Navedeno se postiže nakon dodavanja AH i ESP polja i tretiranja celog paketa kao korisnog sadržaja novog spoljnog IP paketa sa novim IP zaglavljem.

U režimu tunelovanja ESP šifruje i opciono autentifikuje ceo unutrašnji IP paket, uključujući unutrašnje IP zaglavlje, dok AH u režimu tunelovanja autentifikuje ceo unutrašnji IP paket i odabrane delove spoljnog IP zaglavlja [1].

Redosled postupaka sa paketima za rad ESP u transportnom i tunel režimu, je sledeći:

– U transport režimu blok podataka koji se sastoji od segmenta transportnog sloja sa dodatim ESP završnim blokom se šifruje, sa dodatim zaglavljem za autentifikaciju (opciono);

– U režimu tunelovanja, ESP se koristi za šifrovanje celog IP paketa, ESP zaglavlje ide ispred paketa i šifruje se paket zajedno sa ESP završnim blokom.



Sl. 1 Opseg ESP šifrovanja u transportnom i tunel modu [6]

Sl. 1 prikazuje format jednog ESP paketa. Indeks bezbednosnih parametara (SPI) definiše jednu bezbednosnu asocijaciju kojom se određuje algoritam šifrovanja i autentifikacije, ključevi, inicijalizacione vrednosti, životni vek ključeva i vezani parametri koji se koriste uz ESP. Broj sekvence je vrednost brojača paketa kojom se sprečava ponavljanje paketa. Sadržaj paketa je promenljive dužine i predstavlja segment transportnog sloja (transport režim) ili IP paket (tunel režim). U tunel modu, celom paketu se dodaje novo IP zaglavlje koje ima dovoljno informacija za rutiranje, ali ne i za analizu saobraćaja [6].

Važan deo IPSec koji se odnosi na upravljanje ključevima obuhvata određivanje i distribuciju ključeva. Dokumentom RFC 4301 definisane su dve vrste upravljanja ključevima:

– Ručno (administrator definiše sistem sopstvenim ključevima i ključevima drugih sistema sa kojima komunicira);

– Automatizovano (omogućava generisanje ključeva za bezbednosnu asocijaciju na zahtev koji je pogodan za velike sisteme sa rastućom konfiguracijom) [7].

Protokol koji se koristi za automatizovano upravljanje ključevima za IPSec je ISAKMP (Internet Security

 ISBN 978-86-7466-930-3

Association and key Management Protocol) i definisan je dokumentom IETF RFC 2408. ISAKMP definiše procedure za kreiranje i upravljanje bezbednosnim asocijacijama, tehnike generisanja ključeva, ublažavanje pretnji (npr. od DDoS napada) [8].

ISAKMP ne nalaže konkretan algoritam za razmenu ključeva, već se sastoji od jednog skupa tipova poruka koje omogućavaju upotrebu raznovrsnih algoritama za razmenu ključeva.

Karakteristike IKE određivanja ključeva su:
- Osujećenje DDoS napada;
- Omogućava razmenu ključeva za pregovaranje oko grupe ključeva;
- Obezbeđuje od napada ponavljanjem korišćenjem jednokratnih brojeva;
- Omogućava razmenu javnih ključeva;
- Onemogućava napad tipa "čovek u sredini".

IKE potprotokol obezbeđuje dogovaranje protokola, algoritama i ključeva između učesnika u komunikaciji, proverava autentičnost učesnika koji učestvuju u postupku dogovaranja, omogućava razmenu podataka na osnovu kojih će se generisati ključevi i upravljati razmenom ključeva. IKE potprotokol obavlja se u dve faze [9].

U prvoj fazi dva učesnika uspostavljaju bezbedni komunikacioni kanal kojim će se obaviti dogovaranje bezbednosnih parametara i razmena ključeva. Dogovaranje parametara i razmena ključeva, odnosno uspostava bezbednosne asocijacije obavlja se u drugoj fazi. Za sprovođenje postupka koriste se tri načina razmene informacija, dva za prvu fazu i jedan za drugu fazu IKE potprotokola:
- Osnovni način;
- Agresivni način;
- Brzi način.

Osnovni način razmene informacija (engl. Main mode) koristi se u prvoj fazi IKE potprotokola i služi da bi se uspostavio bezbednosni komunikacioni kanal kojim će se obaviti razmena podataka potrebnih za kasniju komunikaciju AH ili ESP potprotokolima.

Agresivni način razmene, slično kao i osnovni, koristi se u prvoj fazi IKE potprotokola i služi za uspostavljanje sigurnog komunikacionog kanala za dogovor učesnika i ne obavlja se kroz bezbedni kanal. Agresivni način koristi samo tri poruke u razmeni i nešto je jednostavniji i brži od osnovnog načina, ali se dokazivanje identiteta ne vrši kroz bezbedan kanal.

Nakon uspostave bezbednog kanala primenom osnovnog ili agresivnog načina razmene, započinje druga faza IKE potprotokola. Druga faza koristi se brzim načinom razmene ključeva koja služi za dogovaranje bezbednosnih parametara komunikacije AH ili ESP potprotokolom i za razmenu tajnih simetričnih ključeva.

III. JEDNA REALIZACIJA ZAŠTITE PAKETSKOG TELEFONSKOG SAOBRAĆAJA UPOTREBOM TEHNOLOGIJE VIRTUELNIH PRIVATNIH MREŽA

U uvodu rada predstavljena je podela VPN prema tome ko je odgovoran za uspostavu zaštićene komunikacije (provajder ili korisnik), kao i koja vrsta servisa se ostvaruje (sloj 2 ili 3 OSI referentnog modela). Predloženi model koje će u nastavku biti prikazan omogućava realizaciju jedne VPN koja bi predstavljala primer uspostave zaštite VoIP putem VPN za koju je „odgovoran" provajder, na OSI sloju 3 i da se primenom programskog alata „Wireshark" snimi i analizira ostvareni saobraćaj. Za navedene potrebe je uspostavljena mrežna topologija prikazana na Sl. 2.

Ruteri predstavljaju periferne rutere provajdera na kojima se vrši konfigurisanje VPN konekcije, po modelu "oblast – oblast". Na ruteru 1 su konfigurisani i uspostavljeni VPN, VoIP i DHCP server.

VPN server je konfigurisan sledećim parametrima:
- ISAKMP razmena kripto ključeva (policy 10);
- Kripto algoritam 3DES;
- Algoritam za autentifikaciju MD5 [10];
- Rad u tunel modu.

VoIP server je određen sledećim parametrima:
- Konekcija SIP protokolom;
- Kodek g711 ulaw.

Za razumevanje rada VoIP, ukratko će biti objašnjen prenos signalizacije i kontrola saobraćaja u VoIP prenosnim sistemima, u kojima se najčešće koriste H.323 i SIP protokoli.

H.323 preporuka je deo familije ITU-T preporuka sa zajedničkom oznakom H.32x koje se odnose na multimedijalne komunikacije preko različitih mreža. H.323 definiše protokole zadužene za usluge multimedijalnih komunikacija preko mreža zasnovanih na komutaciji paketa. H.323 se najčešće koristi kao signalizacioni i kontrolni protokol u VoIP i za video konferencije, a bio je prvi standard koji je koristio RTP protokol (Real-time Transfer Protocol) za konkretni prenos audio i video signala preko mreže.

H.323 je standard koji omogućava multimedijalnu komunikaciju preko različitih mreža (usko pojasne ISDN, širokopojasne B-ISDN, lokalne računarske mreže, mreže na bazi komutacije kola). Cilj je postizanje interoperabilnosti sa različitim mrežama za prenos multimedijalnih informacija, kroz upotrebu zajedničkih preporuka, procedura i poruka, kao i uvođenjem komponente mrežnog prolaza. H.323 standard predstavlja skup protokola namenjenih za obavljanje različitih funkcija u okviru H.323 sistema i to: audio kodere i dekodere, video kodere i dekodere, signaliziranje poziva, kontrola poziva, protokol prenosa u realnom vremenu (RTP), protokol kontrole prenosa u realnom vremenu (RTCP), registraciju, pristup i status i ostale protokole za prenos podataka u realnom vremenu [4].

 ISBN 978-86-7466-930-3

SIP je protokol za uspostavljanje, modifikaciju i raskidanje multimedijalnih sesija u paketskim mrežama. SIP u kombinaciji sa drugim protokolima se koristi za opis karakteristika sesije potencijalnim učesnicima [11]. SIP je delo IETF (Internet Engineering Task Force) i razvijen je kao mehanizam za uspostavljanje raznovrsnih sesija, a može se koristiti za unicast i multicast komunikaciju. SIP je peer-to-peer protokol, što znači da nije centralizovan, već je servisna inteligencija izmeštena prema krajevima mreže ka krajnjim korisnicima, kao kod računarskih mreža. U okviru SIP poruka se najčešće prenosi SDP (Session Description Protocol), mada standard ostavlja otvorenim i druge mogućnosti [12].

H.323 i SIP su dva konkurentska protokola za multimedijalne komunikacije na paketskim mrežama.

SIP se odlikuje sledećim prednostima:

– Fleksibilnost (omogućava korišćenje sa različitim transportnim i drugim protokolima);
– Arhitektura i osobine mu se prirodno uklapaju Internet okruženje, dok H.323 ima neke osobine protokola fiksne telefonije;
– Poseduje mnoga proširenja potrebna za različite sisteme ličnih komunikacija (prisutnost, instant poruke, posredno upravljanje pozivom) [13].

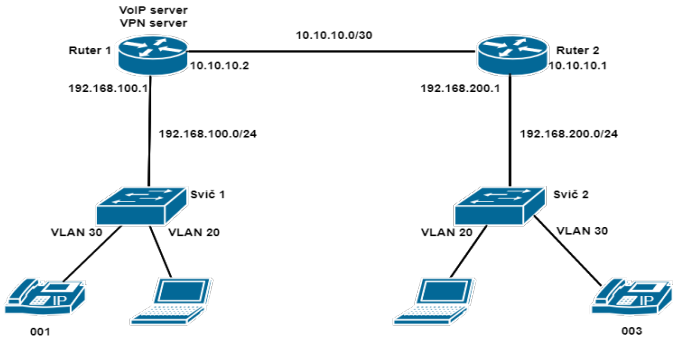
DHCP server je uspostavljen za opseg adresa koji obezbeđuje formiranje logički odvojenih mreža sa opsegom adresa 20.20.20.0 i 30.30.30.0 u različitim virtuelnim lokalnim mrežama. Prema mrežnoj topologiji formirane su dve LAN mreže u okviru kojih je izvršeno razdvajanje saobraćaja (telefonskog i podaci), uspostavom dve VLAN na OSI sloju L2, konfigurisanjem svičeva 1 i 2 (VLAN 20 i 30), dok je po jedan interfejs na svičevima konfigurisan kao trank interfejs [14]. Nakon provere konektivnosti, uspostavljen je paketski telefonski saobraćaj bez zaštite pomoću virtuelne privatne mreže i realizovano snimanje saobraćaja, upotrebom programskog alata “Wireshark”. Rezultat i procesi analize nezaštićenog saobraćaja prikazani su u Tabeli 1. Uspostavljena je VPN sesija između dva rutera i realizovano snimanje saobraćaja u zaštićenom modu. Procesi u toku uspostave zaštićenog paketskog telefonskog saobraćaja prikazani su u Tabeli 2.

Za realizaciju mrežne topologije na slici 2 korišćena je sledeća mrežna oprema:

– CISCO 2900 ruter…….2 kom;
– CATALIST 3650 svič ……..2 kom;
– Računar sa ETH mrežnim interfejsom…. 2 kom;
– IP telefoni……..2 kom.

Procesi prikazani u Tabeli 1, a koji se odnose na uspostavu i održavanje nezaštićenog telefonskog saobraćaja prikazuju proces uspostavljanja prisutnosti uređaja u mreži i utvrđivanja mrežnih usluga (SSDP), uspostavu logičke topologije mreže i saobraćaja protokola za sprečavanje petlji (STP). Pozivanjem jednog korisnika od strane drugog korisnika ustanovljava se IP adresa pozvanog korisnika kroz broadcast upit od strane pozivajućeg korisnika (ARP proces). Kroz DNS proces se povezuju IP adresa i ime domena pozvanog korisnika. Istovremeno se šalje poruka radi utvrđivanja dostupnosti korisnika (ICMP poruka). U toku uspostave VoIP komunikacije šalje se “hello poruka” u OSPF procesu, radi konstruisanja putanje između dva rutera. Kroz SIP signalizaciju vrši se pozivanje jednog od strane drugog korisnika, nakon čega se ostvaruje TCP sesija kroz proces “trostrukog rukovanja”. U sklopu SIP procesa razlikuju se faze (traying, ringing i OK), u kojima se mogu uočiti status procesa pozivanja, koji se na kraju završava uspešnom uspostavom komunikacije. Ceo proces je praćen slanjem kontrolnih TCP poruka (ACK), kojima se određuje broj bajtova koji se može poslati pre dobijanja sledeće dozvole za slanje, kao i slanjem poruka kojima se vrši sinhronizacija uspostave TCP sesije (SYN). Poseban segment kontrole komunikacije predstavlja kontrola konektivnosti na L2 nivou (LOOP). Dalji proces komuniciranja je praćen razmenom paketa u realnom vremenu (RTP) koji nose govorni informacioni sadržaj.

Procesi prikazani u Tabeli 2, a koji se odnose na uspostavu i održavanje zaštićenog telefonskog saobraćaja, pored na početku navedenih procesa u nezaštićenom prenosu, sadrži proces razmene kripto ključeva u procesu uspostave IPSec tunela (ISAKMP). Specifično za proces prenosa u zaštićenom modu je uspostavljen trank između dva sviča na kojima su konfigurisane dve VLAN (DTP).



Sl. 2 Mrežna topologija za potrebe jedne realizacije zaštite paketskog telefonskog saobraćaja upotrebom tehnologije virtuelnih privatnih mreža

Prikazana realizacija mrežne topologije, snimanje i analiza mrežnog saobraćaja omogućava detaljno razumevanje uspostave VPN tunela, uz upotrebu simulacionog softvera, kao i praćenje procesa razmene informacionog sadržaja u realnom vremenu. Tokom procesa prenosa informacionog sadržaja u realnom vremenu, periodično se uočavaju procesi uspostavljanja prisutnosti uređaja u mreži (SSDP), obaveštavanja o nedostupnosti uređaja u slučaju prekida konektivnosti (ICMP poruke) i razmena “hello” poruka u OSPF procesu. Navedena realizacija stoga omogućava potpun uvid u sve procese u toku prenosa paketskog telefonskog saobraćaja, što može poslužiti u edukaciji i o procesima prenosa multimedijalnih informacionih sadržaja.

 ISBN 978-86-7466-930-3

TABELA I
POZIV UČESNIKA 001 KA UČESNIKU 003 BEZ VPN ZAŠTITE

R.br.	Vreme [s]	Izvorišna adresa	Odredišna adresa	Protokol	Veličina [B]
1.	0,00	Cisco_e1:aa:81	Spanning_tree	STP	60
2.	0,17	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
3.	1,99	Cisco_e1:aa:81	Spanning_tree	STP	60
4.	2,73	ASUSTEkC_9d:86:8a	Cisco_00:e3:e1	ARP	42
5.	2,74	Cisco_00:e3:e1	ASUSTEkC_9d:86:8a	ARP	60
6.	3,433	40.40.40.3	192.168.100.1	DNS	75
7.	3,434	40.40.40.1	40.40.40.3	ICMP	70
8.	4,00	Cisco_e1:aa:81	Spanning_tree	STP	60
9.	4,18	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
10.	5,32	40.40.40.1	224.0.0.5	OSPF	90
11.	5,99	Cisco_e1:aa:81	Spanning_tree	STP	60
12.	6,56	40.40.40.3	192.168.100.1	DNS	75
13.	6,561	40.40.40.1	40.40.40.3	ICMP	70
14.	6,83	40.40.40.3	30.30.30.1	SIP/SDP	922
15.	7,16	40.40.40.3	30.30.30.1	TCP	590
16.	7,16	30.30.30.1	40.40.40.3	TCP	60
17.	7,16	40.40.40.3	30.30.30.1	TCP	386
18.	7,16	30.30.30.1	40.40.40.3	SIP	418
19.	7,18	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
20.	7,36	40.40.40.3	30.30.30.1	TCP	54
21.	7,99	Cisco_e1:aa:81	Spanning_tree	STP	60
22.	8,30	40.40.40.3	30.30.30.1	TCP	58
23.	8,32	Cisco_e1:aa:81	Cisco_e1:aa:81	LOOP	60
24.	8,37	30.30.30.1	40.40.40.3	TCP	590
25.	8,37	30.30.30.1	40.40.40.3	SIP	124
26.	8,37	40.40.40.3	30.30.30.1	TCP	54
27.	9,99	Cisco_e1:aa:81	Spanning_tree	STP	60
28.	10,18	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
29.	11,97	10.10.10.2	40.40.40.3	RTP	214
30.	11,97	30.30.30.1	40.40.40.3	TCP	590
31.	11,97	30.30.30.1	40.40.40.3	SIP/SDP	424
32.	11,97	40.40.40.3	30.30.30.1	TCP	54
33.	11,98	40.40.40.3	10.10.10.2	RTP	55
34.	11,98	40.40.40.3	10.10.10.2	TCP	66
35.	11,98	10.10.10.2	40.40.40.3	TCP	60
36.	11,98	40.40.40.3	10.10.10.2	TCP	54
37.	11,98	40.40.40.3	10.10.10.2	SIP	459
38.	11,98	10.10.10.2	40.40.40.3	TCP	60
39.	11,99	10.10.10.2	40.40.40.3	RTP	214
40.	11,99	Cisco_e1:aa:81	Spanning_tree	STP	60
41.	12,01	10.10.10.2	40.40.40.3	RTP	214
42.	12,03	10.10.10.2	40.40.40.3	RTP	214
43.	12,25	40.40.40.3	10.10.10.2	RTP	214

TABELA II
POZIV UČESNIKA 001 KA UČESNIKU 003 SA VPN ZAŠTITOM

R.br.	Vreme [s]	Izvorišna adresa	Odredišna adresa	Protokol	Veličina [B]
1.	0,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
2.	0,03	Cisco_e1:aa:81	Spanning_tree	STP	60
3.	0,73	40.40.40.3	10.10.10.2	ISAKMP	126
4.	0,73	10.10.10.2	40.40.40.3	ISAKMP	126
5.	2,03	Cisco_e1:aa:81	Spanning_tree	STP	60
6.	4,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
7.	4,03	Cisco_e1:aa:81	Spanning_tree	STP	60
8.	6,03	Cisco_e1:aa:81	Spanning_tree	STP	60
9.	6,88	Fe80:e010:c683:5106:f3e8	Ff02::1:2	DHCPv6	149
10.	7,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
11.	8,03	Cisco_e1:aa:81	Spanning_tree	STP	60
12.	8,67	Cisco_e1:aa:81	Cisco_e1:aa:81	LOOP	60
13.	10,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
14.	10,03	Cisco_e1:aa:81	Spanning_tree	STP	60
15.	10,88	Fe80:e010:c683:5106:f3e8	Ff02::1:2	DHCPv6	154
16.	12,03	Cisco_e1:aa:81	Spanning_tree	STP	60
17.	13,30	Cisco_e1:aa:81	CDP/VTP/DTP	DTP	60
18.	13,30	Cisco_e1:aa:81	CDP/VTP/DTP	DTP	90
19.	14,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
20.	14,03	Cisco_e1:aa:81	Spanning_tree	STP	60
21.	16,03	Cisco_e1:aa:81	Spanning_tree	STP	60
22.	17,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
23.	18,03	Cisco_e1:aa:81	Spanning_tree	STP	60
24.	18,67	Cisco_e1:aa:81	Cisco_e1:aa:81	LOOP	60
25.	19,32	Cisco_e1:aa:81	ASUSTEkC_9d:86:8a	ARP	60
26.	19,32	ASUSTEkC_9d:86:8a	Cisco_e1:aa:81	ARP	42
27.	20,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
28.	20,03	Cisco_e1:aa:81	Spanning_tree	STP	60
29.	20,51	40.40.40.3	10.10.10.2	TCP	66
30.	22,03	Cisco_e1:aa:81	Spanning_tree	STP	60
31.	23,51	40.40.40.3	10.10.10.2	TCP	66
32.	24,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
33.	24,03	Cisco_e1:aa:81	Spanning_tree	STP	60
34.	25,22	ASUSTEkC_9d:86:8a	Cisco_00:e3:e1	ARP	42
35.	25,23	Cisco_00:e3:e1	ASUSTEkC_9d:86:8a	ARP	60
36.	26,03	Cisco_e1:aa:81	Spanning_tree	STP	60
37.	27,00	Fe80:e010:c683:5106:f3e8	Ff02::c	SSDP	208
38.	27,50	40.40.40.3	10.10.10.2	SIP/SDP	1095
39.	27,50	10.10.10.2	40.40.40.3	SIP	284
40.	28,30	40.40.40.3	10.10.10.2	SIP	831
41.	30,25	10.10.10.2	40.40.40.3	RTP	214
42.	31,50	40.40.40.3	10.10.10.2	RTP	214
43.	31,75	10.10.10.2	40.40.40.3	RTP	214

 ISBN 978-86-7466-930-3

IV. ZAKLJUČAK

Rezultat jedne realizacije zaštite paketskog telefonskog saobraćaja, predstavljen u ovom radu prikazuje da uspostava i održavanje VPN tunela kao načina zaštite paketskog telefonskog saobraćaja podrazumeva primenu niza protokola, specifično dizajniranih za prenos informacionih sadržaja u realnom vremenu (SIP, SSDP, RTP) kao i protokola za obezbeđenje zaštite u toku prenosa (ISAKMP, IPSec). Specifično za predstavljenu realizaciju predstavlja upotreba linka za prenos različitih servisa i time korišćenje protokola kojima se omogućava konvergencija servisa u prenosu (DTP, SSDP).

LITERATURA

[1] W.Stallings, Osnove bezbednosti mreža: Aplikacije i standardi, Računarski fakultet, Beograd, 2014.
[2] M.Stojanović, V.Aćimović-Raspopović, Savremene IP mreže: Arhitekture, tehnologije i protokoli, Akademska misao, Beograd, 2012.
[3] A.Smiljanić, Osnove i primena Interneta, Elektrotehnički fakultet Univerziteta u Beogradu, Beograd, 2015.
[4] D.Nemec, D.Vukobratović, V.Crnojević, Č.Stefanović, Tehnologija VoIP sistema, Fakultet tehničkih nauka, Novi Sad, 2007.
[5] Security in the Internet Architecture, RFC: 1636, jun 1994, https://datatracker.ietf.org/doc/html/rfc1636
[6] IP Encapsulating Security Payload, RFC: 4303, decembar 2005, https://datatracker.ietf.org/doc/html/rfc4303
[7] Security Architecture for the Internet protocol, RFC: 4301, decembar 2005, https://datatracker.ietf.org/doc/html/rfc4301
[8] Internet Security Association and key Management Protocol, RFC: 2408, novembar 1998, https://datatracker.ietf.org/doc/html/rfc2408
[9] Internet Key Exchange Protocol Version 2 (IKEv2), RFC: 5996, septembar 2010, https://datatracker.ietf.org/doc/html/rfc5996
[10] B.Scheiner, Primenjena kriptografija, Mikro knjiga, Beograd, 2007.
[11] R.Swale, D.Collins, Carrier Grade Voice Over IP, McGraw Hill Professional, 2004.
[12] M.Jevtović, Komunikacioni protokoli Interneta, Akademska misao, Beograd, 2011.
[13] I.Bašičević, "Prilog razvoju arhitekture za obezbeđivanje usluga u računarskim mrežama nove generacije", doktorska disertacija, Fakultet tehničkih nauka, Novi Sad, 2008.
[14] S.Gajin , Principi konfigurisanja računarskih mreža, Akademska misao, Beograd, 2018.

ABSTRACT

The research presented in this paper addresses an example of how to execute packet telephone traffic protection using virtual private network technology through configuring servers for packet telephone traffic and also demonstrates the secure transfer with the use of virtual private network technology in tunnel mode by applying the appropriate protocols for privacy protection, authentication, integrity protection and crypto key exchange. Traffic recording and analysis has been performed using the Wireshark software in both secure and non-secure transfer. The results obtained can help better understand the complex process of setting up a tunnel through the use of simulation software in education.

Packet Telephone Traffic Transfer Protection Using Technology of Virtual Private Networks

Mićo Živanović, Jovan Bajčetić, Ivan Tot

 ISBN 978-86-7466-930-3

LDPC dekoderi sa reinicijalizacijama koji objedinjuju tvrde odluke i razmenu mekih poruka

Predrag Ivaniš, *Senior Member, IEEE*, Srđan Brkić, *Member, IEEE* i Bane Vasić, *Fellow, IEEE*

***Apstrakt* — U ovom radu predloženi su postupci koji kombinuju dve klase iterativnih algoritama koje se uobičajeno koriste za dekodovanje zaštitnih kodova sa proverama parnosti male gustine (eng. Low Density Parity Check, LDPC). Prva strategija zasnovana je na bit-fliping algoritmu male kompleksnosti, a predložena modifikacija omogućava značajno poboljšanje performansi dekodera uz zadržavanje male prosečne računske kompleksnosti. Druga strategija zasnovana je na algoritmu sa propagacijom verodostojnosti i za cilj ima dodatno poboljšanje korekcionih sposobnosti dekodera, naročito kada se koriste kodovi sa malom dužinom kodne reči.**

***Ključne reči* — Bit-fliping algoritam, belief-propagation algoritam, iterativno dekodovanje, kodovi sa proverama parnosti male gustine.**

I. UVOD

KODOVI sa proverama parnosti male gustine (eng. Low Density Parity Check, LDPC) predloženi su od strane Roberta Galagera 1960. godine [1], a posebno veliku popularnost su doživeli u poslednjoj deceniji XX veka [2], [3]. Njihova sposobnost da dostignu kapacitet kanala uz relativno nisku kompleksnost dekodovanja [4] učinila ih je poželjnim rešenjem za obezbeđivanje pouzdanog prenosa podataka u bežičnim telekomunikacionim sistema novije generacije. Ovi kodovi se danas koriste u digitalnoj televiziji (eng. Digital Video Broadcasting, DVB) [5], bežičnim lokalnim mrežama (eng. Wireless Local Area Networks, WLAN) [6], a od skora su uvršteni i u standard za petu generaciju širokopojasnih mobilnih mreža (5G) [7].

Poboljšanje performansi sistema usled primene LDPC kodova pre svega je diktirano algoritmom koji se koristi za njihovo dekodovanje. Ovi algoritmi su po pravilu iterativni i implementiraju se nad Tanerovim grafom [8], koji je u potpunosti određen matricom provere parnosti koja definiše kod. U najvećem broju slučajeva, optimalne performanse se postižu ako se pri dekodovanju koristi tzv. algoritam propagacije verodostojnosti (eng. belief-propagation, BP), kod koga čvorovi u grafu razmenjuju poruke na osnovu kojih se procena kodne reči po pravilu poboljšava u svakoj narednoj iteraciji. Verzija BP algoritama pogodna za primenu u dekoderima LDPC kodova poznata je i pod nazivom algoritam sumiranja i množenja (eng. Sum Product Algorithm, SPA), a više detalja vezanih za njegovu efikasnu implementaciju može se naći u radu [9].

Predrag Ivaniš – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: predrag.ivanis@ etf.bg.ac.rs).

Srđan Brkić – Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: srdjan.brkic@ etf.bg.ac.rs).

Bane Vasić – Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA (e-mail: vasic@ece.arizona.edu).

Mana BP algoritma je velika složenost implementacije. Računske operacije koje se izvode u čvorovima Tanerovog grafa mogu biti veoma kompleksne, a pritom se podrazumeva da su poruke koje se razmenjuju realni brojevi. Pri hardverskoj implementaciji poželjno je da se operacije u čvorovima pojednostave i da se poruke predstave sa konačnom preciznošću (implementacija sa fiksnim zarezom). Stoga se u većini postojećih sistema koriste podoptimalne verzije BP algoritma, kao što su razne vrste min-sum algoritma [10]. Drugi značajan nedostatak BP algoritma je vezan za činjenicu da je on optimalan samo ako Tanerov graf ima oblik stabla. Ako se u grafu pojavljuju ciklusi male dužine, performanse BP algoritma su daleko od performansi koje bi bilo moguće dobiti primenom algoritma koji postiže maksimalnu verodostojnost pri odlučivanju (eng. Maximum Likelihood, ML). Ovo je posebno izraženo za kratke kodove. U literaturi se mogu naći rešenja koja obezbeđuju bolje performanse od BP algoritma za isti broj iteracija [11], a koja sa povećanjem broja iteracija obezbeđuju performanse bliske ML granici [12].

Sa druge strane, algoritam koji u svakoj iteraciji invertuje određene bitove kodne reči (eng. bit-flipping, BF) predstavlja postupak izuzetno niske kompleksnosti. Iako originalni BF algoritam ima prilično loše performanse po pitanju ispravljanja grešaka, isto se ne može reći i za nedavno predloženu modifikaciju BF algoritma nazvanu bit fliping algoritam sa gradijentnim spustom (eng. Gradient Descent Bit Flipping, GDBF) [13]. U našem prethodnom radu [14] predložen je probabilistički GDBF algoritam (PGDBF), kod koga primena slučajne sekvence pri dekodovanju omogućava ispravljanje kombinacije grešaka koje klasičan GDBF algoritam ne može korigovati. Dalje unapređenje ove ideje je razmotreno u radu [15]. Kombinovanje PGDBF algoritma i slučajnih reinicijalizacija može u velikoj meri poboljšati performanse dekodera (odgovarajući algoritam MUDRI je izložen u radu [16]). Analiza data u radovima [17]-[20] pokazuje da dekoderi bazirani na tvrdom odlučivanju uz slučajne reinicijalizacije obezbeđuju performanse bliske ML granici, ukoliko se može dozvoliti veliki broj iteracija pri dekodovanju.

Nedavno je pokazano da se performanse GDBF algoritma mogu približiti performansama BP algoritma, ako se destimuliše invertovanje istog bita u nekoliko uzastopnih iteracija [21]. Kombinacija tog pristupa sa determinističkim reinicijalizacijama ima potencijal da omogući čak i bolje performanse od BP algoritma, uz znatno manju računsku kompleksnost [22].

U nastavku će biti razmotreno nekoliko strategija za kombinovanje BP i GDBF algoritma, sa ciljem da se postigne veća pouzdanost odlučivanja ili manja kompleksnost implementacije, koristeći prednosti jednog i drugog pristupa.

 ISBN 978-86-7466-930-3

II. MODEL SISTEMA I PREGLED TIPIČNIH ALGORITAMA ZA DEKODOVANJE LDPC KODOVA

U ovom radu posmatramo binarne LDPC kodove kod kojih je informaciona reč $\boldsymbol{i}$ dužine k, dok je odgovarajuća kodna reč $\boldsymbol{x}$ dužine n. Ovi kodovi se označavaju sa (n,k), a njihov kodni količnik iznosi $R=k/n$.

LDPC kod se opisuje matricom provere parnosti $\boldsymbol{H}$, dimenzija $n\times m$, koja je po pravilu retka (broj jedinica u ovoj matrici je znatno manji od broja nula). U ovom radu će biti razmotreni kodovi kod kojih u svakoj koloni matrice $\boldsymbol{H}$ ima tačno γ jedinica, dok svaka vrsta matrice $\boldsymbol{H}$ sadrži tačno ρ jedinica (regularni kodovi). Parametri γ i ρ nazivaju se težina kolona i težina vrsta, respektivno. Odgovarajući bipartitni graf $\mathcal{G}$ sastoji se od skupa varijabilnih čvorova $\boldsymbol{V} = \{v_1, v_2, ..., v_n\}$ i skupa čvorova provere parnosti $\boldsymbol{C} = \{c_1, c_2, ..., c_m\}$. Čvorovi v_i i c_j su susedi ako su povezani na grafu, tj. ako je ispunjeno $\boldsymbol{H}_{ij}=1$. Skup suseda čvora v_i je označen sa N_{v_i}, dok je skup suseda čvora c_j označen sa N_{c_j}.

Nakon što LDPC koder od informacione reči $\boldsymbol{i}$ formira kodnu reč $\boldsymbol{x}$, ona se prenosi kroz binarni simetrični kanal (BSC). Ovaj kanal unosi nasumične, tj. vremenski nekorelisane, greške sa verovatnoćom p. Odgovarajuća sekvenca greške $\boldsymbol{e}$ može se dobiti uz pomoć generatora slučajne promenljive sa Bernulijevom raspodelom $B(1, p)$. U ovom slučaju, primljena reč $\boldsymbol{y}$ na izlazu kanala formira se tako što se poslata kodna reč $\boldsymbol{x}$ sabere bit-po-bit sa sekvencom greške $\boldsymbol{e}$, tj. $\boldsymbol{y}=\boldsymbol{x}\oplus\boldsymbol{e}$.

Primljena reč se dovodi na ulaz dekodera koji ima zadatak da verno rekonstruiše poslatu kodnu reč, odnosno poslatu informacionu sekvencu (koju treba dostaviti odredištu). Pošto je izdvajanje informacione sekvence prilično jednostavno ako se ispravno rekonstruiše poslata kodna reč, pri analizi dekodera se teži da procena kodne reči bude što je moguće pouzdanija. Drugim rečima, treba minimizovati verovatnoću da se procena poslate kodne reči, označena sa $\hat{\boldsymbol{x}}$, razlikuje od poslate reči $\boldsymbol{x}$. Ova verovatnoća se obično naziva verovatnoća greške po kodnoj reči (eng. Word Error Rate, WER). Iterativni dekoder procenu poslate kodne reči po pravilu poboljšava iz iteracije u iteraciju, čime se omogućava da i za prilično duge kodne reči kompleksnost dekodera ostane relativno mala [10].

Dekoder u svakoj iteraciji proverava da li su sve provere parnosti zadovoljene. Kada je ovaj uslov ispunjen, proces dekodovanja se prekida (tada procena odgovara kodnoj reči). Obično se zadaje maksimalan broj iteracija koje se mogu iskoristiti za dekodovanje, označen sa L.

Neka je procena kodne reči u l-toj iteraciji označena sa $\hat{\boldsymbol{x}}^{(l)}$, $l=1,2,...,L$. Odgovarajući sindrom $\boldsymbol{S}^{(l)}=\hat{\boldsymbol{x}}^{(l)}\boldsymbol{H}^T$ ima sve komponente ravne nuli samo u slučaju da su sve provere parnosti zadovoljene. Ako dekoder tokom L uzastopnih iteracija ne formira procenu za koju je $\mathbf{S}^{(l)}=\mathbf{0}$, dekodovanje se proglašava neuspešnim. Ukoliko se dekodovanje prekine za $l_0\le L$ i ako je $\hat{\boldsymbol{x}}^{(l_0)}=\boldsymbol{x}$, smatra se da je ta kodna reč uspešno dekodovana nakon l_0 iteracija.

U današnjim telekomunikacionim sistemima i sistemima za zapis podataka obično se za dekodovanje LDPC kodova koriste dve klase algoritama. Jedna je zasnovana na BP algoritmu, koga odlikuju visoka pouzdanost pri dekodovanju i velika složenost implementacije. Druga klasa algoritama je zasnovana na BF postupku, koji je daleko jednostavniji ali uz nešto veću verovatnoću greške pri odlučivanju. Osnovne karakteristike dve pomenute klase algoritama, kao i njihovih varijanti, navedene su u nastavku.

A. *Algoritmi sa propagacijom verodostojnosti*

Poznato je da se LDPC kodovi mogu uspešno dekodovati korišćenjem algoritama kod kojih se poruke iterativno razmenjuju između povezanih čvorova u grafu (eng. message-passing). Poruke koje se razmenjuju mogu biti veoma jednostavne, kao kod Galager-A/B algoritma (binarne poruke), ali se znatno bolji rezultati dobijaju ako ove poruke imaju oblik realnih brojeva.

Osnovna ideja BP algoritma će biti ilustrovana na grafu prikazanom na slici 1, gde su ispravno primljeni biti kodne reči označeni belim varijabilnim čvorovima, dok pogrešno primljenim bitima odgovaraju crni čvorovi. Zadovoljene provere parnosti označene su belim, a nezadovoljene crnim kvadratima. U BP algoritmu poruke koje varijabilni čvorovi inicijalno šalju ka povezanim proverama parnosti zavise samo od verovatnoće greške u binarnom kanalu

$$m^{(1)}_{v_i\to c_j}(v_i=0\,|\,y_i)=P(v_i=0\,|\,y_i)=\begin{cases}p, & y_i=1,\\ 1-p, & y_i=0,\end{cases} \qquad (1)$$

a pošto su u pitanju verovatnoće, jasno je da se može pisati $m^{(1)}_{v_i\to c_j}(v_i=1\,|\,y_i)=1-m^{(1)}_{v_i\to c_j}(v_i=0\,|\,y_i)$. Intenzitet ovih poruka srazmeran je uverenosti čvora v_i da bi mogao da ima vrednost $v_i=a$, kada mu je poznata primljena vrednost y_i.

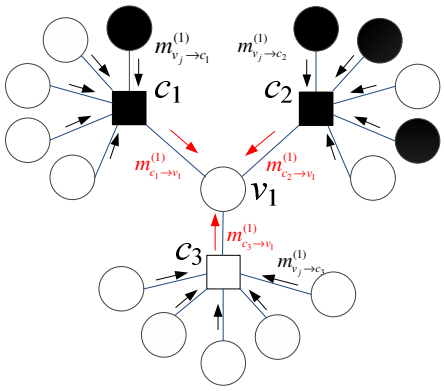
Imajući u vidu da vrednosti varijabilnih čvorova treba da odgovaraju bitima validne kodne reči (za koju su sve provere parnosti ravne nuli), u narednim iteracijama su poruke koje se razmenjuju između čvorova na Tanerovom grafu određene izrazima [10]

$$m^{(l)}_{c_j\to v_i}(v_i=0)=\frac{1}{2}\left(1-\prod_{v_k\in N_j\setminus v_i}\left(2m^{(l)}_{v_k\to c_j}(v_k=1)-1\right)\right), \qquad (2)$$

$$m^{(l+1)}_{v_i\to c_j}(v_i=0)=P(v_i=0\,|\,y_i)\prod_{c_z\in N_i\setminus c_j}m^{(l)}_{c_z\to v_i}(v_i=0), \qquad (3)$$

a jasno je da i dalje važi $m^{(l)}_{c_j\to v_i}(v_i=1)=1-m^{(l)}_{c_j\to v_i}(v_i=0)$, $m^{(l+1)}_{v_i\to c_j}(v_i=1\,|\,y_i)=1-m^{(l+1)}_{v_i\to c_j}(v_i=0\,|\,y_i)$. Verovatnoća da je v_i=0 se sada ažurira na osnovu uverenja povezanih čvorova

$$P^{(l)}(v_i=0)=P(v_i=0\,|\,y_i)\prod_{c_z\in N_i}m^{(l)}_{c_z\to v_i}(v_i=0). \qquad (4)$$



Sl 1. Ilustracija BP algoritma na grafu koji odgovara regularnom kodu sa parametrima γ=3, ρ=5.

 ISBN 978-86-7466-930-3

Na ovaj način se uverenost čvora o vrednostima koje na njega povezani čvorovi treba da imaju, pod datim uslovima, po pravilu poboljšava iz iteracije u iteraciju, što obično rezultuje sve pouzdanijim odlukama. Da bi se ublažili numerički problemi, BP algoritam se obično implementira u logaritamskom domenu, tako što se u svakoj iteraciji određuje logaritamski količnik verodostojnosti [8]

$$LLR_i^{(l)} = \log\left(P^{(l)}(v_i=1)/P^{(l)}(v_i=0)\right). \quad (5)$$

Ipak, uvek treba imati na umu da je BP algoritam optimalan samo u slučaju da je graf oblika stabla. Ovo nije ispunjeno kada u grafu postoje ciklusi male dužine, a posebno je kritično kada je dužina kodne reči mala.

B. *Bit-fliping algoritmi*

Bit-fliping algoritam je jednostavan postupak iterativnog dekodovanja LDPC kodova, koji je predložio Galager [1]. Algoritam je zasnovan na određivanju broja nezadovoljenih provera parnosti koje su povezane na svaki varijabilni čvor. Ovo izračunavanje se izvodi nad trenutno dostupnom procenom kodne reči, pa tako čvoru v_i u l-toj iteraciji odgovara energetska funkcija oblika

$$\Lambda_{BF}^{(l)}(v_i) = \sum_{c_j \in N_{v_i}} \bigoplus_{v_k \in N_{c_j}} \hat{x}_k^{(l)}. \quad (4)$$

Ukoliko je ova veličina za i-ti bit (i=1,2,…,n) veća od unapred definisanog praga T, vrši se invertovanje tog bita, čime se formira njegova procena u narednoj iteraciji

$$\hat{x}_i^{(l+1)} = \begin{cases} \hat{x}_i^{(l)} \oplus 1, & \Lambda^{(l)}(v_i) \geq T, \\ \hat{x}_i^{(l)}, & \Lambda^{(l)}(v_i) < T. \end{cases} \quad (5)$$

Polazeći od primljenog vektora ($\hat{\boldsymbol{x}}^{(0)} = \boldsymbol{y}$), na ovaj način se u svakoj iteraciji po pravilu formiraju sve pouzdanije procene poslate kodne reči. Ovaj algoritam često greši, pa bi za tipičnu vrednost $T = \gamma/2$ u primeru sa slike 1 bila doneta pogrešna odluka da čvor v_1 treba invertovati.

GDBF algoritam uvodi dve modifikacije – u energetsku funkciju se dodaje korelacioni član koji zavisi od odgovarajućeg bita primljene reči i invertuju se samo vrednosti onih varijabilnih čvorova koje u toj iteraciji imaju maksimalnu vrednost energetske funkcije [13, 14]:

$$\Lambda_{GDBF}^{(l)}(v_i) = y_i \oplus \hat{x}_i^{(l)} + \sum_{c_j \in N_{v_i}} \bigoplus_{v_k \in N_{c_j}} \hat{x}_k^{(l)}, \quad (6)$$

$$\hat{x}_i^{(l+1)} = \begin{cases} \hat{x}_i^{(l)} \oplus 1, & \Lambda^{(l)}(v_i) = \max_i \Lambda^{(l)}(v_i), \\ \hat{x}_i^{(l)}, & \Lambda^{(l)}(v_i) < \max_i \Lambda^{(l)}(v_i). \end{cases} \quad (7)$$

Kod GDBF algoritma sa momentumom (GDBF-w/m) energetska funkcija ima oblik [21]

$$\Lambda^{(l)}(v_i) = \alpha\left(y_i \oplus \hat{x}_i^{(l)}\right) + \beta \sum_{c_j \in N_{v_i}} \bigoplus_{v_k \in N_{c_j}} \hat{x}_k^{(l)} + \mu_i^{(l)}, \quad (8)$$

gde se korelacionom članu i broju nezadovoljenih provera parnosti pridružuju težinski koeficijenti (označeni sa α i β, respektivno), a na čitavu funkciju se dodaje novi član $\mu_i^{(l)}$ (momentum). Vrednost momentuma zavisi od broja iteracija koje su protekle od prethodnog invertovanja tog bita (koji je označen sa w_i) i može uzeti vrednosti definisane momentum vektorom $\mu = [\mu(1), \mu(1), ..., \mu(w_{\max})]$, pri čemu je $\mu(w_i) = 0$ ako je $w_i > w_{\max}$. Pokazano je da ovaj algoritam za neke kodove ima performanse uporedive sa BP algoritmom, naročito za male vrednosti parametra p [21].

III. STRATEGIJE ZA KOMBINOVANJE BF I GDBF ALGORITMA

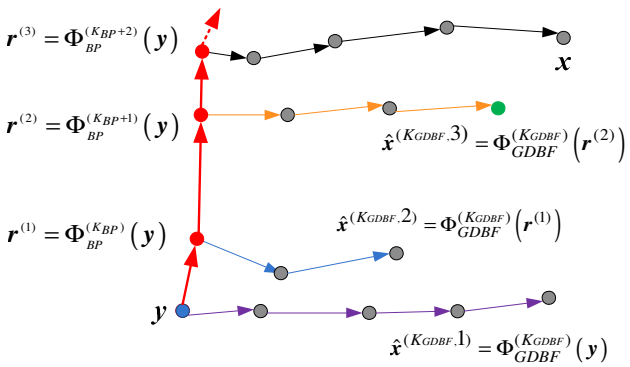
Jasno je da dve prethodno opisane klase algoritama rade na osnovu različitih principa, pa se nameće pitanje da li bi se mogle kombinovati tako da se objedini "najbolje iz dva sveta". Prvi korak u tom smeru učinjen je u radu koji smo nedavno objavili [22], pri čemu se GDBF-w/m algoritam modifikuje samo u nekim iteracijama.

Za razliku od prethodnih istaraživanja, u ovom radu je predlažemo kombinovanje BP algoritma i GDBF algoritma, pri čemu se jedan od njih koristi kao osnovno rešenje, dok se elementi drugog algoritma koriste za reinicijalizaciju ulaza dekodera. Dve takve strategije opisane su u nastavku.

A. *GDBF-w/m kao osnovni algoritam*

U ovoj varijanti se primljena reč dovodi na ulaz GDBF-w/m dekodera i za datu vrednost parametra p posmatra se kako se menja WER u zavisnosti od L. Formalno se može zapisati da GDBF-w/m dekoder nakon l iteracija preslikava $\boldsymbol{y}$ u procenjenu vrednost $\hat{\boldsymbol{x}}^{(l)} = \Phi_{GDBF}^{(l)}(\boldsymbol{y})$. Kada postane jasno da GDBF-w/m u dodatnim iteracijama neće značajno smanjiti WER, dekodovanje se prekida. Broj iteracija nakon koga se dekodovanje prekida je K_{GDBF} i određuje se empirijski. Primljena reč se zatim dovodi na ulaz BP dekodera, koji nakon K_{BP} iteracija na svom izlazu formira procenu $\boldsymbol{r}^{(1)}$, tj. $r^{(1)} = \Phi_{BP}^{(l)}(\boldsymbol{y})$, koja se koristi kao ulaz GDBF-w/m dekodera u novoj rundi. Po potrebi, postupak se ponavlja na način ilustrovan slikom 2, tako što se referenca $\boldsymbol{r}^{(q)}$ dobija kao izlaz BP dekodera nakon K_{BP}+q-1 iteracija, a energetska funkcija se računa na osnovu izraza

$$\Lambda^{(l)}(v_i) = \alpha\left(r_i^{(q)} \oplus \hat{x}_i^{(l)}\right) + \beta \sum_{c_j \in N_{v_i}} \bigoplus_{v_k \in N_{c_j}} \hat{x}_k^{(l)} + \mu_i^{(l)}.$$

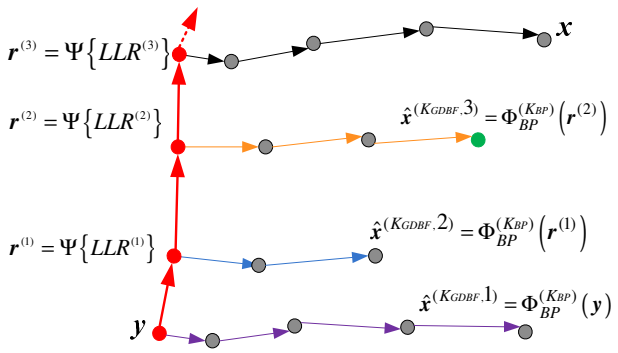


Sl 2. Ilustracija prve strategije, osnova je GDBF w/m algoritam, dok procena nakon odgovarajuće iteracije BP algoritma određuje referencu.

B. *BP kao osnovni algoritam*

U ovom scenariju se primljena reč dovodi na ulaz BP dekodera. Nakon svake iteracije pronađu se čvorovi v_i kod kojih je $LLR_i^{(l)}$ najveći po apsolutnoj vrednosti, uz uslov da bi taj bit bio invertovan kada bi se odluka donela u toj iteraciji, tj. $LLR_i^{(l)} \times LLR_i^{(0)} < 0$. Ove pozicije se redom zapisuju u odgovarajuću memoriju i čine skup kritičnih čvorova. Ako BP algoritam tokom K_{BP} iteracija ne obavi dekodovanje, postupak se ponavlja za novi ulaz dekodera. Odgovarajuća referenca r_q se formira tako što se u reči $\boldsymbol{y}$ invertuje samo jedan od kritičnih čvorova, pa se može formalno pisati $\boldsymbol{r}^{(q)} = \Psi\{LLR^{(q)}\}$ (videti sliku 3).

 ISBN 978-86-7466-930-3



Sl 3. Ilustracija prve strategije, osnova je GDBF w/m algoritam, dok procena nakon odgovarajuće iteracije BP algoritma određuje referencu.

IV. NUMERIČKI REZULTATI

Efekat dve predložene strategije će biti ilustrovan za dva regularna koda, koji su konstruisani koristeći različite postupke, pri čemu oba imaju težinu kolona γ=3 i dužinu najkraćeg ciklusa jednaku g=8. Kodovi imaju male dužine kodnih reči, a poznato je da tada BP algoritam predstavlja podoptimalno rešenje.

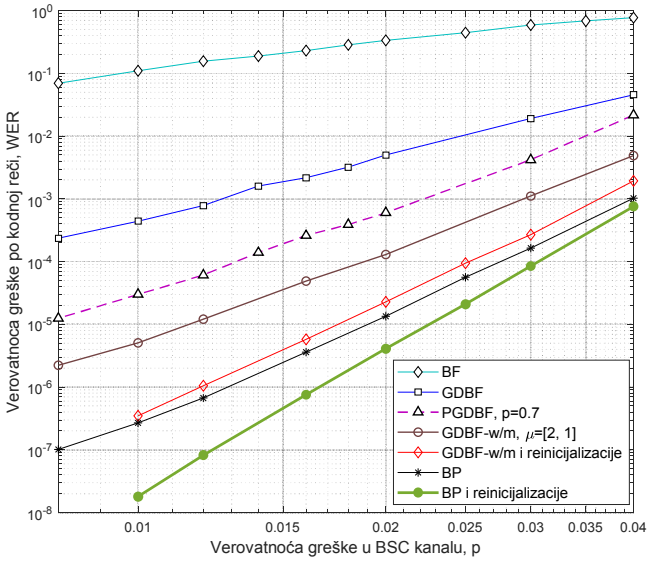
Verovatnoća greške po kodnoj reči određena je Monte Karlo simulacijom, pri čemu se generišu kodne reči dužine n i na svaku od njih se superponira slučajno generisan vektor greške iste dužine, generisan u skladu sa zadatom verovatnoćom greške u BSC. Kako bi bili sigurni da je simetrija dekodera zadovoljena, ne šalje se uvek kodna reč "sve nule", već se na izlaz kodera emituju različite kodne reči. U slučaju kada se dekodovanje završi za najviše L iteracija, porede se poslata i procenjena kodna reč, da bi bili sigurni da procena ne odgovara kodnoj reči koja je različita od poslate kodne reči. U slučaju kada se tokom L uzastopnih iteracija ne formira procena kodne reči kod koje su sve provere parnosti jednake nuli, dekodovanje se proglašava neuspešnim. Simulacija za jedan skup parametara se zaustavlja kada se detektuje 200 pogrešno primljenih kodnih reči, što određuje jednu tačku na grafiku.

Performanse dekodera će prvo biti prikazane za regularni Tanerov kod (155,64), koji ima težinu kolona γ=3 i težinu vrsta ρ=5. U pitanju je kvazi-ciklični LDPC kod, a odgovarajući metod konstrukcije predložen je u radu [23]. Ovaj kod se često koristi za testiranje algoritama dekodovanja [11, 24, 25].

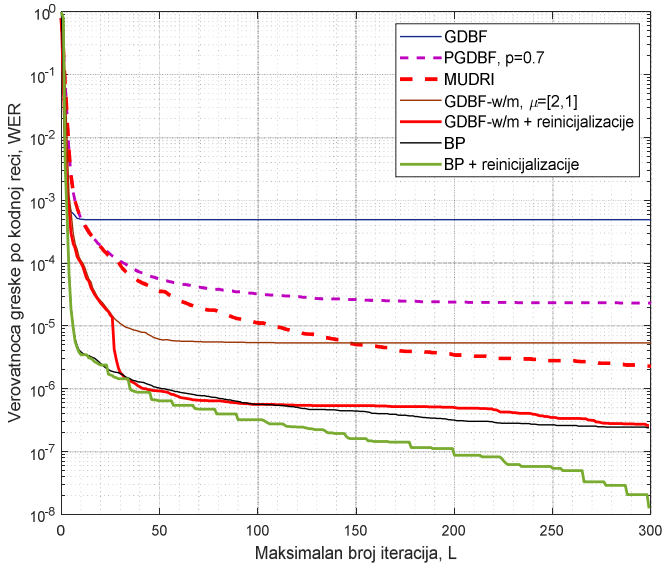
Zavisnost verovatnoće greške po kodnoj reči od verovatnoće greške u BSC prikazana je na slici 4. U slučaju kada BP algoritam određuje samo reinicijalizovane reference od kojih GDBF-w/m počinje dekodovanje u svakoj rundi, kriva je prikazana crvenom punom linijom. Svaka runda traje po K_{GDBF}=25 iteracija i ako se za to vreme ne detektuje da su sve provere parnosti jednake nuli, referenca se reinicijalizuje (sa K_{BP}=5) i ponovo se pokreće GDBF-w/m algoritam za nešto izmenjenu ulaznu sekvencu. Zelena kriva odgovara slučaju kada se koristi samo BP algoritam, ali se u svakoj rundi od K_{BP}=10 iteracija kao ulaz dekodera koristi reinicijalizovana referenca. Za male vrednosti parametra p, prva strategija obezbeđuje performanse uporedive sa BP algoritmom dok drugoj strategiji odgovaraju superiorne performanse.

Verovatnoća greške po kodnoj reči u zavisnosti od maksimalnog dozvoljenog broja iteracija pri dekodovanju ilustrovana je na slici 5. Vidi se da već nakon prve reinicijalizacije (L>30) performanse modifikovanog GDBF-w/m dekodera postaju uporedive sa performansama BP dekodera. Sa druge strane, modifikovani BP dekoder za bilo koju vrednost parametra L ima bolje performanse od ostalih razmatranih rešenja.

Dok strategija br 1. obezbeđuje verovatnoću greške po kodnoj reči koja je praktično ista kao u slučaju primene znatno složenijeg BP algoritma, jasno je da strategija br. 2 obezbeđuje dodatno poboljšanje performansi u odnosu na BP algoritam. Ovo je rezultat reinicijalizacija kod kojih se invertuju vrednosti najkritičnijih varijabilnih čvorova, što je princip sličan onom koji se primenjuje u GDBF algoritmu.



Sl. 4. Zavisnost verovatnoće greške po kodnoj reči od verovatnoće greške u BSC, za Tanerov kod (115,64) i L=300. Prikazane su performanse za GDBF-w/m i BP, kao i dve strategije njihovog kombinovanja.



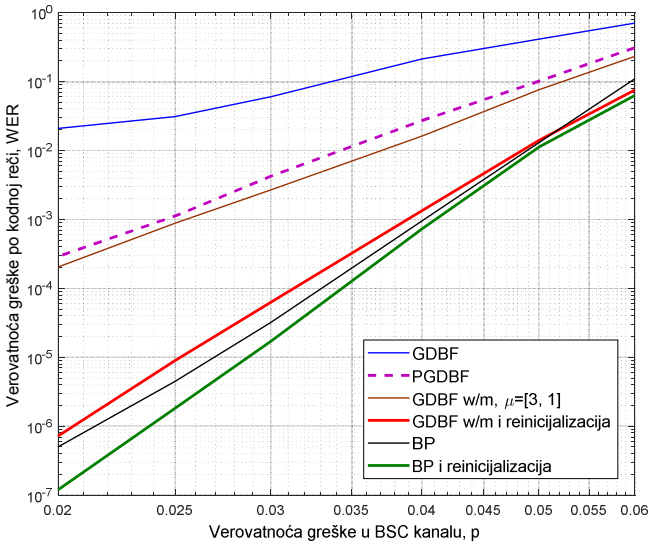
Sl. 5. Zavisnost verovatnoće greške po kodnoj reči od broja iteracija, za Tanerov kod (115,64) i p=0.01. Prikazane su performanse za GDBF-w/m i BP, kao i dve strategije njihovog kombinovanja.

 ISBN 978-86-7466-930-3

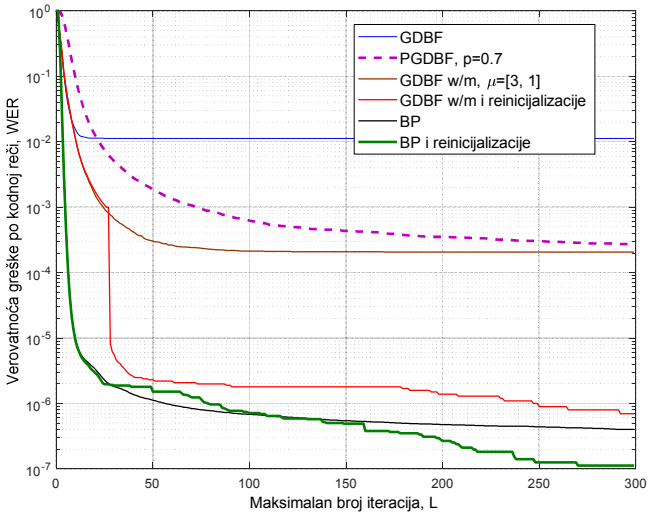
Numerički rezultati za regularni PEG kod (504,252), koji ima težinu kolona γ=3 i težinu vrsta ρ=6. U pitanju je kod formiran pomoću tehnike sa progresivnim formiranjem ivica grafa (eng. Progressive Edge Growth, PEG), koja je prvobitno predložena u radu [26].

Zavisnost verovatnoće greške po kodnoj reči od verovatnoće greške u BSC prikazana je na slici 6. Crvena linija odgovara slučaju kada se BP algoritam koristi samo za formiranje reinicijalizovane reference (sa K_{BP}=10), dok osnovu dekodera predstavlja GDBF-w/m algoritam (opisano u odeljku III-A). Kao i u prethodnom slučaju, svaka runda traje po K_{GDBF}=25 iteracija. Zelena puna linija odgovara slučaju kada se u svakoj rundi koristi samo BP algoritam, a kao ulaz dekodera koristi se reinicijalizovana referenca (na način opisan u odeljku III-B). U ovom slučaju, prva runda se završava nakon K_{BP}=25 iteracija, dok se nakon prve reinicijalizacije ova vrednost smanjuje na K_{BP}=10. U oba slučaja postižu se performanse uporedive sa BP algoritmom, dok druga strategija rezultuje osetnijim poboljšanjem performansi u tzv. *error floor* regionu.

Uticaj parametra L na performanse dekodera za PEG kod (504,252), kada je p=0.02, prikazan je na slici 7. Performanse dekodera u kome je primenjena strategija opisana u odeljku III-B su za malu vrednost parametra L uporedive sa performansama BP dekodera. Neznatna degradacija performansi u intervalu $25<L<75$ može da posluži kao indikacija da bi se još bolji rezultati mogli dobiti ako se u prvoj rundi dekodovanja parametar K_{BP} dodatno poveća. Sa druge strane, može se uočiti da se sa povećanjem maksimalnog broja iteracija performanse značajno poboljšavaju u odnosu na BP algoritam. Ovo je posledica reinicijalizacija u svakoj rundi dekodovanja. Taj efekat smo prethodno uočili u radu [16], za slučaj kada se koriste slučajne reinicijalizacije. U radu [17] pokazano je da se sa dovoljnim povećanjem parametra L performanse PGDBF dekodera asimptotski približavaju ML granici. U radovima [19] i [20] pokazano je da se sličan efekat postiže ako se slučajne reinicijalizacije kombinuju sa najjednostavnijim algoritmom koji koristi massage-passing princip za dekodovanje LDPC kodova (Gallager-B algoritam).



Sl. 6. Zavisnost verovatnoće greške po kodnoj reči od verovatnoće greške u BSC, za Tanerov kod (115,64) i L=300. Prikazane su performanse za GDBF-w/m i BP, kao i dve strategije njihovog kombinovanja.



Sl. 7. Zavisnost verovatnoće greške po kodnoj reči od broja iteracija, za PEG kod (504,252) i p=0.02. Prikazane su performanse za GDBF-w/m i BP, kao i dve strategije njihovog kombinovanja.

Rezultati prikazani na slikama 4-7 pokazuju da se značajno poboljšanje performansi može postići i kada su reinicijalizacije deterministčke. Za razliku od pristupa izloženog u radu [22], gde se opisan prilično složen postupak deterministčke reinicijalizacije, ovde je predložen koncept zasnovan na kombinaciji dva algoritma:

- U prvoj strategiji, reč od koje GDBF-w/m dekoder započinje dekodovanje u svakoj rundi dobija se na osnovu procena u pojedinim iteracijama BP algoritma.
- U drugoj strategiji se metoda bliska GDBF algoritmu (flipovanje najkritičnijeg bita) koristi za reinicijalizaciju, dok se BP koristi za dekodovanje u svakoj rundi. U prethodnim primerima pretpostavljeno je da se u procesu reinicijalizacije invertuju samo oni čvorovi v_i za koje je zadovoljeno $LLR_i^{(l)} = \max_i \left\{ LLR_i^{(l)} \right\}$. Ipak, preliminarni rezultati pokazuju da se za kodove sa dužim kodnim rečima bolje performanse postižu ako se ovaj skup proširi čvorovima za koje je ispunjena relacija $LLR_i^{(l)} \geq 0.9 \times \max_i \left\{ LLR_i^{(l)} \right\}$.

Treba zapaziti da se u prvoj strategiji dekodovanje najčešće završi primenom GDBF algoritma. Ukoliko je $p=10^{-2}$, prosečan broj iteracija potrebnih za dekodovanje reči Tanerovog koda (155,64) pomoću GDBF-w/m algoritma sa reinicijalizacijama iznosi $\bar{l} = 1.85$, dok je za $p=3\times10^{-2}$, prosečan broj iteracija $\bar{l} = 3.59$. Ako je $p=10^{-2}$, verovatnoća da GDBF-w/m tokom prvih K_{GDBF}=25 iteracija ne završi dekodovanje iznosi $\mathrm{WER}(25) \approx 1.2\times10^{-5}$ (videti sliku 5), iz čega se može zaključiti da se BP algoritam veoma retko pokreće, naročito u error-floor regionu, gde performanse predloženih rešenja posebno dolaze do izražaja.

Direktna posledica je da postupak predložen u prvoj strategiji obezbeđuje znatno manju prosečnu računsku kompleksnost u odnosu na slučaj kada se koristi BP dekoder. Ovaj postupak je baziran na GDBF-w/m algoritmu, a u radu [27] i doktorskoj disertaciji [28] navedeno je da u slučaju koda dužine n=1296 implementacija GDBF algoritma obezbeđuje šest puta veći prosečan protok i približno devet puta manju površinu na čipu u odnosu na najefikasniju poznatu implementaciju min-sum algoritma (koji predstavlja podoptimalnu varijantu BP algoritma, pogodnu za implementaciju zasnovanu na fiksnom zarezu).

 ISBN 978-86-7466-930-3

V. ZAKLJUČAK

Dve predložene strategije kombinovanja BP i GDBF-w/m algoritama pogodne su u slučaju kada su u prijemniku dostupna oba tipa dekodera. Dobijeni rezultati pokazuju da se kombinovanjem dva uobičajena pristupa mogu dobiti značajno poboljšane performanse ili performanse uporedive onima koje ima BP dekoder, ali uz smanjenu kompleksnost.

U ovom radu su parametri dekodera određeni empirijski. Sigurni smo da se dodatno poboljšanje performansi može dobiti ako se izvrši optimizacija pojedinih parametara, kao što su momentum vektor ili pragovi za odlučivanje kod GDBF-w/m algoritma, kao i broj iteracija koji odgovara pojedinim rundama dekodovanja. Ovo će biti tema naših budućih istraživanja.

ZAHVALNICA

Rezultati objavljeni u ovom radu delom su dobijeni u okviru saradnje ostvarene kroz i program Saradnje srpske nauke sa dijasporom Fonda za nauku Republike Srbije (br. ugovora 6462951), kao i kroz ERASMUS+ KA2 program saradnje Univerziteta u Beogradu i Univerziteta u Arizoni. Angažovanje Predraga Ivaniša i Srđana Brkića podržano je od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije. Angažovanje Baneta Vasića podržano je od strane NSF u okviru projekata CIF-1855879, CCF-2106189, CCSS-2027844 i CCSS-2052751 i NASA-SURP.

LITERATURA

[1] R. G. Gallager, *Low Density Parity Check Codes*, MIT Press, Cambridge, Mass., 1963.

[2] D. J. C. MacKay and R. M. Neal, "Near Shannon Limit Performance of Low Density Parity Check Codes," *Electronics Letters*, vol. 32, no. 18, pp. 1645-1646, Aug. 1996.

[3] D. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

[4] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, 47(2):619–637, 2001.

[5] ETSI Digital Video Broadcasting (DVB). *Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and other Broadband Satellite Applications; Part 2: DVB-S2 Extensions (DVB-S2X)*, ETSI EN 302307-2 V.1.1.1 (2014-10); ETSI: Sophia Antipolis, France, 2014.

[6] IEEE Standard for Information Technology—*Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,* IEEE Standard 802.11-2016; IEEE: New York, NY, USA, 2016.

[7] 3rd Generation Partnership Project. *Technical Specification Group Radio Access Network; NR; Multiplexing and Channel Coding (Release 16)*, 3GPP TS 38.212 V16.5.0 (2021-03); 3GPP: Valbonne, France, 2021.

[8] L. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, Vol. 27, no. 5, pp. 533-547, Sep. 1981.

[9] X.-Y. Hu, E. Eleftheriou, D.-M. Arnold and A. Dholakia, "Efficient implementations of the sum-product algorithm for decoding LDPC codes," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM 2001),* San Antonio, USA, November 25-29 2001, vol.2, pp. 1036-1036E.

[10] D. Declercq, M. Fossorier, and E. Biglieri, *Channel Coding: Theory, Algorithms, And Applications*. Academic Press Library in Mobile and Wireless Communications, Elsevier, 2014.

[11] S. K. Planjery, D. Declercq, L. Danjean, and B. Vasić, "Finite alphabet iterative decoders, Part I: Decoding beyond belief propagation on the binary symmetric channel," *IEEE Trans. Commun*, vol. 61, no. 10, pp. 4033–4045, Nov. 2013.

[12] D. Declercq, E. Li, B. Vasić, and S. Planjery, "Approaching maximum likelihood decoding of finite length LDPC codes via FAID diversity," in *Proc. IEEE Information Theory Workshop*, Lausanne, Switzerland, Sep. 3–7 2012, pp. 487–491.

[13] T. Wadayama, K. Nakamura, M. Yagita, Y. Funahashi, S. Usami and I. Takumi, "Gradient descent bit flipping algorithms for decoding LDPC codes," *IEEE Trans. Commun*, vol. 58, no. 6, pp. 1610-1614, June 2010.

[14] O.-A. Rasheed, P. Ivanis, and B. Vasić, "Fault-tolerant probabilistic gradient-descent bit flipping decoders," *IEEE Commun. Letters*, vol. 18, no. 9, pp. 1487 – 1490, Sep. 2014.

[15] H. Cui, J. Lin, Z. Wang, "An Improved Gradient Descent Bit-Flipping Decoder for LDPC Codes," *IEEE Trans. Circuits and Systems I: Regular Papers*, vol. 66, no. 8, pp. 3188-3200, Aug. 2019.

[16] P. Ivanis, O.-A. Rasheed, and B. Vasić, "MUDRI: A fault-tolerant decoding algorithm," in *Proc. IEEE International Conference on Communications (ICC 2015)*, London, UK, June 8–12 2015, pp. 4291– 4296.

[17] B. Vasić, P. Ivaniš, D. Declercq, and K. LeTrung, "Approaching maximum likelihood performance of LDPC codes by stochastic resonance in noisy iterative decoders," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2016, pp. 1–9.

[18] D. Declercq, C. Winstead, B. Vasic, F. Ghaffari, P. Ivanis, and E. Boutillon, "Noise-Aided Gradient Descent Bit-Flipping Decoders approaching Maximum Likelihood Decoding", in *Proc 9th International Symposium on Turbo Codes & Iterative Information Processing (ISTC 2016)*, Special Session: Noisy Error Correction, Brest, France, 5-9 September 2016, pp. 300-304.

[19] P. Ivaniš, B. Vasić, D. Declercq, "Performance Evaluation of Faulty Iterative Decoders using Absorbing Markov Chains", *in Proc. IEEE International Symposium on Information Theory (ISIT 2016)*, Barcelona, Spain, July 10-15 2016, pp. 1566-1570.

[20] P. Ivaniš and B. Vasić, "Error Errore Eicitur: A Stochastic Resonance Paradigm for Reliable Storage of Information on Unreliable Media," *IEEE Trans. Commun*, vol. 64, no. 9, pp. 3596-3608, Sep. 2016.

[21] V. Savin, "Gradient Descent Bit-Flipping Decoding with Momentum," in *Proc. International Symposium on Topics in Coding (ISTC 2021)*, Montreal, Canada, 30 August - 3 September 2021.

[22] P. Ivaniš, S. Brkić, B. Vasić, "Suspicion Distillation Gradient Descent Bit-Flipping Algorithm," *Entropy*, vol. 24, no. 4, Article No. 558, Apr. 2021.

[23] R. M. Tanner, D. Sridhara, T. Fuja, "A class of group-structured LDPC codes," In *Proc. ISTA*, Ambleside, UK, 2001.

[24] S. Zhang and C. Schlegel, "Controlling the Error Floor in LDPC Decoding," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3566-3575, Sep. 2013.

[25] R. Asvadi, A. H. Banihashemi and M. Ahmadian-Attari, "Lowering the Error Floor of LDPC Codes Using Cyclic Liftings," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2213-2224, Apr. 2011.

[26] X.-Y. Hu, E. Eleftheriou and D.-M. Arnold, "Progressive edge-growth Tanner graphs,". in *Proc IEEE Global Telecommunications Conference (GLOBECOM'01)*, November 2001, vol.2, pp. 995-1001.

[27] K. Le, F. Ghafari, D. Declercq and B. Vasić, "Efficient Hardware Implementation of Probabilistic Gradient Descent Bit-Flipping," *IEEE Trans. Circuits and Systems I: Regular* Papers, vol. 64, no. 4, pp. 906-917, Apr. 2016.

[28] L. T. Khoa, *New Direction on Low Complexity Implementation of Probabilistic Gradient Descent Bit-Flipping Decoder*, PhD Thesis, École Nationale Supérieure de l'Électronique et de ses Applications, Université de Cergy Pontoise, Cergy Pontoise, France, 2017.

ABSTRACT

In this paper we propose the approaches that combine two types of iterative decoding algorithms that are usually used for decoding of low density parity check codes (LDPC). One strategy is based on a low-complexity bit-flipping algorithm, and the proposed modification enable significant performance improvement, with no significant increase of the average computing complexity. The other strategy is based on belief propagation decoder, and the resulting decoder has improved error correction capabilities for the codes with short codeword length.

LDPC decoders with re-initializations based on synergy of hard decision and message passing principles

Predrag Ivaniš, Srđan Brkić, Bane Vasić

 ISBN 978-86-7466-930-3

Analiza performansi kooperativnog diverziti sistema u kompozitnom fedingu modelovanom odnosom α-μ i gama raspodela

Edis Mekić, Irfan Fetahović, i Edin Dolićanin

***Apstrakt*— U ovom radu je izvedeno novo opšte, jednostavno rešenje u zatvorenom obliku za funkciju gustine verovatnoće odnosa proizvoda slučajnih promenljivih predstavljenih α-μ i Gama raspodelama i slučajne promenjive predstavljene Gama raspodelom. Ova rešenja se primenjuju u analizi performansi komunikacionih sistema sa kooperativnim diverziti sistemom koji se koristi za poboljšanje prijema signala na čije anvelope utiče brzi i spori feding, dok na anvelopu kokanalne interference utiče samo brzi feding.**

***Ključne reči*— α-μ raspodela; Gama raspodela; feding; kooperativni diverziti**

I. UVOD

Proizvodi i odnosi slučajnih promenljivih, kao i rešenja funkcija gustina verovatnoća (PDF) i kumulativnih funkcija (CDF) u zatvorenoj formi su u žiži interesovanja telekomunikacionih istraživanja.

Rešenja u zatvorenom obliku za PDF i CDF se mogu koristiti u analizi verovatnoće otkaza multihop kognitivnih mreža [1]. Kao posebno značajni u analizi otkaza i ergodičnog kapaciteta mreža i u prisustvu fedinga su se pokazali odnosi i proizvodi anvelopa signala koje su modelovane α-μ raspodelom [2]. Analiza odnosa signal interference (SIR) datog kao odnos proizvoda dve k- μ slučajne promenljive i Nakagami-m slučajne promenljive, kao i odnos proizvoda dve Rayleigh-ove slučajne promenljive i Rayleigh-ove slučajne promenljive je evaluirana u radovima [3,4] i korišćena za analizu bežičnih relejnih komunikacionih sistema koji se sastoje od dve sekcije, u prisustvu fedinga i kokanalne interferencije. Jednostavno matematičko rešenje u slučaju kada se koriste samo α-μ slučajne promenljive je dato u [5]. Dobijeno rešenje se koristi za modelovanje fedinga u multihop sistemima, u prisustvu kokanalne interference

Slučajna promenljiva može se koristiti za modelovanje anvelope signala u prisustvu fedinga. Feding je pojava da kvalitet prenosa signala opada zbog refleksije i refrakcije signala o objekte koji se nalaze na putanji istog (brzi feding) ili zbog velikih prepreka na putanji signala (spori feding). Osim ovoga, na anvelopu signala utiče i kokanalna interferenca koja se javlja zbog višestrukog korišćenja frekvencija. Anvelopa signala u prisustvu fedinga se može modelovati proizvodom dve slučajne promenljive [6], dok se kokanalna interferenca modeluje odnosom slučajnih promenljivih [7].

Proizvod α-μ i lognormalne slučajne promenljive daju odlične rezultate u modelovanju realnih efekata sporog fediga, međutim primena lognormalne raspodele ne daje rešenja u zatvorenom obliku za PDF anvelope signala na prijemu [8]. Kao alternativa lognormalnoj rapodeli može se koristiti Gama raspodela [9].

Da bi ublažili efekte fedinga na anvelopu signala mogu se koristiti diverziti tehnike koje se sastoje od većeg broja antena [10].

U ovom radu ćemo analizirati bežičnu mrežu sa kooperativnim diverziti protokolom za ublažavanje efekta fedinga [11].

Brzi feding i njegovo delovanje na signal između predajnika i prijemnika u slučaju idealnog koherentnog prenosa u poludupleks modu modelovaćemo kao odnos α-μ promenljivih, dok ćemo efekat sporog fedinga modelovati kao proizvod ovog odnosa sa Gama slučajnom promenljivom. Dobijena rešenja za PDF biće primenjena u modelu kooperativne diverziti mreže, pošto su multihop relejni sistemi specijalni slučajevi ove diverziti tehnike. Validnost rezulata ćemo pokazati na primeru jednostavnog dual hop sistema.

II. FIZIČKI MODEL BEŽIČNE MREŽE SA KOOPERATIVNIM DIVERZITI PROTOKOLIMA

Primer bežične mreže koja se sastoji iz niza releja je dat na Sl.1. Terminali TS_1, TS_2,…TS_N prenose signal do terminala TR_1, TR_2,…,TR_M , respektivno. Ovako definisan sistem može da se koristi kao osnova za modelovanje različitih bežičnih sistema. Recimo, možemo da predstavimo sistem mobilne telefonije ako pretpostavimo da su TS_1, TS_2,…,TS_N mobilni uređaji, a $TR_1= TR_2=…=TR_M$ bazne stanice. Model možemo da koristimo za reprezentaciju LAN mreže, bilo da je u pitanju ad-hoc ili mreža zasnovana na pristupnim tačkama. Ako pretpostavimo da su terminali $TR_1= TR_2=…=TR_M$, onda imamo LAN mrežu sa pristupnom tačkom, a ako pretpostavimo da je $TR_1\neq TR_2\neq…\neq TR_M$, onda imamo ad-hoc mrežu. Najzad, ako se svi terminali fokusiraju na prenos informacija od TS_1 do TR_M, dok su prelazni koraci prenosa redom od TR_1 do TS_2, od TS_2 do TR3, … TS_N do TR_M, onda imamo multihop sistem.

Poslednji navedeni model ćemo analizirati tako što ćemo

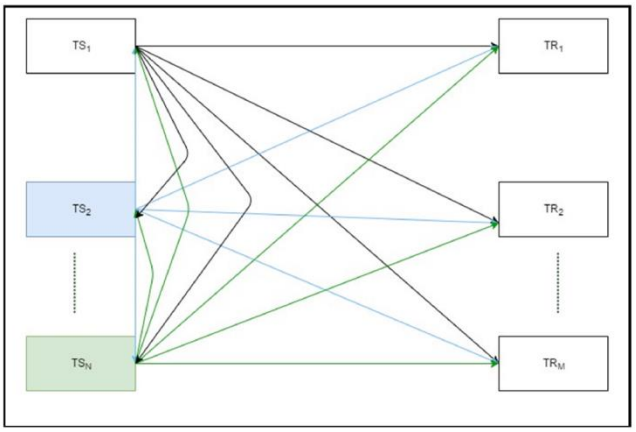
Edis Mekić – Departman za tehničko-tehnološke nauke, Državni univerzitet u Pazaru, Vuka Karadžića 9, 36300, Novi Pazar, Srbija (e-mail: emekic@np.ac.rs).

Irfan Fetahović – Departman za tehničko-tehnološke nauke, Državni univerzitet u Pazaru, Vuka Karadžića 9, 36300, Novi Pazar, Srbija (e-mail: ifetahovic@np.ac.rs).

Edin Dolićanin – Departman za tehničko-tehnološke nauke, Državni univerzitet u Pazaru, Vuka Karadžića 9, 36300, Novi Pazar, Srbija (e-mail: edin@np.ac.rs).

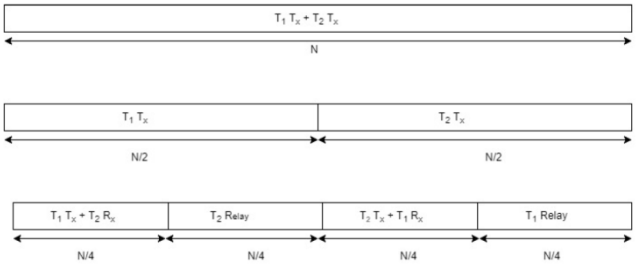
 ISBN 978-86-7466-930-3

ga ograničiti na mod poludupleksa, i to na takav način da svaki od releja pojačava primljeni signal. Radi jednostavnijeg računa pretpostavićemo da imamo idealni koherentni prenos, gde se primljeni signal u potpunosti obnavlja na transmisionom releju.



Sl.1. Bežična mreža zasnovana na sistemu releja

Kanale između releja ćemo podeliti na ortogonalne podkanale, pri čemu će svaki od terminala koristiti i/n stepena slobode u kanalu (Sl.2).



Sl.2. Alokacija kanala

Sa obzirom na to da je prenos između predajnika i prijemnika pod uticajem i brzog i sporog fedinga, simetrija kanala će nam omogućiti da stavimo fokus za početak na jedan par predajnik-terminal.

III. Statistička analiza proizvoda odnosa α-μ i Gama slučajnih promenljivih

Anvelope polaznog signala su modelovane slučajnim promeljivim a[n], gde je n=1,2,…,N/2, dok drugi terminal prenosi vrednosti n=N/2+1,…N. Terminal može da koristi samo polovinu stepena slobode. Anvelopu primljenog signala modelovaćemo kao proizvod odnosa slučajnih promenljivih.

$$\varphi[n] = \frac{a[n]\,c[n]}{b[n]} \tag{1}$$

Slučajne promenljive a[n] i b[n] su modelovane kao α-μ dok je c[n] modelovana gama slučajnom promenljivom. PDF proizvoda odnosa izračunaćemo tako što za početak računamo PDF odnosa dve α-μ slučajne promenljive.

$$\phi[n] = \frac{a[n]}{b[n]} \tag{2}$$

$$p_\phi(\phi[n]) = \int_0^\infty |J| p_\phi(\phi[n]a[n]) p_b(b[n]) db \tag{3}$$

Vrednost |J| je Jakobijan defnisan na sledeći način:

$$|J| = \left|\frac{da[n]}{d\phi[n]}\right| = b[n] \tag{4}$$

PDF združene promenljive φ[n] može se izračunati korišćenjem sledećeg izraza:

$$p_\varphi(\varphi[n]) = \int_0^\infty |J| p_\phi\left(\frac{\varphi[n]}{c[n]}\right) p_c(c[n]) dc \tag{5}$$

Odabir odgovarajuće kooperativne ili nekooperativne akcije je zasnovan na merenju odnosa signal-interferenca (SIR). Na prijemnoj strani merimo vrednosti anvelope signala. Možemo primeniti adaptiran mod prenosa, u skladu sa izmerenom vrednošću. Ako izmerena vrednost anvelope pada ispod određenog praga, ponavljamo prenos, a ako je iznad ovog praga onda prenosimo signal, ovo je korišćenje pojačaj i prosledi tehnike da bi se postigao diverziti.

Za kooperativni diverziti sistem, gde signal prolazi kroz niz releja, PDF slučajne promenljive može biti izračunat na sledeći način:

$$\varphi[n] = \min(\varphi[n]_1, \varphi[n]_2, \dots, \varphi[n]_N) \tag{6}$$

$$p_\varphi(\varphi[n]) = \sum_{m=1}^{N} |J| p_{\varphi_m}(\varphi[n]) \prod_{\substack{k=1 \\ k\neq m}}^{N} (1 - F_{\varphi_k}(\varphi[n])) \tag{7}$$

Sada ćemo modelovati efekte, brzog, sporog fedinga i interference. Za modelovanje anvelope na koju deluje brzi feding i interference koristimo α-μ raspodelu. Vrednost μ predstavlja broj klastera prenosa, dok koeficijent α predstavlja nelinearnost okoline. Radi jednostavnije analize slučajnu promenljivu između bilo kojih releja *a[n]*, predstavljamo kao promenljivu a.

Slučajna promenljiva α-μ raspodele je opisana sledećom jednačinom:

$$p_a(a) = \alpha \left(\frac{\mu_a}{\Omega_a}\right)^{\mu_a} \frac{\gamma^{\alpha\mu_a - 1}}{\Gamma(\mu_a)} e^{-\frac{\mu_a a^\alpha}{\Omega_a}} \tag{8}$$

U datoj jednačini snaga anvelope signala je $\Omega_a = \varepsilon\langle a^\alpha\rangle$, α predstavlja nelinearnost okoline, μ_a je inverzna vrednost normalizovane varijanse *α, ($\mu_a \geq 0.5$)*.

Spori feding modelovaćemo sledećom Gama slučajnom promenljivom:

$$p_a(a) = \frac{a^{k-1} e^{-\frac{a}{\Omega_a}}}{\Gamma(a)\Omega_a} \tag{9}$$

Vrednost sporog fedinga definisana je promenljivom k, pri čemu niža vrednost k znači da je veći uticaj sporog fedinga.

Da bi izračunali odgovarajući PDF slučajne promenljive φ dat jednačinom (5), prvo moramo da izračunamo PDF odnosa dve slučajne α-μ promenljive $\phi=a/b$, rešavanjem jednačine (3). Ovaj međukorak nam omogućuje da izračunamo združenu gustinu verovatnoće zaobilazeći kompleksnije Mellinove transformacije.

Primenom jednačina [12 (3.461 i 6.631), 13 (26)] združena gustina verovatnoće se može prikazati preko Meijer's G

 ISBN 978-86-7466-930-3

funkcija.

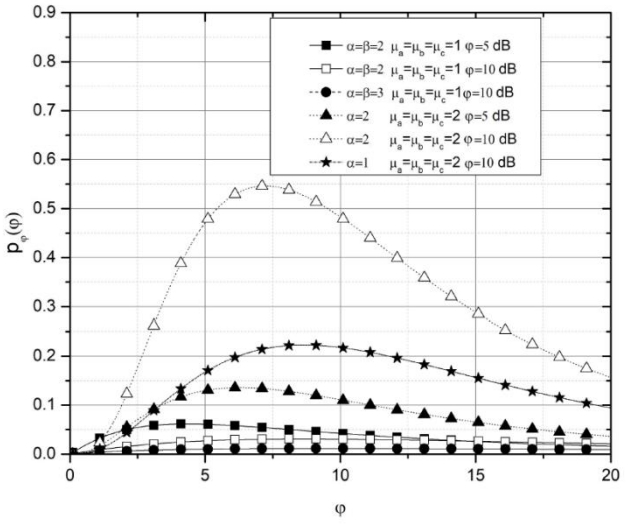
$$p_\phi(\phi) = \alpha\left(\frac{\mu_a\Omega_b}{\mu_b\Omega_a}\right)^{\mu_a}\frac{\phi^{\alpha\mu_a-1}}{\Gamma(\mu_a)\Gamma(\mu_a)}G_{1,1}^{1,1}\left(\frac{1}{\varphi^\alpha}\frac{\mu_a\Omega_b}{\mu_b\Omega_a}\middle|\begin{matrix}1-\mu_a-\mu_b\\0\end{matrix}\right) \quad (10)$$

PDF promenljive φ se izvodi primenom izraza (5). Korišćenjem izraza [14, 15] dobijamo izraz za PDF u zatvorenom obliku preko Meijer's G funkcija.

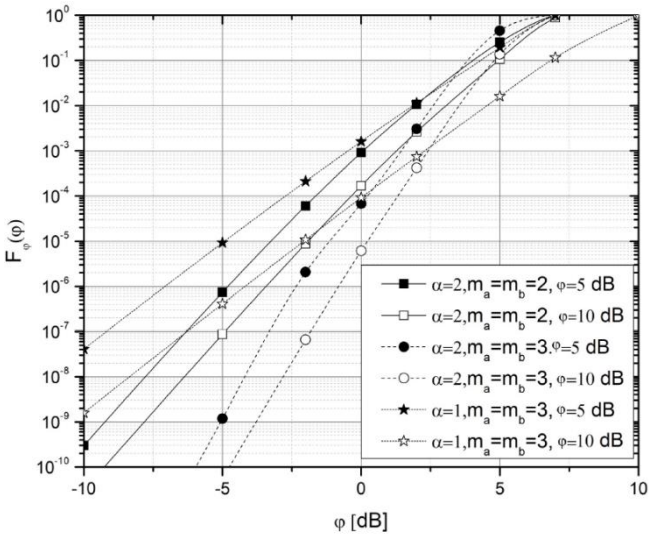
$$p_\varphi(\varphi) = \left(\frac{\mu_a\Omega_b}{\mu_b\Omega_a}\right)\left(\frac{1}{\Omega_c}\right)^{\alpha\mu_a}\frac{\alpha^{k-\alpha\mu_a-1}\varphi^{\alpha\mu_a-1}}{(2\pi)^{\frac{\alpha-1}{2}}\Gamma(\mu_a)\Gamma(\mu_c)\Gamma(k)}\times \times G_{1+\alpha,\ 1}^{1,\ 1+\alpha}\left(\frac{1}{\varphi^\alpha}\frac{\mu_b\Omega_a\Omega_c}{\mu_a\Omega_b}\middle|\begin{matrix}\frac{1-k+\alpha\mu_a}{2},\dots,\frac{1-k+\alpha\mu_a}{2},1\\ \mu_a+\mu_b\end{matrix}\right) \quad (11)$$

CDF se može izračunati primenom sledećeg izraza

$$F_\varphi(\varphi) = \int_0^\varphi p_\varphi(s)ds \quad (12)$$



Sl.3. PDF za različite vrednosti parametra α-μ promenljive



Sl.4. CDF za različite vrednosti parametra α-μ promenljive

Primenom dobijenih izraza, a menjajući parametar α-μ u izvedenim izrazima, dobijen je matematički model koji pokazuje da povećanje snage anvelope signala smanjuje verovatnoću otkaza sistema. Na Sl.3. je dat PDF združene gustine verovatnoće anvelope modelovanog signala, dok je verovatnoća otkaza data kroz numeričko izračunavanje CDF anvelope koja je modelovana novom združenom gustinom verovatnoće (Sl.4), dok povećanje dubine fedinga povećava verovatnoću otkaza.

IV. Zaključak

U radu je predstavljena analiza kooperativnog diverziti sistema kao jednog od pristupa za smanjivanje uticaja brzog i sporog fedinga na prenos signala u prisustvu interference. Uticaj efekata brzog fedinga i kokanalne interference smo modelovali odnosom dve α-μ slučajne promenljive, dok je efekat sporog fedinga modelovan proizvodom navedenog odnosa i Gama slučajnom promenljivom. Izračunat je izraz za PDF u zatvorenom obliku. Međutim, za kompletnu analizu neophodno je izračunati i CDF u zatvorenom obliku. To bi omogućilo brzu i efikasnu simulaciju ne samo navedenog slučaja, već i velikog broja drugih slučajeva koji se mogu izračunati kao specijalni slučajevi predstavljenog.

Literatura

[1] Y. A. Rahama, M. H. Ismail, M. S. Hassan, "On the distribution of the product and ratio of products of EGK variates with applications", vol. 68, no. 2, pp. 231-238 Telecommunication Systems, 2018.

[2] P. N. Rathie, A. K. Rathie, L. C. Ozelim, "The product and the ratio of\(\alpha-\mu\) random variables and outage, delay-limited and ergodic capacities analysis,", vol. 4, no. 1, pp. 100-108, Physical Review and Research International, 2014.

[3] D. Krstic, I. Romdhani, M.M.B. Yassein, S. Minic, G. Petkovic, P. Milacic, "Level crossing rate of ratio of product of 2 two ku random variables and Nakagami-m random variable," In: Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), IEEE International Conference on 1620-162, 2015.

[4] D. Krstic, M. Stefanovic, S. Minic, M. Peric, "Analysis of Ratio of One and Product of Two Rayleigh Random Variables and its Application in Telecommunications", no. 3, International Journal of Communications, 2018.

[5] E. Mekic, N. Sekulovic,M. Bandjur, M. Stefanovic, P. Spalevic, "The distribution of ratio of random variable and product of two random variables and its application in performance analysis of multi-hop relaying communications over fading channels," vol. 8, no. 7A, pp. 133-137, Przeglad Elektrotechniczny, 2012.

[6] G. E. Corazza and F. Vatalaro, "A statistical model for land mobile satellite channels and its application to nongeostationary orbit systems," vol. 43, pp. 738-741, IEEE Transactions Vehucilar Technologies, 1994.

[7] J.D. Parsons, *The Mobile Radio Propagation Channels*. 2nd ed. Wiley, 2002.

[8] P.G. Babalis, C.N. Capsalis. "Impact of the combined slow and fast fading channel characteristics on the symbol error probability for multipath dispersion less channel characterized by a small number of dominant paths," vol. 47, no. 5, pp. 653-657, 1999.

[9] T.A. Tsiftsis, "Performance of wireless multihop communications systems with cooperative diversity over fading channels,"vol. 21, no. 5, pp. 559-565. International Journal of Communication Systems, 2008.

[10] N. Dimitriou, A. Polydoros, A. Barnawi, "Cooperative schemes for path establishment in mobile ad-hoc networks under shadow-fading," vol. 11, no. 8, pp. 2556-2566, Ad Hoc Networks, 2013

[11] D. da Costa, M. Yacoub, G. Fraidenraich, "Second-order statistics of equal-gain and maximal-ratio combining for the α-μ (generalized gamma) fading distribution," IEEE 9th International Symposium on Spread Spectrum Techniques and Applications, pp. 342–346, 2006.

[12] I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, 7th edition. New York, Academic, 2007.

[13] V.S. Adamchik, O.I. Marichev," Algorithm for calculating integrals of hypergeometric type functions and its realization in reduce system" in Proceedings of International Symposium on Symbolic and Algebraic Computation (ISSAC '90), pp. 212–224, 1990.

[14] http://functions.wolfram.com/07.34.17.0012.01.

[15] http://functions.wolfram.com/07.34.21.0088.01.

 ISBN 978-86-7466-930-3

ABSTRACT

In this paper, we present novel general, simple, exact and closed-form expressions for the probability density function (PDF) and of ratio of product of random α-μ distributed variable and generalized Gamma distributed variable and random α-μ distributed variable. These results have application in performance analysis of cooperative diversity communication systems in different fading transmission environments where envelope of the signal is affected by fast and slow fading and envelope of co-channel interference by fast fading.

Performance Analysis of the Cooperative Diversity Communication Systems in Composite Fading Scenario Modelled as Ratio of α-μ Variates and Gamma Variate

Edis Mekić, Irfan Fetahović i Edin Dolićanin

 ISBN 978-86-7466-930-3