JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

2/2021

(*Contents Continued on Back Cover*)

on-line: ISSN 1899-8852

Paper

Digital Predistortion of Wideband Signals with Reduced Complexity Based on Feedback Wiener System

Tayeb H. C. Bouazza, Smail Bachir, and Claude Duvanaud

XLIM Laboratory UMR-CNRS 7252, Institute of Technology of Angoulême, University of Poitiers, Angoulême, France

https://doi.org/10.26636/jtit.2021.144520

Abstract—Digital predistortion (DPD) using baseband signals is commonly used for power amplifier linearization. This paper is devoted to this subject and aims to reduce DPD complexity. In this study, we propose a structure that allows to decrease the number of DPD parameters by using multiple blocks, with each one of them dedicated to characterizing the non-linear behavior and/or memory effects. Such a structure is based on the feedback Wiener system, involving a FIR filter used as a feedback path to reproduce the PA inverse dynamics. A memory polynomial block (MP) is inserted as the final element to minimize the modeling errors. A relevant model identification method, based on an iterative algorithm, has been developed as well. The proposed architecture is used for the linearization of a commercial class-AB LDMOS RF PA by NXP Semiconductors, in wideband communication systems. Comparison of performance with the conventional generalized memory polynomial model (GMP) shows that the proposed model offers similar results, with its advantage consisting in the reduced number of parameters.

Keywords—digital predistortion, feedback Wiener model, GMP mode, parameter identification, power amplifier.

1. Introduction

The key challenge in the design of radio frequency (RF) power amplifiers (PA) is to achieve high efficiency characteristics by using transistors at their near-to-saturation point [1]. Under such operating conditions, PA non-linearities and memory effects create significant signal distortions in both time and frequency domains, such as, for instance, scattered constellations and asymmetries in spectral regrowth [2]–[5]. These effects are more pronounced in the case of high-power fluctuations in multi-band and multi-carrier signals [6], [7]. So, such a behavior degrades the transmitter's efficiency and decreases transmission quality. One solution relied upon to minimize these effects, while simultaneously respecting spectral masks and without compromising efficiency, is to apply linearization techniques.

Several linearization techniques have been developed to mitigate PA non-linearities at high levels, and consequently to improve PA linearity versus power efficiency trade-off. Predistortion methods have been proposed as a solution with high potential to overcome nonlinear effects [8]–[10]. These techniques aim to introduce inverse non-linearities that compensate the PA gain, as well as phase and memory effects distortions [11], [12]. Depending on the position of the predistorter and on the provided signals, three types of predistortion techniques may be distinguished: those applied in RF [11], in intermediate frequencies (IF) [13] and in baseband (BB) [14]. From all linearization techniques referred to above, baseband digital predistortion (DPD) receives the most attention. It is widely deployed in modern wireless systems, as it allows to achieve good linearization performance through the use of reduced sampling frequency, without additional RF elements, and is, therefore, more cost effective.

In DPD and due to the complexity of the PA behavior, non-linear mathematical functions are required to sufficiently describe the inverse of PA characteristics [15]. In the state-of-the-art, the commonly used models are derived from the Volterra series [16]–[18]. Among them, one may distinguish the memory polynomial (MP) model [19], [20], the generalized memory polynomial (GMP) model [21] or the non-linear auto-regressive moving average (NARMA) model [22]. Other models, such as the block-oriented non-linear system (for instance Hammerstein and Wiener) [23], vector-switched models [24], decomposed vector rotation models [25], and neural network models [14] are used as well.

The use of a large number of terms is suitable for making the DPD more accurate, but unfortunately, this comes at the cost of a complicated implementation and long lead times required to estimate the coefficients. In this study, we focus on reducing the number of the model's parameters and we propose to study and use the feedback Wiener (FW) model [23] as a predistorter. To generate PA memory effects, a filter block with time delays is used in its

feedback path. A low order MP model may also be cascaded with the FW block for modeling errors in wideband applications [26]. The discussed DPD function using the proposed structure, referred to, in this paper, the feedback Wiener with memory polynomial (FWMP), and its identification algorithm are presented and tested using a 2-stage 20 W class-AB LDMOS RF PA by NXP Semiconductors. Studies concerned with model complexity and focusing on optimizing the number of model coefficients and drawing comparisons with the performance of the GMP model show a good compromise between linearization accuracy and model complexity.

This paper is organized as follows. In Section 2, we introduce a new, less complex, block-oriented model based on the feedback Wiener system. The identification process of a predistorter using the proposed FWMP model is described in Section 3. Linearization performance experimental tests using the proposed structure and a comparison with the MP and GMP models are described in Section 4. Finally, conclusions and some perspectives are given in Section 5.

2. Block-oriented Model Description

A block-oriented model will be used in this study for the implementation of the predistorter (Fig. 1). It has been established by relying on the circuit-based approach, allowing to take into account the fundamental non-linear properties, memory effects and the bilateral behavior of the active devices [26].

As shown in Fig. 1, this structure is based on a combination of two blocks: a feedback Wiener system which models the main PA behavior, i.e. interaction between non-linearities and memory effects, and an MP model for the remaining modeling errors. The FW block itself is made up of two sub-blocks: a feed-forward memory-less non-linearity and a feedback finite impulse response (FIR) filter, where q^{-1} is the unit time delay.

The main signals of the FWMP model may be formulated as:

$x(n)$ is the output of the FW system, such as:

$$x(n) = \sum_{p=1}^{P} c_p \cdot w(n)^p = \sum_{p=1}^{P} c_p \cdot \Big[g_0 \cdot M_{in}(n) - d(n) \Big]^p , \quad (1)$$

where c_p are the non-linear terms of the non-linearity function, g_0 is the complex gain and P is the non-linearity order. M_{in} is the model input. $F(\omega)$ is a FIR filter and its output $d(n)$ may be formulated as:

$$d(n) = \sum_{m=1}^{M} b_m \cdot x(n-m) , \quad (2)$$

where M is the memory depth of the FIR filter. Signal $x(n)$ is used as an input for the MP model:

$$M_{out}(n) = \sum_{p=0}^{P_a-1} \sum_{m=0}^{M_a-1} a_{pm} \cdot x(n-m) \cdot |x(n-m)|^p , \quad (3)$$

where M_{out} is the model output, while P_a and M_a are the non-linearity order and memory depth, respectively.

Note that, since non-linearities and memory effects are treated separately in the FW block, the proposed model has the advantage of an additive evolution in its first block, meaning that after incrementing a parameter in the FW block, only a single increment in the number of model coefficients occurs. As a result, the total number of coefficients is given by:

$$N_{FWMP} = \overbrace{(P+M)}^{FW} + \overbrace{(P_a \times M_a)}^{MP} . \quad (4)$$

3. DPD using FWMP Model

The DPD is based on the estimation of the inverse PA characteristics to compensate its static and dynamic non-linearities. In the case of the proposed model, Fig. 2 shows the principle of the off-line DPD estimation process based

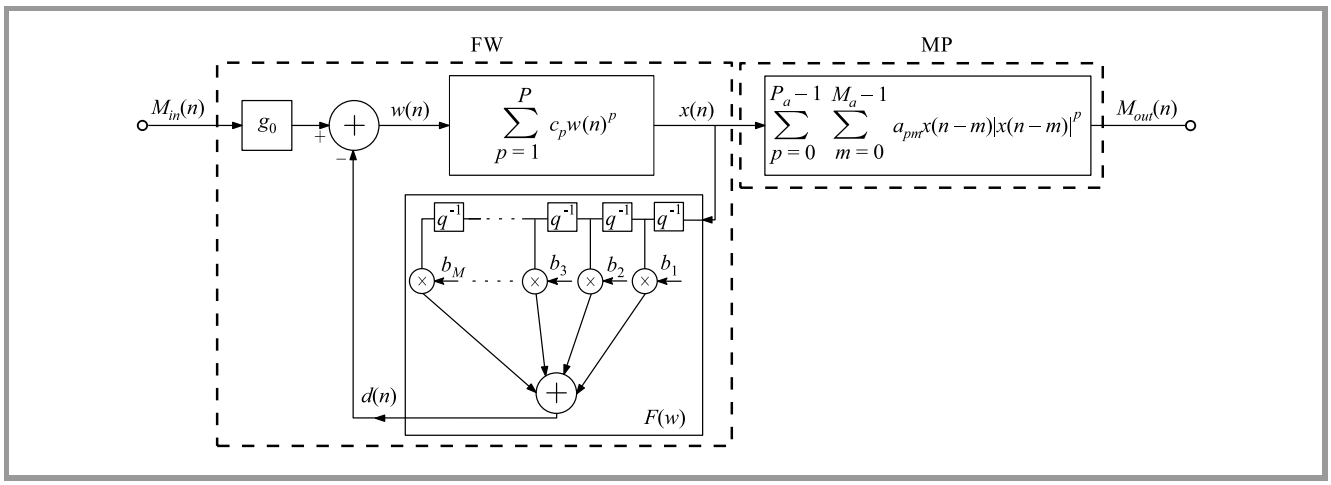


Fig. 1. Feedback Wiener memory polynomial model.

on the minimization of the quadratic criterion (cost function) according to the output errors – Eq. (5). Note that G is the PA linear gain used for output normalization.

$$J = \sum_{n=1}^{N} \varepsilon_i(n)^2 + \varepsilon_Q(n)^2 \,, \tag{5}$$

where N is the length of the signal used (the number of samples).

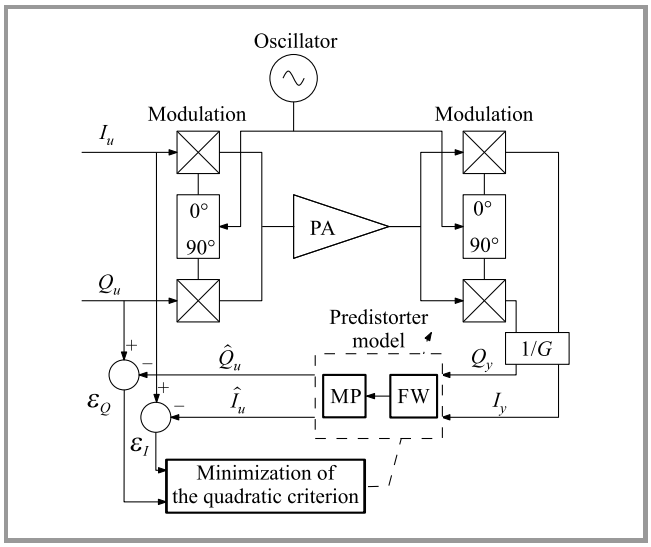


Fig. 2. Offline DPD identification.

As mentioned before, the FWMP model shown in Fig. 1 is composed of two separated blocks, so the identification of such a model class is complex, because of the intermediate unmeasured signals. In other words, the identification of one block requires the simulation of the previous block. In our case, the unmeasured signals are $w(n)$, $x(n)$ and $d(n)$. The key-term separation principle is a good solution for the identification of this model by separating the model into non-linear static and linear dynamic blocks [23]. So, the identification process will be performed in two phases, starting with identification of the FW block followed by the simulation of the intermediate signals. Then, characterization of the MP block will follow.

3.1. FW Block Identification

The use of a feedback loop in the FW block renders its one-step identification impossible. We propose, in Fig. 3, an

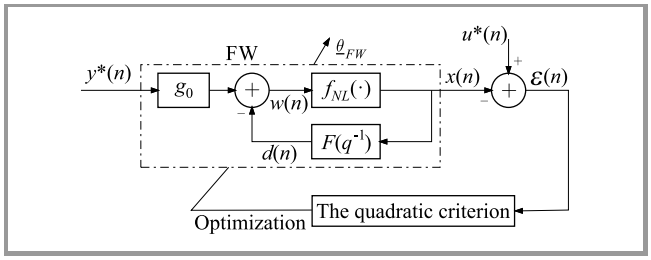


Fig. 3. Identification of the FW block.

iterative process to estimate it using the measured PA output and input complex envelope noted by $y^*(n)$ and $u^*(n)$, respectively. So, in the case of DPD, the measured output data y^* will become the input of the FW model ($M_{in} = y^*$), and the measured input data u^* will become its output. The FW vector of coefficients $\underline{\theta}_{FW}$ is estimated as:

$$\underline{\theta}_{FW} = [g_0 \quad b_1 \cdots b_M \quad c_1 c_2 \cdots c_P] \,. \tag{6}$$

The FW model for N samples, based on Eqs. (1) and (2), can be rewritten as:

$$x(n) = \varphi_{FW}^T(n, iter) \cdot \underline{\theta}_{FW} \,, \tag{7}$$

with:

$$\varphi_{FW}(n, iter) = \big[y^*(n) \quad -x(n-1) \\ \cdots -x(n-M) \quad w(n)\, w(n)^2 \cdots w(n)^P\big] \,.$$

During this iterative process, and in order to avoid the problem of overparametrization [23], we set the first coefficient of the non-linear function c_1 in Eq. (1) to 1. Note that during the first iteration of this process, the memory-less function is off. We choose to start the identification process with the estimation of the FIR filter, due to its stability. The iterative identification process of the FW block (Fig. 3) is:

1. Initialization of $w(n)$, $x(n)$ and the FW coefficients – Eq. (6), as:
   $w(n) = y^*(n)$,
   $x(n) = u^*(n)$,
   $\underline{\theta}_{FW} = [1 \quad 0 \cdots 0 \quad 1\, 0 \cdots 0]$;
2. Identification of the new feedback filter $F(q^{-1})$ coefficient b_m and the complex gain g_0:
   $\underline{\theta}_{FW}(1) = [g_0 \quad b_1 \cdots b_M]$,
   $\varphi_{FW}(n, 1) = [y^*(n) \quad -x(n-1) \cdots -x(n-M)]$,
   $\underline{\theta}_{FW}(1)$ can be obtained from:

$$\underline{\theta}_{FW}(1) = \left(\phi^H \cdot \phi\right)^{-1} \phi^H \cdot x \,, \tag{8}$$

where $\phi = [\varphi_{FW}^T(1,1)\ \varphi_{FW}^T(2,1)\ \cdots\ \varphi_{FW}^T(N,1)]$.

A QR decomposition function (qrd) is used to avoid matrix inversion problems in $\left(\phi^H \cdot \phi\right)$ as:

$$\mathrm{qrd}(\phi^H \cdot \phi) = Q \cdot R \,, \tag{9}$$

where Q is an orthogonal unit vectors and R is an upper triangular matrix. Equation (8) becomes:

$$\underline{\theta}_{FW}(1) = \left(R^{-1} \cdot Q^H\right) \phi^H \cdot u^* \,. \tag{10}$$

Based on Eqs. (1)–(2) and using the FW model obtained, we simulate the new intermediate signals noted as $d(n)$, $w(n)$ and $x(n)$;

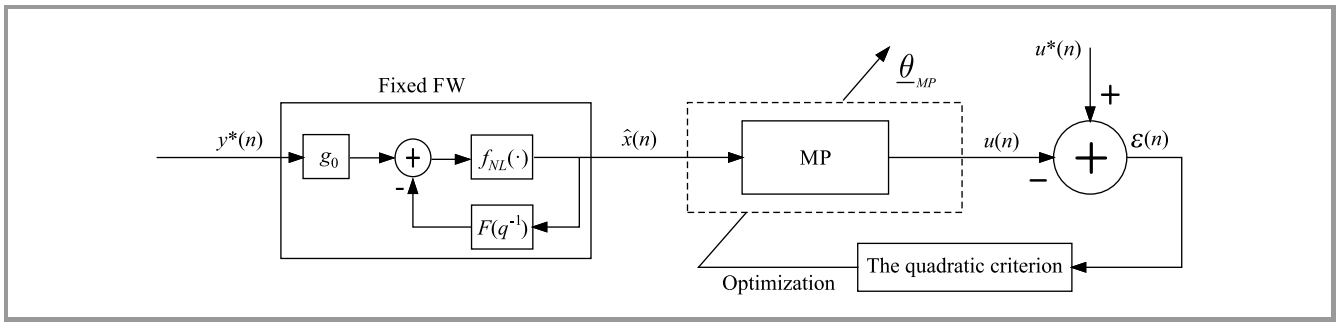


Fig. 4. Identification of the MP block.

3. Estimation of all FW block coefficients $\underline{\theta}_{FW}$ – Eq. (6) using previously simulated intermediate signals, based on Eqs. (7) and (10);

4. Minimization of the quadratic criterion $J = \sum_{n=1}^{N} \varepsilon^2(n)$;

5. Repetition of steps 3-4 until a desirable normalized mean square error (NMSE) is reached, or until the adding of another iteration does not reduce the cost function J, where:

$$\mathrm{NMSE}_{dB} = 10\log_{10}\left[\frac{\sum_{n=1}^{N}\left|u^*(n)-x(n)\right|^2}{\sum_{n=1}^{N}\left|u^*(n)\right|^2}\right]. \quad (11)$$

with $x(n)$ being the model output obtained at the last iteration.

3.2. MP Block Identification

After the convergence of the FW model, it will be fixed, as shown in Fig. 4, and the simulated output signal $x(n)$ will be used as an input of the MP block.
The MP model coefficients a_{pm} can be obtained using the least squares (LS) algorithm [27] and the system regression derived from relation (3). Thus, for a set of N samples, the optimal coefficients vector $\underline{\theta}_{MP}$ is obtained by solving the following optimization problem:

$$\min_{\underline{\theta}_{MP}}(J) \quad \text{where} \quad J = \frac{1}{N}\sum_{n=1}^{N}\left|u^*(n)-u(n)\right|^2. \quad (12)$$

$u(n) = \varphi_{MP}^T(n,\underline{\theta}_{MP})\cdot\underline{\theta}_{MP}$ is the model output with:

$$\varphi_{MP}(n,\underline{\theta}_{MP}) = \Big[x(n)\ \cdots\ x(n-m)|x(n-m)|^p\ \cdots\ x(n-M_a)|x(n-M_a)|^{P_a}\Big]$$

$$\underline{\theta}_{MP} = \Big[a_{00}\cdots a_{pm}\cdots a_{(P_a-1)(M_a-1)}\Big].$$

The offline estimation of the $\underline{\theta}_{MP}$ vector is:

$$\underline{\theta}_{MP} = \left(\phi^H\cdot\phi\right)^{-1}\phi^H\cdot u^*, \quad (13)$$

where $\phi = \Big[\varphi_{MP}^T(1,\underline{\theta}_{MP})\,\varphi_{MP}^T(2,\underline{\theta}_{MP})\cdots\varphi_{MP}^T(N,\underline{\theta}_{MP})\Big]$.

4. Experiments and Results

4.1. Experimental Setup and Signal Acquisition

In this section, we present the experimental validation of the proposed FWMP model and the comparison of its linearization performance with that of a GMP model. The test bench used is shown in Fig. 5. A 2-stages 20 W class-AB LDMOS RF power amplifier by NXP Semiconductors has been used to validate the proposed model. It has a linear gain of 28 dB and its 1 dB compression point is around 41.7 dBm, corresponding to an output power back-off (OBO) of 0 dB. A vector signal generator (SMBV100A by Rohde & Schwarz) is used for up-converting the base-band signal that was generated beforehand using Matlab

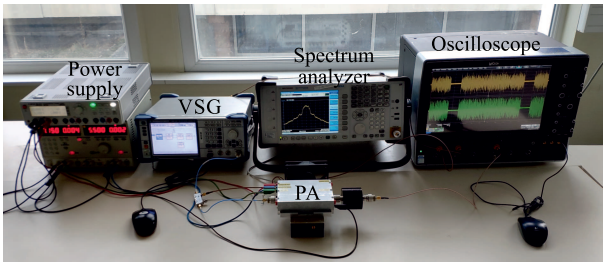


Fig. 5. Experimental setup.

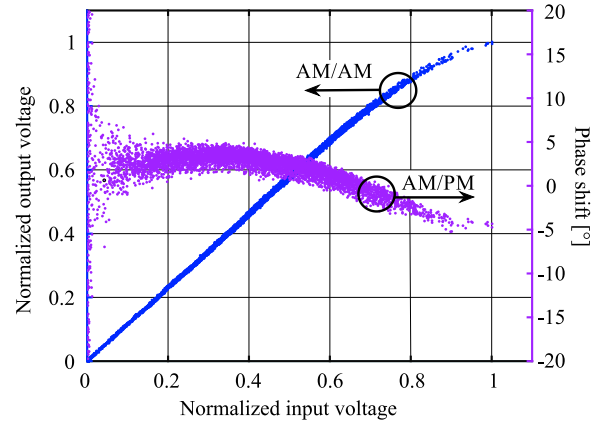


Fig. 6. Instantaneous AM/AM and AM/PM characteristics for a 64-QAM input signal with 7.5 dB PAPR.

and uploaded to the VSG, with a carrier frequency of 3.7 GHz. The sequence used is a filtered 10 MHz 64-QAM modulated signal with a peak-to-average power ratio (PAPR) of approximately 7.5 dB.

At the PA output, the RF signal was acquired at a sampling frequency of 40 GHz and then numerically down-converted and demodulated using a 4-channels oscilloscope (LeCroy WaveMaster 816Zi-A). The input and output baseband signals are synchronized in the time domain using Matlab.

Non-linearities and memory effects of the used PA may be observed from the dynamical AM/AM and AM/PM functions in Fig. 6. We can see that the gain is compressed by PA when the input level increases.

4.2. Experimental Results

In our study, the merit value refers to the NMSE criterion given by Eq. (11) which translates the modeling accuracy. So, to determine the DPD structure using the FWMP model in terms of the trade-off between performances and complexity, an exhaustive search is performed in Fig. 7.

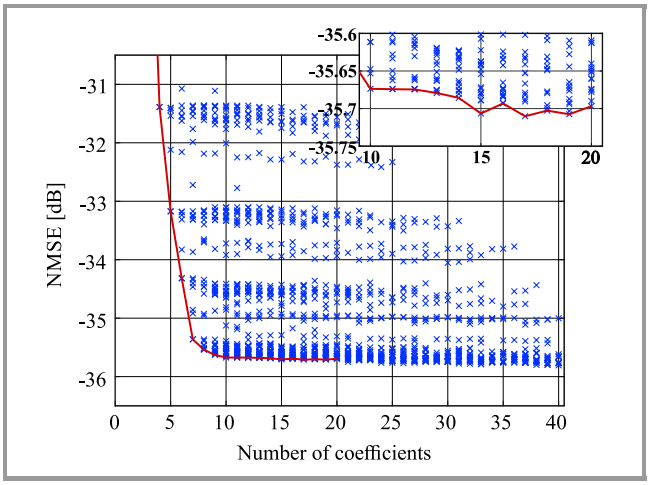


Fig. 7. Determination of the most relevant FWMP model structure.

We mapped NMSE in the time-domain versus the number of coefficients by testing all possible structures. A single set of input and output signals measured is used in this mapping and is the same during the process of identifying all structures. Each point in the map corresponds to an NMSE value using a set of FWMP parameters P, M, P_a and M_a – see Eqs. (1)–(3). Here, 1715 combinations were tested with P = 1:7, M = 0:4, P_a = 1:7, and M_a = 1:7. During the optimization phase, we determined that it takes less than 10 iterations for the model to converge to the lowest NMSE, so for each structure, we iterated the model 10 times to obtain the final coefficients.

As shown in this NMSE map, an increase in the number of coefficients allows to improve estimation performance. It may also be noticed that for the same number of coefficients, several values of NMSE may be identified. For example, for structures with 10 coefficients, the lowest NMSE of approx. -35.6 dB, is obtained with the model orders $(P = 7, M = 2, P_a = 1, M_a = 1)$, while the worst result of approx. -22 dB is obtained for $(P = 1, M = 2, P_a = 1, M_a = 7)$. These results show the importance of an offline DPD evaluation determining the best FWMP structure, i.e. using a minimum number of coefficients for a given NMSE requirement.

Figure 8 shows the parameter orders of the FWMP models, with different numbers of coefficients (up to 20 coefficients), which ensure the best performance in terms of NMSE.

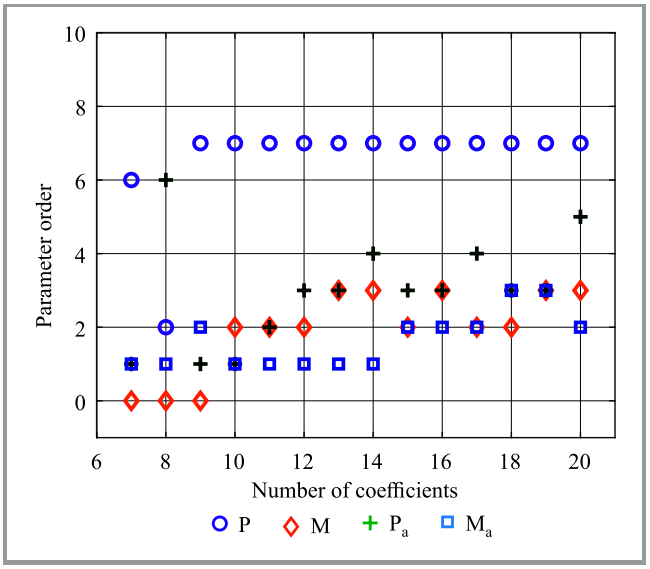


Fig. 8. Composition of the best structures obtained with different numbers of coefficients.

We can see from Fig. 8 that the FW block is the one that contributes the most to the description of the DPD function. An FW non-linearity order P of 7 is sufficient to describe the non-linear behavior of the PA used. Also, for structures with up to 10 coefficients, we realize that it is more relevant to use only the FW block, while starting from 11 coefficients, the deployment of the MP block provides better performance.

Table 1
Comparison of complexity (number of FLOP) and performance of different model structures

	P	M	NMSE [dB]	No. of FLOPs	
				Add.	Multipl.
NL	7	0	–35.53	6	28
	8	0	–35.57	7	36
	9	0	–35.64	8	45
	10	0	–35.64	9	55
	11	0	Unstable	10	66
FW	7	2	–35.67	8	31

To show the importance of using the feedback loop (FIR filter), in Table 1 a comparison between complexity (num-

ber of FLOPs) and performance of the individual cases is presented, in which:

- only the memoryless polynomial is used,
- only the FW is used.

A floating point operation (FLOPs) describes the arithmetic operation on floating point numbers.
We can see from Table 1 that the best NMSE with only the NL function, close to the one obtained with the complete FW block, is obtained with a non-linearity order of 10, but at the cost of a significant increase in the number of FLOPs. Moreover, the NL function becomes unstable starting from the non-linearity order of 11. It also needs to be noted that the introduction of the feedback filter helps improve the NMSE, with only a slight increase in the number of FLOPs.
By tracking the evolution of the lowest NMSE in each column (red line in Fig. 7), we can note that there is no significant enhancement of the NMSE after 10 coefficients. Zooming on the area between 10 and 20 coefficients enables a precise measuring, which helped us choose a structure with 15 coefficients as a point of reference for our study. This structure, obtained with ($P = 7,\ M = 2,\ P_a = 3,\ M_a = 2$), offers a good trade-off between modeling accuracy and model complexity.

4.3. Comparison of the FWMP Structure with the MP and GMP Models

To show the importance of the FW block in the proposed cascaded model, we performed a set of experiments - both with and without the FW block. Thus, partial mapping was performed using the MP model only (Eq. (3)) with 10 to 20 coefficients (see red triangles in Fig. 9).

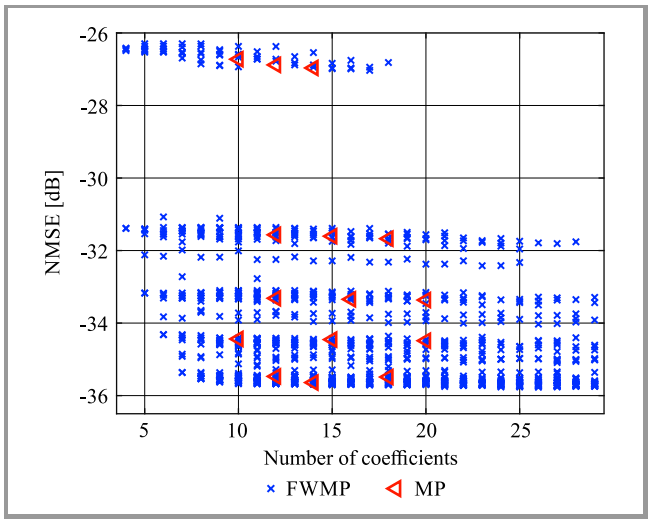


Fig. 9. Comparison of FWMP and MP models.

As one may notice in Fig. 9, introduction of the FW block in the general case allows to reach a lower NMSE. So, whatever the number of coefficients is, the FWMP model allows to improve NMSE. For example, in the reference case of a model with 15 coefficients, the FW introduces additional dynamics that improve the NMSE by approx. 1 dB.
Below, we compare FWMP with a GMP model [21].
As a reminder, the GMP model is given by the following relation:

$$\begin{aligned} M_{out}(n) = &\sum_{p=0}^{P_a-1}\sum_{m=0}^{M_a-1} a_{pm}\cdot M_{in}(n-m)\cdot|M_{in}(n-m)|^p \\ &+\sum_{p=1}^{P_b}\sum_{m=0}^{M_b-1}\sum_{l=1}^{L_b} b_{pml}\cdot M_{in}(n-m)\cdot|M_{in}(n-m-l)|^p \\ &+\sum_{p=1}^{P_c}\sum_{m=0}^{M_c-1}\sum_{l=1}^{L_c} c_{pml}\cdot M_{in}(n-m)\cdot|M_{in}(n-m+l)|^p, \end{aligned} \quad (14)$$

and its number of coefficients N_{GMP} may be obtained by:

$$N_{GMP} = P_a\cdot M_a + P_b\cdot M_b\cdot L_b + P_c\cdot M_c\cdot L_c\,. \quad (15)$$

In order to compare FWMP with the GMP model, we deployed them under the same conditions, and plotted the obtained NMSE with structures comprising from 1 to 50 coefficients. The purple dots show the results for the GMP model (Fig. 10). In this case, 345945 combinations were tested with P_a = 1:7, M_a = 1:4, P_b = 1:7, M_b = 1:7, L_b = 1:5, P_c = 1:7, M_c = 1:7, L_c = 1:5.

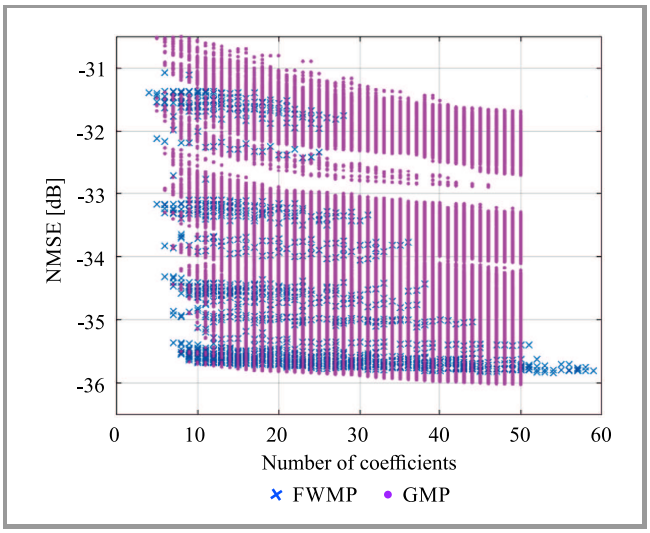


Fig. 10. Comparison of FWMP and GMP models.

The first remark concerns the greater number of potential structures to be tested, observed in the case of the GMP model. In fact, and as expressed in Eq. (15), there are 8 sizing parameters applicable to the GMP model: non-linearity orders (P_a, P_b and P_c), memory depths (M_a, M_b, and M_c), and lagging and leading delay tap lengths (L_b and L_c). That is the major drawback of the GMP model, where the number of combinations increases rapidly along with the model orders. In the case of FWMP from Eqs. (1)–(3), we reduce the number of sizing parameters to 4: non-linearity orders (P and P_a), filter order M and memory depth M_a.

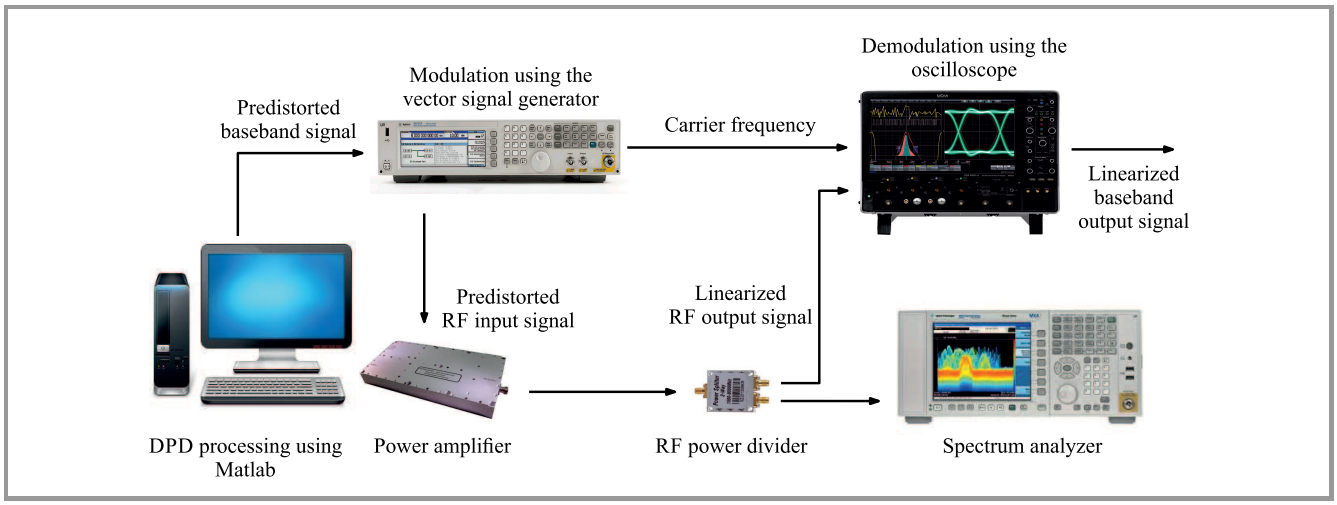


Fig. 11. Linearization system scheme.

The second remark concerns the performance of both models. As can be concluded from Fig. 10, for structures with up to 27 coefficients, both models offer almost the same performance, with the FWMP model having the advantage of low complexity. Beyond that, the GMP model shows a slight improvement, for example of 0.23 dB in the case of a structure with 50 coefficients.

Based on these comparisons, the best GMP and FWMP structures with 15 coefficients, allowing the lowest NMSE, have been extracted and used as a DPD function to linearize the PA behavior. The number of arithmetic operations of each model is presented in Table 2.

Table 2
Comparison in terms of complexity (number of FLOPs) of the two used models

Model	Model structure	FLOPs per block		Total	
		Add.	Multipl.	Add.	Multipl.
FWMP	$P = 7, M = 2$	8	31	13	43
	$P_a = 3, M_a = 2$	5	12		
GMP	$P_a = 7, M_a = 1$	6	28	12+2	48
	$P_b = 2, M_b = 2, L_b = 1$	3	10		
	$P_c = 2, M_c = 1, L_c = 2$	3	10		

As shown in Table 2, and in comparison with the GMP model, the proposed FWMP model is characterized by a lower number of arithmetic operations.

Figure 11 illustrates the experimental process in which the input signal is predistorted using Matlab software and then uploaded to VSG, which provides the RF predistorted signal to be injected to the PA. Both time and frequency experiments are performed for FWMP and GMP models under the same conditions. Review of these results allows to determine the contribution of the proposed model.

In the time-domain, the linearized AM/AM and AM/PM characteristics of the LDMOS PA used, obtained using the two models, namely FWMP and GMP, are shown in Figs. 12 and 13, respectively. We can see that the FWMP model offer good linearization performance, similar to that of the GMP structure. No significant differences between the studied models are noticed and the obtained results show the robustness and the effectiveness of both solutions.

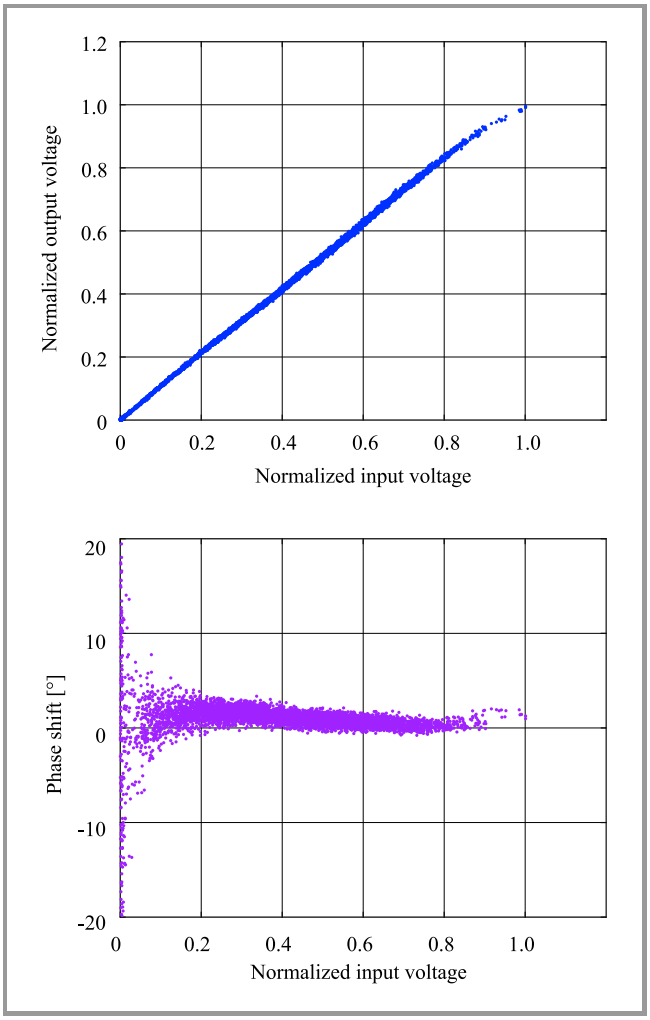


Fig. 12. Linearized AM/AM and AM/PM characteristics using FWMP model.

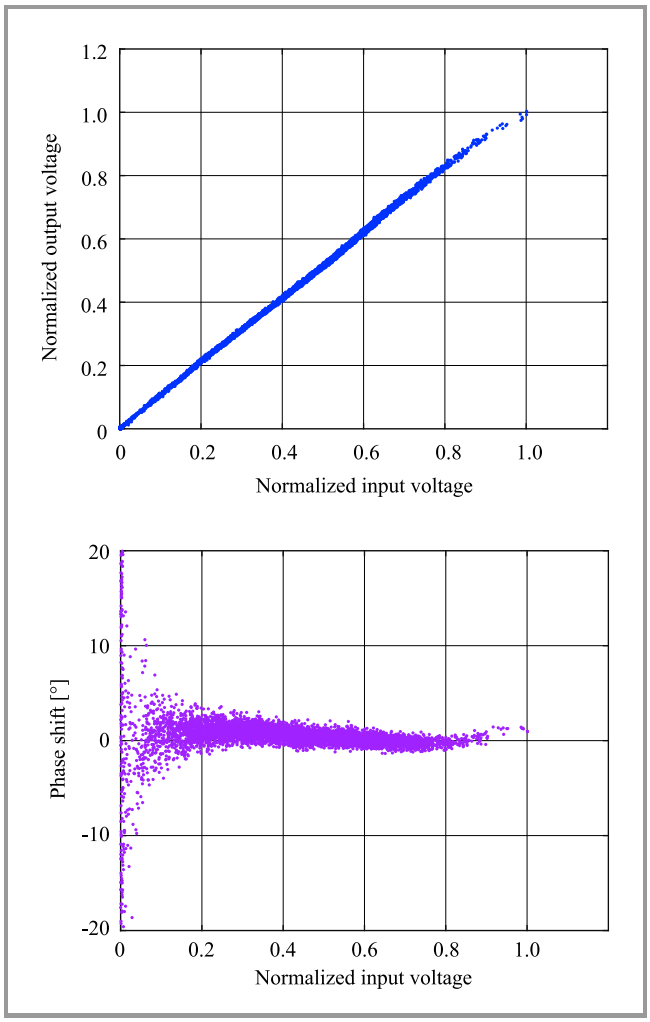


Fig. 13. Linearized AM/AM and AM/PM characteristics using GMP model.

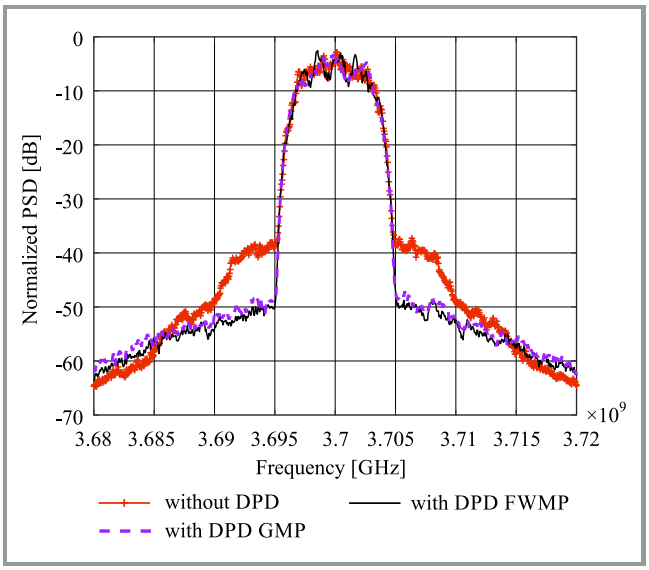


Fig. 14. Output spectra measured using FWMP and GMP linearizer.

Table 3
Comparison of ACPR achieved

ΔW [MHz]	ACPR [dB]	Original	DPD GMP	DPD FWMP
5	Lower	–32.82	–42.37	–43.77
5	Upper	–32.59	–41.66	–42.84
10	Lower	–35.52	–43.84	–45.16
10	Upper	–35.31	–43.45	–44.56

Results shown in Table 3 confirm those obtained in Fig. 14, where the performance of GMP and FWMP was similar, and allow for a 10 dB ACPR improvement compared with the original signal. These results confirm the contribution of the FWMP structure in sizing the optimal linearizer.

4.4. Results in Frequency Domain

It is known that the PA non-linearities create spectral regrowth distortions and adjacent channel noise [28]. To determine the performance of FWMP in the frequency domain, in Fig. 14, we present the measured PA output spectra, both with and without DPD, using the two models. According to the obtained spectra, we may see in Fig. 14 that the linearization performance of both models is similar and that the suppression of sideband noises caused by PA non-linearities and memory effects is effective.

Table 3 presents a comparison between the adjacent channel power ratio (ACPR) obtained using the two models and different adjacent channel bandwidths $\Delta W_{L/U}$ (5 and 10 MHz). As a reminder, ACPR at the PA output is based on the discrete Fourier transform $Y(\omega)$ of $y(n)$ and is used to evaluate out-of-band distortion, provided that L and U are the lower and the upper adjacent channel frequencies, respectively. M is the main channel frequency.

$$ACPR_{dBc} = \frac{\int_{\Delta W_M} Y(\omega)\mathrm{d}\omega}{\int_{\Delta W_{L/U}} Y(\omega)\mathrm{d}\omega}, \qquad (16)$$

5. Conclusion

Power amplifier non-linearities and memory effects have been discussed in this paper, and a new cascaded structure has been proposed as a low complexity linearizer. This model, based on the feedback Wiener system, ensures the correction and an overall improvement of the spectral distortions over a wide range of frequency bands.

The measurement system and the identification method were presented as well. Experimental results obtained for a commercial class-AB LDMOS PA by NXP Semiconductors demonstrated that the performance of the FWMP structure is better than that of the MP model and similar to that of the GMP model. Spectral analysis also shows an improvement in the out-of-band emission by up to 10 dB of ACPR, which may increase the power efficiency of the transmitter. The proposed model contributes to reducing the number of parameters, which is a considerable gain in term of the number of combinations to be tested during the model identification process.

We can finally conclude that the FWMP model may outperform other DPD structures in reducing spectral regrowth

or ACPR, without increasing model complexity. However, since the proposed model is a 2-stage cascaded DPD, identification of its coefficients is more complicated. To deal with this drawback, a new way to identify its complex coefficients, based on iterative estimation, is proposed. Future work will focus on using the DPD FWMP model in on-line identification for a reconfigurable PA, where different learning architectures may be used for stage-by-stage identification.

Acknowledgement

Research activities described in this publication are part of the APOGEES project supported by BPI France, region Occitanie and region Nouvelle Aquitaine. The APOGEES project has been labeled by pole Aerospace Valley, pole I&R and pole Alpha-RLH in the framework of the French FUI22 research program.

References

[1] P. Reynaert and M. Steyaert, "Mobile communication systems and power amplification", in *RF Power Amplifiers for Mobile Communications*. Springer Science & Business Media, 2006, pp. 9–64 (ISBN: 9781402051166).

[2] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers", *IEEE Trans. on Commun.*, vol. 29, no. 11, pp. 1715–1720, 1981 (DOI: 10.1109/TCOM.1981.1094911).

[3] J. Vuolevi, T. Rahkonen, and J. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 49, no. 8, pp. 1383–1389, 2001 (DOI: 10.1109/22.939917).

[4] H. Ku, M. D. Mckinley, and J. Kenney, "Quantifying memory effects in RF power amplifiers", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 50, no. 12, pp. 2843–2849, 2002 (DOI: 10.1109/TMTT.2002.805196).

[5] E. Ngoya, C. Quindroit, and J. M. Nebus, "On the continuous-time model for nonlinear memory modeling of RF power amplifiers", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 57, no. 12, pp. 3278–3292, 2009 (DOI: 10.1109/TMTT.2009.2033297).

[6] C. S. Aitchison, M. Mbabele, M. R. Moazzam, D. Budimir, and F. Ali, "Improvement of third-order intermodulation product of RF and microwave amplifiers by injection", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 49, pp. 1148–1154, 2001 (DOI: 10.1109/22.925508).

[7] Y. Aimer, B. S. Bouazza, S. Bachir, and C. Duvanaud, "Interleaving technique implementation to reduce PAPR of OFDM signal in presence of non-linear amplification with memory effects", *J. of Telecommun. and Inform. Technol.*, no. 3, 2018 (DOI: 10.26636/jtit.2018.123517).

[8] M. Vaskovic, "Compensation of nonlinear distortion in RF amplifiers for mobile communications", Ph.D. thesis, University of Westminster, London, England, 2014 [Online]. Available: https://westminsterresearch.westminster.ac.uk/item/8yvxv/compensation-of-nonlinear-distortion-in-rf-amplifiers-for-mobile-communications

[9] S. Bachir, C. E. Nicusor, and C. Duvanaud, "Linearization of RF power amplifiers using adaptive Kalman filtering algorithm", *J. of Circ., Syst., and Computers*, vol. 20, no. 6, pp. 1001–1018, 2011 (DOI: 10.1142/S0218126611007724).

[10] C. Nader, P. N. Landin, W. Van Moer, N. Bjorsell, and P. Handel, "Performance evaluation of peak-to-average power ratio reduction and digital pre-distortion for OFDM based systems", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 59, no. 12, pp. 3504–3511, 2011 (DOI: 10.1109/TMTT.2011.2170583).

[11] I. Teikari, "Digital predistortion linearization methods for RF power amplifiers", Ph.D. thesis, Aalto University of Technology, Espoo, Finland, 2008 [Online]. Available: http://lib.tkk.fi/Diss/2008/isbn9789512295463/isbn9789512295463.pdf

[12] M. A. Hussein, Y. Wang, G. Peyresoubes, B. Feuvrie, and S. Toutain, "LUT/parametric digital predistortion approach for the linearization of power amplifiers characteristics", in *Proc. 38th Eur. Microwave Conf. EuMC 2008*, Amsterdam, Netherlands, 2008, pp. 571–574 (DOI: 10.1109/EUMC.2008.4751516).

[13] P. Banelli and G. Baruffa, "Mixed BB-IF predistortion of OFDM signals in non-linear channels", *IEEE Trans. on Broadcast.*, vol. 47, pp. 137–146, 2001 (DOI: 10.1109/11.948266).

[14] X. Feng, "Efficient baseband digital predistortion techniques for linearizing power amplifier by taking into account nonlinear memory effect", Ph.D. thesis, University of Nantes, Nantes, France, 2015 [Online]. Available: https://hal.archives-ouvertes.fr/tel-01206266

[15] L. Ding, "Digital predistortion of power amplifiers for wireless applications", Ph.D. Thesis, Georgia Institute of Technology, Georgia, USA, 2004 [Online]. Available: https://smartech.gatech.edu/bitstream/handle/1853/5184/ding_lei_200405_phd.pdf

[16] V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations*". London: Blackie & Son Ltd, 1930.

[17] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited Volterra series-based digital predistortion for wideband RF power amplifiers", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 60, pp. 4198–4208, 2012 (DOI: 10.1109/TMTT.2012.2222658).

[18] H. E. Hamoud, T. Reveyrand, S. Mons, and E. Ngoya, "A comparative overview of digital predistortion behavioral modeling for multi-standards applications", in *Proc. Int. Worksh. on Integr. Nonlin. Microw. and Millim.-wave Circ. INMMIC 2018*, Brive La Gaillarde, France, 2018 (DOI: 10.1109/INMMIC.2018.8430010).

[19] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory", *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, 2001 (DOI: 10.1049/el:20010940).

[20] S. Amin, P. Landin, P. Händel, and D. Rönnow, "2D extended envelope memory polynomial model for concurrent dual-band RF transmitters", *Int. J. of Microw. and Wirel. Technol.*, vol. 9, no. 8, pp. 1619–1627, 2017 (DOI: 10.1017/S1759078717000277).

[21] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers", *IEEE Trans. on Sig. Process.*, vol. 54, no. 10, pp. 3852–3860, 2006 (DOI: 10.1109/TSP.2006.879264).

[22] P. L. Gilabert, A. Cesari, G. Montoro, E. Bertran, and J. Dilhac, "Multi-lookup table FPGA implementation of an adaptive digital predistorter for linearizing RF power amplifiers with memory effects", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 56, no. 2, pp. 372–384, 2008 (DOI: 10.1109/TMTT.2007.913369).

[23] F. Guo, "A new identification method for Wiener and Hammerstein systems", Ph.D. thesis, University of Karlsruhe, Germany, 2003 [Online]. Available: https://digbib.ubka.uni-karlsruhe.de/volltexte/fzk/6955/6955.pdf

[24] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 60, no. 4, pp. 1166–1174, 2012 (DOI: 10.1109/TMTT.2012.2184295).

[25] A. Zhu, "Decomposed vector rotation-based behavioral modeling for digital predistortion of RF power amplifiers", *IEEE Trans. on Microw. Theory and Techniq.*, vol. 63, no. 2, pp. 737–744, 2015 (DOI: 10.1109/TMTT.2014.2387853).

[26] T. H. C. Bouazza, S. Bachir, and C. Duvanaud, "Behavioral blocks model for complexity-reduced modeling of RF power amplifiers", in *Proc. of the IEEE Int. Symp. on Circ. and Syst. ISCAS 2019*, Sapporo, Japan, 2019 (DOI: 10.1109/ISCAS.2019.8702517).

[27] B. Friedlander and M. Morf, "Least squares algorithms for adaptive linear-phase filtering", *IEEE Trans. on Acoust., Speech, and Sig. Process.*, vol. 30, no. 3, pp. 381–390, 1982 (DOI: 10.1109/TASSP.1982.1163903).

[28] A. Cheaito, "Analytical analysis of in-band and out-of-band distortions for multicarrier signals: Impact of non-linear amplification, memory effects and predistortion", Ph.D. thesis, INSA Rennes, University of Bretagne Loire, Rennes, France, 2017 [Online]. Available: https://www.theses.fr/2017ISAR0001.pdf

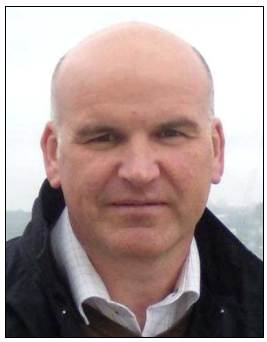
Tayeb H. C. Bouazza received his B.Sc. and M.Sc. degrees in Electronics and Telecommunications from the University of Saïda, Algeria in 2011 and 2013, respectively. Since 2017, he has been a Ph.D. student focusing on electronics, microelectronics, nanoelectronics and microwaves at the XLIM laboratory, Department of Smart Networks and Systems, University of Poitiers, France. His current research interests are in modeling non-linear systems, signal processing and wireless communication.

https://orcid.org/0000-0003-3873-0087

E-mail: tayeb.habib.chawki.bouazza@univ-poitiers.fr
XLIM Laboratory UMR-CNRS 7252
Institute of Technology of Angoulême
University of Poitiers
4 avenue de Varsovie
16000 Angoulême, France

Smail Bachir received B.Sc. and M.Sc. degrees in Signal Theory from the Polytechnic School of Algeria in 1997. He joined the scientific department of Leroy Somer Society and the University of Poitiers in France, where he received his Ph.D. degree in Automatic and Electrical Engineering in 2002 and the habilitation degree (HDR) in 2015. He is presently an Associate Professor at the University of Poitiers and a researcher at XLIM laboratory with the Department of Smart Networks and Systems. His research interests include signal processing, nonlinear systems parameter identification, electronic devices and wireless circuits.

E-mail: smail.bachir@univ-poitiers.fr
XLIM Laboratory UMR-CNRS 7252
Institute of Technology of Angoulême
University of Poitiers
4 avenue de Varsovie
16000 Angoulême, France



Claude Duvanaud received his Ph.D. in Electronics and Communication Engineering from the University of Limoges, France, in 1993 and the habilitation degree (HDR) from the University of Poitiers in 2003. Currently, he is an Associate Professor at the University of Poitiers and XLIM laboratory, France. His research interests include modeling, simulation and design of non-linear power amplifiers and communication systems.

E-mail: claude.duvanaud@univ-poitiers.fr
XLIM Laboratory UMR-CNRS 7252
Institute of Technology of Angoulême
University of Poitiers
4 avenue de Varsovie
16000 Angoulême, France

Paper

Performance Analysis of Filtered OFDM Based Downlink and Uplink NOMA System over Nakagami-*m* Fading Channel

Shaika Mukhtar and Gh. Rasool Begh

National Institute of Technology (NIT) Srinagar, Jammu and Kashmir, India

https://doi.org/10.26636/jtit.2021.148020

Abstract—Efficient consumption of available resources and fulfillment of increasing demands are the two main challenges which are addressed by exploring advanced multiple access schemes along with efficient modulation techniques. To this end, non-orthogonal multiple access (NOMA) is discussed as a promising scheme for future 5G traffic. NOMA enables the users to share same resource block, permitting certain level of interference. In this paper, we propose filtered OFDM (F-OFDM) as a transmission waveform for NOMA systems, as it offers all the advantages of OFDM with the additional provision of sub-band filtering to satisfy the diverse services of the users. We examine F-OFDM in both downlink and uplink NOMA systems. Error-related performances of both downlink and uplink F-OFDM NOMA systems are analyzed and compared with conventional OFDM NOMA system over Nakagami-*m* fading channel. The results show that the error performance of F-OFDM NOMA is better than that of OFDM NOMA. An improvement of about 2 dB and 1 dB in bit error rate is achieved in downlink and uplink F-OFDM NOMA, respectively. Monte Carlo simulations are conducted for different values of fading parameter *m*, supporting the obtained analytical results.

Keywords—bit error rate, orthogonal multiple access, out-of-band emission, successive interference cancellation.

1. Introduction

The 5G era is envisioned to consist of huge number of users, requiring low latency and high speed connectivity. To support such a scenario, there has been an evolution in the manner of sharing limited resources among the users [1]. The traditional orthogonal multiple access (OMA) schemes, such as TDMA, ensure absolute orthogonality among the users, leading to negligible inter-user interference [2]. This level of non-interference helps in easy extraction of message signals at the receiver. However, these OMA schemes are not sufficiently capable of supporting large number of users because of the limited availability of resources. To overcome this hurdle, the idea of non-orthogonal multiple access has been introduced [3], [4], wherein users share their resources. In NOMA, freedom degrees (time, frequency, code) are exploited in a non-orthogonal manner, using superposition coding that leads to a certain level of interference among the users. To overcome the effect of this interference, successive interference cancellation (SIC) is used at the receiver. Non-orthogonal resource sharing in NOMA offers diverse advantages, for example improved spectral efficiency and throughput, unbiased fairness among the users, reduction in user scheduling, and reduction in latency [3]. These useful features of NOMA motivate the incorporation of NOMA in future communication systems. To improve the performance of NOMA, the choice of waveform is a critical issue. Orthogonal frequency division multiplexing is one of the well-known waveform designs for NOMA. However, OFDM suffers from high peak to average power ratio (PAPR) and out-of-band emission (OOBE) [6], [7]. Larger out-of-band emissions occur due to disturbing transitions from one block to another. Such an undesirable feature causes interference to adjacent channel users [8], [9]. One simple way of reducing OOBE in OFDM is to shape the transmitting signal properly, in order to achieve better spectral confinement.

In this paper, we exploit the idea of the sub-band filtering of OFDM waveform, leading to the generation of an F-OFDM waveform. In a multiuser scenario, such sub-band splitting and filtering is beneficial, as optimized numerology may be used to satisfy service diversity-related requirements in 5G networks [9], [10]. In addition to the advantages of the foundational OFDM waveform, F-OFDM alleviates the use of the guard band, providing better spectral efficiency. It is also MIMO-friendly, ensuring higher compatibility with other technologies as well [11]. Based on all these desirable features, F-OFDM evolves as a suitable choice for NOMA systems.

2. Literature Review

Over the past few years, NOMA has been extensively studied worldwide. In [5], a unified network for NOMA

transmission has been described and its standardization has been presented. Considering the choice of waveforms for NOMA, the OFDM NOMA model is proposed in [6], with the different issues affecting this approach addressed. In [9], the authors compared each of the available 5G waveforms with F-OFDM and concluded that F-OFDM is a promising waveform for 5G. As far as service diversity is concerned, [10] analyzes OFDM-based waveforms in the presence of mixed numerologies, and concludes that F-OFDM shows better spectral performance and robustness than other OFDM-based waveforms. In [12], various pulse shaping filter designs for filtered OFDM are discussed, wherein semiequiripple filter, equiripple filter and windowed sinc filters are analyzed.

In [13], the authors have proposed generalized fast convolution-based filtered OFDM, wherein highly selective sub-band filtering is performed. Being simpler in concept, F-OFDM is gaining attention in low latency based 5G communication, as discussed in [14], where the authors analyzed a polar coded filtered OFDM system. In [15], SINR analysis of OFDM and F-OFDM for machine type communications are researched, with the authors observing that F-OFDM has a potential to mitigate undesirable distortions leading to better SINR results. For asynchronous uplink 5G communications, the usefulness of F-OFDM is considered in [16] and, accordingly, closed form expressions are derived for intercarrier interference (ICI).

3. Research Contribution

Based on the available literature regarding NOMA, we observe that the incorporation of F-OFDM in NOMA system has not been discussed yet. Considering this research gap, we propose to explore the utility of F-OFDM in the NOMA system.

In this paper, OFDM NOMA signals are spectrally shaped by using a sinc filter and Hann windowing. Actually, the impulse response of the sinc filter stretches to infinity on both ends. To exploit the advantages of the sinc filter used as a pulse shaping filter, truncation of impulse response is carried out using window functions [9]. The basic idea behind windowing is to obtain the product of the impulse response of filter $h(t)$ and window function $w(t)$:

$$H_w(t) = h(t).w(t) \ . \tag{1}$$

This soft truncation leads to better impulse response of the pulse shaping filter, ensuring lower inter symbol interference (ISI). These specifically designed filters are efficient enough to balance between frequency and time localization [17]. The effect of applying pulse shaping and windowing is shown in Fig. 1, where the power spectral densities of OFDM and F-OFDM are compared. It is evident that the filter has shaped a basic OFDM signal so that the gain of 80 dB is achieved in side lobe attenuation. This improvement of spectral characteristics is reflected in the constellation diagram of the F-OFDM NOMA signal, as shown in Fig. 2. Upon comparing the constellation diagram of the F-OFDM NOMA signal with the constellation

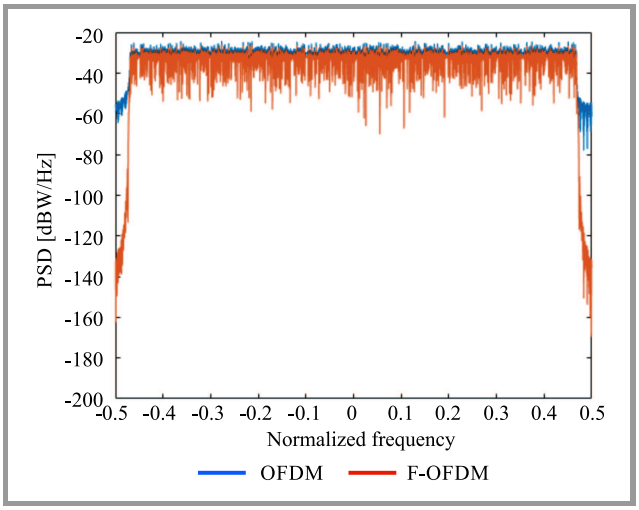


Fig. 1. Power spectral density comparison of OFDM and F-OFDM.

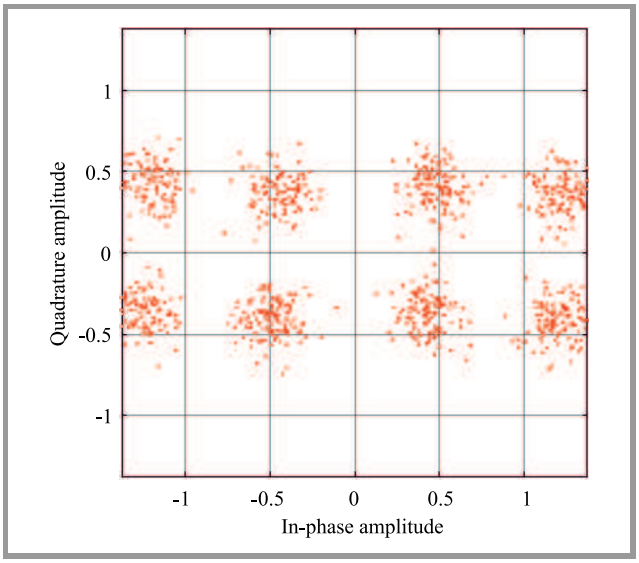


Fig. 2. Constellation diagram of F-OFDM NOMA signal.

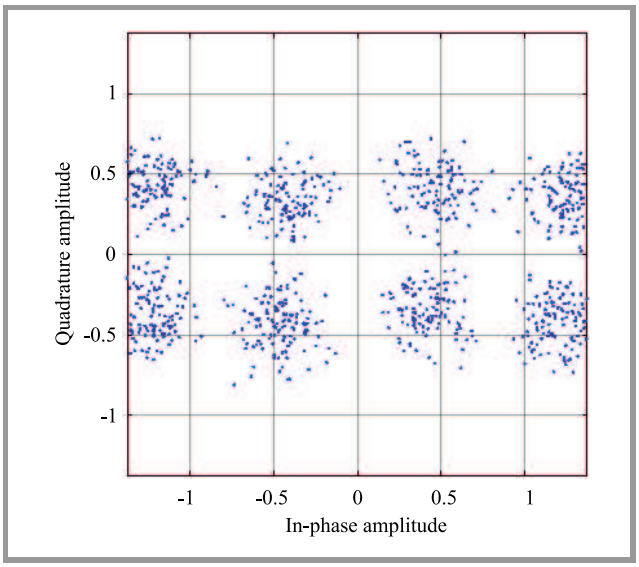


Fig. 3. Constellation diagram OFDM NOMA signals.

diagram of a simple OFDM-NOMA signal shown in Fig. 3, it is obvious that F-OFDM constellation points are more concentrated than OFDM signal points, due to better SNR. To the best of our knowledge, this is the first paper which considers BER performance of F-OFDM based downlink and uplink NOMA systems. We also consider the generic Nakagami-m fading channel to analyze the behavior of the system, as this model presents both strong and weak fading scenarios by altering the value of fading parameter *m* [18]. The main contributions of this paper are:

- We propose Filtered OFDM based downlink and uplink NOMA models with windowed sinc filters for spectral shaping.
- We derive and examine the exact closed form BER expressions of downlink Filtered OFDM NOMA users over Nakagami-*m* fading channel.
- For uplink mode, the analysis usually becomes intractable, so we derive and analyze approximate closed form BER expressions of uplink Filtered OFDM NOMA users over Nakagami-*m* fading channel.
- The derived expressions are presented in compact form which are easily implementable in common software packages. The obtained simulations for different values of fading parameter *m* support the validity of these derived analytical expressions.

The rest of the paper is organized as follows. Section 4 describes the architecture of the proposed downlink and uplink F-OFDM NOMA system. BER performance assessment of the proposed downlink and uplink F-OFDM NOMA models is shown in Sections 5 and 6, respectively. The obtained results are simulated and discussed in Section 7, with conclusions presented in Section 8.

4. F-OFDM NOMA System Model

First, we consider the downlink F-OFDM NOMA system consisting of a base station (BS) and two users – near user (NUE) and far user (FUE). We assume that the channel conditions of NUE are better than those of FUE. So, QPSK modulation is used for NUE and BPSK modulation is used for FUE to achieve better spectral efficiency [4]. After modulation, the modulated signals are multiplied by their respective allotted power levels such that symbol energies are $E_{NUE} = \alpha P_s$ and $E_{FUE} = (1-\alpha)P_s$ with α being the power coefficient and P_s being the total transmit power. Depending on channel conditions, more power is allotted to the far user than the near user, keeping the total transmitted power equal to a unity. This step ensures user fairness in the NOMA system. After this fractional transmit power allocation (FTPA), the signals are superposed [6]. The generated superimposed signal is given by:

$$x = \sqrt{E_{NUE}}x_{NUE} + \sqrt{E_{FUE}}x_{FUE} \ , \tag{2}$$

where x_{NUE} and x_{FUE} are QPSK and BPSK modulated signals of NUE and FUE, respectively. The signal given by Eq. (2) is fed to the IFFT block to generate OFDM symbols s_{ofdm}. A cyclic prefix is appended to s_{ofdm} to mitigate the effect of ISI. The OFDM signal generated in this manner undergoes pulse shaping by passing through a low pass filter $f(n)$. We realize this low pass filter by considering sinc filter as a pulse shaping filter, truncated using the Hann window. These soft truncated filters have a sharp transition band to reduce guard bands [14]. The output of the filter is given as:

$$s_f = s_{ofdm} * f(n) \ , \tag{3}$$

where '$*$' represents the convolution operation. This F-OFDM signal is transmitted by the BS. The signals received at NUE and FUE, after passing through their respective AWGN channels, are:

$$\begin{aligned} y_{NUE} &= s_f + w_{NUE} \ , \\ y_{FUE} &= s_f + w_{FUE} \ , \end{aligned} \tag{4}$$

where w_{NUE} and w_{FUE} represent AWGN noise (zero mean, $N_0/2$ variance) of NUE and FUE channels, respectively.
At FUE receiver, the received signal y_{FUE} is first passed through the matched filter $f^*(-n)$, which maximizes the SNR of the received signal and avoids interference from neighboring users [17]. After this, the cyclic prefix is removed and FFT operation is carried out. Then, BPSK demodulation provides the message signal of FUE. No SIC is performed at FUE, leading to a simpler receiver. On the other hand, the decoding process at NUE involves execution of SIC. NUE first decodes the FUE signal. Then, SIC block eliminates interference caused by the far user signal. After removal of far user symbols, NUE decodes its own message signal.
Furthermore, the proposed uplink FOFDM NOMA model is described in the following manner. The transmitter side of uplink F-OFDM NOMA consists of two users which are communicating with the BS. The near user (NUE) modulates its signal and feeds its QPSK modulated signal to the IFFT block to generate OFDM symbols to which a cyclic prefix (CP) is appended. After this, pulse shape filtering and windowing is performed, leading to the formation of F-OFDM signal of NUE, x_{nue}, which is transmitted to BS. Simultaneously, FUE generates and transmits its F-OFDM signal x_{fue}. These NUE and FUE F-OFDM signals are received at BS with different SNRs, depending on their respective transmit power and channel conditions [19]. The BS performs all the decoding in the uplink scenario. The signal received at the BS using uplink F-OFDM NOMA is given as:

$$y = \sqrt{P_N}\,x_{nue} + \sqrt{P_F}\,x_{fue} + w \ , \tag{5}$$

where P_N and P_F are transmit power levels of NUE and FUE, respectively, and w is AWGN noise. BS first performs SIC and decodes the NUE signal. After subtraction of NUE signals from the received signal, the remaining signal is used to decode FUE signals. In this way, message signals of both users are decoded by BS [20].

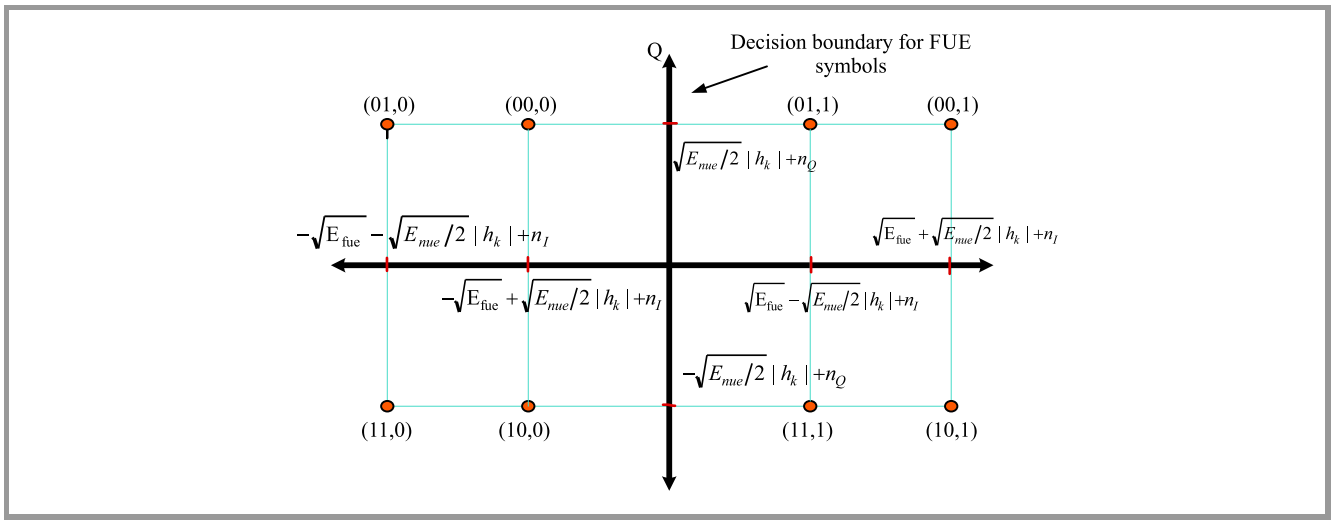


Fig. 4. Signal space representation of superposed F-OFDM signal received by users in the downlink mode.

5. Performance Evaluation of Downlink F-OFDM NOMA

We derive the BER expressions of F-OFDM based NOMA users over Nakagami-*m* fading channel. We incorporate the filter block acting as a spectrum shaper with the spectrum shaping factor p. As described in the model, at each user, the received signal, after passing through the matched filter and the OFDM demodulator, is sent to the decoder block. In the downlink mode, the signal space received by both users (NUE and FUE) is shown in Fig. 4.

Here, the superposed symbols are represented in the form of (y_1, y_2, y_3) with y_1 and y_2 bits representing the first and the second bit of QPSK modulated NUE symbols, and y_3 representing BPSK modulated bits of FUE. The decision logic is based on the concept that an error in any superposed symbol may occur, when the in-phase AWGN noise is strong enough to shift the received symbol to another region across the decision boundary [21]. The decoding is done using the maximum likelihood (ML) detector over the superposed symbols. The probability of error is given by the sum of the probability of error of each symbol multiplied by their respective prior probabilities [23]. Here, prior probabilities of all superposed symbols are assumed to be equal.

5.1. Error probability of Downlink Filtered Far User

Here, we consider the decoding of FUE symbols by taking NUE symbols as noise. Using ML detection at FUE, we have [22]:

$$P_{fue}(e) = \frac{1}{2}P\left(n_I \geq \sqrt{E_{fue}} + \sqrt{E_{nue}/2}\right) + \frac{1}{2}P\left(n_I \geq \sqrt{E_{fue}} - \sqrt{E_{nue}/2}\right) . \quad (6)$$

Here, $E_{fue} = p.E_{FUE}$ and $E_{nue} = p.E_{NUE}$, where p is the shaping factor which depends on the properties of the designed filter (as shown in Appendix B). Also, E_{fue} and E_{nue} are respective filtered symbol energies of the far and near NOMA user. In terms of Gaussian Q function, we have:

$$P_{fue}(e) = \frac{1}{2}\left[Q(\sqrt{\gamma_1}) + Q(\sqrt{\gamma_2})\right] , \quad (7)$$

where γ_1 and γ_2 represent filtered SNRs of outer four and inner four constellation points of the signal space diagram given in Fig. 4:

$$\gamma_1 = \left(\sqrt{\frac{2E_{fue}}{N_o}} + \sqrt{\frac{E_{nue}}{N_o}}\right)^2 ,$$
$$\gamma_2 = \left(\sqrt{\frac{2E_{fue}}{N_o}} - \sqrt{\frac{E_{nue}}{N_o}}\right)^2 . \quad (8)$$

Equation 7 gives the expression for instantaneous error probability of F-OFDM modulated downlink FUE NOMA user in AWGN channel. Now, we consider the fading environment which leads to the inclusion of channel coefficients in the derivation of BER expressions:

$$P_{fue}(e) = \frac{1}{2}\left[Q(\sqrt{\gamma_3}) + Q(\sqrt{\gamma_4})\right] , \quad (9)$$

where γ_3 and γ_4 represent faded SNRs of outer four and inner four constellation points from Fig. 4, where:

$$\gamma_3 = \left(\sqrt{\frac{2E_{fue}}{N_o}} + \sqrt{\frac{E_{nue}}{N_o}}\right)^2 |h_{fue}|^2 ,$$
$$\gamma_4 = \left(\sqrt{\frac{2E_{fue}}{N_o}} - \sqrt{\frac{E_{nue}}{N_o}}\right)^2 |h_{fue}|^2 . \quad (10)$$

Evaluating the average BER over fading channel:

$$\overline{P_{fue}(e)} = \frac{1}{2}\left[\underbrace{\int_0^\infty Q(\sqrt{\gamma_3})\, f_{\gamma_3}(\gamma_3)\mathrm{d}\gamma_3}_{I_3} + \underbrace{\int_0^\infty Q(\sqrt{\gamma_4})\, f_{\gamma_4}(\gamma_4)\mathrm{d}\gamma_4}_{I_4}\right] . \quad (11)$$

Considering Nakagami-*m* fading distribution, the integrals I_3 and I_4 are given by [24]:

$$I_3 = \frac{1}{2}\left[1 - \sqrt{\frac{\frac{\overline{\gamma_3}}{2}}{m+\frac{\overline{\gamma_3}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_3}}{2}}{m+\frac{\overline{\gamma_3}}{2}}}{4}\right)^k\right],$$

$$I_4 = \frac{1}{2}\left[1 - \sqrt{\frac{\frac{\overline{\gamma_4}}{2}}{m+\frac{\overline{\gamma_4}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_4}}{2}}{m+\frac{\overline{\gamma_4}}{2}}}{4}\right)^k\right], \quad (12)$$

where m is fading parameter ranging from 0.5 to ∞ and:

$$\overline{\gamma_3} = \left(\sqrt{\frac{2E_{fue}}{N_o}} + \sqrt{\frac{E_{nue}}{N_o}}\right)^2 E\left[|h_{fue}|^2\right],$$

$$\overline{\gamma_4} = \left(\sqrt{\frac{2E_{fue}}{N_o}} - \sqrt{\frac{E_{nue}}{N_o}}\right)^2 E\left[|h_{fue}|^2\right]. \quad (13)$$

On substituting Eq. (12) in Eq. (11), we have:

$$\overline{P_{fue}(e)} = \frac{1}{4}\left[1 - \sqrt{\frac{\frac{\overline{\gamma_3}}{2}}{m+\frac{\overline{\gamma_3}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_3}}{2}}{m+\frac{\overline{\gamma_3}}{2}}}{4}\right)^k\right] + \frac{1}{4}\left[1 - \sqrt{\frac{\frac{\overline{\gamma_4}}{2}}{m+\frac{\overline{\gamma_4}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_4}}{2}}{m+\frac{\overline{\gamma_4}}{2}}}{4}\right)^k\right]. \quad (14)$$

Hence, Eq. (14) gives the average error probability of spectrally shaped OFDM downlink FUE over the Nakagami-*m* fading channel for integer values of *m*. On substituting *m*=1 in Eq. (14), we obtain the average probability of error for FUE over the Rayleigh fading channel, which is given by:

$$\overline{P_{fue}(e)} = \frac{1}{4}\left[\left(1 - \sqrt{\frac{\frac{\overline{\gamma_3}}{2}}{m+\frac{\overline{\gamma_3}}{2}}}\right) + \left(1 - \sqrt{\frac{\frac{\overline{\gamma_4}}{2}}{m+\frac{\overline{\gamma_4}}{2}}}\right)\right]. \quad (15)$$

The closed form average BER expressions given in Eqs. (14)–(15) are compact presenting the behavior of downlink FUE over Nakagami-*m* and Rayleigh fading channels, respectively.

5.2. Error Probability of Downlink Filtered Near User

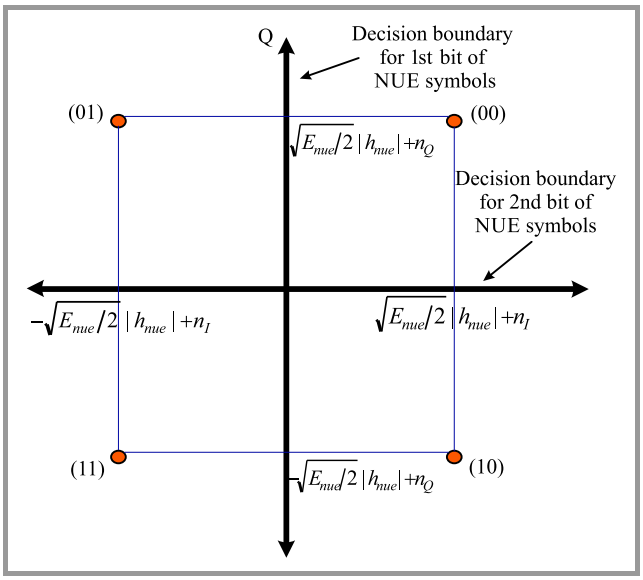


Fig. 5. Signal space representation after correct decoding and subtraction of filtered FUE symbols.

The detection of NUE symbols depends on the correct detection of FUE symbols. The signal space remaining after correct detection and subtraction of FUE symbols is given in Fig. 5. Here, the prior probabilities given by Eq. (6) are used to calculate the error probability of NUE. Considering ML detection, the probability of error of NUE symbols is obtained by taking the average of error probabilities of two bits [22]. Using [21] along with our proposed pulse shaping of the transmitting signals, we have:

$$P_{nue} = \frac{1}{4}\left[Q(\sqrt{\gamma_5}) \times \{4 - Q(\sqrt{\gamma_6}) - Q(\sqrt{\gamma_7})\} - Q(\sqrt{\gamma_6})\right], \quad (16)$$

where the filtered SNRs are given as:

$$\gamma_5 = \frac{E_{nue}}{N_o}|h_{nue}|^2,$$

$$\gamma_6 = \left(\sqrt{\frac{2E_{fue}}{N_o}} + \sqrt{\frac{E_{nue}}{N_o}}\right)^2 |h_{nue}|^2,$$

$$\gamma_7 = \left(\sqrt{\frac{2E_{fue}}{N_o}} - \sqrt{\frac{E_{nue}}{N_o}}\right)^2 |h_{nue}|^2. \quad (17)$$

Now, for evaluating average BER over the Nakagami-*m* fading channel, we solve each Q-function term of Eq. (16) in the similar manner as we solved Eq. (9) for FUE in the downlink mode. After the results so obtained are substituted in Eq. (16), we have:

$$\overline{P_{nue}(e)} = \frac{1}{8}\left[P_5 \times \{8 - P_6 - P_7\} - P_6\right] , \qquad (18)$$

with

$$P_5 = 1 - \sqrt{\frac{\frac{\overline{\gamma_5}}{2}}{m+\frac{\overline{\gamma_5}}{2}}} \sum_{k=0}^{m-1} \binom{2k}{k} \left(\frac{1-\frac{\frac{\overline{\gamma_5}}{2}}{m+\frac{\overline{\gamma_5}}{2}}}{4}\right)^k ,$$

$$P_6 = 1 - \sqrt{\frac{\frac{\overline{\gamma_6}}{2}}{m+\frac{\overline{\gamma_6}}{2}}} \sum_{k=0}^{m-1} \binom{2k}{k} \left(\frac{1-\frac{\frac{\overline{\gamma_6}}{2}}{m+\frac{\overline{\gamma_6}}{2}}}{4}\right)^k ,$$

$$P_7 = 1 - \sqrt{\frac{\frac{\overline{\gamma_7}}{2}}{m+\frac{\overline{\gamma_7}}{2}}} \sum_{k=0}^{m-1} \binom{2k}{k} \left(\frac{1-\frac{\frac{\overline{\gamma_7}}{2}}{m+\frac{\overline{\gamma_7}}{2}}}{4}\right)^k , \qquad (19)$$

and

$$\overline{\gamma_5} = \frac{E_{nue}}{N_o} E\left[|h_{nue}|^2\right] ,$$

$$\overline{\gamma_6} = \left(\sqrt{\frac{2E_{fue}}{N_o}} + \sqrt{\frac{E_{nue}}{N_o}}\right)^2 E\left[|h_{nue}|^2\right] ,$$

$$\overline{\gamma_7} = \left(\sqrt{\frac{2E_{fue}}{N_o}} - \sqrt{\frac{E_{nue}}{N_o}}\right)^2 E\left[|h_{nue}|^2\right] . \qquad (20)$$

In this way, Eq. (18) gives the average probability of error of NUE in a downlink F-OFDM NOMA system over the Nakagami-*m* fading channel for integer values of *m*. On substituting *m*=1 in Eq. (18), we obtain error probability of NUE over the Rayleigh channel.

6. Performance Evaluation of Uplink F-OFDM NOMA

In the uplink mode, the signal space of the received signal at BS is shown in Fig. 6. The received signal is a combination of user signals and noise. In the uplink mode, BS performs SIC to decode NUE signals and then, after subtraction, FUE signals are decoded. In this way, the decoding order in the uplink mode is the reverse of the process in the downlink F-OFDM NOMA [19].

6.1. Error Probability of Uplink Filtered Near User

Assuming that x_{NUE} and x_{FUE} symbols have equal prior probabilities and using the concept of ML detection presented in Eq. (6), the probability of a NUE error is given as [21]:

$$P_{nue}(e) = \frac{1}{2}P\left(n_Q \geq \sqrt{E_{nue}/2}h_{nue}\right) + \frac{1}{4}\left\{P\left(n_I \geq \sqrt{E_{nue}/2}h_{nue} + \sqrt{E_{fue}/2}h_{fue}\right) + P\left(n_I \geq \sqrt{E_{nue}/2}h_{nue} - \sqrt{E_{fue}/2}h_{fue}\right)\right\} . \qquad (21)$$

A simplified form of the equation is:

$$P_{nue}(e) = \frac{1}{2}\left[Q(\sqrt{\gamma_8}) + \frac{1}{2}Q(u) + \frac{1}{2}Q(v)\right] , \qquad (22)$$

where:

$$\gamma_8 = \frac{E_{nue}}{N_o}|h_{nue}|^2 ,$$

$$u = \sqrt{\frac{E_{nue}}{N_o}}h_{nue} + \sqrt{\frac{2E_{fue}}{N_o}}h_{fue} ,$$

$$v = \sqrt{\frac{E_{nue}}{N_o}}h_{nue} - \sqrt{\frac{2E_{fue}}{N_o}}h_{fue} . \qquad (23)$$

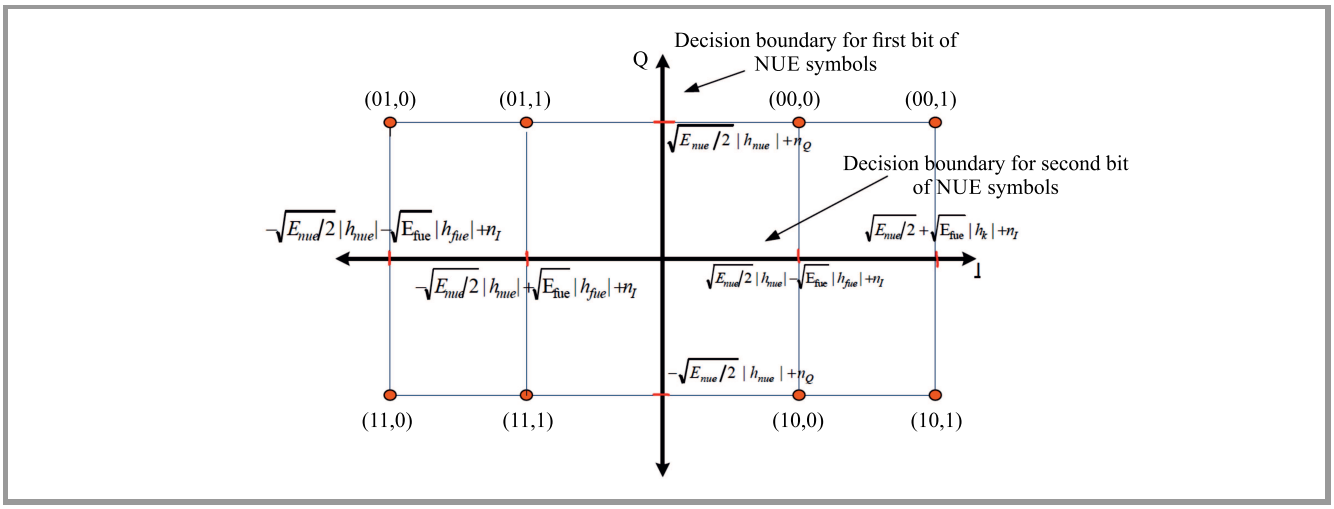


Fig. 6. Signal space representation of the received signal at BS in uplink F-OFDM NOMA.

Here, variables u and v represent the sum of and the difference between two random variables, respectively. We assume that two users in uplink NOMA transmit independently and face the same type of fading channel. So, we consider the sum of and the difference between two independent and identical distributed random variables (h_{nue} and h_{fue}). Further, we know that the value of Gaussian Q function of the sum of two quantities is lower than the difference between the same quantities. Considering this fact, we have $Q(u) << Q(v)$. Ignoring the sum term, the error probability of F-OFDM modulated NUE in the uplink mode is:

$$P_{nue}(e) = \frac{1}{2}\left[Q(\sqrt{\gamma_8}) + \frac{1}{2}Q(v)\right] . \tag{24}$$

Evaluating average BER over the fading channel, we have:

$$\overline{P_{nue}(e)} = \frac{1}{2}\left[\underbrace{\int_0^\infty Q(\sqrt{\gamma_8}) f_{\gamma_8}(\gamma_8)\mathrm{d}\gamma_8}_{I_8} + \frac{1}{2}\underbrace{\int_0^\infty Q(v) f(v)\mathrm{d}v}_{I_v}\right] . \tag{25}$$

The integrals in Eq. (25) can be easily implemented in common software packages, but considering the need of having compact expressions, we derive approximate closed-form average BER expressions in terms of well-known functions, such as Beta and Gauss hypergeometric functions.

Integral I_8 is one of the integrals of Eq. (25) which is easily solved by following the analytical steps performed in the downlink scenario for the Nakagami-*m* fading channel:

$$I_8 = \frac{1}{2}\left[1 - \sqrt{\frac{\frac{\overline{\gamma_8}}{2}}{m+\frac{\overline{\gamma_8}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_8}}{2}}{m+\frac{\overline{\gamma_8}}{2}}}{4}\right)^k\right] , \tag{26}$$

where:

$$\overline{\gamma_8} = \frac{E_{nue}}{N_o}E[|h_{nue}|^2] . \tag{27}$$

Now, by considering the fact that h_{nue} and h_{fue} are independent, integral I_v may be rewritten as [21]:

$$I_v = \int_0^\infty\int_0^\infty Q\left(a_1 h_{nue} - a_2 h_{fue}\right) f_{h_{nue}}(h_{nue}) f_{h_{fue}}(h_{fue})\mathrm{d}h_{nue}\mathrm{d}h_{fue} , \tag{28}$$

where:

$$a_1 = \sqrt{E_{nue}/N_o} , \quad a_2 = \sqrt{2E_{fue}/N_o} . \tag{29}$$

We solve integral I_v by considering Chernoff bound of Q function [26]:

$$Q(x) \cong \frac{1}{2}\mathrm{e}^{\frac{x^2}{2}} . \tag{30}$$

This approximation simplifies the analysis over fading channels, so we have:

$$I_v \cong \int_0^\infty\int_0^\infty \mathrm{e}^{\frac{(a_1 h_{nue} - a_2 h_{fue})^2}{2}} f_{h_{nue}}(h_{nue}) f_{h_{fue}}(h_{fue})\mathrm{d}h_{nue}\mathrm{d}h_{fue} \tag{31}$$

After solving this integral as shown in Appendix A, we have:

$$I_v \cong N\times\left[\frac{B(2m,\frac{1}{2})}{(a_1a_2)^{2m}}{}^2F_1(m,m+0.5;2m+0.5;0.5-2G) + \frac{\Gamma(m+0.5)\sqrt{\pi}}{\Gamma(m)}\frac{a_1a_2}{(AC)^{m+0.5}}{}^2F_1(m+0.5,m+0.5;1.5;\frac{K}{C})\right] , \tag{32}$$

where $B(.,.)$ and ${}^2F_1(.,.;.;.)$ are Beta and Gauss hypergeometric functions given in [25]. Here:

$$N = \frac{4}{(\Gamma m)^2}\left(\frac{m}{\Omega}\right)^{2m}\frac{\Gamma 2m}{2^{2m}},\ G = \frac{E}{F^2},\ E = \left(\frac{a_2^2}{2}+\frac{m}{\Omega}\right) - \frac{a_1^2a_2^2}{8A},\ F = \frac{a_1a_2}{\sqrt{2A}},\ A = \left(\frac{a_1^2}{2}+\frac{m}{\Omega}\right),\ K = \frac{a_1^2a_2^2}{4A},\ C = \left(\frac{a_2^2}{2}+\frac{m}{\Omega}\right) . \tag{33}$$

Using Eq. (26) and Eq. (32) in Eq. (25), we have:

$$\overline{P_{nue}(e)} = \frac{1}{4}\left\{\left[1 - \sqrt{\frac{\frac{\overline{\gamma_8}}{2}}{m+\frac{\overline{\gamma_8}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_8}}{2}}{m+\frac{\overline{\gamma_8}}{2}}}{4}\right)^k\right] + N\times\left[\frac{B(2m,\frac{1}{2})}{(a_1a_2)^{2m}}{}^2F_1(m,m+0.5;2m+0.5;0.5-2G)\right] + \frac{\Gamma(m+0.5)\sqrt{\pi}}{\Gamma(m)}\frac{Na_1a_2}{(AC)^{m+0.5}} {}^2F_1\left(m+0.5,m+0.5;1.5;\frac{K}{C}\right)\right\} . \tag{34}$$

This is an expression of the approximate error probability of NUE in the uplink scenario over the Nakagami-*m* fading channel, in terms of Beta and Gauss hypergeometric functions, for integer values of m. Furthermore, the error probability of NUE in the uplink scenario over the Rayleigh fading channel is obtained by substituting $m = 1$ in Eq. (34) leading to:

$$\overline{P_{nue}(e)} = \frac{1}{4}\left[1-\sqrt{\frac{\overline{\gamma_8}}{2+\overline{\gamma_8}}}\right] + \frac{1}{4\Omega^2}\left[\frac{B(2,0.5)}{(a_1a_2)^2} \times {}^2F_1\left(1,1.5;2.5;0.5-2Z\right) + \frac{\Gamma(1.5)\sqrt{\pi}a_1a_2}{4(PQ)^{1.5}}\left(1-\frac{Y}{Q}\right)^{-1.5}\right], \quad (35)$$

where:

$$Z=\frac{R}{S^2},\ R=\left(\frac{a_2^2}{2}+\frac{1}{\Omega}\right)-\frac{a_1^2a_2^2}{8P}, F=\frac{a_1a_2}{\sqrt{2P}},$$
$$P=\left(\frac{a_1^2}{2}+\frac{1}{\Omega}\right),\ Y=\frac{a_1^2a_2^2}{4P},\ Q=\left(\frac{a_2^2}{2}+\frac{1}{\Omega}\right). \quad (36)$$

We observe that the derived approximate average BER expressions over Nakagami-*m* and Rayleigh fading channels are described by well-known functions which are usually used in wireless communication. These functions are readily available in common software packages which help in providing better comparison with the simulations.

6.2. Error Probability of Uplink Filtered Far User

After decoding NUE symbols, BS subtracts these symbols from the main signal and performs decoding of FUE symbols. The signal space representation after subtraction of decoded NUE symbols is shown in Fig. 7. Error probabilities of each symbol within the constellation may be calculated by considering the decision boundary [21]:

$$P_{fue}(e) = \frac{1}{2}P\left(\sqrt{E_{fue}}h_{fue} \le n_I \le \sqrt{E_{nue}/2}h_{nue} + \sqrt{E_{fue}h_{fue}}\right) + \frac{1}{2}P\left(n_I \le -\sqrt{E_{fue}}\right). \quad (37)$$

In terms of Q function, we have:

$$P_{fue}(e) = Q\left(\sqrt{\gamma_9}\right) - \frac{1}{2}Q(u), \quad (38)$$

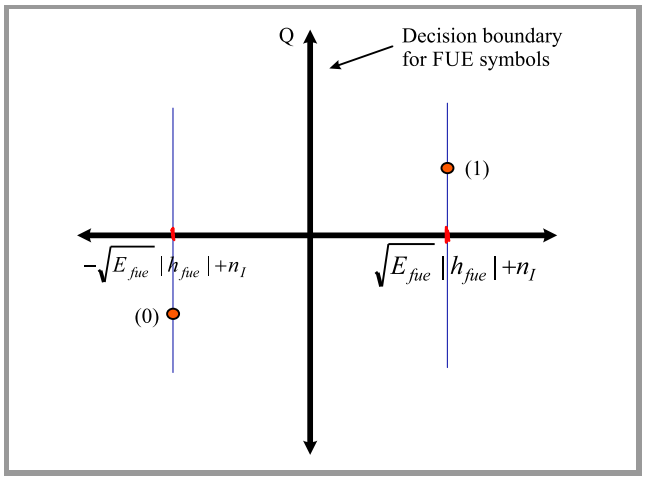


Fig. 7. Representation of signal space remaining at BS after subtraction of correctly decoded NUE symbols.

where:

$$\gamma_9 = \frac{2E_{fue}}{N_o}|h_{fue}|^2, \qquad \overline{\gamma_9} = \frac{2E_{fue}}{N_o}E[|h_{fue}|^2],$$
$$u = a_1h_{nue} + a_2h_{fue}\ . \quad (39)$$

For average probability of error of uplink FUE over fading channel:

$$\overline{P_{fue}(e)} = \underbrace{\int_0^\infty Q\left(\sqrt{\gamma_9}\right)f_{\gamma_9}(\gamma_9)\mathrm{d}\gamma_9}_{I_9} - \frac{1}{2}\underbrace{\int_0^\infty Q(u)f(u)\mathrm{d}u}_{I_u}\ . \quad (40)$$

Using the Nakagami-*m* fading model analysis [22]:

$$I_9 = \frac{1}{2}\left[1-\sqrt{\frac{\frac{\overline{\gamma_9}}{2}}{m+\frac{\overline{\gamma_9}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_9}}{2}}{m+\frac{\overline{\gamma_9}}{2}}}{4}\right)^k\right]. \quad (41)$$

Integral I_u is given by:

$$I_u = \int_0^\infty Q(u)f(u)\mathrm{d}u\ . \quad (42)$$

For the Nakagami-*m* fading environment, integral I_u is solved similarly to integral I_v for uplink NUE. Following the manipulations shown in Appendix A, we have:

$$I_u \cong N\times\left[\frac{B(2m,\frac{1}{2})}{(a_1a_2)^{2m}}{}^2F_1\left(m,m+0.5;2m+0.5;0.5-2G\right)\right]. \quad (43)$$

Upon substituting the expressions of I_9 and I_u given by Eq. (41) and Eq. (43), respectively, in Eq. (40):

$$\overline{P_{fue}(e)} = \frac{1}{2}\left[1-\sqrt{\frac{\frac{\overline{\gamma_9}}{2}}{m+\frac{\overline{\gamma_9}}{2}}}\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\frac{\frac{\overline{\gamma_9}}{2}}{m+\frac{\overline{\gamma_9}}{2}}}{4}\right)^k\right] - \frac{1}{2}\left[N\times\frac{B(2m,\frac{1}{2})}{(a_1a_2)^{2m}}{}^2F_1\left(m,m+0.5;2m+0.5;0.5-2G\right)\right]. \quad (44)$$

For the Rayleigh fading channel, we take $m=1$ in Eq. (44):

$$\overline{P_{fue}(e)} = \frac{1}{2}\left[1-\sqrt{\frac{\overline{\gamma_9}}{2+\overline{\gamma_9}}}\right] - \frac{1}{2\Omega^2}\left[\frac{B(2,0.5)}{(a_1a_2)^2}{}^2F_1\left(1,1.5;2.5;0.5-2Z\right)\right]. \quad (45)$$

The approximate closed-form BER expressions for FUE in the uplink mode, derived with the use of the proposed model are described by well-known functions. Such compact expressions help to follow the dependence of BER on different fading parameters.

7. Results and Discussions

We simulate the proposed F-OFDM NOMA model using MATLAB software. We consider the following simulation parameters:

- IFFT/FFT points = 1024,
- cyclic prefix length = 64,
- filter length = 513,
- sinc filter – prototype,
- Hann windowing.

For spectrum shaping, the length of the sinc filter is assumed to equal 513, which is greater than the cyclic prefix, thus leading to better OOBE reduction [17]. F-OFDM ensures 80 dB suppression in side lobe power, resulting in lower OOBE. This helps mitigate interference between adjacent users. Hence, we propose F-OFDM as a better modulation technique for NOMA systems. While simulating the obtained analytical results, we take pulse shaping factor of $p = 1.5$ (see Appendix B), considering moderate SNR improvement achieved by using the simulated pulse shaping filter.

7.1. Performance of Downlink F-OFDM NOMA

We evaluate BER performance of both NOMA user types in AWGN, as well as Rayleigh and Nakagami-*m* channels, using the proposed downlink F-OFDM NOMA model. The power coefficients assigned to NUE and FUE in the downlink mode are 0.2 and 0.8, respectively. Monte Carlo simulations are performed to observe variations in BER with SNR changes. The curves so obtained are plotted, such that the solid lines represent the analytical results obtained and the line markers (*) represent the simulation results.

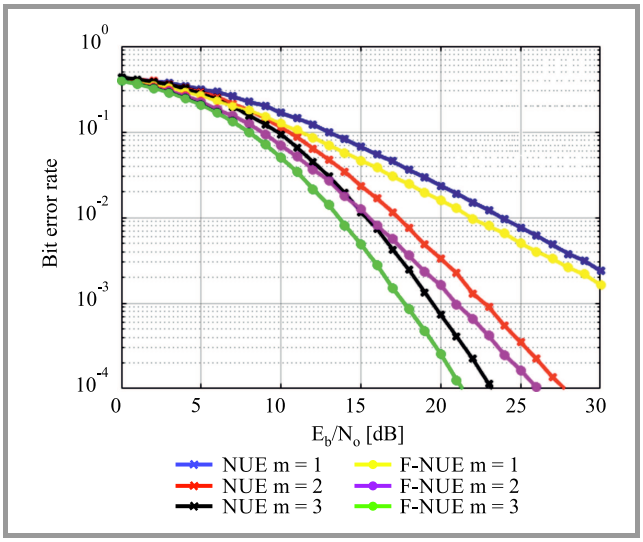


Fig. 9. BER performance of NUE in downlink F-OFDM NOMA over the Nakagami-*m* channel.

Interpretation of the curves shows that the proposed spectrally shaped NUE and FUE offer better performance in all three channels (AWGN, Rayleigh and Nakagami-*m*), with a considerable BER improvement of approximately 2 dB. Such an improvement is quite motivating to implement F-OFDM in downlink NOMA. We observe that fading leads to higher BERs, as is visible from Rayleigh BER curves in Figs. 8 and 10. Furthermore, over the Nakagami-*m* fading channel, user behaviors for different values of *m* (1, 2, and 3) are observed and shown in Figs. 9 and 11. It is evident that an increase in fading parameter *m* leads to

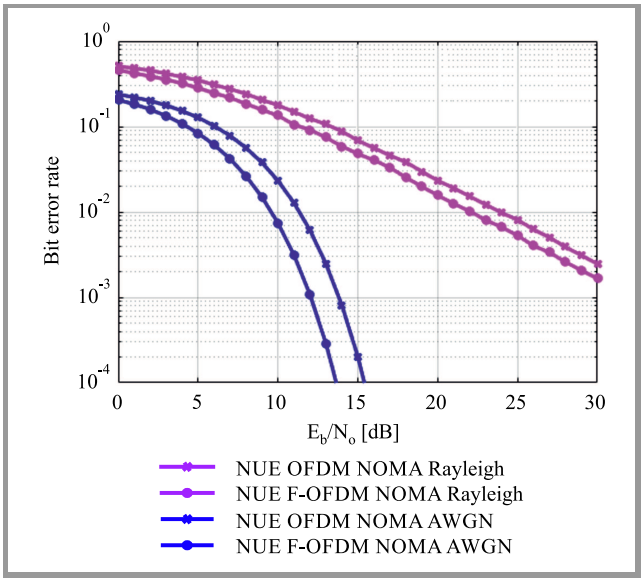


Fig. 8. BER performance of NUE in downlink F-OFDM NOMA in AWGN and over the Rayleigh channel.

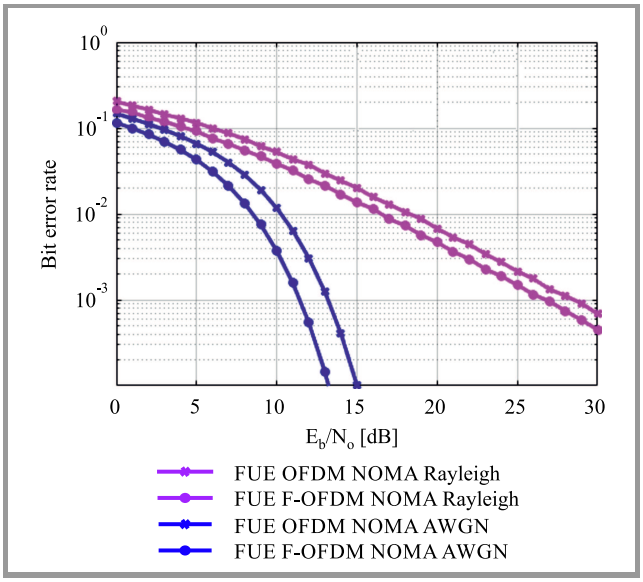


Fig. 10. BER performance of FUE in downlink F-OFDM NOMA in AWGN and over the Rayleigh channel.

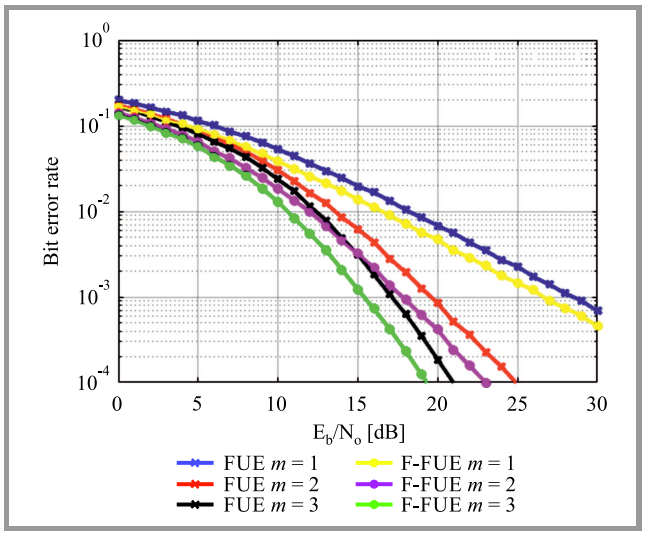


Fig. 11. BER performance of FUE in downlink F-OFDM NOMA over the Nakagami-*m* channel.

better BER performance. This is attributed to the fact that higher *m* means lower fading leading to better BER results [24].

7.2. Performance of Uplink F-OFDM NOMA

In the proposed uplink F-OFDM NOMA, we evaluate BER performance of both NUE and FUE in AWGN, Rayleigh and Nakagami-*m* channels. The simulations are plotted to present differences between BER and SNR. It is observed from the BER curves that the proposed spectrally shaped NUE and FUE show better performance in AWGN, Rayleigh and Nakagami-*m* channels, with a BER improvement of approximately 1 dB. Such a BER improvement is lower than in the proposed downlink F-OFDM NOMA.

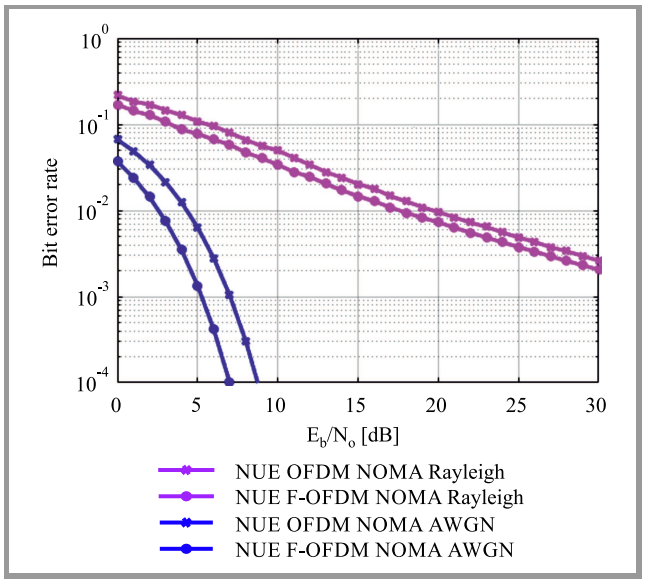


Fig. 12. The BER performance of NUE in uplink F-OFDM NOMA in AWGN and Rayleigh channel.

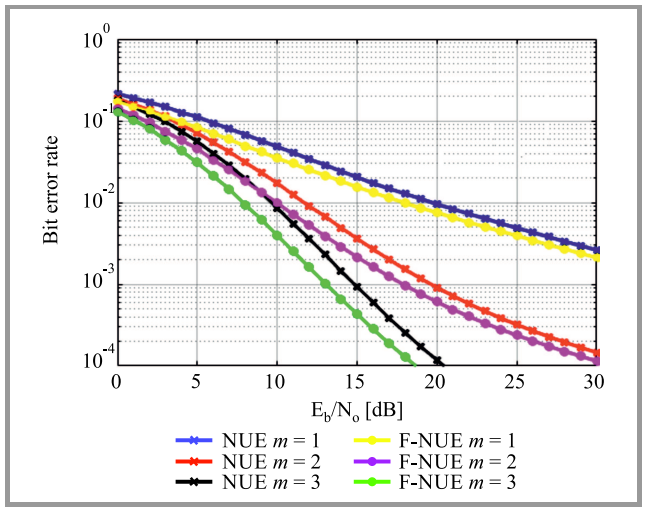


Fig. 13. The BER performance of NUE in uplink F-OFDM NOMA over Nakagami-*m* channel.

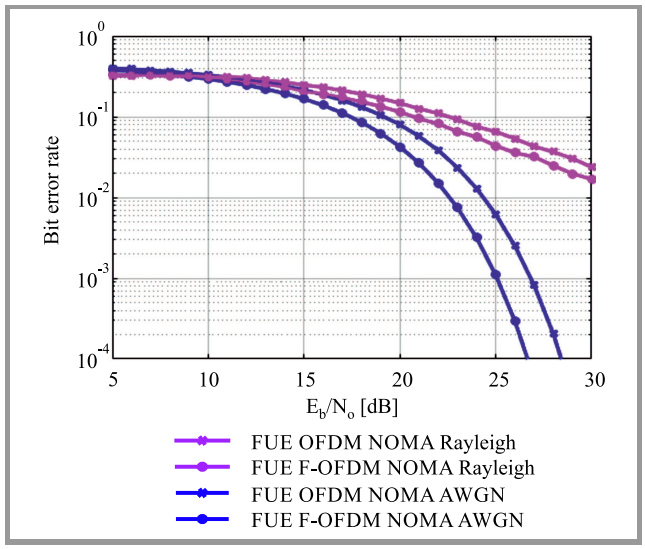


Fig. 14. The BER performance of FUE in uplink F-OFDM NOMA in AWGN and Rayleigh channel.

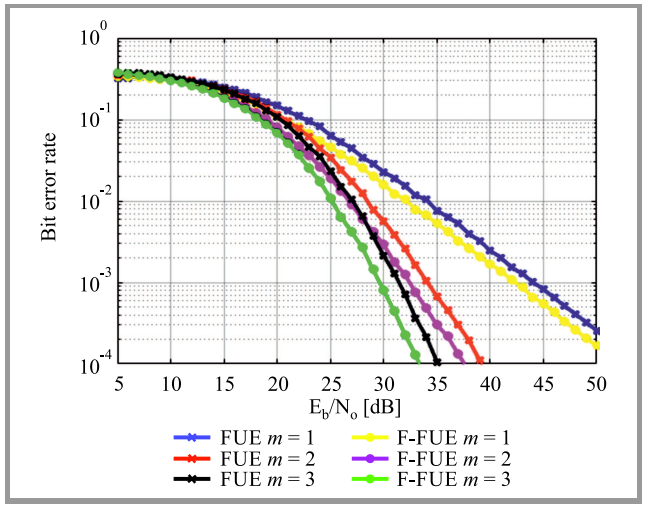


Fig. 15. The BER performance of FUE in uplink F-OFDM NOMA over Nakagami-*m* channel.

Just as in the downlink scenario, Rayleigh curves show higher BER than AWGN curves, as shown in Figs. 12 and 14. Furthermore, we observe the user behaviors over the Nakagami-*m* channel for different values of *m* (1, 2, and 3), as shown in Figs. 13 and 15. It is evident that BER decreases with an increase in *m*, following the behavior shown by the users in the downlink mode over the Nakagami-*m* fading channel.

8. Conclusions

NOMA is a promising multiple access technique capable of meeting the diverse needs of the increasing number of users. Selection of an efficient modulation technique determines the performance of the NOMA system. In this paper, we propose the use of filtered OFDM as an efficient modulation technique for NOMA. F-OFDM has the potential to support the diverse nature of services rendered in NOMA environments. In this paper, simple sinc filtering along with Hann windowing is performed to generate F-OFDM NOMA signals characterized by better spectral confinement. We evaluate the performance of F-OFDM based downlink and uplink NOMA systems. For the downlink scenario, we derive compact, closed-form BER expressions of near and far NOMA users over the Nakagami-*m* fading channel. For the uplink mode, we derive approximate BER expressions for both users, described by easily implementable functions, over the Nakagami-*m* fading channel. Monte Carlo simulations are carried out to validate the analytical results obtained. Simulations are performed for different values of fading parameter *m*. Using the proposed model, BER improvement of nearly 2 dB and 1 dB is achieved by NOMA users in the downlink and uplink scenarios, respectively. Such behavior of F-OFDM NOMA encourages the application of F-OFDM in NOMA systems intended for 5G.

Appendix A

Integral I_v is given by:

$$I_v \cong \int_0^\infty [I_{inner1}]\, \mathrm{e}^{-\frac{a_2^2 h_{fue}^2}{2}} f_{h_{fue}}(h_{fue})\mathrm{d}h_{fue} \ , \qquad (46)$$

where I_{inner1} is:

$$I_{inner1} = \int_0^\infty \mathrm{e}^{-\frac{a_1^2 h_{nue}^2}{2}} \mathrm{e}^{a_1 a_2 h_{nue} h_{fue}} f_{h_{nue}}(h_{nue})\mathrm{d}h_{nue} \ . \qquad (47)$$

I_{inner1} is solved by considering Nakagami-*m* distribution with *m* as fading parameter and $\Omega = E[|h_{nue}|^2]$:

$$I_{inner1} = \frac{2}{\Gamma m}\left(\frac{m}{\Omega}\right)^m \int_0^\infty \mathrm{e}^{-A h_{nue}^2 + M h_{nue}} h_{nue}^{2m-1}\mathrm{d}h_{nue} \ , \qquad (48)$$

where:

$$A = \frac{a_1^2}{2} + \frac{m}{\Omega}, \quad M = a_1 a_2 h_{fue} \ . \qquad (49)$$

Using the standard solution given in [25], the solution of Eq. (48) is given as:

$$I_{inner1} = \frac{2}{\Gamma m}\left(\frac{m}{\Omega}\right)^m \frac{\Gamma 2m}{(2A)^m} \mathrm{e}^{\frac{M^2}{8A}} D_{-2m}\frac{-M}{\sqrt{2A}} \ , \qquad (50)$$

where $D_v(.)$ is the parabolic cylinder function [25]. Substituting the expression for $D_v(-z)$ given in [28] in Eq. (50):

$$I_{inner1} = I_x + I_y \ , \qquad (51)$$

where:

$$I_x = \frac{2}{\Gamma m}\left(\frac{m}{\Omega}\right)^m \frac{\Gamma 2m}{(2A)^m} \mathrm{e}^{\frac{M^2}{8A}} D_{-2m}\left(\frac{M}{\sqrt{2A}}\right),$$
$$I_y = \frac{2}{\Gamma m}\left(\frac{m}{\Omega}\right)^m \frac{\Gamma 2m}{(2A)^m} \frac{2^{-m+0.5}(2m)\sqrt{\pi}}{\Gamma m + 1} \frac{a_1 a_2 h_{fue}}{\sqrt{2A}}$$
$$\times {}^1F_1\left(m + 0.5; 1.5; \frac{a_1^2 a_2^2 h_{fue}^2}{4A}\right) \ , \qquad (52)$$

where ${}^1F_1(.;.;)$ is the confluent hypergeometric function [25].

We substitute the expression of I_{inner1} in terms of I_x and I_y in Eq. (46) along with the Nakagami-*m* distribution of h_{fue}. Using [29] and [30], the final expression of I_v is obtained. Similarly, we have integral I_u:

$$I_u \cong \int_0^\infty [I_{inner2}]\, \mathrm{e}^{\frac{-a_2^2 h_{fue}^2}{2}} f_{h_{fue}}(h_{fue})\mathrm{d}h_{fue} \ , \qquad (53)$$

where I_{inner2} is:

$$I_{inner2} = \int_0^\infty \mathrm{e}^{-\frac{a_1^2 h_{nue}^2}{2}} \mathrm{e}^{-a_1 a_2 h_{nue} h_{fue}} f_{h_{nue}}(h_{nue})\mathrm{d}h_{nue} \ . \qquad (54)$$

Integral I_{inner2} is solved in the similar manner as Eq. (48) leading to:

$$I_{inner2} = \frac{2}{\Gamma m}\left(\frac{m}{\Omega}\right)^m \frac{\Gamma 2m}{(2A)^m} \mathrm{e}^{\frac{M^2}{8A}} D_{-2m}\left(\frac{M}{\sqrt{2A}}\right) \ , \qquad (55)$$

where $D_v(.)$ is the parabolic cylinder function [25]. Now, we substitute the expression I_{inner2} in Eq. (53) along with the Nakagami-*m* distribution of h_{fue}. Using [30], the final expression of I_u is obtained.

Appendix B

We introduce shaping factor *p* as a parameter which reflects the effect of filtering that is relied upon in the BER analysis of the F-OFDM NOMA signal. In this regard, the value of parameter *p* is chosen depending on the nature of the simulated filter. Since the filter improves SNR of the filtered signal, we define *p* as the ratio between SNR of the filtered OFDM signal and SNR of the unfiltered OFDM signal. Using the simulation parameters given in Section 7, the value of *p* is evaluated as 1.5. All derived expressions consider this shaping factor as a scaling parameter of filtered SNR.

Table 1
BER comparison of F-OFDM NOMA at SNR= 10 dB

Value of p	Simulation BER	Analytical BER
$p = 0.5$	0.0037	0.04
$p = 1$	0.0037	0.011
$p = 1.5$	0.0037	0.00371
$p = 2$	0.0037	0.0012

We also confirm the value of p iteratively, by considering the general AWGN downlink F-OFDM NOMA case of the far user, as shown in Table 1. It is clear that for $p = 1.5$, there is a close match between BER simulations and analytical results.

References

[1] A. Benjebbour *et al.*, "NOMA: From concept to standardization", in *IEEE Conf. on Standards for Commun. and Networking (CSCN)*, Tokyo, 2015, pp. 18–23 (DOI: 10.1109/CSCN.2015.7390414).

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005 (ISBN: 9780511807213).

[3] M. Aldababsa, "A tutorial on non-orthogonal multiple access (NOMA) for 5G and beyond", *Wireless Commun. and Mobile Comput.*, vol. 2018, 2018 (DOI: 10.1155/2018/9713450).

[4] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges", in *IEEE Commun. Surveys and Tut.*, vol. 19, no. 2, 2017, pp. 721–742 (DOI: 10.1109/COMST.2016.2621116).

[5] Y. Chen *et al.*, "Toward the standardization of non-orthogonal multiple access for next generation wireless networks", in *IEEE Commun. Mag.*, vol. 56, no. 3, 2018, pp. 19–27 (DOI: 10.1109/MCOM.2018.1700845).

[6] K. Arslan and S. Y. Shin, "Linear precoding techniques for OFDM-based NOMA over frequency-selective fading channels", *IETE J. of Research*, vol. 63, no. 4, 2017, pp. 536–551 (DOI: 10.1080/03772063.2017.1299045).

[7] V. K. Trivedi *et al.*, "Enhanced OFDM-NOMA for next generation wireless communication: A study of PAPR reduction and sensitivity to CFO and estimation errors", *AEU-Int. J. of Electron. and Commun.*, vol. 102, 2019, pp. 9–24 (DOI: 10.1016/j.aeue.2019.01.009).

[8] Y. Cai *et al.*, "Modulation and Multiple Access for 5G Networks", *IEEE Commun. Surveys and Tutorials*, vol. 20, no. 1, pp. 629–646, 2018 (DOI: 10.1109/COMST.2017.2766698).

[9] X. Zhang, M. Jia, L. Chen, J. Ma, and J. Qiu, "Filtered-OFDM – enabler for flexible waveform in the 5th generation cellular networks", in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, 2015, pp. 1–6 (DOI: 10.1109/GLOCOM.2015.7417854).

[10] P. Guan *et al.*, "5G Field Trials: OFDM-based waveforms and mixed numerologies", *IEEE J. on Selected Areas in Commun.*, vol. 35, no. 6, pp. 1234–1243, 2017 (DOI: 10.1109/JSAC.2017.2687718).

[11] F. A. P. de Figueiredo, N. F. T. Aniceto, J. Seki, I. Moerman, and G. Fraidenraich, "Comparing F-OFDM and OFDM performance for MIMO systems considering a 5G scenario", in *Proc. IEEE 2nd 5G World Forum (5GWF)*, Dresden, Germany, 2019, pp. 532–535, (DOI: 10.1109/5GWF.2019.8911702).

[12] J. J. L. Quispe and L. G. P. Meloni, "Pulse shaping filter design for filtered OFDM transceivers", in *Proc. of the 3rd Brazilian Technol. Symp.*, 2019, pp. 131–143 (DOI: 10.1007/978-3-319-93112-8-15).

[13] J. Yli-Kaakinen, T. Levanen, A. Palin, M. Renfors, and M. Valkama, "Generalized fast-convolution-based filtered-OFDM: techniques and application to 5G new radio", *IEEE Transac. on Signal Process.*, vol. 68, pp. 1213–1228, 2020 (DOI: 10.1109/TSP.2020.2971949).

[14] T. B. Deepa, "Performance evaluation of polar coded filtered OFDM for low latency wireless communications", *Wireless Personal Commun.*, 2020 (DOI: 10.1007/s11277-020-07777-2).

[15] K. C. Hu and A. G. Armada, "SINR analysis of OFDM and F-OFDM for machine type communications", in *IEEE 27th Annual Int. Symp. on Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, Valencia, Spain, 2016, pp. 1–6 (DOI: 10.1109/PIMRC.2016.7794702).

[16] S. Wang, J. S. Thompson, and P. M. Grant, "Closed-form expressions for ICI/ISI in filtered OFDM systems for asynchronous 5G uplink", *IEEE Transac. on Commun.*, vol. 65, no. 11, pp. 4886–4898, 2017 (DOI: 10.1109/TCOMM.2017.2698478).

[17] J. Abdoli, M. Jia, and J. Ma, "Filtered OFDM: A new waveform for future wireless systems", in *Proc. IEEE 16th Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2015, pp. 66–70 (DOI: 10.1109/SPAWC.2015.7227001).

[18] M. Nakagami, "The m-distribution - a general formula of intensity distribution of rapid fading", in *Proc. Statistical Methods in Radio Wave Propag.*, Los Angeles, CA, USA, 1960, pp. 3–36 (DOI: 10.1016/B978-0-08-009306-2.50005-4).

[19] H. Tabassum, M. S. Ali, E. Hossain, M. J. Hossain, and D. I. Kim, "Uplink vs. downlink NOMA in cellular networks: challenges and research directions", in *Proc. IEEE 85th Vehic. Technol. Conf.*, Sydney, 2017, pp. 1–7 (DOI: 10.1109/VTCSpring.2017.8108691).

[20] Y. Neng *et al.*, "Uplink nonorthogonal multiple access technologies toward 5G: A survey", *Wireless Commun. and Mobile Comput.*, 2018 (DOI: 10.1155/2018/6187580).

[21] K. Ferdi and H. Kaya, "BER performances of downlink and uplink NOMA in the presence of SIC errors over fading channels", *IET Commun.*, 2018 (DOI: 10.1049/iet-com.2018.5278).

[22] M. Jain *et al.*, "Performance analysis at far and near user in NOMA based system in presence of SIC error", *AEU-Int. J. of Electronics and Commun.*, 2020, (DOI: 10.1016/j.aeue.2019.152993).

[23] J. Proakis, M. Salehi, and G. Bauch, *Contemporary communication systems using MATLAB*. Nelson Education, 2012 (ISBN: 9780495082514).

[24] M. K. Simon and M. S. Alouini, *Digital Communication Over Fading Channels*. New York: Wiley, 2005 (ISBN: 9780471649533).

[25] I. S. Gradshteyn, I. M. Ryzhik, *Table of integrals, series, and products*. San Diego: Academic Press, 2007 (ISBN: 9780123736376).

[26] N. Kapucu and M. Bilim, "Analysis of analytical capacity for Fisher-Snedecor F fading channels with different transmission schemes", *Electronics Letters*, vol. 55, no. 5, pp. 283–285, 2019 (DOI: 10.1049/el.2018.7813).

[27] Hypergeometric function, *Wolfram mathworld*, 2017 [Online]. Available: https://functions.wolfram.com/PDF/Hypergeometric2F1.pdf

[28] Parabolic cylindrical function, *Wolfram mathworld*, 2017 [Online]. Available: http://functions.wolfram.com/07.41.16.0006.01.

[29] Parabolic cylindrical function *Wolfram mathworld*, 2017 [Online]. Available: https://functions.wolfram.com/PDF/ParabolicCylinderD.pdf

[30] Kummer confluent hypergeometric function *Wolfram mathworld*, 2017 [Online]. Available: http://functions.wolfram.com/07.20.21.0012.01

Shaika Mukhtar received her B.E. and M.Tech. degrees in Electronics and Communication Engineering in 2012 and 2015, respectively. Currently, she is pursuing Ph.D. at the National Institute of Technology, Srinagar. She works as a Senior Research Fellow at the Advanced Communication Lab, NIT Srinagar. Her areas of interest include wireless communication, non-orthogonal multiple access (NOMA), high speed networks and next generation networks. Her research aims at understanding the different aspects of NOMA for future communication networks. She is a student member of IEEE.
E-mail: shaika.mukhtar@yahoo.com
Advanced Communication Lab
Department of Electronics and Communication Engineering
National Institute of Technology Srinagar
Jammu and Kashmir
India

Gh. Rasool Begh received his Ph.D. degree from the National Institute of Technology, Srinagar, India. He works as an Associate Professor at the Department of Electronics and Communication Engineering, NIT Srinagar. His areas of interest include cognitive radios, OFDM, MIMO, cooperative communications, error control coding and security. He is a member of IEEE MTTS.
E-mail: grbegh@nitsri.ac.in
Advanced Communication Lab
Department of Electronics and Communication Engineering
National Institute of Technology Srinagar
Jammu and Kashmir
India

Paper

LEES: a Hybrid Lightweight Elliptic ElGamal-Schnorr-Based Cryptography for Secure D2D Communications

Javeria Ambareen, M. Prabhakar, and Tabassum Ara

School of Computing and Information Technology, REVA University, Bengalore, India

https://doi.org/10.26636/jtit.2021.146020

Abstract—Device-to-device (D2D) communications in 5G networks will provide greater coverage, as devices will be acting as users or relays without any intermediate nodes. However, this arrangement poses specific security issues, such as rogue relays, and is susceptible to various types of attacks (impersonation, eavesdropping, denial-of-service), due to the fact that communication occurs directly. It is also recommended to send fewer control messages, due to authenticity- and secrecy-related prevailing requirements in such scenarios. Issues related to IoT applications need to be taken into consideration as well, as IoT networks are inherently resource-constrained and susceptible to various attacks. Therefore, novel signcryption algorithms which combine encryption with digital signatures are required to provide secure 5G IoT D2D communication scenarios in order to protect user information and their data against attacks, without simultaneously increasing communication costs. In this paper, we propose LEES, a secure authentication scheme using public key encryption for secure D2D communications in 5G IoT networks. This lightweight solution is a hybrid of elliptic curve ElGamal-Schnorr algorithms. The proposed scheme is characterized by low requirements concerning computation cost, storage and network bandwidth, and is immune to security threats, thus meeting confidentiality, authenticity, integrity and non-repudiation-related criteria that are so critical for digital signature schemes. It may be used in any 5G IoT architectures requiring enhanced D2D security and performance.

Keywords—5G networks, authentication, D2D communication, IoT, lightweight cryptography.

1. Introduction

Device-to-device (D2D) communication is a novel technology available in 5G networks, allowing two devices located nearby to communicate without approaching the base station. It is a boon for areas with low or no coverage. Smartphones and IoT devices may act as small base stations, providing all connectivity-related benefits of a 5G network to nearby devices, thus enhancing coverage. Two types of D2D communications are possible, as shown in Fig. 1, namely inband and outband. The inband scenario uses the licensed spectrum and may be divided into non-overlapping portions of D2D (overlay) or may not be divided at all (underlay). Outband communication uses the unlicensed spectrum and helps eliminate interference caused by such devices as Wi-Fi, Bluetooth, etc. It is further divided into controlled (where D2D communication is controlled by the network) or autonomous (where D2D control is left to users) varieties.

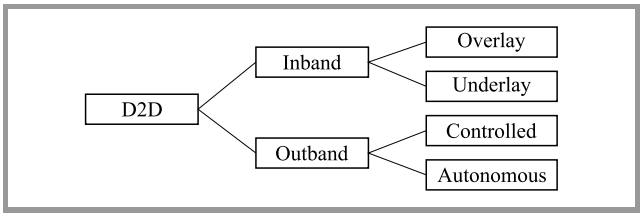


Fig. 1. Types of D2D communication.

In D2D communication, the devices act as relays. They may either be a transparent relay (TR), simply amplifying and forwarding the message, or a non-transparent relay (NTR), decoding and forwarding the message. Problems appear when these NTR-type relays become rouge and jeopardize the entire network, resulting in various security attacks. The situation gets aggravated since these devices are resource-constrained, with very limited computational and storage power. So, any new information security algorithms developed for the purpose of this scenario would need to be computationally lightweight. Authenticated encryption algorithms and digital signatures need to be used in any data transfers to secure these against common attacks and to maintain confidentiality, privacy and authenticity of the data involved.

Public key encryption and digital signatures are the key pillars of modern cryptography. In a real-world scenario, when two parties are communicating over a wireless communication channel which may be insecure, the encryption

algorithms data confidentiality of data and the digital signatures provide data authentication. Elliptic curve cryptography (ECC) was developed in 1985 and is one of the most widely used public key cryptography schemes. Finite keys are one of the main features in the algebraic structure of the solution. Its key sizes and the security level that it offers make it more popular compared to other algorithms. ElGamal is also a public key cryptographic technique but is based on the Diffie-Hellman key exchange. There are a few approaches to applying ECC combined with either ElGamal or Schnorr [1], both offering promising results, but no references are available in connection with combining elliptic curve-based ElGamal with Schnorr.

Taking this forward, in this paper we present LEES – a lightweight authentication scheme for secure D2D communications in 5G IoT networks. It is a hybrid implementation of ECC ElGamal encryption and the Schnorr digital signature scheme.

The remaining sections of the paper are organized as follows. Section 2 discusses the related work and is followed by a discussion on security considerations in D2D communication, presented in Section 3. Section 4 discusses the encryption and various digital signature scheme preliminaries along with the notations used in this paper. Section 5 presents the system model, while Section 6 presents the implementation schema. Section 7 discusses the results, while the overall conclusion is presented in Section 8.

2. Related Work

Verifying user identity as part of the authentication process, hiding sensitive information to ensure anonymity, preserving data confidentiality and integrity through the use of encryption, hash functions or message authentication, as well as optimized and cost-effective implementation methods are the topics outlined in the security-related considerations of the authors o of [2]. Paper [3] proposes a secure service-oriented authentication framework, where fog nodes that are responsible for forwarding data in a 5G network use a slice selection mechanism that ensures preservation of privacy. The Diffie-Hellman based [1] present a security analysis of two schemes: the Huang-Chang convertible signcryption scheme (which serves as a basis for the Schnorr signature) and the Kwak-Moon group signcryption scheme. The results show that both schemes are insecure. The Huang-Chang scheme fails to ensure confidentiality, while the Kwak-Moon scheme does not satisfy the properties of unforgeability, coalition-resistance, and traceability in its current form.

Paper [4] identifies the keyescrow problem of major cryptographic schemes and proposes a certificate-less signature scheme (CLS) for lightweight devices in Industrial Internet of Things (IIoT). In this procedure the keys are retained during the post-decryption phase and an authorized entity has access to these, but under a few predefined conditions. The scheme enables the selection of keys based on exponentials and ensures integrity through the use of hash functions. This pairing-based scheme is proven to be secure against type I and type II adversaries under the extended bilinear strong Diffie-Hellman (EBSDH) and bilinear strong Diffie-Hellman (BSDH) assumptions.

Article [5] proposes that validity of each participating user equipment (UE) be authenticated by 5G authentication and key agreement (AKA) only once during its life time, and that these checks be performed before generating a D2D token. The base stations communicate their public key through elliptic-curve digital signature algorithm (ECDSA) to generate the D2D token. The D2D communication process comprises three stages [6], the first one consists in discovering the device that identifies nodes in its proximity by sending out a request message in the broadcast mode. A nearby node responds with a D2D token and UE identity subscription concealed identifier (SUCI) in an encrypted form. The link setup phase comes next, where each node sends SUCI and the D2D token to the base station for verification. After verification, the secret keys are exchanged using the elliptic-curve Diffie-Hellman (ECDH) algorithm. The last step is the secure data transmission stage that relies on the authenticated encryption with associated data (AEAD) cipher to encrypt the data using the D2D token, before transmitting the data. However, 5G AKA has limitations and is susceptible to replay attacks. Hence, alternatives need to be looked at, and this is precisely the goal of this paper.

3. Security Considerations

Digital signature schemes are one of the most important cryptographic primitives enabled by public-key cryptography. These methods allow messages to be authenticated through the use of asymmetric encryption systems in which the sender and the receiver are not required to share a common but secret key. Digital signatures are cryptographic primitives which play a fundamental role in ensuring entity authentication, data origin authentication, data integrity and non-repudiation. Functionalities provided by a digital signature can be summarized as follows:

- **Authentication**. A private key of the sender is used to sign the message, thus authenticating the source of the message. The private key is not shared with anyone. It is known to the sender only. It can be verified by anyone by decrypting it with the use of the sender's public key.
- **Integrity**. Integrity is the most important property of the message, as it ensures that the data has not been compromised. Integrity may be ensured by digitally signing the message. The signature is generated with respect to the data contained in the message. If the message is altered, the receiver may easily determine that at the time of verification. It is extremely challenging to alter the message or its signature without the knowledge of the private key. Hence, the data is unaltered during the transmission.

- **Non-repudiation**. This characteristic ensures the integrity of data and guarantees that a third party will be able to verify the source of data and its integrity. It ensures that the sender cannot deny sending of the message/data. And this is basically supported by digitally signing the message with the help of a private key of the sender itself. When the receiver performs verification using the public key of the sender, a proof is obtained that the message/data has been sent by the same sender. Hence, denying the message becomes impractical.

In general, each algorithm used for signing a message will comprise two different key processes: one for signing and the other for verifying the message at the other end. Below, various security threats that loom large over D2D communications are enumerated.

3.1. Attacks in D2D Communications

Security threats in D2D communications are primarily related to the fact that the process is based on radio transmissions [5]. The most common types of attacks include the following:

- **Eavesdropping** – in this attack the intruder is able to listen-in without the actual participating devices (PDs) being aware of that fact. If confidentiality of cryptographic data is maintained, this attack may be thwarted.
- **Impersonation** – the intruder comes across as a valid participating device – or worse the base station (BS) – and steals data. If cryptographic authentication is enforced, this attack may be thwarted.
- **Forgery** – the intruder may send forged or malicious content to all participating devices, thereby confusing the entire system. If cryptographic data integrity via digital signature is enforced, this attack may be thwarted.
- **Control data** – the attacker may change the control data itself. Cryptographic techniques, such as authentication, confidentiality and integrity, are required to thwart this attack.
- **Denial of service** (DoS) – this kind of attack may render a service unavailable. Cryptographic actions, such as authentication, confidentiality and integrity are needed to thwart this attack.

3.2. Attack Resiliency

To cope with the attacks listed above and to secure D2D communications, it is worth looking at some attack resiliency requirements suggested in [4], [7], [8]:

- **Authentication** – identity check of participating devices performed on a frequent basis.
- **Data confidentiality** – data sent between participating devices should be encrypted.
- **Data integrity** – data sent with the use of authenticated devices should be verified to ensure it has not been tampered with.
- **Privacy** – all confidential information of participating devices must be kept secret, e.g. number, location, etc.
- **Traceability** – the source of malicious messages should be traceable.
- **Anonymity** – identity of participating devices should not be disclosed to neighboring devices or to intruders.
- **Non-repudiation** – a digital signature is an effective solution for both transmission and reception non-repudiation, wherein one can stop the participating devices from saying no to transmitting or receiving a message.
- **Revocability** – revoking privileges of participating devices in the event of a malicious D2D service.

4. Preliminaries and Notations

4.1. Encryption and Digital Signature Schemes

There are many digital signature schemes, with ElGamal, elliptic curve and Schnorr algorithms being the most popular of them. The elliptic curve digital signature algorithm (ECDSA) is based on the modified digital signature algorithm (DSA). It works on elliptic curves that are defined over a mathematical group and discrete logarithmic problems for its key formats. The smaller footprints and efficiency of elliptic curve cryptography have led to its widespread adoption.

ElGamal combines the discreet logarithmic problem and the algebraic properties of modular exponentiation. At the core of the algorithm is a key pair which includes a private and a public key. When the sender sends a message, a digital signature is generated for it by the private key. Verification of the signature is carried out using the public key of the signer. The three key properties that a digital signature is supposed to offer, i.e. authentication, integrity, and nonrepudiation, are ensured by the digital signature in this case. The ElGamal signature algorithm is rarely used in practice. The ECDSA and other variants are used on a much wider scale. However, it is worth mentioning that ElGamal encryption, i.e. an asymmetric key encryption algorithm, is widely used in public key cryptography. The Diffie-Hellman key exchange forms the basis of this scheme.

The Schnorr signature offers numerous advantages over ECDSA and ElGamal signatures. It is an amalgamation of these schemes that is much simpler and faster. Its proven

security record is another of its advantages, provided that a random hash function with sufficient entropy is used with a sufficiently hard elliptic curve discrete logarithm problem (ECDLP). There is no security proof for ECDSA, however there is a definitive proof for Schnorr, according to which breaking the Schnorr algorithm implies breaking the discrete logarithmic problem. The linearity property is another key advantage of the Schnorr signature.
The parameters and notations used in the paper are described in Table 1.

Table 1
Notations and descriptions used

Parameters	Description
K	Private key
R	Random nonce
G	Generator point on elliptic curve
msg	Message
H'	Hash(msg)
P	Public key ($P = K \times G$)
S	Signature
CA	Certifying authority
PD	Participating device
BS	Base station

In ElGamal and ECDSA, the need to find the signature requires a division of the random nonce. Since it is a modulo operation, performance suffers, as an extended Euclidean algorithm or Fermat's little theorem may be required, calling for plenty of multiplication operations to be performed. The Schnorr signature is linear with no modulo division, thus making the process simpler and faster. ECDSA relies on a modulo division over different random nonces, making it difficult to add up ECDSA signatures. Schnorr's linear property makes it feasible to add up Schnorr signatures. This linearity makes it possible for multiple participating entities to collaboratively produce a signature for the sum of their public keys.
Considering the above as a motivation, lightweight elliptic ElGamal Schnorr-based authentication scheme is proposed (LEES).

5. System Model

The proposed model comprises the participating devices (PD) or a relay which provides coverage and connectivity to nearby devices (Fig. 2). A base station (BS) is a high computation node deployed by the mobile network service provider. There is a certification authority (CA) which is responsible for issuing certificates to all communicating nodes. As a pre-requisite for communication, the nodes (including PD and BS) have to register with the CA being responsible for whitelisting the public keys of all the nodes in the network. Thus, to avoid the key escrow problem, PD and BS generate their private and public keys. This generation of keys is not shouldered by CA.

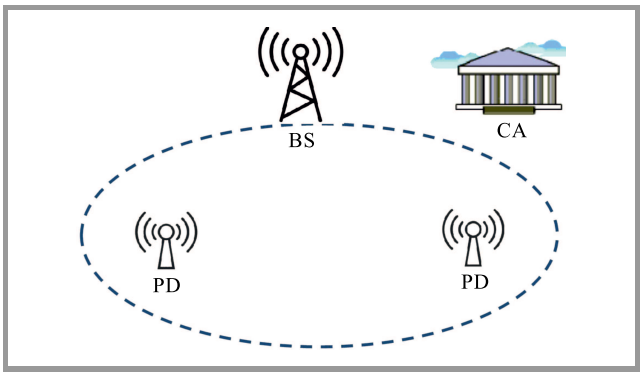


Fig. 2. System model diagram of the proposed solution.

6. Implementation

The proposed signcryption scheme is a hybrid implementation of the elliptic curve-based ElGamal encryption algorithm fused with the Schnorr digital signature scheme for enhanced security. The choice is based on the proven secure nature of the ElGamal cryptosystem and on the security of the Schnorr signature scheme. However, user identity is sent in plain text in this technique. Later it was refined with encrypted ID, but it still remains susceptible to replay attacks.
The proposed system may be split into four stages:

Stage 1 – yoke stage,

Stage 2 – entity detection stage,

Stage 3 – corroboration stage/trust establishment stage,

Stage 4 – secure data communication stage.

We assume that the 5G-AKA+ authentication and privacy preserving protocol is used prior to this stage or in accordance with this stage, so as to establish a foundation for secure communications ahead. We use the ElGamal public key as the encryption and decryption algorithm. The parameters generated at this stage are listed in Table 2.

Table 2
Parameters generated at the setup stage

Parameter	Description
a	A huge prime number
b	A huge prime factor of $(a-1)$
c	An integer which is of the order b mod a
$h()$	A secure one-way hash function
KH	One-way hash function with a key K
(E,D)	E encipher and D decipher

Here, two keys using public key infrastructure (PKI) are generated initially by the base station and participating devices and certified by the certifying authority. This stage can then be split as shown below:

KeyGen PD. Let K_{PD} and P_{PD} be the private and public key of the PD (sender) certified by CA:

$$PD = (K_{PD}, P_{PD}) , \tag{1}$$

where P_{PD}= C^{-K_PD} mod a.

Initially, when PD wants to communicate with BS, PD computes points on the elliptic curve (EC) which are then broadcast with base point F after periodic intervals. To compute K_{PD} from a field J_N $(1 < K_{PD} < N) J_N$, the public K_{PD} is computed as:

$$P_{PD} = K_{PD} \cdot G . \tag{2}$$

KeyGen BS. Let K_{BS} and P_{BS} be the private and public key of the BS (receiver) certified by CA:

$$BS = (K_{BS}, P_{BS}) , \tag{3}$$

where P_{BS}= $C^{-K_{BS}}$ mod a.

The signcryption stage. Calculate:

$$K = h\left(P_{BS}^{K_{BS}}\right) \bmod a . \tag{4}$$

Divide K into K_1 and K_2 of suitable length and:

$$x = KH_{K1}(msg) , \tag{5}$$

$$y = r + (d \cdot K_{PD}) \bmod b , \tag{6}$$

$$z = E_{K1}(msg) , \tag{7}$$

which is the ElGamal encryption of the plaintext with K_1 key.

A time stamp (TS) is added and the PD sends (x, y, z, TS) to the BS.

The unsigncryption stage. To recover the plaintext *msg* from (x, y, z, TS), the base station calculates the hash function:

$$K = \mathrm{hash}\left(C^s \cdot P_{PD}^d\right)^{K_{BS}} \bmod a . \tag{8}$$

Split K in K_1 and K_2, compute:

$$msg = DK_1(z) \tag{9}$$

where *msg* is assumed to be a valid message if $KH_{K2}(msg) = x$.

As mentioned above, for encryption and decryption algorithms, we use the ElGamal public key. For public key cryptography needs, this is an asymmetric key encryption algorithm. It has two other advantages. Firstly, since it is based on solving the difficult discrete logs in a large prime modulus, its security is tight. Secondly, its encryption is probabilistic, which ensures that the same plaintext produces a new cipher text every time encryption occurs. The algorithm may be divided into three key elements: key generator, encipher, and decipher.

- Key generation. Select a random x from $1, \ldots, b$ and determine $h = c^x$.
- Encipher. Select y from $1, \ldots, b-1$, determine $C_1 = C^x \bmod a$ nd $A = P_{PD}^x \bmod b$. Change the secret message *msg* into a factor msg' of B. Determine $C_2 = (A.msg) \bmod b$. The cipher text is (C_1, C_2).
- Decipher. Calculate the shared secret $S = C_1^x$ and compute $msg' = C_2 S^{-1}$, where S^{-1} is the inverse of S in group B.

The message is then converted back into plaintext message *msg* by the BS.

In the event of any dispute, the BS just needs to change a valid cipher text into a signature which can be verified publicly to satisfy a third-party certifying authority that the cipher text is, as a matter of fact, generated by the participating device only.

7. Result and Analysis

The proposed scheme scores well on both computation cost and memory consumption. This is due to the fact that LEES uses ECC, as this operation forms an Abelian group due to it being performed in a finite field. Hence, both addition and multiplication of points are different and faster from normal multiplication. It is lightweight and does not suffer from a lower security level in spite of ECC keys being shorter compared to RSA/DH, as shown in Fig. 3. Therefore, LEES usage leads to lower computational overheads and eases the handling of keys, since the number of bits required is lower compared to RSA/DH. This leads us to conclude that even memory consumption and network traffic will be reduced significantly as a lower number of bits is sent.

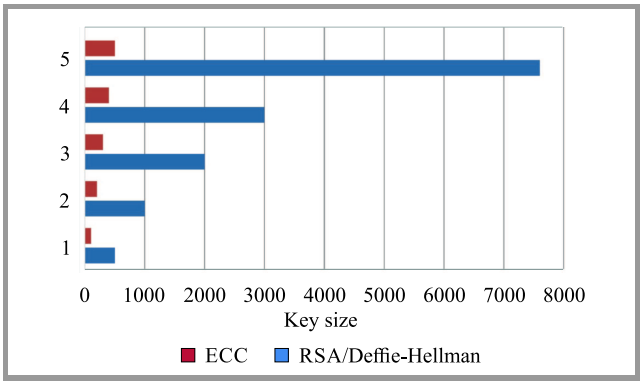


Fig. 3. Comparison of key size of ECC and RSA/DH.

In the proposed scheme, we used 5G-AKA+ for primary authentication of a PD before initiating the D2D setup stage. This framework, available in 5G networks, offers unlinkability and also satisfies both efficiency and design drawbacks in 5G-AKA. Moving forward, secondary authentication is performed by the CA based on the private and public keys generated by the PD and BS. This is followed by signcryption and unsigncryption stages which encrypt/decrypt messages using the ElGamal algorithm and verify

the Schnorr digital signature before the receiver receives any data. Thus, authentication takes place at each stage, thereby ensuring secure D2D communications.

Except for PD, any attacker (including BS) cannot forge a valid cipher text (x, y, z) for any message *msg*, such that the verification equations mentioned above are all satisfied. Also, except for the designated receiver, i.e. the BS, no third-party can derive the message *msg* from the cipher text (x, y, z).

5G networks inherently provide encrypted identity and thus anonymity to PD. In addition, the private and public key of PD and BS is unique and certified by CA. Further data sent by PD is encrypted with private key of BS, which offers further anonymity to PD.

Once BS reveals a triple (msg, y, z), anyone can verify that (y, z) is PD's signature. Hence, an authority may settle any potential disputes between PD and BS.

LEES authentication and data encryption processes are designed using a lightweight cryptographic protocol. It is lightweight, since it uses a ECC-based public key cryptosystem which utilizes only a 256 bit key compared to a 1024 bit RSA key offering the same level of security.

Since all 5G IoT devices will be resource-constrained, the lightweight cipher used in LEES works efficiently by ensuring data confidentiality, integrity and authentication.

7.1. Analysis Based on Security Against Key Attacks

Impersonation attack. LEES is robust against impersonation attacks. This is a common problem in D2D communications with devices acting as relays. As per the proposed scheme, this is avoided in a two-step approach. Firstly, the PD will encrypt the message by its private key. The receiving BS fetches the public key P_{PD} from the whitelist maintained by the CA. This does not allow it to open the message and decryption fails. Thus, LEES is secure against impersonation attacks.

Baby step and giant step (BSGS) method. The solution proposed by Shank helps solve the DLP problem by focusing on collisions and by minimizing complexity at approx. by $\sqrt{N}$ times, which is almost 50% of the original size. If a 192-bit curve is used, considering $\sqrt{N}$, we get 10^{16} points, which will require 10^{21} bytes for storing the hash – a result that is much lower compared to the proposed scheme which works out to the order of 10^{156} attempts, thus making it impossible to guess the key.

Brute force attack. An intruder may lay its hands on the public key of PD and BS and, say, also the base point F on the elliptic curve EC. If it obtains access to the private key of BS (P_{BS}), then the entire network is compromised. If P_{PD} is accessed, the small network that PD is serving gets compromised. However, since LEES uses ECC for key selection, is based on DLP and the key size is over 384 bits, the network is secured against such attacks.

Pollard's Rho method. It is a lightweight attack. It performs two operations called parallelization and random walk. It attempts to reduce to the square root of the attack to find the secret key. Since LEES uses ECC, it needs to be mentioned that the key size of ECC is ~ 571 bits, which is equal to around 15360 bits of RSA as shown in Table 3. Therefore, considering the first key and taking its square root, the problem becomes unfeasible to solve. As seen in Fig. 3, the size of RSA increases significantly compared to ECC, which increases moderately.

Table 3
LEES vs. RSA/DH key comparison

LEES [bits]	RSA/DH [bits]
112	512
224	2048
571	15360

Relay attack. Here, the intruder saves the accessed message and sends it at some other time intervals, leading to great losses. To avoid such attacks, a time stamp is introduced in (x, y, z, TS) during the signcryption stage. Based on the session, the time to live (TTL) is calculated and, if the difference between the current time and the message time is greater than TTL, the message is considered to be a fresh message. Otherwise, it is a stale message.

7.2. Analysis of Authentication Overhead

Generally, four steps are involved in a normal authentication process to ensure a secure transmission. However, in the proposed scheme, communication may be of the occur in one-step only, or in two-steps maximum, to deliver the cipher text and to perform authentication. Therefore, a comparison between LEES and other contemporary schemes shows that the communication overhead and, thereby, network bandwidth are highly reduced, by up to four times, without compromising security. Hence, the solution is lightweight. Also, a comparison between LEES and other contemporary designs, such as the ultra-lightweight mutual authentication protocol (ULMAP) and the session initiation protocol (SIP), shows that the proposed scheme outperforms the above solutions in term of the number of messages exchanged with almost the same level of trust and security, as shown in Fig. 4.

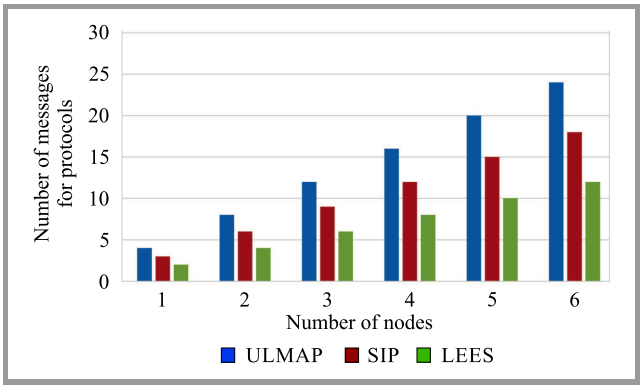


Fig. 4. Comparison of authentication message overhead.

8. Conclusion

The scheme developed has been analyzed in terms of its key parameters: computational overhead, security, key attacks and authentication overhead. It is observed that LEES requires less computational resources and eases the handling of keys, since the number of bits required is lower by a factor compared to RSA/DH. This leads us to conclude that even memory consumption and network traffic will be significantly reduced, as lower number of bits are sent. The analysis of the proposed scheme focusing on attack resiliency shows that it offers, when implemented, good levels of authentication, data confidentiality, anonymity and efficiency. In terms of protection against attacks, it has been determined as being secure against most attacks. Finally, a comparison between LEES and other algorithms showed that the communication overhead and, thereby, network bandwidth are highly reduced (by as much as four times) with LEES, without compromising security.

References

[1] G. Wang, R. H. Deng, D. Kwak, and SangJae Moon, "Security analysis of two signcryption schemes", in *Information Security 7th International Conference, ISC 2004, Palo Alto, CA, USA, September 27-29, 2004, Proceedings*, K. Zhang and Y. Zheng, Eds. *LNCS*, vol. 3225, pp. 123–133. Berlin: Springer, 2004 (DOI: 10.1007/978-3-540-30144-8_11).

[2] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in LTE-advanced networks: A survey", *IEEE Commun. Surv. and Tutor.*, vol. 17, no. 4, pp. 1923–1940 2015 (DOI: 10.1109/COMST.2014.2375934).

[3] L. M. Theobald *et al.*, "Device-to-device discovery for proximity-based service in LTE-advanced system", *IEEE J. on Selected Areas in Communication*, vol. 33, no. 1, pp. 55–66, 2015 (DOI: 10.1109/JSAC.2014.2369591).

[4] Y.-S. Shiu, S.-Y. Chang, H.-C. Wu, S. C.-H. Huang, H.-H. Chen, "Physical layer security in wireless networks; A tutorial", *IEEE Wireless Communication*, vol. 18, no. 2, pp. 66–74, 2011 (DOI: 10.1109/MWC.2011.5751298).

[5] A. S. Khan, Y. Javed, J. Abdullah, J. M. Nazim, and N. Khan, "Security issues in 5G device to device communication", *Int. J. of Comp. Sci. and Netw. Secur.* (*IJCSNS*), vol. 17 no. 5, pp. 366–375 [Online]. Available: http://paper.ijcsns.org/07_book/201705/20170550.pdf

[6] W. Stallings, *Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud*. Boston: Addison-Wesley Professional, 2015 (ISBN: 9780367378158).

[7] H. Lazrag, H. Chaibi, S. Rachid, and M. D. Rahmani, "An optimal and secure routing protocol for wireless sensor networks", in *Proc. of 6th Int. Conf. on Multime. Comput. and Systems ICMCS 2018*, Rabat, Morocco, 2018 (DOI: 10.1109/ICMCS.2018.8525911).

[8] M. Alenezi, K. Almustafa, and M. Hussein, "On virtualization and security-awareness performance analysis in 5G cellular networks", *J. of Engin. Sci. and Technol. Rev.*, vol. 11, no. 1, pp. 199–207, 2018 (DOI: 10.25103/jestr.111.24).

Javeria Ambareen, a passionate security techie, has over a decade of academic and industry experience. Having an M.Tech. degree, her key areas of interest include IoT security, application security, automation and machine learning.

E-mail: javeriaster@gmail.com
School of Computing and Information Technology
REVA University
Bangalore, India

M. Prabhakar received his M.Sc. and Ph.D. degrees in Computer Engineering from Anna University, Chennai. He has 21 years of teaching experience and is currently working as an Associate Professor at the School of Computing & Information Technology, REVA University, Bangalore, India. His areas of research interest include adhoc networks and cybersecurity.
E-mail: prabhakar.m@reva.edu.in
School of Computing and Information Technology
REVA University
Bangalore, India

Tabassum Ara is a Research Scholar in Computer Science at Reva University, Bangalore, Karnataka, India. With 20 years of academic experience, she holds B.Eng., M.Sc. and M.Tech. degrees and is an Assistant Professor at the Department of Computer Science at HKBK College of Engineering in Bangalore, India. Her research interests focusing on IoT, security and WSN.
E-mail: tabuara@gmail.com
School of Computing and Information Technology
REVA University
Bangalore, India

Paper

Residual Energy-Aware Clustering Transformation for LEACH Protocol

P. Ullas[1] and K. S. Shivaprakasha[2]

[1] *P. E. S. College of Engineering, Mandya, India*
[2] *N. M. A. M. Institute of Technology, Nitte, India*

https://doi.org/10.26636/jtit.2021.147420

Abstract—Energy is one of the crucial performance parameters in wireless sensor networks (WSNs). Enhancement of network lifetime is an important consideration as well. Low energy-aware clustering hierarchy (LEACH) is one of the protocols proposed for WSNs. In this paper, a residual energy-aware clustering transformation protocol for LEACH (REACT-LEACH), enhancing performance of LEACH by introducing a clustering mechanism, is proposed. The proposed cluster head (CH) rotation and cluster reformation processes are more effective in REACT-LEACH, as residual energy is considered to be one of the metrics. Performance of REACT-LEACH is validated based on simulations.

Keywords—clusters formation, cluster head, residual energy, wireless sensor network.

1. Introduction

Wireless sensor networks (WSNs) have been attracting special interest from the research community in recent years due to their applicability in almost all fields of research. As the nodes of WSNs are battery-powered, effective utilization of energy available in these networks is a primary goal that needs to be achieved in order to ensure long network lifetime. Since frequent recharging or replacement of batteries is infeasible in some WSN applications, energy efficiency has to be assured through proper protocol design [1], [2]. Routing is a process of establishing a path from the source node to the sink. Optimal route selection depends also on the placement of nodes within WSNs. If the nodes are deployed densely, the chance of establishing alternative paths reaching the destination is higher. Thus, better energy efficiency is assured. On the contrary, if the nodes are deployed sparsely, the number of alternate paths leading to the destination will be lower, which limits the options of selecting the most energy efficient path. It also needs to be borne in mind that nodes located close to the base station (BS) are usually generally overloaded compared with other nodes in the network [3].

Cluster-based protocols have proved to offer better energy efficiency. However, cluster reformation process drains batteries. The scenario is even less favorable when the nodes are mobile, as this means that frequent topology changes are encountered. Most protocols consider reforming all clusters whenever the residual energy of any of the cluster heads (CHs) falls below the threshold value. In this paper, the REACT-LEACH protocol is proposed, relying on rotating the CH within the cluster, without actually modifying the existing cluster. The reformation of the cluster is performed only when the nodes are out of the communication range.

The paper is organized as follows. Section 2 presents the literature review. Section 3 describes the LEACH protocol. Section 4 gives an insight into the proposed REACT-LEACH protocol. Section 5 presents the results and an analysis thereof, while Section 6 concludes the paper.

2. Related Work

To overcome problems with energy, different routing protocols have been proposed by researchers [4]. For example, paper [1] proposes an energy efficient direction-based PDORP routing protocol for WSNs, where the fixed cooperative-based clustering approach is used. This protocol is designed for fixed WSNs. A comparison with LEACH is made and the protocol proves to behave well in homogeneous distributed clustering. Selection of the CH is an issue of key importance in the clustering algorithm.

The authors of [5] present a survey concerned with routing protocols for wireless sensor networks. They showcase, inter alia, a delay aware routing protocol. In this case, sensor nodes are homogeneous and static in nature with very limited mobility. This protocol is designed to manage neighboring nodes. Each node maintains a forwarding table to select an efficient path for transmitting data between the node and the BS. If the node is within BS range, data is transferred directly. Else, multi-hop communication takes place.

In article [6], the position responsive routing protocol (PRRP) is presented, showcasing a novel approach to CH selection. Here, the nodes are distributed in the form of a grid – and approach that is required in such applications as disaster management, forest fire surveillance sys-

tems, etc. Energy consumption is directly proportional to the communication distance. Each processing round consists of four phases.

In the energy-aware routing protocol [7], energy optimization is performed by relying on hybrid algorithms, i.e. the dynamic source routing (DSR) protocol is used and is found to be more suitable in terms of low energy density. The goal is to identify dead nodes and choose a different path suitable for the transmission.

Article [8] introduces LEACH, an application-specific protocol architecture for wireless micro-sensor networks. The mathematical formulas of all the probabilities that are considered for the process of cluster formation and CH selection in the LEACH protocol are clearly presented in this paper.

3. LEACH Protocol

The LEACH protocol aims to form clusters in a network based on such parameters as the level of the radio signal received, the number of times the node has been a CH, and the average network energy level. In fact, LEACH incorporates an algorithm which is distributed and each node decides, autonomously, whether to become a CH or not depending on a random value. This value is a probability and covers many parameters, such as distance of the node from BS, the number of times the node has been a CH, residual energy, etc. This randomly chosen value is compared with a defined threshold. If the chosen value is below the threshold, then the node becomes a CH. Figure 1 shows communication between the nodes and BS during the cluster formation process.

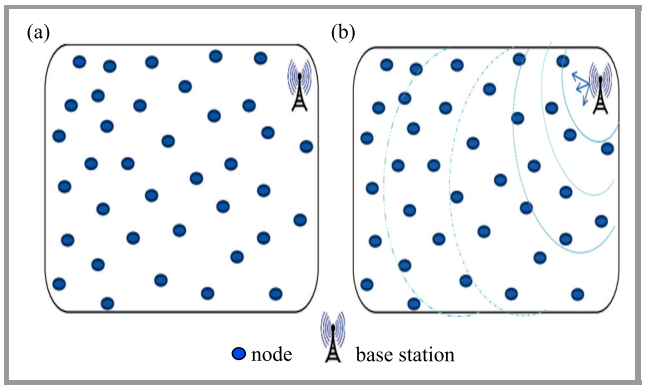


Fig. 1. Distribution of nodes (a) and nodes communicating with BS for CH selection process (b).

Once the CHs are formed, they broadcast "hello" messages. The CH nodes may receive hello messages from more than one CH. If such messages are received, they decide to join the appropriate cluster based on their distance from the respective CH. This process is as shown in Fig. 2. The algorithm runs multiple times, and different nodes assume the role of CHs in each round. The threshold value is calculated in each round using the following formula [9]:

$$T_n = \frac{p}{1 - P \cdot \left(r \bmod \frac{1}{p}\right)}, \tag{1}$$

where T_n is the threshold value for n rounds, P is the percentage of CHs in the network, r is the current round number, and G stands for the set of nodes that were not acting as CHs in the $(\frac{1}{p})$ preceding rounds.

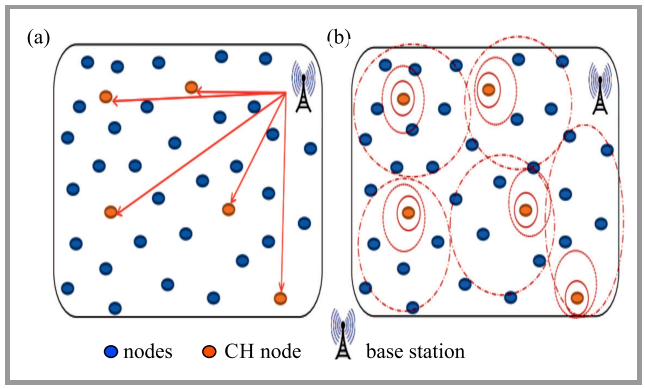


Fig. 2. CH selection (a) and advertisement from a CH to nearby nodes to form a cluster (b).

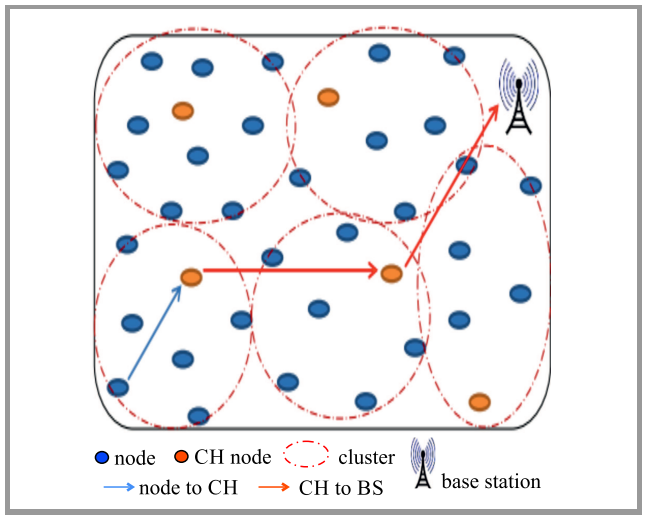


Fig. 3. Communication between nodes and BS through CH.

Data sent from any node to BS passes through the CHs. Cluster members convey the information to the respective CHs, and the CHs then communicate with the BS. The process is shown in Fig. 3, where the control flow is indicated. Various improvements to LEACH protocol have been proposed by researchers, making it more suitable for specific applications [10]. The following aspects may be addressed to ensure further performance enhancements:

- a node that is located farther away from the CH dies faster when compared to a node located close by,
- CH selection is not dependent on the residual energy and is a random process,

- energy losses are incurred during the reformation of clusters and CHs, in each round,
- there is chance of having a maximum number of CHs concentrated in one part of the network, which leads to some nodes not having any CHs in their vicinity,
- the routing table must be updated for each round,
- selection of an efficient path to the BS from every CH should be performed in each round.

Based on those potential areas of development, a decent number of variants of the LEACH protocol has been proposed in the literature [11]. In most of them, the criterion for the rotation of CHs was addressed better, with the residual energy of nodes taken into consideration [12], [13].

4. Residual Energy-Aware Clustering Transformation for LEACH

As discussed in Section 3, CH reformation taking place in LEACH is incorporated in each round. This process requires a lot of control packet exchanges and, thus, is energy-intensive. The proposed REACT-LEACH approach may be a good solution to overcome this obstacle. REACT-LEACH also operates in rounds with the initial round being exactly the same as in LEACH [14]. The initial cluster formation and CH selection procedures are depicted in Figs. 1–3. During the first round, REACT- LEACH is similar to the LEACH protocol, but later on the new algorithm depends purely on the residual energy of the current CHs for their reselection and for the cluster reformation process.
In REACT-LEACH, the process of CH reselection and cluster reformation depends on residual energy and distance. Once CHs have been selected, clusters are set based on the distance between the nodes and CH. Next, the nodes within the cluster transfer data to the respective CHs the CHs transfer the same to the BS – meaning that the process is similar to that relied upon in LEACH. The CH reselection and cluster reformation process is only performed if:

1. residual energy of the current CH falls below the threshold value,
2. the CH has moved far away from the within its cluster,
3. the nodes have moved far away from their CH.

In scenario 1, the CH collects residual energy data from all its cluster members [15]. The node with the maximum residual energy level is selected to be the new CH. The newly appointed CH reforms the cluster based on its communication range. The new CH broadcasts an advertisement packet within its transmission range.
In scenario 2, due to mobility-related considerations, there may be a chance that the CH moves away from the cluster members. When the CH moves away from any nodes, then the distance ones attempt to connect to their nearby clusters by sending a request message to the respective CHs. This process occurs in each round, thus no node is left out within the network.
During the selection of a new CH within the cluster, the current CH should collect residual energy data from all cluster nodes. This process is shown in Fig. 4. The control flow is indicated by arrows.

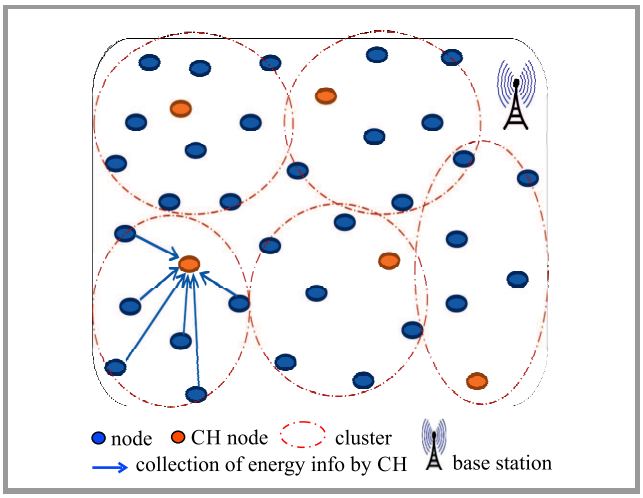


Fig. 4. Residual energy information flow from nodes to CH for new CH selection.

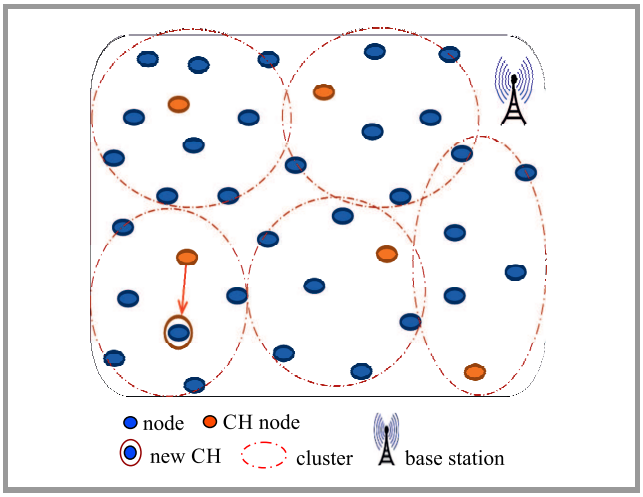


Fig. 5. New CH selection in REACT-LEACH.

The new CH is selected based on the highest residual energy level amongst the nodes within a given cluster. Then, the cluster will be reformed, with the location of the new CH taken into consideration, depending on the distance between cluster nodes. During this CH reselection process, only the nodes within the current clusters may become a CH. Nodes that are within the communication range of the new CH form a new cluster [16]. Any node that is left out during the reformation of clusters may have to join the nearest CH. This process is depicted in Fig. 5.

In scenario 3, similarly to scenario 2, mobility may lead to a node moving away from its cluster. The node that is left out of the cluster tries to communicate with the nearest CH and joins the respective cluster. Figure 6 shows a detailed flow of the REACT-LEACH protocol.

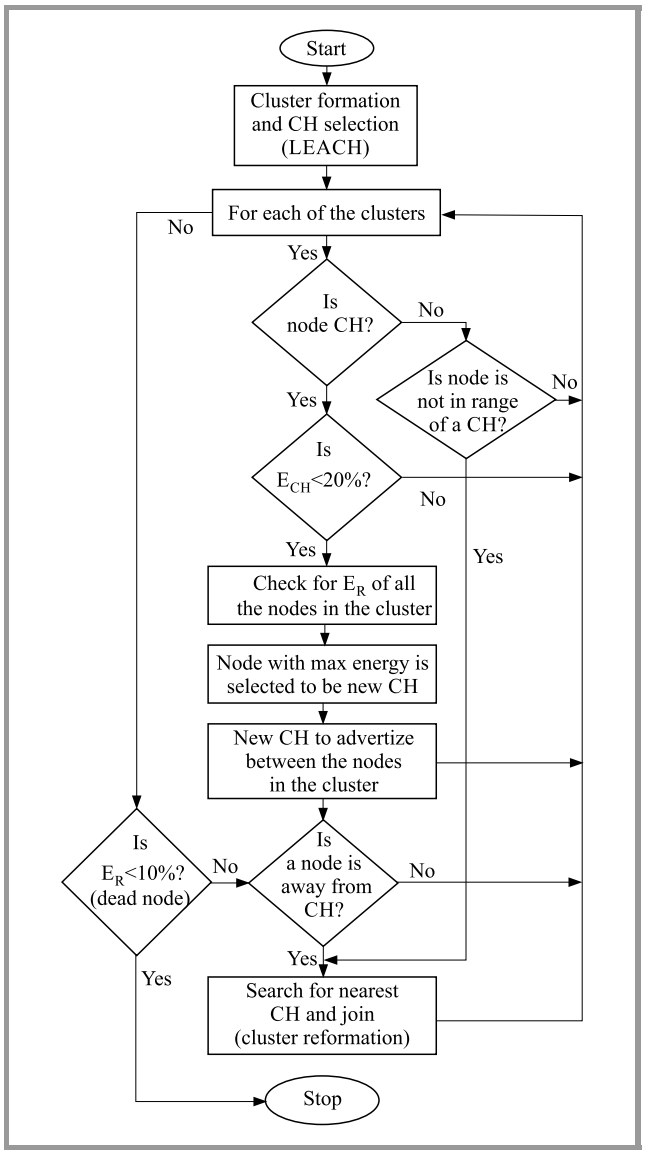


Fig. 6. Flow chart of REACT-LEACH.

5. Simulation Results and Analysis

Performance of the proposed REACT-LEACH approach is compared with the traditional LEACH protocol based on key performance parameters. Table 1 shows the parameters have been taken and the average values are considered to verify considered while validating performance using the NS-3 software simulator.

In this paper, performance of REACT-LEACH and LEACH protocols is compared through simulations. 5 trails were taken and the average values are considered to verify overall network performance levels.

Table 1
Parameters used in simulations

Simulation parameters	Specifications
Simulation environment	NS-3
Number of nodes	70
Antenna type	Omnidirectional
Network area	400×400 m
Deployment of nodes	Random
Simulation rounds	2000 rounds
Initial energy of nodes	0.5 J
Mobility pattern	Random

The comparison shown in Fig. 7 proves that the REACT-LEACH protocol is characterized by a constant number of CHs during the entire simulation cycle [17], [18]. Despite the fact that CHs change depending on residual energy levels, the total number of CHs remains constant. Once the number of CHs is selected by the random procedure in LEACH, it is maintained for all subsequent operations.

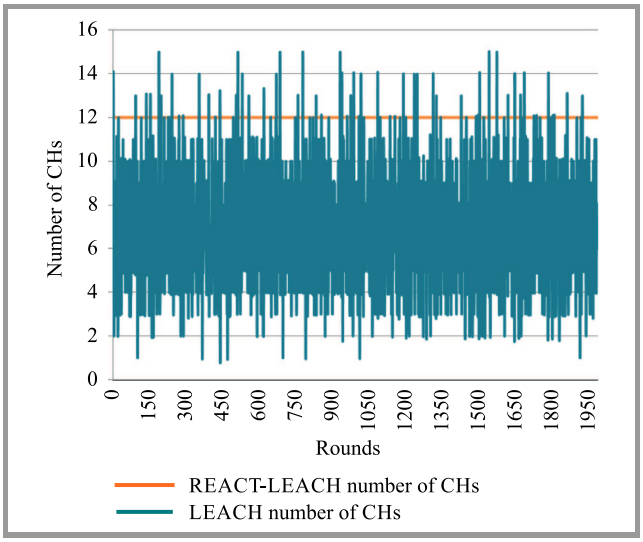


Fig. 7. Comparison of number of CHs that are selected in each round.

Table 2 compares performance of the proposed REACT-LEACH approach with the traditional LEACH protocol in terms of the average energy level at each node and the total energy within the network.

The average energy level of each node, after individual operation rounds, may be observed in Fig. 8. The REACT-LEACH approach seems to be more energy-efficient when compared to the LEACH protocol. The drainage of energy experienced during cluster formation in each round in LEACH is reduced in REACT-LEACH, because cluster reformation takes place only when there is a need. After completion of the simulation, the average percentage of residual energy in the network is found to equal 90% of the initial value in REACT-LEACH, whereas in LEACH it amounts to 28% only.

Table 2
Comparison of LEACH and REACT-LEACH

Round	REACT-LEACH				LEACH			
	Avg energy [J]	Total energy [J]	Number of CHs	Percent of CHs	Avg energy [J]	Total energy [J]	Number of CHs	Percent of CHs
1	0.999663	34.9882			0.999666	34.9883	13	0.19
200	0.989061	34.6171			0.928034	32.4812	4	0.06
400	0.978459	34.2461			0.856415	29.9745	7	0.10
600	0.967857	33.875			0.784272	27.4495	5	0.07
800	0.957255	33.5039	12	0.26	0.712611	24.9414	8	0.11
1000	0.946653	33.1329			0.641585	22.4555	8	0.11
1200	0.936051	32.7618			0.570115	19.954	6	0.09
1400	0.925449	32.3907			0.497864	17.4252	10	0.14
1600	0.914847	32.0196			0.426424	14.9248	6	0.09
1800	0.904028	31.641			0.354799	12.418	3	0.04
1999	0.893322	31.2663			0.283827	9.93396	6	0.09

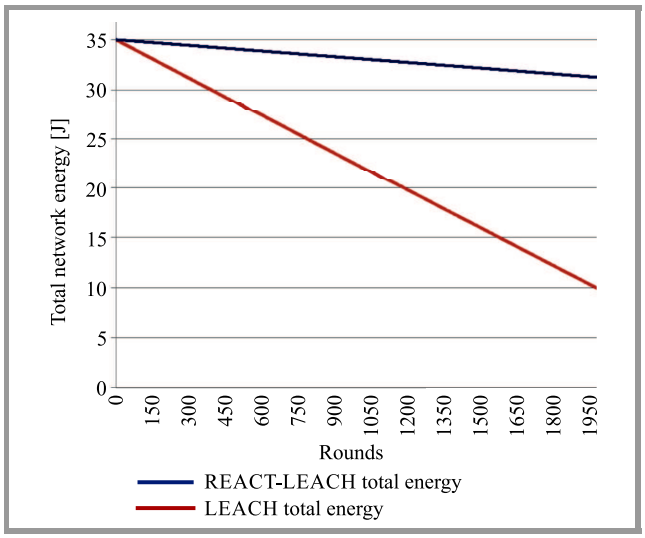


Fig. 8. Comparison of node energy levels [%].

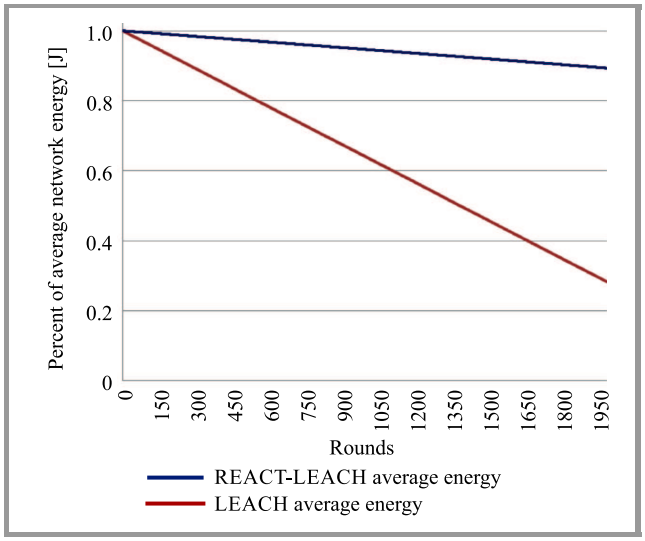


Fig. 9. Comparison of total energy of the network.

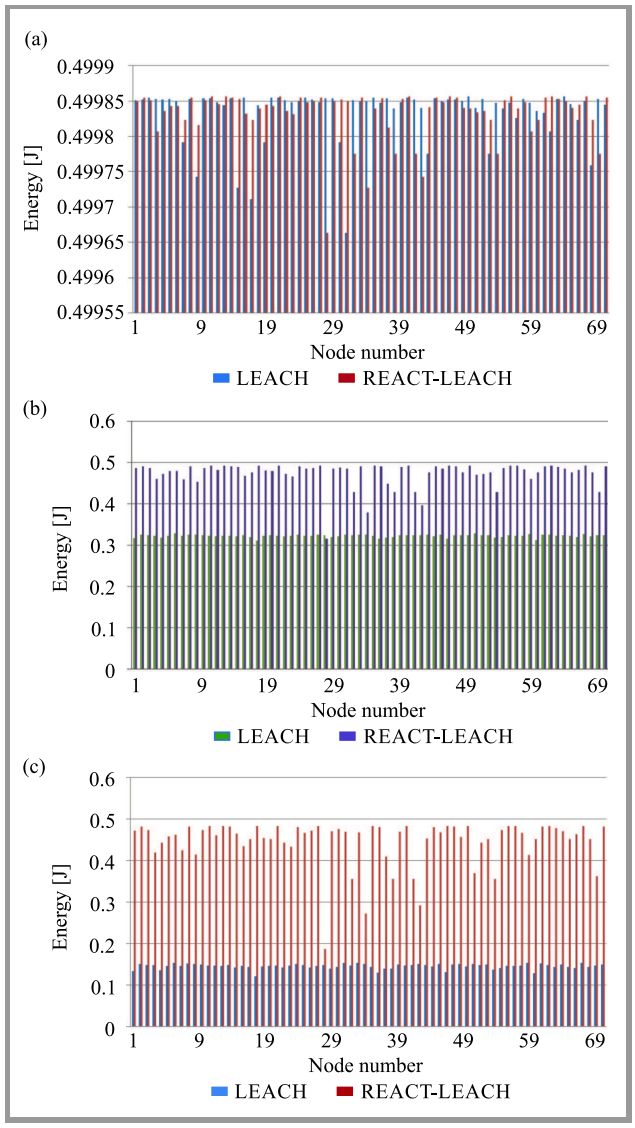


Fig. 10. Residual energy of each nodes after: (a) one round, (b) 1000 rounds, and (c) 2000 rounds.

The total energy consumed by the network (Fig. 9) is directly dependent on the nodes' energy and, thus, the total energy drain of the network is comparable to the graph shown in Fig. 8. Energy consumption in reduced significantly in REACT-LEACH.
Figure 10 shows the residual energy level of each node after different operation rounds. It is proved that the REACT-LEACH approach assures a better energy balance between all nodes, as each of the nodes serves as a CH at some point of time. The REACT-LEACH protocol ensures also that higher residual energy levels are maintained in all nodes, and thus leads to an enhancement in network lifetime.

6. Conclusion

The REACT-LEACH protocol offers better energy efficiency when compared with the well-known LEACH approach. The average residual energy at each node in REACT-LEACH is around 90% after 2000 rounds of simulation, versus 28% in the case of LEACH. This is mainly because of the fact that multiple cluster formation and CH selection processes have been completely eliminated in REACT-LEACH. In the proposed protocol, control-related communication required in cluster formation and CH selection processes occurs only when there is a need, instead of taking place in each round. Validation is performed based on simulations, with low mobility of the nodes considered.

References

[1] G. S. Brar *et al.*, "Energy efficient direction-based PDORP routing protocol WSN", *IEEE Access*, vol. 4, pp. 3182–3194, 2016 (DOI: 10.1109/ACCESS.2016.2576475).

[2] Sh. Chen, J. Yao, and Y. Wu, "Analysis of the power consumption for wireless sensor network node based on Zigbee", *Procedia Engin.*, vol. 29, pp. 1994–1998, 2012 (DOI: 10.1016/j.proeng.2012.01.250).

[3] Akila and U. Maheswari, "A survey on recent techniques for energy efficient routing in WSN", *Int. J. of Sensors and Sensor Netw.*, vol. 6, no. 1, pp. 8-15, 2018 (DOI: 10.11648/j.ijssn.20180601.12).

[4] M. Elshrkawey, S. M. Elsherif, and M. Elsayed Wahed, "An enhancement approach for reducing the energy consumption in wireless sensor networks", *J. of King Saud Univer. – Comp. and Inform. Sciences*, vol. 30, no. 2, pp. 259–267, 2018 (DOI: 10.1016/j.jksuci.2017.04.002).

[5] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks", *Ad Hoc Networks*, vol. 3, no. 3, pp. 325–349, 2005 (DOI: 10.1016/j.adhoc.2003.09.010).

[6] B. Bhuyan and N. Sarma, "Performance comparison of a QoS aware routing protocol for wireless sensor networks", *Commun. and Netw.*, vol. 8, no. 1, pp. 45–55, 2016 (DOI:10.4236/cn.2016.81006).

[7] N. Zaman, L. T. Jung, and M. M. Yasin, "Enhancing energy efficiency of wireless sensor network through the design of energy efficient routing protocol", *J. of Sensors*, vol. 2016, Article ID 9278701, pp. 1–16, 2016 (DOI: 10.1155/2016/9278701).

[8] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks", *IEEE Trans. on Wirel. Commun.*, vol. 1, no. 4, pp. 660–670, 2002 (DOI: 10.1109/TWC.2002.804190).

[9] A. S. Raghuvanshi, S. Tiwari, R. Tripathi, and N. Kishor, "Optimal number of clusters in wireless sensor networks: An FCM approach", in *Proc. of In. Conf. on Comp. and Commun. Technol. ICCCT 2010*, Allahabad, India, 2010 (DOI: 10.1109/ICCCT.2010.5640391).

[10] S. El Khediri, N. Nasri, and A. Kachouri, "Fixed competition-based clustering approach wireless sensor network", in *Proc. of The Int. Conf. on Adv. Wirel., Inform., and Commun. Technol. AWICT 2015*, Sousse, Tunisia, 2015 [Online]. Available: https://cedric.cnam.fr/fichiers/art_3549.pdf

[11] Ma. Tabassum *et al.*, "An energy aware event-driven routing protocol for cognitive radio sensor networks", *Wireless Networks*, vol. 22, pp. 1523–1536, 2016 (DOI: 10.1007/s11276-015-1043-8).

[12] A. Panchal, L. Singh, and R. K. Singh, "RCH-LEACH: Residual energy based cluster head selection in LEACH for wireless sensor networks", in *Proc. of Int. Conf. on Elec. and Electron. Engin. ICE3 2020*, Gorakhpur, India, 2010, pp. 322–325 (DOI: 10.1109/ICE348803.2020.9122962).

[13] P. Ullas and K. S. Shivaprakasha, "Smart clustering LEACH (SC-LEACH) protocol for improved network survivability in wireless sensor networks", *Int. J. of Recent Technol. and Engin*. (*IJRTE*), vol. 8, no. 2, pp. 6106-6110, 2019 (DOI: 10.35940/ijrte.B3858.078219).

[14] M. J. Handy, M. Haase, and D. Timmermann, "Low energy adaptive clustering hierarchy with deterministic cluster-head selection", in *Proc. of 4th Int. Worksh. on Mob. and Wirel. Commun. Netw.*, Stockholm, Sweden, 2002 (DOI: 10.1109/MWCN.2002.1045790).

[15] M. B. Yassein, Z. Jaradat, N. Hijazi, and C. Mavromoustakis, "New load balancing algorithm for LEACH protocol (F-VCH LEACH)", *Sensors & Transducers*, vol. 145, no. 10, pp. 172–182, 2012 [Online]. Available: http://www.scopus.com/inward/record.url?scp=84873976856

[16] T. S. Panaga and J. S. Dhillon, "Dual head static clustering algorithm for wireless sensor networks", *AEU – Int. J. of Electron. and Commun.*, vol 88, pp. 148–156, 2018 (DOI: 10.1016/j.aeue.2018.03.019).

[17] B. A. Sabarish and R. Lavanya, "Modified LEACH protocol for wireless sensor network", *Int. J. of Comp. Appl.*, vol. 62, no. 3, 2013 (DOI: 10.5120/10057-4648).

[18] K. S. Shivaprakasha, M. Kulkarni, and N. Joshi, "Improved network survivability using multi-threshold adaptive range clustering (M-TRAC) algorithm for energy balancing in wireless sensor networks", *J. of High Speed Networks*, vol. 19, no. 2, pp. 99–113, 2013 (DOI: 10.3233/JHS-130466).

[19] M. Pandey, L. K. Vishwakarma, and A. Bhagat, "An energy efficient clustering algorithm for increasing lifespan of heterogeneous wireless sensor networks", in *Smart and Innovative Trends in Next Generation Computing Technologies. NGCT 2017*, P. Bhattacharyya, H. Sastry, V. Marriboyina, and R. Sharma, Eds. *Communications in Computer and Information Science*, vol. 828. Singapore: Springer, 2018 (DOI: 10.1007/978-981-10-8660-1_20).

[20] W. Jerbi, A. Guermazi, and H. Trabelsia, "A routing protocol Orphan-LEACH to join orphan nodes in wireless sensor network", *Comp. Sci. & Informa. Technol.* (*CS & IT*), pp. 135–147, 2016 (DOI:10.5121/csit.2016.60412).

[21] K. S. Shivaprakasha and M. Kulkarni, "Energy efficient routing protocols for wireless sensor networks: A survey", *Int. Rev. on Comp. and Software* (*I.R.E.C.O.S.*), vol. 6, no. 6, pp. 944–949, 2011.

[22] M. Revanesh, V. Sridhar, and J. M. Acken, "Secure coronas based zone clustering and routing model for distributed wireless sensor networks", *Wirel. Pers. Commun.*, vol. 112, pp. 1829–1857, 2020 (DOI: 10.1007/s11277-020-07129-0).

[23] M. Carlos-Mancilla, E. López-Mellado, and M. Siller, "Wireless Sensor Networks Formation: Approaches and Techniques", *J. of Sensors*, vol. 2016, Article ID 2081902, 2016 (DOI: 10.1155/2016/2081902).

[24] A. Belghith and M. S. Obaidat, "Wireless sensor networks applications to smart homes and cities", in *Smart Cities and Homes*, M. S. Obaidat and P. Nicopolitidis, Eds. Morgan Kaufmann, 2016, pp. 17–40 (ISBN: 9780128034545).

[25] K. Haseeb *et al.*, "RCER: Reliable cluster-based energy-aware routing protocol for heterogeneous wireless sensor networks", *PLoS One*, vol. 14, no. 10, 2019 (DOI: 10.1371/journal.pone.0224319).

P. Ullas received his B.E. in Electronics and Communication Engineering from Bahubali College of Engineering, Visvesvaraya Technological University, Karnataka, and M.Tech. in Digital Communication and Networking from Oxford College of Engineering, Visvesvaraya Technological University, Karnataka, in 2012 and 2014, respectively. He is currently pursuing his Ph.D. in wireless sensor networks. Ullas is working as an Assistant Professor at the Department of Electronics and Communication Engineering, PES College of Engineering, Mandya, Karnataka. His areas of research interest include wireless sensor networks, as well as mobile ad-hoc networks.

https://orcid.org/0000-0002-8266-7217
E-mail: upk345@gmail.com
Department of E&CE
PES College of Engineering
Mandya, KAR, India

K. S. Shivaprakasha received his B.E. in Electronics and Communication Engineering from Bahubali College of Engineering, Visvesvaraya Technological University, Karnataka, and M.Tech. in Digital Electronics and Communication Systems from Malnad College of Engineering, Visvesvaraya Technological University, Karnataka, in 2004 and 2007, respectively. He completed his Ph.D. in Wireless Sensor Networks at the National Institute of Technology Karnataka (NITK), Surathkal, Karnataka, in 2015. Currently, he is a Professor at the Department of Electronics and Communication Engineering, N. M. A. M. Institute of Technology, Nitte, Karnataka, India. His areas of research interest include wireless sensor networks, mobile ad-hoc networks, information coding theory and cryptography.

https://orcid.org/0000-0002-5078-6078
E-mail: shivaprakasha.ks@gmail.com
Department of E&CE
N. M. A. M. Institute of Technology
Nitte, KAR, India

Paper

Political and Economic Contexts of Implementing 5G in Poland and in Selected European Countries

Urszula Soler

John Paul II Catholic University, Lublin, Poland

https://doi.org/10.26636/jtit.2021.149720

Abstract—The technology race to achieve the position of an economic leader is a phenomenon that has been taking place all over the world. The 5G technology has become a vital component of this race over the recent years. The technical capabilities it offers and the role it may play in the economy have become the subject of political debate and are at the very core of the "war for technology" between two superpowers: China and the United States. The European Union is aware of the fact that the position Europe enjoys on the international arena depends, to a large extent, on how quickly European countries will develop and implement 5G. Are individual European member states capable of seamlessly implementing the assumptions of strategies and plans concerned with the development of 5th generation technologies? Will the security of 5G networks be ensured in Europe? These are just some of the issues that are analyzed in this article with their economic and political context taken into consideration. A broader perspective is presented, with primary focus on the global geopolitical situation and on the conflict between China and the United States. The study was conducted by relying on an in-depth analysis of strategic state documents, reports drawn up by institutions tasked with implementing and monitoring the development of 5G technology, as well as literature on the subject and online resources.

Keywords—5G technology, cybersecurity, economy, Europe, mobile networks, telecommunications.

1. Introduction

Europe, as well as the rest of the world, is currently implementing fifth generation (5G) cellular technologies. After its technical standards have been agreed upon, the solution is now in the deployment and commercialization phase. 5G is a term that used to describe the fifth generation of mobile network systems supporting mobile voice and data transmission, with Internet access included [1].

5G is not a new concept. It is neither the first nor the last generation of mobile networks. It is merely an update of the technological solutions dating back to the early 1980s, first having the form of analog mobile telephony (1G) and then transformed, in 1991, into digital cellular telephony with the added feature of short text messages (SMS) (2G). The next step in the development process took place at the beginning of the 21st century, when the third generation (3G) telephony offering fast (by the then-standards) data transmission and Internet access (from 14 to 28 Mb/s). Nearly a decade later, in 2009, the 4G version enabled data to be transferred with the speeds of up to 300 Mb/s. The 5G version being implemented currently will not only speed up the transmission of data by up to 60 times (up to 20 Gb/s), but it will also support more devices per square kilometer (up to 1 million for 5G, and up to 100,000 for 4G) and will reduce transmission delays within the radio network (from 50 ms in 4G to 1 ms in 5G) [2]. While 5G is a regular technological evolution, it is said to be a breakthrough technology [3], since the qualitative change it brings does not concern its technical capabilities, but the role that the 5G may play in the ecosystem of connected devices.

The strategic character of 5G networks may be one of the reasons why this technology has become the cause of the so-called *cold war on tech* [4] and the subject of information warfare that is largely based on misinformation. For the first time in the history of the telecommunications sector, communication technology has become the subject of trade wars and geopolitical games conducted on such a wide scale. Some researchers believe that the term "cold war on tech" is abused in this context [5], as one of the conditions for the cold war to be fought is the formation of blocks, and these are not present here. However, one cannot fully agree with this statement, and the entire political and economic situation is much more complex and highly dynamic, as discussed later in the article.

Along with the start of the public discussion on 5G, fueled by disinformation activities, social protests are mounting in Europe. Protesters demand a full ban on 5G technology, and their actions are not limited to verbal objections, but extend to destroying critical infrastructure, e.g. setting fire to cellular masts.

The purpose of the article is to take a broader look at the current situation related to the deployment of 5G in Poland and in selected European countries. The analysis will focus on two contexts: economic and political. Specific phenomena related to 5G will be presented here, while a detailed

description of its rollout in individual European countries will not be provided. Examples pertaining to the individual countries are also relied upon to depict specific trends that were not observed during the deployment of previous generations of telecommunication technologies. The study was conducted by relying on an in-depth analysis of strategic state documents, reports drawn up by institutions tasked with implementing and monitoring the development of 5G technology, as well as literature on the subject and online resources.

2. Literature Review

The review of literature on political and economic contexts of implementing the 5G technology in Poland and in selected European countries aims to present the current findings in this area and to identify research limitations. Based on current reports and literature, one may determine the approach to this technology adopted by government agencies and the practices they apply while implementing the proposed strategies for regulating wireless connectivity in the context of 5G. The outcomes of the review will identify the results of the work performed in this field.

Because 5G is relatively new, there is little literature on the implementation of the technology itself, especially with an emphasis placed on the political and economic context. Most of the papers are concerned with technical issues [6]–[11]. These articles are dealing with the deployment of networks, compatibility, and other similar technical aspects. The authors of [12] have looked at 5G in more detail and undertook an analysis of the technological change taking place and of its impact on the society. Transition from 4G LTE to 5G is an archetypal example of technological change. In their analysis, they provide a complementary scenario-based assessment of 5G infrastructure strategies in relation to mobile traffic growth and potential development of the Internet of Things (IoT), Smart Cities or other technology developments (services) that rely on digital connectivity. The experience of the United Kingdom, the Netherlands and India is given as an example [13]–[15].

According to Fettweis and Alamouti [16], the 5G technology is a critical step in the development of wireless connectivity. The authors of [17] argue that 5G cellular communications will be another paradigm shift redefining our future and impacting our societies in ways which cannot be foreseen. A Qualcomm report [17] describes the benefits of 5G and focuses on the commercial benefits achieved by implementing this technology. As noted by the authors of the report, this network standard will significantly expand the potential of various spectra, and its high-band properties will help increase capacity in various areas.

Additional information about the economic context of 5G implementation is offered by a report drawn up by Deloitte [18]. It describes the current competitive trends in the market and the innovative networks that are necessary for the transition from 4G to 5G. The increase in potential profits that wireless carriers may generate thanks to the launch of the new system is a motivation to seek quick implementation of the plan [18]. As a result, the pace of the transition process may differ in individual countries and the scope of the work that needs to be performed may be different as well [19]. Therefore, a need arises to design and introduce specific laws and regulations required to implement 5G.

Some authors focus also on security-related aspects of 5G [20]. In recent years, many authors have been focusing on Huawei and on China's position in the technology race (with particular emphasis placed on cybersecurity). The authors of [21] are concerned with standardization and examine China's standard-related initiatives undertaken on the international scene, perceiving these from the point of view of techno-nationalism and of China's specific interests in Europe. In [22], Kowalski examines the Czechia–China relations in the 2010s, focusing on the theoretical framework of relationalism – an approach, adopted by the Czech ruling and financial elite in an attempt to gain economic benefits from the partnership with China. It is important to know that Prague was selected as the center of European operations of the powerful – at least until recently – CEFC China Energy corporation.

As far as the economic and political context is concerned, three articles are of particular importance. Lemstra [23] asks the following question: "What explains the success of 2G GSM and how can it be applied to create success with 5G in the European Union?" In an attempt to answer this question, the article presents two images of potential 5G futures, titled "evolution" and "revolution", serving as an input to the political debate on 5G leadership options. These images reflect two extremes in terms of potential 5G futures. Evolution follows the pattern of previous generations and current trends. Revolution clearly breaks away from these trends and from the path to leadership with 5G, as it leverages the power of standard APIs to build specific services, enabling network virtualization as the architectural backbone of 5G.

The problem of Chinese technological presence in Europe is described by Kavalski in [24]. He concentrated his attention on CEE countries. Kavalski offers a brief overview of the history of this relationship by focusing on the "17+1" mechanism. The article asks whether there is anything other than an instrumental economic justification for CEE countries' willingness to partner with China. This is also one of the topics of this article. The problem of information and trade war between the US and China is also raised by Longtin [25]. Focusing on the geopolitical context, he explores the innovation policies of Nokia and Huawei to understand how the Chinese company was able become the leader of that sector.

The number of publications on 5G is immense. However, it is difficult to identify any papers focusing on political and economic contexts of implementing the technology. Social, technical and economic challenges related to 5G are described by Pandey *et al.* [26], but in the context of India.

Insights concerning the UAE are given by AlRaeesi and Habibur [27] who describe the proactive approach to the deployment of 5G. Its comparison with the European approach and the results achieved may be a very interesting subject for future studies.

3. Political Context – Europe and Member States

It is assumed in literature that the features of 5G networks listed in the introduction, such as higher bandwidth, reduction of delays and increased number of connected devices (up to 1 million per square kilometer), will contribute to boosting competitiveness and innovative nature of the economy [28]. The use of sensor networks or autonomous vehicles and robots will accelerate the adoption of Industry 4.0. 5G networks and the services offered based thereon will make peoples' lives easier. Uneven pace of and delays in the deployment of these networks will create inequalities and may result in digital re-exclusion of large areas of Europe.

The deployment of 5G in Europe is based on two strategic documents: 5G for Europe: *An Action Plan and Connectivity for a Competitive Digital Single Market: Towards a European Gigabit Society* [29]. In these documents, two main objectives were adopted: enabling 5G connectivity as a fully developed commercial service in at least one major city by 2020, and uninterrupted and secure access to the 5G network in all urban areas and on all major terrestrial transport paths by 2025. 5G networks are to be developed in Europe by relying on the 700 MHz [30], 3400–3800 MHz [31] and 26 GHz [32] frequency bands. According to European Union documents, the Member States were to develop their own strategic documents for the deployment of 5G networks.

According to the European 5G Observatory [33], despite the recommendations of the European Commission, only 11 EU countries have published their national 5G development strategies. These include: Austria, Germany, Italy and Estonia. Some countries have also published national broadband plans. In total, various forms of 5G deployment plans were announced by 27 European countries (26 in the EU and the United Kingdom).

In Poland, the *National Broadband Plan* [34] has been drawn up and was amended by the government on. March 10, 2020 the *5G Strategy for Poland* that was made available to the public for consultation in January 2018 was to be another of the key documents, but the consultations have not been completed as of the end of June 2020 and the document has not been officially adopted [35]. These documents are based, to a considerable extent, on their European counterparts and on the goals, time frames and tools set and developed for the deployment of 5G. Problems with refarming the 700 MHz band for 5G systems that were caused, inter alia, by the delay in concluding a cross-border agreement with Russia [36] and by the need to adapt the digital terrestrial television network [37] accordingly, resulted in the postponement in issuing the applicable clearance until the second half of 2020 [38]. Consultations concerning the 3.4–3.8 GHz frequency band auction documentation were in progress until the end of February 2020. Ultimately, the auction has been scheduled for March 7, 2020, but due to the COVID pandemic, it was postponed and then canceled altogether. Confusion concerning the 3400–3800 MHz band auction delayed the Polish launch of 5G services relying on those frequencies by at least six months, despite the fact that all operators call for those frequencies to be made available to them as soon as practicable [38].

To accelerate the deployment of 5G in Poland, it is imperative that a tender or an auction be held and that bands be made available. Legal regulations required to facilitate investment processes and construction of 5G network infrastructure need to be passed as well. It was only at the beginning of 2020 that standards setting the permissible electromagnetic field levels were adjusted to EU requirements [39], [40]. Polish EMF standards in effect previously were originally introduced in 1984 and were based on Soviet Union specifications from an era when no cellular networks existed. They were extremely restrictive and contributed to slowing down the expansion of 4G networks [41].

The specific traits of 5G networks deployed in urban environments require the installation of a dense network of small transmitters (3.4–3.8 GHz or 26 GHz). Therefore, in order to offer 5G coverage in urban areas, it is necessary to install a great number of small antennas on urban infrastructure (buildings, road signs, power line posts, etc.) [41, p. 35]. On June 30, 2020, the European Commission adopted an *Implementing Regulation on small-area wireless access points or small antennas*, ordering the installation of such small base stations in all member states based on a permit-exempt deployment regime [42]. In Poland, under the *Act on supporting the development of telecommunications networks and services*, the obligation to obtain a building permit has been partially abolished and replaced with an obligation to notify given project. Fees due for access to vertical infrastructure have been waived as well [41, p. 36]. At the same time, however, the Ministry of Environment listed base stations as undertakings that may exert a significant impact on the environment, thus doing away with the facilitating measures introduced previously [43]. According to the Polish Chamber of Information Technology and Telecommunications (PIIT) and the Polish Chamber of Commerce for Electronics and Telecommunications, this will definitely hinder the construction of 5G networks. Regulation of the Minister of Digitization on the minimum technical and organizational measures and methods that telecom companies are required to adopt to ensure the security or integrity of networks or services, will enter into effect by the end of the year [44]. The said Regulation requires operators rendering 5G services to comply with 5G cybersecurity standards related, inter alia, to threat

identification and prevention mechanisms. It also ensures competition between suppliers and calls for conducting security audits. Regardless of the said Regulation, special requirements will also be laid out in the decisions pursuant to which right to use the 3400–3800 MHz band will be awarded in an auction or a tender, but the wording of these requirements still remains unknown.

4. Economic Context. Development of Commercial 5G Networks

As far as the technology relied upon in designing 5G networks and user terminals is concerned, one may state that its implementation is already underway in Europe. However, contrary to previous predictions by experts, it is not the industrial sector, but users of high-speed Internet who have turned out to be the first customer group. The search for European "unicorns" harnessing the potential of 5G networks is still ahead of us.

The first 5G smartphones were introduced in Europe in the second and third quarter of 2019. The first commercial deployments focus on offering enhanced mobile broadband (eMBB) services, with increased bitrate for data transmission customers. Solutions involving a significant reduction in delays, ultra reliable low-latency communication (URLLC), and an increased number of devices (massive IoT) have not yet been implemented commercially.

By the end of March 2020, 5G commercial services were deployed in 10 countries: Austria, Finland, Germany, Hungary, Ireland, Italy, Latvia, Romania, Spain and the United Kingdom [45]. Tests are still underway in other EU countries. By the end of 2019, over 181 tests were announced, and 5G networks were launched commercially in 130 EU cities. So far, Italy, Germany, France, Spain and the United Kingdom have conducted the highest number of tests in Europe.

In Poland, due to administrative delays in making the harmonized European spectrum available to operators (caused by the reasons described in the part of the paper dealing with the political context), the implementation of the first commercial 5G services, forced by commercial demand and strong competition among Polish operators, occurs with the use of frequencies other than those comprising the European harmonized bands. Polkomtel has launched a commercial service using the 2.6 GHz band, with Play and T-Mobile Polska relying on the 2.1 GHz band. Orange Polska announced that these services would be launched in July 2020, under reserve that this deadline may be postponed to the end of the year [46].

As far as tests of the 5G technology relying on the EU-wide harmonized bands are concerned, the first tests of 5G networks operating in urban areas were performed in September 2018 by Orange Polska [47]. In June 2019, another mobile operator, Play, cooperating with the Office of Electronic Communications, Łódź University of Technology, Ericsson and the Łódź Special Economic Zone, signed an agreement to join the *S5 – Akcelerator* pilot programm. It aims to foster innovative solutions relying on 5G technology [47]. Therefore, Łódź has become Poland's first city mentioned in the *5G for Poland Strategy* to pilot the solution. 5G devices have already been tested by T-Mobile and Polkomtel as well. The aforementioned tests are of technical nature and are subject to restrictive territorial limitations, with the services offered not being available commercially. Unlike in many other EU countries, no research or development programs concerning 5G tests and pilots promoting the technology (i.e. identifying economic unicorns and innovative services) have been conducted or established in Poland.

The cost of investing in 5G networks is huge, and the expenditures may be even higher due to reckless regulatory decisions taken in light of the trade war discussed below. Therefore, profitability may only be ensured by finding users who, by offering innovative services for industry or agriculture, will foster demand for telecom services that are more expensive than telephony or regular (even fast) Internet access. Simultaneously, it is important to ensure that the cost of building these networks remains as low as possible. These barriers in Poland were to be resolved by the *Polish 5G* agreement [48] proposed in 2019 at the initiative of Exatel, the Polish Development Fund Group and four operators: Orange Polska, T-Mobile Polska, Polkomtel, and Play. Its goal was to create a business (wholesale) model for building a common infrastructure for the 700 MHz band. The state would have a majority shareholding in the project and would contribute the 700 MHz frequency band, while the operators would make passive infrastructure and financial contributions [49]. However, no specific solutions have been reached so far.

On the other hand, research programs have been introduced by governments of other countries: willing to stimulate the country's economic growth, the UK government allocated 200 million pounds for testbeds and trials of new applications [50], Germany provided 26 million euros in funding to support three research projects focusing on cities, the medical sector and university campuses [51]. The Czech government also organized a competition for cities that are eligible to obtain. 2 million Czech crowns for testing their 5G networks [52]. The authorities of Vienna came up with a city-led initiative allocating 20 million euros to the development of the 5G network in the Austrian capital [53].

Some governments have set up special organizations, such as *Invest in Finland* – an entity operating in Nokia's motherland [54]. It is a government organization dedicated to financing innovation and promoting trade, travel, and investment. One of its goals is to implement "communication of the future" systems. It was the Finnish mobile operator Elisa [55] that launched the world's first 5G network. Hungary implemented an interesting solution by launching the *5GC project – the Hungarian 5G Coalition (5GK – Magyarországi 5G Koalíció)* [56], comprising representatives of the government, market players, and academia

members. Its purpose is to plan and coordinate the implementation of the network in Hungary. Thanks to this initiative, Hungary has become one of the first European centers for 5G development (alongside Austria, Germany and Estonia). The first switchboard in Hungary was commissioned in July 2018 at the Magyar Telekom headquarters in Budapest [57].

All European countries need to invest in constructing 5G networks die to two reasons: economic competitiveness and citizens' access to next-generation digital services. Such investments will also translate directly into attracting modern, technologically-advanced projects, i.e. the so-called Industry 4.0, thus increasing the number of jobs on the market. According to the European Commission, 5G networks will be one of the most important components of the digital economy in the next decade. It is estimated that global revenues related to the development of 5G-based services will amount to 225 billion euros in 2025 [28, p. 21]. With such data taken into consideration, the trade war on tech fought by the world's superpowers is of great financial significance as well.

5. Network Security in the Context of a Global Conflict of Superpowers

The projected economic and geopolitical importance of 5G results in many countries paying close attention to the option of interfering with the free market and the freedom of economic activity by introducing specific rules to ensure the security of 5G networks [58]. So far, operators have been enjoying full independence in choosing suppliers of their network components, remaining responsible for the security of their networks and for, ensuring confidentiality or compliance with regulations set forth in the General Data Protection Regulation (GDPR). The initial lack of specific European regulations regarding 5G networks resulted in the approaches taken by different countries not being uniform. Consequently, due to its economic dependence, Europe has become a venue of a direct "trade war on tech" between the US and China. As a result, network security has become a part of the political agendas of European countries deliberating their economic interventions policies.

Globally, the most controversial steps (justified by the need to ensure the security of the network and of the economy), were taken by the US. By means of an executive order, the country banned the use of telecommunications equipment manufactured by companies recognized as a threat to the national security [59]. The order concerned mainly manufacturers from China, as it was this country that was identified as a "strategic competitor" in the US national security strategy [60].

China is developing the so-called Digital Silk Road (DSR). The project assumes that undertakings in the field of technology, 5G networks and e-commerce will be implemented in cooperation with selected countries. The goal of DSR is to promote China's own technology standards. The rivalry with the U.S. and the importance of the digital sector during the COVID-19 pandemic have exerted even more pressure on China [61].

Huawei Technologies, a Chinese company founded in Shenzhen in 1987, is at the very center of the dispute. It is one of the leaders in the 5G technology market, is present in 170 countries around the world and has been active on the European market since 2000. The controversy around the company stems, inter alia, from the escalating dispute between the US administration and China. Both of these countries are engaged in growing competition for global technology leadership. The US has made multiple allegations against the Chinese tech tycoon, accusing it of stealing trade secrets of American companies, violating international bans and supplying Iran with equipment that allowed it to monitor anti-government demonstrations in Tehran in 2009, and of attempting to conceal Huawei's business exchanges with North Korea, taking place despite the economic sanctions imposed on this country [62]. Allegations of corporate espionage have been made as well [63]. The company disputes the said allegations, while the Chinese Foreign Ministry accuses the US government of "economic bullying behavior" and of misusing security issues to "suppress Chinese enterprises with unwarranted charges" [63]. In addition, Huawei filed a lawsuit in New Orleans in December 2019 challenging a recent FCC decision that prohibits U.S. operators from using federal subsidies to purchase Huawei equipment [64]. At the same time, the US calls on other allied countries to boycott the implementation of Huawei's 5G technology, using NATO structures for this purpose as well [65].

In Europe, attitudes toward this technology war vary from one country to another. At the initial phase of the network security debate, Washington secured the support of Romania, Poland, Estonia, Latvia, and the Czech Republic, all of which signed joint 5G security statements or memorandums with the US government [66]–[70] declaring their intention to provide access to their 5G networks to "trusted suppliers" only. The statements themselves, however, are of little value. They must be backed up by laws forcing telecom companies to abide by their terms. In Poland, the government fast-tracked regulations on 5G security, canceling the 5G spectrum auction that was already in progress and demanding the telecommunications market regulator (President of the Office of Electronic Communications) to include specific obligations in the terms of the repeated auction [71]. As of the end of June 2020, this has only resulted in delaying the process of making the 3400–3800 MHz spectrum band available to Polish 5G layers by at least six months.

The Estonian Parliament passed amendments to the Electronic Communications Act, requiring operators to coordinate with the national communications authority on 5G deployment [72]. This means that national security and intelligence agencies will be able to interfere in the process by imposing restrictions on supplier selection. The Czech Republic turned out to be a strong European supporter of the

American approach, relying on the *Prague principles* [73] that have allowed to reach a Western consensus on Chinese suppliers [74].

The German government is skeptical of Washington's allegations against Huawei. German operators (Deutsche Telekom and Vodafone) operate 4G networks that rely (to over 50%) on Huawei hardware [74]. Germany also fears that a ban on purchasing equipment from Chinese suppliers will seriously damage its economic ties with China – the country's largest trade partner. Deutsche Telekom has presented an analysis in which the exclusion of Huawei from the 5G market is labeled as the "Armageddon scenario" expected to generate additional 3 billion euros in costs [75]. Discussions are ongoing, however. Under the pressure of the parliament and Angela Merkel's smaller coalition partners, the German Federal Ministry of the Interior proposed, in May 2020, a draft law [76] that would enhance the security requirements binding upon 5G providers, simultaneously providing the ministry with new powers to block non-trusted suppliers. It remains known how the government plans to assess the trustworthiness of suppliers, as unspecified certification mechanisms are being considered.

France has adopted a different approach, as it has attempted to assume a leading role in 5G security in Europe. National security checks were carried out in the country, focusing on cybersecurity policies in effect at the individual operators, as well as on their suppliers choices. In 2019, the government, operating via the Cybersecurity Agency, was authorized to block base stations (RAN) used by service providers if their operation would pose a threat to national security. The said right was added to the package of regulations binding upon telecom companies [74]. The entire telecommunications infrastructure has been recognized as being of critical importance for the state – an approach that paves the way for deep interference with economic freedoms. However, no final decisions to exclude Huawei from the process of building 5G in France have been made so far. The Chinese are trying to influence the decision-making process in Paris by promising to invest millions in constructing Europe's largest manufacturing base [77]. However, the outbreak of the pandemic has put that process on halt.

The decisions of Italy, Belgium and the Netherlands also seem to be important for both the US and China. In 2019, the government of Italy – the EU's third largest economy – adopted Legislative Decree No. 64/2019 (DL 64/2019), which amends the law known as the *Golden Power Legislation (Legislazione sul potere d'oro)*, governing the state's powers to intervene in transactions involving enterprises operating in the defense, national security, communications, energy and transport sectors ("strategic sectors") [78]. The change allows the government to block contracts between operators and equipment suppliers [79]. At the end of 2019, new cybersecurity "perimeter" regulations were passed, which would impose new requirements on telecommunications and IT services used in "strategic sectors" [80]. The government is now finalizing the list of enterprises, sectors and government organizations that would be subject to the stricter regime.

Belgium is a strategic country, as it is a host for the NATO headquarters and the key EU institutions. The country's intelligence services have recommended the government to limit the use of "non-trusted suppliers". The administration is working on law amendments that are to restrict the participation of Chinese suppliers, at least in the so-called "core" of the network [81].

The Dutch government, traditionally close to the US in regards to cybersecurity and intelligence, also adopted a new law in December 2019 [82]. It allows to ban the sales of goods if there is a suspicion that they may sabotage the telecoms network, or that their suppliers have close ties with or legal obligations towards foreign governments that could pose a security threat. The Netherlands has previously indicated that this means a ban on the use of equipment from high-risk suppliers in the so-called "core" of the network [74].

Those European countries which are undecided, with their governments continuing to consult telecom companies, intelligence services and market regulators on the proposed legal amendments that would eliminate high-risk suppliers, continues to prevail as of mid-2020. These include, inter alia, Spain, Portugal, Luxembourg, Sweden, Austria, and Finland.

The Spanish economy minister announced, in February 2020, that she was working on legal acts regulating network security issues. However, the COVID-19 epidemic caused delays in legislative work and the spectrum auction was postponed to a later date [83]. At the same time, Spain's largest operator, Telefónica, declared that it would reduce the share of Huawei's equipment in the modernized networks, but would continue to use it nonetheless. The second largest operator, Orange, confirmed its cooperation with ZTE, Huawei and Ericsson on the Spanish market [84]. It seems that Spain has adopted a liberal approach focusing on the overriding goal of development and mass-scale deployment of 5G [28].

The above was confirmed on the 10th of October 2020, during the XXXI Spanish-Portuguese summit. Spanish Prime Minister Pedro Sanchez announced that Telefonica guaranteed that 5G coverage in Spain would reach 75% by the end of 2020. This will be done in cooperation with Huawei, although he did not exclude the possibility of cooperating with "other foreign entrepreneurs". Moreover, the governments of both countries announced the adoption of a "Joint Cross-Border Development Strategy". They also planned to work on specific infrastructure projects, with the construction of AVE (high-speed rail) between Madrid-Lisbon and the implementation of a new 5G technology reality, known as "Atlantic Corridor", being the most important of them [85].

No final decisions have been made by Sweden (the home of Huawei's European competitor – Ericsson) and Finland (the home of another European competitor – Nokia) [86]. In Finland, Huawei equipment is widely used in 4G networks. At the same time, the US – a country which is

falling behind in 5G and does not have any domestic suppliers offering 5G networks, is considering purchasing shares in Ericsson and Nokia through its economic tycoons, such as Cisco or Google [87]. Simultaneously, American companies, such as AT&T, Dell, Microsoft, Intel and Infineon, closely support the development of the Open RAN standard. Korea's Samsung and China's ZTE are also involved in the work on that standard, and the participating operators include also China Mobile, Deutsche Telekom, and Orange [88]. However, this solution is not ready yet. The work is ongoing and it its results will most likely be harnessed while focusing on the 6G network in 5 to 10 years' time.

An approach similar to that taken by the French (but only in the field of 5G) is considered by Denmark which is preparing legislations classifying the entire 5G network as critical infrastructure [89]. As a result, China, referring to the violation of the *GATT Free Trade Agreement*, threatened to terminate a free trade agreement between the Faroe Islands (an autonomous part of Denmark) and China if Denmark failed to sign the 5G agreement with Huawei [90].

After initial hesitation, Great Britain joined the coalition with the United States and proposed an alliance of 10 countries to reduce reliance on China in the process of introducing 5G technology [91]. Such an alliance would be made up of Australia, South Korea, India and the G7 countries (France, Japan, Germany, USA, UK, Italy, and Canada). The initiative is a result of concerns about Huawei's and ZTE's domination in the process of deploying 5G in Europe. Interestingly, the UK government previously approved Huawei's participation in the construction of the 5G network in the country, but at the same time imposed a cap of 35% on Huawei equipment's market share [92].

As far as European countries open to all 5G equipment suppliers are concerned, one may list Austria whose chancellor has publicly announced that his country is "fundamentally technologically neutral" and does not intend to succumb to US pressure in the area of excluding Huawei from 5G deployment [93]. The prime ministers of Slovakia [94] and Hungary [95] adopted a similar attitude.

The president of Russia has assumed a strong stance with this regard. The deal between the Russian company of Mobile TeleSystems and Huawei, concerning the deployment of the 5G network, was signed in an almost ceremonial manner, in the presence of the presidents of China and Russia. According to CNN, "Russia does not share the US security concerns and even suggests that the deal is a try-on for an internet iron curtain" [96]. This stance shows that the notion of a cold war on tech [97] has some justification, although not everyone agrees. Kaan Sahin claims [98] that one of the conditions for the Cold War to come into existence is the emergence of blocks, as it was the case in the second half of the 20th century. According to him, these are absent in this trade war. However, the US administration's policy of polarization and the emergence of a Chinese-Russian block seem to contradict this optimistic assumption. On the worldwide scene, Australia has also introduced a ban on the use of Chinese equipment in 5G networks. The division into countries supporting the US and defending themselves against pressure to eliminate Huawei from the market becomes ever more apparent in Europe as well.

We shall see, in the coming years, whether such a division will be gaining in importance, considering the fact that the world is moving towards disintegration of the Internet (China, Russia, Iran, etc.). Will the division of the Internet into separated and filtered regional blocks result in the division of technology providers serving separate regions? It seems that such a thought is becoming ever more prevailing among the important politicians of the superpowers. But does it make any sense? To judge this, one needs to take a step back in their considerations and look at security from the technical perspective.

6. Conclusions

The article reviews the current situation related to the deployment of 5G in Poland and in some other European countries. The rollout of 5G in Europe has been set in an economic and political context, and is presented from a broader point of view, with the global geopolitical situation and the conflict between China and the US taken into consideration. 5G-related phenomena, such as deployment delays observed in some European countries, are also presented. Examples of specific European countries were likewise used to present trends that were non-existent during the implementation of previous generations of telecommunication technologies.

After a critical analysis of reports drawn up by institutions monitoring the development of 5G, as well as strategic documents of the individual states, literature of the subject and online sources, one may conclude that two main reasons exist as to why investment projects concerned with the construction of 5G networks in Europe are of such great importance. The first of those reasons is Europe's desire to ensure its economic competitiveness in its relations with the rest of the world. The other is the citizens' access to next-generation digital services. These two aspects are directly related to the global desire to take advantage of the benefits offered by Industry 4.0 – a phenomenon that is most likely to boost economic growth and create new jobs. Europe's position on the international arena depends on how quickly European countries will develop and deploy 5G. Each country pursues its own economic policy and struggles with other internal problems, meaning that the progress in developing and deploying 5G networks may not be even in different European states.

Network security is also crucial, especially in the context of the global conflict between two superpowers (US and China) over dominance in the deployment of 5G. Due to China racing ahead and the US lagging behind, this conflict is escalating to other countries, including those in Europe. The pressure to exclude the Chinese company of Huawei from the market, exerted by the US, has resulted in divisions, also in the European Union. The dependency of

specific countries on a single 5G vendor will result in the lack of diversity in the area of devices and solutions used. Consequently, the pace of innovation may slow down both at national and European level (in the absence of competition), and the 5G infrastructure may become more vulnerable from the security point of view, especially if multiple operators rely on one vendor only. Such a supplier may come under commercial pressure, be sanctioned, or simply fail commercially. Another key point is that a limited number of vendors may lower the market incentive to develop more secure products. It is not clear whether this higher financial cost will result in greater security, as China will not be eliminated from the supply chain of European producers within the next decade.

5G will be crucial in the context of ensuring economic growth of European countries. At the same time, it is also an element of political discussions centered around technological security, the "war on tech" and information warfare, fueled by disinformation. Therefore, social protests are mounting, also in Europe, demanding a total ban on 5G. The protests are not limited, in an increasing number of cases, to verbal dissatisfaction, involve but also destruction of critical infrastructure. In recent months, Europe has witnessed a wave of antenna towers being set on fire. Unfortunately, in mid-2020, in its documents serving as a foundation for the rollout of 5G in Europe, the European Commission failed to elaborate on how to win social acceptance for 5G and did not propose any actions aimed at expanding the technical knowledge of European societies. This lack of social acceptance, caused mostly by ignorance and fear, could affect the pace of 5G development, thus slowing down Europe's economic growth and depriving it of its position in the international economic race.

References

[1] U. Soler and M. Busiło, "Education of society as a tool to counteract disinformation in implementing new technologies. On the example of 5G mobile telecommunications network and Warsaw sewage system", in *Proc. of the Int. Conf. Applications of Electromag. in Modern Engin. and Medicine*, Janów Podlaski, Poland, 2019, pp. 210–214 (DOI: 10.23919/PTZE.2019.8781728).

[2] ITU-R Recommendation M.2083-0, "IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond", 2015 [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf

[3] Instytut Kościuszki, "Sieć 5G będzie kręgosłupem gospodarki cyfrowej" [Online]. Available: https://ik.org.pl/siec-5g-bedzie-kregoslupem-gospodarki-cyfrowej/ (accessed on 15.06.2020) [in Polish].

[4] A. Segal, "Year in review 2019: the U.S.-China tech cold war deepens and expands", *Council on Foreign Relations*, 2019 [Online]. Available: https://www.cfr.org/blog/year-review-2019-us-china-tech-cold-war-deepens-and-expands (accessed on 26.06.2020).

[5] K. Sahin, "The tech cold war illusion", *Berlin Policy Journal*, 2020 [Online]. Available: https://berlinpolicyjournal.com/the-tech-cold-war-illusion/ (accessed on 15.06.2020).

[6] S. Z. Asif, *5G Mobile Communications: Concepts and Technologies*. Florida: CRC Press, 2018 (ISBN: 9781498751551).

[7] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G Mobile and Wireless Communications Technology*. New York: Cambridge University Press, 2016 (ISBN: 9781107130098).

[8] J. Rodriguez, *Fundamentals of 5G Mobile Networks*. Wiley, 2015 (ISBN: 9781118867525).

[9] B. Rao, A. J. Harrison, and B. Mulloth, *Defense Technological Innovation: Issues and Challenges in an Era of Converging Technologies*. Business, 2020 (ISBN: 9781789902099).

[10] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2020 (ISBN: 9780128143230).

[11] K. David and H. Berndt, "6G vision and requirements: is there any need for beyond 5G?", *IEEE Vehicular Technol. Mag.*, vol. 13, no. 3, pp. 72–80, 2018 (DOI: 10.1109/MVT.2018.2848498).

[12] E. J. Oughton, Z. Frias, T. Russell, D. Sicker, and D. D. Cleevely, "Towards 5G: Scenario-based assessment of the future supply and demand for mobile telecommunications infrastructure", *Technological Forecasting and Social Change*, vol. 133, pp. 141–155, 2018 (DOI: 10.1016/j.techfore.2018.03.016).

[13] V. Weerakkody and G. Dhilon, "Moving from E-Government to T-Government: a study of process reengineering challenges in a UK local authority context", *Int. J. of Electronic Government Research*, vol. 4, no. 4, pp. 1–16, 2008 (DOI: 10.4018/jegr.2008100101).

[14] E. J. Oughton, Z. Frias, S. Van der Gaast, and R. Van der Berg, "Assessing the capacity, coverage and cost of 5G infrastructure strategies: Analysis of the Netherlands", *Telematics and Informatics*, vol. 37, pp. 50–69, 2019 (DOI: 10.1016/j.tele.2019.01.003).

[15] S. Saxena, "Enhancing ICT infrastructure in public services, Factors influencing mobile government (m-government) adoption in India", *The Bottom Line*, vol. 30 no. 4, pp. 279–296, 2017 (DOI: 10.1108/BL-08-2017-0017).

[16] G. Fettweis and S. Alamouti, "5G: Personal mobile Internet beyond what cellular did to telephony", *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 140–145, 2014 (DOI :10.1109/MCOM.2014.6736754).

[17] Qualcomm, "Spectrum for 4G and 5G", 2020 [Online]. Available: https://www.qualcomm.com/media/documents/files/spectrum-for-4g-and-5g.pdf (accessed on 18.06.2020).

[18] Deloitte, "5G: The chance to lead for a decade", 2018 [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-tmt-5g-deployment-imperative.pdf (accessed on 21.05.2020).

[19] W. Lemstra, M. Cave, and M. Bourreau, "Towards the successful deployment of 5G in Europe: What are the necessary policy and regulatory conditions?", *Center on Regulation in Europe (CERRE)*, 2017 [Online]. Available: https://cerre.eu/publications/towards-successful-deployment-5g-europe-what-are-necessary-policy-and-regulatory/ (accessed on 25.06.2020).

[20] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A survey on security and privacy of 5G technologies: potential solutions, recent advancements, and future directions", *IEEE Commun. Surveys & Tutorials*, vol. 22, no. 1, pp. 196–248, 2020 (DOI: 10.1109/COMST.2019.2933899).

[21] M. Kim, H. Lee, and J. Kwak, "The changing patterns of China's international standardization in ICT under techno-nationalism: A reflection through 5G standardization", *Int. J. of Informat. Management*, vol. 54, 2020 (DOI: 10.1016/j.ijinfomgt.2020.102145).

[22] B. Kowalski, "Central and eastern Europe, China's core interests, and the limits of relational politics: lessons from the Czech Republic in the 2010s", *East European Politics and Societies*, 2020 (DOI: 10.1177/0888325420952142).

[23] W. Lemstra, "Leadership with 5G in Europe: Two contrasting images of the future, with policy and regulatory implications", *Telecommun. Policy*, vol. 42, no. 8, pp. 587–611, 2018 (DOI: 10.1016/j.telpol.2018.02.003).

[24] E. Kavalski, "China in Central and Eastern Europe: the unintended effects of identity narratives", *Asia European J.* vol. 17, no. 4, pp. 403–419, 2019 (DOI:10.1007/s10308-019-00563-1).

[25] F. Longtin, "The race to 5G in China and the European Union: an analysis of Huawei and Nokia's innovation policies", M.Sc. Thesis, Louvain School of Management, Université catholique de Louvain, 2020.

[26] M. K. Pandey, A. Gaurav, and V. Kumar, "Social, technical and economical challenges of 5G technology in Indian prospective: Still 4G auction not over, but time to think about 5G in India", in *Proc. Int. Conf. on Comput. and Computat. Sciences (ICCCS)*, Greater Noida, India, 2015, pp. 157–162, (DOI: 10.1109/ICCACS.2015.7361342).

[27] A. S. A. AlRaeesi and R.M. Habibur, "Proactive approach to the deployment of 5G technology: insights from the UAE", in *Proc. 2nd Europe – Middle East – North African Regional Conf. of the Int. Telecommun. Society (ITS): "Leveraging Technologies For Growth"*, Aswan, Egypt, 2019 [Online]. Available: https://www.econstor.eu/bitstream/10419/201755/1/ITS2019-Aswan-paper-68.pdf

[28] F. Gelici *et al.*, "5G – szanse, zagrożenia, wyzwania", *Instytut Kościuszki*, Kraków, 2020 [Online]. Available: https://ik.org.pl/wp-content/uploads/raport_5g_szanse_wyzwania_zagrozenia.pdf [in Polish].

[29] "5G for Europe: An Action Plan", *European Commission*, 2016.

[30] "Decision (EU) 2017/899 of the European Parliament and of the Council of 17 May 2017 on the use of the 470–790 MHz frequency band in the Union", *Official J. of the European Union*, 2017 [Online]. Available: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32017D0899

[31] "Commission Implementing Decision (EU) 2019/235 of 24 January 2019 on amending Decision 2008/411/EC as regards an update of relevant technical conditions applicable to the 3400–3800 MHz frequency band (notified under document C(2019) 262) (Text with EEA relevance.)", *Official J. of the European Union*, 2019 [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019D0235

[32] "Commission Implementing Decision (EU) 2019/784 of 14 May 2019 on harmonization of the 24.25–27.5 GHz frequency band for terrestrial systems capable of providing wireless broadband electronic communications services in the Union (notified under document C(2019) 3450) (Text with EEA relevance.)", *Official J. of the European Union*, 2019 [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019D0784

[33] European 5G Observatory, "National 5G plans and strategies", 2017 [Online]. Available: https://5gobservatory.eu/public-initiatives/national-5g-plans-and-strategies/#1533565785008-ace4c4ba-1384 (retrieved 19.06.2020).

[34] Ministerstwo Cyfryzacji, "Narodowy Plan Szerokopasmowy", 2017 [Online]. Available: https://www.gov.pl/web/cyfryzacja/narodowy-plan-szerokopasmowy (retrieved 18.06.2020) [in Polish].

[35] "Strategia 5G dla Polski", Ministerstwo Cyfryzacji, 2018 [Online]. Available: https://www.gov.pl/documents/31305/436699/Strategia+5G+dla+Polski.pdf/0cd08029-2074-be13-21c8-fc1cf09629b0 [in Polish].

[36] Telepolis, "5G: jest porozumienie z Rosją w sprawie pasma 700 MHz", 2019 [Online]. Available: https://www.telepolis.pl/wiadomosci/prawo-finanse-statystyki/5g-porozumienie-z-rosja-w-sprawie-700-mhz (accessed on 17.06.2020) [in Polish].

[37] Telepolis, "Zakończył się drugi etap refarmingu 700 MHz dla 5G, 2020 [Online]. Available: https://www.telepolis.pl/wiadomosci/wydarzenia/zakonczyl-sie-drugi-etap-refarmingu-700-mhz-dla-5g (accessed on 17.06.2020) [in Polish].

[38] Telepolis, "Nowa aukcja 5G powinna odbyć się jak najszybciej – tego chcą telekomy", 2020 [Online]. Available: https://www.telepolis.pl/wiadomosci/wydarzenia/nowa-aukcja-5g-jak-najszybciej (accessed on 17.06.2020) [in Polish].

[39] A. Dziermański, "Od dzisiaj obowiązują nowe normy PEM. Co to oznacza?", *whatnext*, 2020 [Online]. Available: https://whatnext.pl/nowe-normy-pem-1-stycznia-co-to-znaczy/ (accessed on 18.06.2020) [in Polish].

[40] "Nowe normy promieniowania w Polsce od 1 stycznia", *CyberDefence 24*, 2019 [Online]. Available: https://www.cyberdefence24.pl/nowe-normy-promieniowania-w-polsce-od-1-stycznia (accessed on 18.06.2020) [in Polish].

[41] P. Gawlicki *et al.*, "Raport sieci 5G w Polsce – szanse i wyzwania", *Accenture*, 2019 [Online]. Available: https://www.accenture.com/_acnmedia/Accenture/Redesign-Assets/DotCom/Documents/Local/2/Accenture-Report-Web-5G-Poland-Chances-Challenges.pdf#zoom=50 (accessed on 18.06.2020) [in Polish].

[42] J. Kruczek, "KE przyjęła ważne rozporządzenie dot. 5G", *SAT Kurier*, 2020 [Online]. Available: https://satkurier.pl/news/193306/ke-przyjela-wazne-rozporzadzenie-dot-5g.html (accessed on 18.06.2020) [in Polish].

[43] Ł. Czyleko, "Kwalifikacja stacji bazowej telefonii komórkowej (instalacji radiokomunikacyjnej) jako przedsięwzięcia mogącego znacząco oddziaływać na środowisko", *LC Consulting*, 2019 [Online]. Available: https://www.czyleko.pl/kwalifikacja-stacji-bazowej-telefonii-komorkowej-instalacji-radiokomunikacyjnej-jako-przedsiewziecia-mogacego-znaczaco-oddzialywac-na-srodowisko/ (accessed on 18.06.2020).

[44] "Rozporządzenie Ministra Cyfryzacji w sprawie minimalnych środków technicznych i organizacyjnych oraz metod, jakie przedsiębiorcy telekomunikacyjni są obowiązani stosować w celu zapewnienia bezpieczeństwa lub integralności sieci lub usług", *Ministerstwo Cyfryzacji*, 2020 [Online]. Available: https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20200001130/O/D20201130.pdf [in Polish].

[45] "Announcements of commercial launches", *European 5G Observatory*, 2020 [Online]. Available: https://5gobservatory.eu/market-developments/5g-services/ (accessed on 19.06.2020).

[46] P. Gajkowski, "Jeśli 5G w Polsce, to u kogo? Przegląd ofert operatorskich", *gsmmaniak.pl*, 2020 [Online]. Available: https://www.gsmmaniak.pl/1147863/jesli-5g-w-polsce-to-u-kogo-przeglad-operatorskich/ (accessed on 13.06.2020).

[47] M. Jaślan, "Polskie testy 5G: priorytet technologii nad biznesem", *telko.in*, 2019 [Online]. Available: https://www.telko.in/polskie-testy-5g-priorytet-technologii-nad-biznesem (accessed on 20.06.2020).

[48] "Polskie 5G nabiera kształtów", *Ministerstwo Cyfryzacji*, 2019 [Online]. Available: https://www.gov.pl/web/cyfryzacja/polskie-5g-nabiera-ksztaltow (accessed on 20.06.2020) [in Polish].

[49] "Exatel: powołanie Polskiego 5G wymaga konsultacji z UE", *Puls Biznesu*, 2019 [Online]. Available: https://www.pb.pl/exatel-powolanie-polskiego-5g-wymaga-konsultacji-z-ue-977903 (accessed on 20.06.2020) [in Polish].

[50] J. Wagstaff, "New $65 million package for 5G trials", *UK 5G*, 2020 [Online]. Available: https://uk5g.org/5g-updates/read-articles/new-65-million-package-5g-trials/ (accessed on 20.06.2020).

[51] "Federal Ministry of Transport and Digital Infrastructure provides funding of 26 million euros to three 5G research projects", *Federal Ministry of Transport and Digital Infrastructure*, 2019 [Online]. Available: https://www.bmvi.de/SharedDocs/EN/PressRelease/2019/080-scheuer-funding-5g-research.html (accessed on 20.06.2020).

[52] "Winners announced of smart cities competition to test 5G technology", *Ministry of Industry and Trade*, 2019 [Online]. Available: https://www.mpo.cz/en/guidepost/for-the-media/press-releases/winners-announced-of-smart-cities-competition-to-test-5g-technology–251502/ (accessed on 20.06.2020).

[53] "Władze Wiednia dopłacą do budowy stacji bazowych 5G", *Speedtest*, 2020 [Online] Available: https://www.speedtest.pl/wiadomosci/5g/wieden-dofinansowanie-budowy-5g/ (accessed on 20.06.2020) [in Polish].

[54] "Invest in Finland", *Business in Finland*, 2020 [Online]. Available: https://www.businessfinland.fi/en/do-business-with-finland/invest-in-finland/take-the-fast-track-to-finland/ (accessed on 19.06.2020).

[55] "Finland's vision for 5G development", *ITU News*, 2018 [Online]. Available: https://news.itu.int/finland-5g-development/?fbclid=IwAR2HSenKydfStg8224NWcnm3POmugp8EpLdPDJn4BKyrDFxWNPcxrDKG_UA (accessed on 19.06.2020).

[56] P. Smejkal, "Az európai 5G-s hálózat kiépítéséért folyó harc állhat a Huawei-botrány hátterében", *Forbes*, 2019 [Online]. Available: https://forbes.hu/uzlet/huawei-kemkedes-botrany-egyesult-allamok-5g-mobil-eu/ (accessed on 20.06.2020) [in Hungarian].

[57] Z. Szabó, "Nagy lépés elõtt Magyarország – mikor jöhet a sebességváltás?", *Napi.hu*, 2019 [Online]. Available: https://www.napi.hu/tech/invitech-halozat-5g-kis-albert-iot.680146.html (accessed on 20.06.2020) [in Hungarian].

[58] I. Albrycht and J. Świątkowska, "Przyszłość 5G czyli Quo Vadis, Europo?", *Instytut Kościuszki*, 2019 [Online]. Available: https://ik.org.pl/wp-content/uploads/ik_brief_programowy_5g-1.pdf (accessed on 21.06.2020) [in Polish].

[59] D. Trump, "Securing the information and communications technology and services supply chain", 2019 [Online]. Available: https://www.federalregister.gov/documents/2019/05/17/2019-10538/securing-the-information-and-communications-technology-and-services-supply-chain (accessed on 21.06.2020).

[60] "Trump labels China a strategic "competitor"", *Financial Times*, 2020 [Online]. Available: https://www.ft.com/content/215cf8fa-e3cb-11e7-8b99-0191e45377ec (accessed on 21.06.2020).

[61] D. Wnukowski, "China's digital silk road: economic and political significance", *The Polish Institute of Int. Affairs, Bulletin*, vol. 229, no. 1659, 2020 [Online]. Available: https://pism.pl/publications/Chinas_Digital_Silk_Road_Economic_and_Political_Significance

[62] "USA przedstawiło kolejne zarzuty wobec Huawei", *wGospodarce.pl*, 2020 [Online]. Available: https://wgospodarce.pl/informacje/75409-usa-przedstawilo-kolejne-zarzuty-wobec-huawei (accessed on 21.06.2020) [in Polish].

[63] "USA oskarżają Huawei o oszustwa bankowe i szpiegostwo przemysłowe. "Rozczarowanie przedstawionymi zarzutami"", *wirtualnemedia*, 2019 [Online]. Available: https://www.wirtualnemedia.pl/artykul/huawei-oskarzenia-o-oszustwa-bankowe-i-szpiegostwo-przemyslowe-dlaczego (accessed on 21.06.2020) [in Polish].

[64] S. Pham, "Huawei sues US government over new FCC restrictions", *CNN*, 2019, [Online] Available: https://edition.cnn.com/2019/12/05/tech/huawei-us-ban-lawsuit/index.html (accessed on 21.06.2020).

[65] M. Gajewski, "USA wypowiedziały wojnę Huawei i namawiają inne kraje do bojkotu firmy", *Spidersweb*, 2018 [Online]. Available: https://spidersweb.pl/2018/11/huawei-zakaz-handlu.html (accessed on 21.06.2020) [in Polish].

[66] "Memorandum of understanding between the Government of Romania and the Government of the United States of America", 2019 [Online]. Available: https://media.hotnews.ro/media_server1/document-2019-11-3-23464357-0-memorandum-5g-romania-sua.pdf (accessed on 21.06.2020).

[67] "U.S.–Poland Joint Declaration on 5G", *Whitehouse.gov*, 2019 [Online]. Available: https://www.whitehouse.gov/briefings-statements/u-s-poland-joint-declaration-5g/ (accessed on 22.06.2020).

[68] "United States–Estonia Joint Declaration on 5G Security", *Whitehouse.gov*, 2019 [Online]. Available: https://www.whitehouse.gov/briefings-statements/united-states-estonia-joint-declaration-5g-security/ (accessed on 22.06.2020).

[69] "Joint Statement on United States – Latvia Joint Declaration on 5G Security", *state.gov*, 2020 [Online]. Available: https://www.state.gov/joint-statement-on-united-states-latvia-joint-declaration-on-5g-security/ (accessed on 22.06.2020).

[70] "Joint Statement on United States – Czech Republic Joint Declaration on 5G Security", *state.gov*, 2020 [Online]. Available: https://www.state.gov/joint-statement-on-united-states-czech-republic-joint-declaration-on-5g-security/ (accessed on 22.06.2020).

[71] "Warsaws telecom powers clash over 5G security rules", *Politico*, 2020 [Online]. Available: https://pro.politico.eu/news/warsaws-telecom-powers-clash-over-5g-security-rules (accessed on 23.06.2020).

[72] "Riigikogu muutis elektroonilise side seadust", *Riigikogu*, 2020 [Online]. Available: https://m.riigikogu.ee/istungi-ulevaated/riigikogu-muutis-elektroonilise-side-seadust/ (accessed on 23.06.2020) [in Estonian].

[73] "Huawei slams western coalition talks in Prague", *pro.politico*, 2020 [Online]. Available: https://pro.politico.eu/news/huawei-slams-western-coalition-talks-in-prague-tktkt (accessed on 23.06.2020).

[74] L. Cerulus, "Trump and friends: Where European countries come down on Huawei", *Politico*, 2020 [Online]. Available: https://www.politico.com/news/2020/05/26/europe-huawei-5g-281701 (accessed on 23.06.2020).

[75] M. Koch and S. Scheuer, "Armageddon – Szenario: Telekom spielt Huawei-Bann durch", *Handelsblatt*, 2020 [Online]. Available: https://www.handelsblatt.com/technik/it-internet/ausschluss-von-netzausruester-armageddon-szenario-telekom-spielt-huawei-bann-durch/25918402.html?ticket=ST-12917777-KnWObMFc2KdxebinQEI9-ap6 (accessed on 23.06.2020) [in German].

[76] "Berlin to tweak network security law raising bar for Huawei", *pro.politico*, 2020 [Online]. Available: https://pro.politico.eu/news/berlin-to-tweak-network-security-law-raising-bar-for-huawei (accessed on 23.06.2020).

[77] L. Cerulus, "How Huawei wields investment to bend EU countries", *Politico*, 2020 [Online]. Available: https://www.politico.eu/article/huawei-dangles-investments-to-european-governments-for-grace/ (accessed on 23.06.2020).

[78] "Italian government acts to strengthen further its "golden powers"", *Hogan Lovells Publications*, 2019 [Online]. Available: https://www.hoganlovells.com/en/publications/italian-government-acts-to-strengthen-further-its-golden-powers (accessed on 23.06.2020).

[79] "Golden Power", *Senato della Repubblica*, 2019 [Online]. Available: https//www.senato.it/service/PDF/PDFServer/BGT/01118751.pdf (accessed on 23.06.2020).

[80] "Legge di bilancio 2019", *Senato della Repubblica*, 2019 [Online]. Available: https://www.senato.it/service/PDF/PDFServer/BGT/01093736.pdf (accessed on 23.06.2020) [in Italian].

[81] "Belgian security services call for caution on Huawei minister", *pro.politico*, 2020 [Online]. Available: https://pro.politico.eu/news/belgian-security-services-call-for-caution-on-huawei-minister (accessed on 23.06.2020).

[82] "Netherlands sets high bar for 5G security", *pro.politico*, 2020 [Online]. Available: https://pro.politico.eu/news/netherlands-sets-high-bar-for-5g-security (accessed on 23.06.2020).

[83] "Calviño señala que la subasta de 5G se celebrará a principios de 2021", *Cinco Días*, 2020 [Online]. Available: https://cincodias.elpais.com/cincodias/2020/05/14/companias/1589479871_667188.html (accessed on 24.06.2020) [in Spanish].

[84] I. Morris, "Orange confirms ZTE as 5G partner in Spain", *Light Reading*, 2020 [Online]. Available: https://www.lightreading.com/5g/orange-confirms-zte-as-5g-partner-in-spain/d/d-id/757867 (accessed on 24.06.2020).

[85] S. de la Cruz, "Sánchez apuesta por el 5G peninsular con Telefónica y Huawei", *La Razon*, 2020 [Online]. Available: https://www.larazon.es/economia/20201012/obvaca4b25cmvag34es3r6zdbm.html (accessed on 12.10.2020) [in Spanish].

[86] "EU stops short of recommending full ban on Huawei from 5G network", *CBC*, 2020 [Online]. Available: https://www.cbc.ca/news/technology/eu-not-recommending-ban-huawei-5g-1.5444158 (accessed on 25.06.2020).

[87] "Nokia and Ericsson remain vulnerable in geopolitical 5G tussle", *Financial Times*, 2020 [Online]. Available: https://www.ft.com/content/1ac1db39-7817-4fe2-ace8-8b67714aafd6 (accessed on 25.06.2020).

[88] "Operator members", *O-RAN Alliance*, 2020 [Online]. Available: https://www.o-ran.org/membership (accessed on 17.06.2020).

[89] "5G Action Plan for Denmark", *Danish Energy Agency*, 2018 [Online]. Available: https://ens.dk/sites/ens.dk/files/Tele/5g_action_plan_for_denmark.pdf (accessed on 25.06.2020).

[90] S. Kruse and L. Winther, "Banned recording reveals China ambassador threatened Faroese leader at secret meeting", *Berlingske*, 2019 [Online]. Available: https://www.berlingske.dk/internationalt/banned-recording-reveals-china-ambassador-threatened-faroese-leader (accessed on 25.06.2020).

[91] M. Druś, "Wielka Brytania proponuje alians 10 państw w celu redukcji zależności w 5G od Chin", *Puls Biznesu*, 2020 [Online]. Available: https://www.pb.pl/wielka-brytania-proponuje-alians-10-panstw-w-celu-redukcji-zaleznosci-w-5g-od-chin-992502 (accessed on 25.06.2020) [in Polish].

[92] "Rząd Wielkiej Brytanii ogranicza udział Huawei w budowie sieci 5G", *Gazeta Prawna*, 2020 [Online]. Available: https://serwisy.gazetaprawna.pl/nowe-technologie/artykuly/1450926,huawei-siec-5g-boris-johnson.html (accessed on 25.06.2020) [in Polish].

[93] S. Palczewski, "Austria nie wykluczy Huawei z budowy 5G", 2020 [Online]. Available: https://www.cyberdefence24.pl/austria-nie-wykluczy-huawei-z-budowy-5g (accessed on 26.06.2020) [in Polish].

[94] "Słowacja: brak dowodów na zagrożenie bezpieczeństwa ze strony Huawei", *Cyber Defence 24*, 2019 [Online]. Available: https://www.cyberdefence24.pl/slowacja-brak-dowodow-na-zagrozenie-bezpieczenstwa-ze-strony-huawei (accessed on 26.06.2020). [in Polish]

[95] G. Szakacs and K. Than, "Hungarian minister opens door to Huawei for 5G network rollout", *Reuters*, 2019 [Online]. Available: https://www.reuters.com/article/us-hungary-telecoms-huawei/hungarian-minister-opens-door-to-huawei-for-5g-network-rollout-idUSKBN1XF12U (accessed on 26.06.2020).

[96] K. Bogacki, "5G w Rosji zbuduje Huawei", *Chip*, 2019 [Online]. Available: https://www.chip.pl/2019/06/5g-w-rosji-zbuduje-huawei/ (accessed on 26.06.2020) [in Polish].

[97] A. Segal, "Year in Review 2019: The U.S.-China tech cold war deepens and expands", *Council on Foreign Relations*, 2019 [Online]. Available: https://www.cfr.org/blog/year-review-2019-us-china-tech-cold-war-deepens-and-expands (accessed on 26.06.2020).

[98] K. Sahin, "The tech cold war illusion", *Berlin Policy Journal*, 2020 [Online]. Available: https://berlinpolicyjournal.com/the-tech-cold-war-illusion/ (accessed on 15.06.2020).

Urszula Soler is a sociologist who explores, in theory and practice, new technologies in the context of social interaction and security. She works at John Paul II Catholic University in Lublin and at the War Studies University. She specializes in issues related to technology assessments, security of technology and sociology of security. She is a member of, inter alia, the Polish Society for Technology Assessment and the Italian Team for Security, Terroristic Issues & Managing Emergencies in Milan. She is an author of many articles and scientific projects. She has been cooperating with Università Cattolica del Sacro Cuore in Milan for many years now.

https://orcid.org/0000-0001-7868-8261
E-mail: urszula.soler@kul.pl
Institute of Sociological Sciences
Faculty of Social Sciences
John Paul II Catholic University
Al. Racławickie 14
20-950 Lublin, Poland

Paper

Ka Band-pass Filter Based on SIW Technology for Wireless Communications

Mehdi Damou[1], Yassine Benallou[1], Mohammed Chetioui[2], Abdelhakim Boudkhil[2], and Redouane Berber[1]

[1] *Dr. Tahar Moulay University of Saida, Saida, Algeria*
[2] *Abu Bakr Belkaid University of Tlemcen, Tlemcen, Algeria*

https://doi.org/10.26636/jtit.2021.150121

Abstract—The paper proposes a new third-order Chebyshev bandpass filter based on the substrate integrated waveguide (SIW) manufacturing technology using an inductive iris and a defected ground structure (DGS) station to resonate in the Ka frequency band, intended for wireless communication applications. All steps that are necessary for designing such a filter have been described in detail based on specific analytical equations harnessed to calculate the different synthesizable parameters of the proposed band-pass filter design, such as the coupling matrix, quality coefficients and initial geometric dimensions. The filter's ideal frequency response is extracted from an equivalent circuit employing localized elements developed with the use of Design Microwave Office Software. Otherwise, HFSS is employed to set the initial parameters of the proposed topology that will not meet the target specifications defined previously. Accordingly, optimization procedures are necessary for different SIW band-pass filter parameters to reach a high frequency response for the proposed design. The detailed results presented show high efficiency of the SIW technology that offers good performance with lower filter volumes. Two topologies have been developed and then optimized to demonstrate the usefulness of EM software.

Keywords—coupling matrix, DGS, filters, Ka band, SIW.

1. Introduction

Operational frequency bands used in satellite wireless systems have become highly saturated these days, especially due to huge demand created by multimedia applications. This requires that the operational frequency range be broadened by considering new design techniques and technologies for different passive and active microwave devices. Accordingly, a thorough investigation has been performed to study the behavior of superconducting front-end planar devices at the Ka band (26–40 GHz) in order to expand the additional range and to meet the future requirements of wireless communications [1]. In fact, with all the prospects associated with this considerable frequency band, it is always desirable to design simple, robust and reliable components, characterized by reduced weight, energy consumption and by low cost. Accordingly, this study investigates primarily the modern technologies relied upon for developing integrated filters operating in the Ka band, notably the SIW technology, in order to demonstrate its high integrability and to achieve good performance parameters.

Traditionally, waveguides were used in designing high-performance filters that claimed a complicated transition to integrated planar devices taking into account their bulky size. The straight forward solution is to combine a rectangular waveguide into the microstrip structure, forming the so-called substrate-integrated waveguide [2] offering good quality factor Q at the input due to the dielectric filling caused by volume reduction [3]. SIW side walls are commonly formed by using metallic via-holes or post walls [4]. Furthermore, many manufacturing techniques may be combined with the substrate integrated waveguide technology to implement different filter topologies. These include, for instance, the DGS technique that is an emerging method relied upon to enhance such filter parameters as narrow operating bandwidth or high weight [5].

This paper presents two design topologies of a novel third pole band-pass filter based on the SIW technology to operate at the Ka band. The first design uses a simple cavity structure. The other, however, employs a new SIW-DGS structure to eliminate losses by avoiding disturbance close to the resonance point and by suppressing higher mode harmonics, thus mastering mutual coupling.

The paper is organized as follows. Section 1 presents the main synthesis steps and coupling matrix calculations for the SIW filter. Section 2 describes, in detail, the process of synthesizing, simulating and optimizing the three pole SIW filter. Section 3 illustrates the design principles of the proposed filter and smiling of the admittance inverter based on perforated, and Section 4 compares the two topologies of the three pole SIW filter.

2. Substrate Integrated Wave Guide Filters

SIW filters with inductive shunt coupling may be designed easily using an air-filled rectangular waveguide. The filter

may be formed by either a shunt inductive post or an iris inductive shunt aperture, as reported in [6]. As indicated in Fig. 1, the first example of the developed SIW band-pass filter model employs the iris inductive shunt. Different resonances are caused due to inductive shunt coupling, in addition to two microstrip-to-SIW transitions at the input and the output. The smallest inductive shunts facilitate input/output coupling, while the biggest ones provide inter-resonator coupling [1]. The filter design procedures are similar to those of an air-filled waveguide filter based on the coupling-matrix technique. Finally, properties of the modeled SIW filter make it possible to manufacture low profile and low-cost microwave filters.

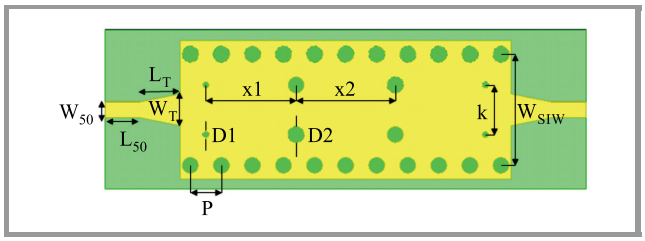


Fig. 1. First SIW band-pass filter model with the iris inductive shunt.

3. Formation of the Coupling Matrix from the Low Pass Prototype

The transmission coefficient of a low pass prototype filter network, defined as the ratio of the transmitted power to the available power, may be represented as [7]:

$$\left|S_{21}(s)\right|^2_{s=\mathrm{j}\omega} = \frac{1}{1+\varepsilon^2\left|K_N(s)\right|^2_{s=\mathrm{j}\omega}}\,, \tag{1}$$

where ε is related to the band-pass ripple and the band-pass return loss R_L via:

$$\varepsilon = \frac{1}{\sqrt{10^{\frac{R_L}{10}}-1}}\,. \tag{2}$$

$K_N(s)$ is the characteristic function of degree N and has the following form:

$$\begin{cases} K_N(\mathrm{j}\omega)\cosh\left(\sum\limits_{n=1}^{N}\cosh h^{-1}(x_n)\right) \\ x_n = \dfrac{\omega-\frac{1}{\omega_n}}{1-\frac{\omega}{\omega_n}} \end{cases}\,, \tag{3}$$

where $s_n = \mathrm{j}\omega_n$ is the location of the n-th transmission zero in the complex s plane. Since the rational polynomial $K_N(s)$ represents the ratio of polynomials $F(s)$ and $P(s)$, it is more appropriate to represent $K_N(s)$ in the following form:

$$K_N(s) = \frac{\frac{F(s)}{P(s)}}{\varepsilon}\,. \tag{4}$$

Polynomials $F(s)$ and $P(s)$ are formed by the zeros of reflection and transmission, respectively, and are assumed to be known with their highest coefficients as unity. Using the conservation of energy formula for a lossless filter, the polynomial $E(s)$ is determined as:

$$\left|E(s)\right|^2 = \left|F(s)\right|^2 + \frac{\left|P(s)\right|^2}{\varepsilon^2}\,. \tag{5}$$

The coupling values M_{ij} ($i=1,2,\ldots,N$ and $j=1,2,\ldots,N$) fully define the filter. Using the Kirchhoff's voltage law, the coupling matrix is derived via an impedance matrix from a set of loop equations.

$$\begin{pmatrix} e_s \\ 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} = \begin{pmatrix} s+R_s & -\mathrm{j}\omega M_{12} & -\mathrm{j}\omega M_{13} & \ldots & \cdot & -\mathrm{j}\omega M_{1N} \\ -\mathrm{j}\omega M_{12} & s & -\mathrm{j}\omega M_{23} & \ldots & \cdot & \cdot \\ -\mathrm{j}\omega M_{13} & -\mathrm{j}\omega M_{23} & s & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ -\mathrm{j}\omega M_{1N} & \cdot & \ldots & -\mathrm{j}\omega M_{N,N-1} & s+R_L & \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ i_3 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ i_N \end{pmatrix}.$$

The loop analysis of the filter yields the following result:

$$[e] = [A][i]\,. \tag{6}$$

For $L = L_1 \ldots = L_i$ [H], $C = C_1 \ldots C_i$ [F], and $s = \mathrm{j}\omega$ and the two-port scattering parameters, the transfer and reflection functions are given by [3]:

$$\begin{aligned} S_{21} &= -2\mathrm{j}\sqrt{R_s R_L}[A]^{-1}_{(n,1)} \\ S_{11} &= 1+2\mathrm{j}R_s[A]^{-1}_{(1,1)}\,. \end{aligned} \tag{7}$$

The general matrix A comprising coupling coefficients m_{ij} and external quality factors $q_{i,ext}$ is presented in [4] as:

$$[A] = \begin{bmatrix} \frac{1}{q_{e1}} & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \frac{1}{q_{en}} \end{bmatrix} + p\begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} - \mathrm{j}\begin{bmatrix} m_{1,1} & m_{1,2} & \ldots & m_{1,n} \\ m_{2,1} & m_{2,2} & \ldots & m_{2,n} \\ \vdots & \vdots & \ldots & \vdots \\ m_{n,1} & m_{n,2} & \ldots & m_{n,n} \end{bmatrix}, \tag{8}$$

$$p = \mathrm{j}\frac{1}{\mathrm{FBW}}\left(\frac{\omega}{\omega_0}-\frac{\omega_0}{\omega}\right),$$

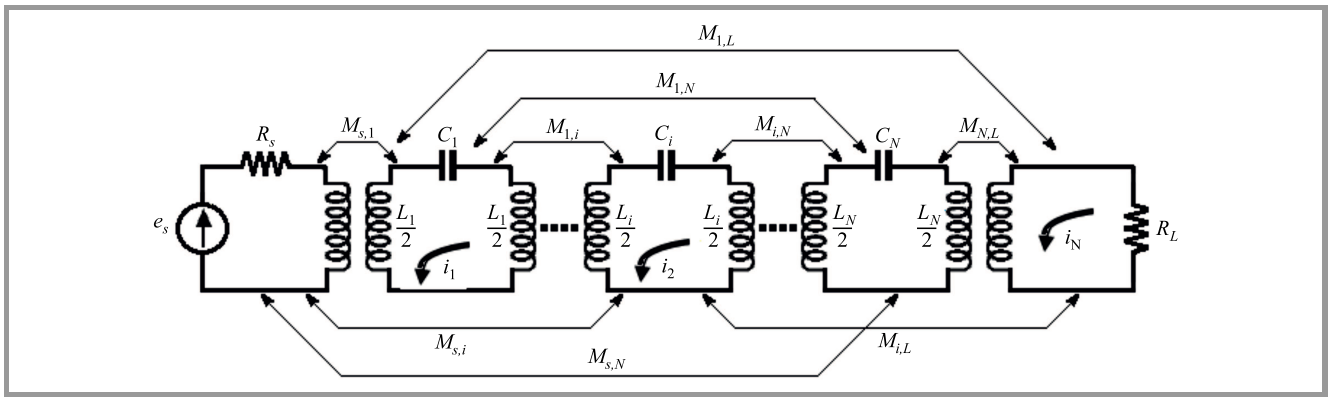


Fig. 2. Generalized model of a cross-coupled resonator filter.

where p is the complex lowpass frequency, ω_0 is the center frequency of the filter, FBW is the fractional bandwidth of the filter, $q_{i,ext}$ $(i = 1, \ldots, n)$ is the scaled external quality factors of the resonator i, and m_{ij} is the normalized coupling coefficients between the resonator i and j [8].

4. SIW Filter Design Using Cross Coupled Resonators

For demonstration purposes, a highly selective third-pole SIW filter with the configuration shown in Fig. 2 has been designed. A cross coupled model is given as an example to illustrate the proposed solution. The target specifications of the SIW band-pass filter are shown in Table 1.

Table 1
Target specifications of the SIW band-pass filter

Center frequency	$f_o = 28.86$ GHz
Bandwidth	BW = 1050 MHz (FBW = 3.63%)
Bandpass ripples	$L_{AR} = 0.0431$ dB
Bandpass return loss	$R_L = 20$ dB
Stopband rejection level	> 40 dB

The element values of the lowpass prototype filter are found to be $g_1 = g_3 = 0.8516,\ g_2 = 1.1032$. The design parameters, i.e. the de-normalized external quality factors, can be determined by means of the following formulas [9]:

$$Q_{eS} = \frac{g_0 g_1}{\text{FBW}}, \quad Q_{eL} = \frac{g_n g_{n+1}}{\text{FBW}}\ , \tag{9}$$

$$M_{i,i+1} = \frac{\text{FBW}}{\sqrt{g_i g_{i+1}}}, \quad i = 1 \text{ to } n-1\ , \tag{10}$$

$$m_{i,i+1} = \frac{M_{i,i+1}}{\text{FBW}}, \quad i = 1, \ldots, n\ . \tag{11}$$

From Eq. (10), in the first step of the design, the lowpass prototype filter elements with the following ideal coupling matrix are evaluated:

$$M = \begin{bmatrix} 0 & 0.0375 & 0 \\ 0.0375 & 0 & 0.0375 \\ 0 & 0.0375 & 0 \end{bmatrix}\ .$$

$$M = Q_{ext\_e} = Q_{ext\_s} = 23.4068\ .$$

Capacitance C_0 and the inductance L_0 are derived from:

$$C_0 = \frac{Q_{e1}}{w_0 \times Z} = 2.5816 \text{ pF}\ , \tag{12}$$

$$L_0 = \frac{Q_{e1}}{w_0 \times Q_{e1}} = 0.0118 \text{ nH}\ . \tag{13}$$

The series impedances Z_{12}, Z_{23}, Z_{34} are:

$$Z_{i,i+1} = \frac{Z}{M_{i-1,i} \times Q_{e1}}\ . \tag{14}$$

The series impedances are: $Z_{01} = 50\ \Omega$, $Z_{12} = 56.91\ \Omega$, $Z_{23} = 56.91\ \Omega$.

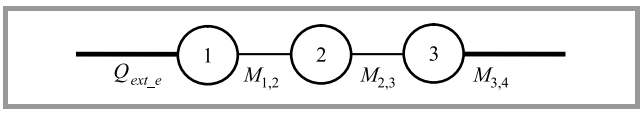


Fig. 3. Generalized model for SIW coupled resonator filter.

Figure 3 shows the corresponding coupling/routing diagram with SIW resonators used to implement the filtering methods that is characteristic of the SIW technology. $M_{i,i+1}$ indicates the direct coupling sequence.
Figure 4 shows an equivalent circuit of the BPF. This circuit consists of three parallel LCR resonators which are separated by wavelength microstrip lines. As shown in Fig. 4, the transmission line $\theta = \pm 90^\circ$ at the center frequency and $w_0 = \frac{1}{\sqrt{L}}$ are employed to represents the coupling coefficient. This circuit model does account for the loss effect. The RLC microstrip transmission line unit section is simulated using Microwave Office (AWR) Software. The proposed filter's ideal frequency response of the equivalent circuit is presented in Fig. 5, which provides a return loss coefficient of less than 25 dB for a bandpass of 1.05 GHz at the center frequency of 28.86 GHz.

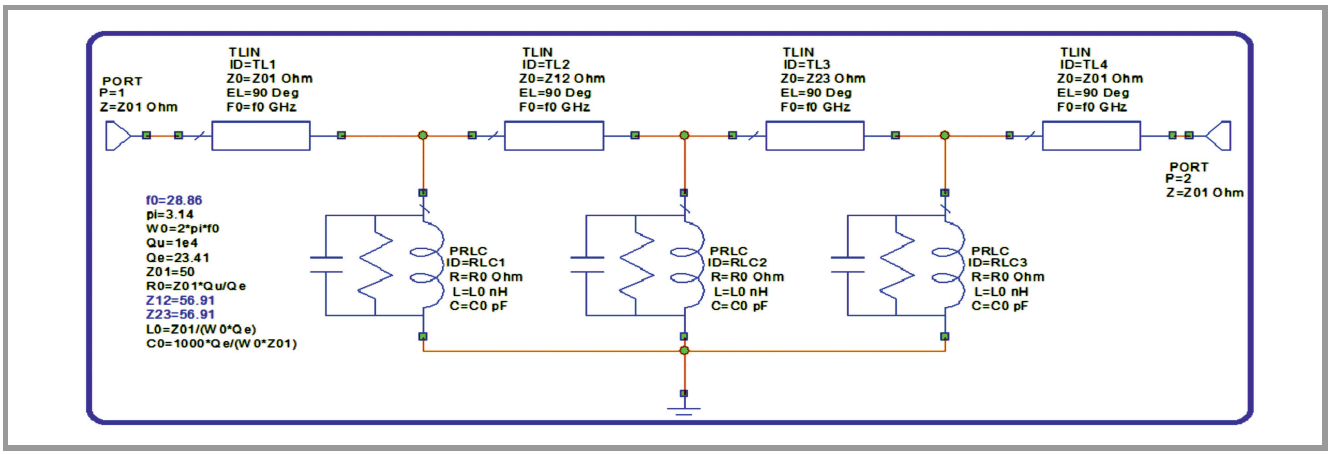


Fig. 4. Coarse model of the third order SIW band-pass filter.

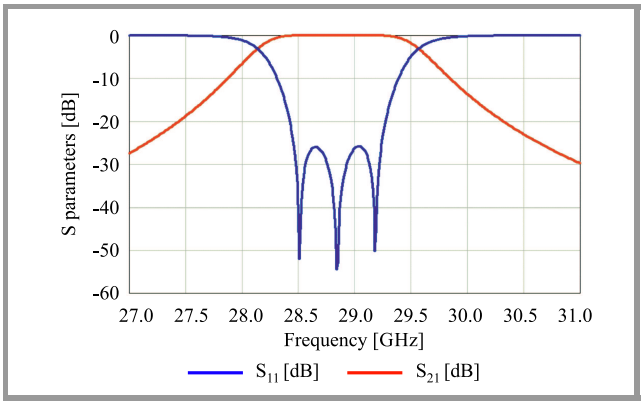


Fig. 5. Ideal response curve of the third order SIW to a band-pass filter based on the CM technique.

5. Simulation of Third Order Band-Pass SIW Filter

In this section, a third order iris SIW band-pass filter is simulated. In the prototype fabrication process, the Arlon Di-Clad 880 substrate with a thickness of 0.25 mm, a dielectric constant of 2.2 and a loss tangent of 0.009 was used. For the desired resonance frequency of 28.86 GHz, diameter d of the metal sights is 0.7 mm, with pitch P of 1.04 mm to prevent radiation leakage [10]:

$$\begin{cases} \lambda_c = \dfrac{c}{f_{101}\sqrt{\varepsilon_r}} = \dfrac{3\cdot 10^8}{28.86\cdot 10^9\sqrt{2.2}} \approx 7 \text{ mm} , \\ \dfrac{d}{\lambda_c} \approx 0.1 \Rightarrow d = 0.7 \text{ mm} , \\ \dfrac{d}{P} \approx 0.67 \Rightarrow P = 1.04 \text{ mm} . \end{cases}$$

Length Leq and width Weq are equal to 5 mm for a single SIW square cavity, as calculated in:

$$f_{101} = \frac{c}{2\pi\sqrt{\varepsilon_r}}\sqrt{\left(\frac{\pi}{Weq}\right)^2 + \left(\frac{\pi}{Leq}\right)^2} = 28.86 \text{ GHz} ,$$

$$Weq = Leq = 5 \text{ mm} .$$

W_{SIW} spacing between the two rows of holes is a relevant physical parameter that is necessary for designing a SIW structure. It may be determined from empirical equations as a function of the equivalent width Weq of the rectangular waveguide, offering the same characteristics in the fundamental mode for the same height and the same dielectric constant:

$$W_{\text{SIW}} = Weq + \frac{d^2}{0.95P} \approx 5.5 \text{ mm} ,$$

$$W_{\text{SIW}} = L = 5.5 \text{ mm (square cavity)} ,$$

$$W_t = 0.4(W_{SIW} - d) \approx 7.304 \text{ mm} ,$$

$$\begin{cases} \dfrac{\lambda}{2} \le L_t \le \lambda , \\ \lambda = \dfrac{c}{f_0\sqrt{\varepsilon_r}} , \\ L_t \approx 13.12 \text{ mm} . \end{cases}$$

The preliminary dimensions of the third-order band-pass filter structure are presented in Table 2, and its topology is shown in Fig. 6. The three resonators are coupled together using an inductive iris, while the two end resonators are coupled to excite and load by 50 Ω microstrip line.
Figure 7 illustrates a very low frequency response when simulating the structure's initial geometric parameters

Table 2
Preliminary dimensions of the third order band-pass filter

Parameters group	Type	Descriptor	Size [mm]
Taper	Width	W_{Taper}	1.92
	Length	L_{Taper}	4.00
SIW guide	Width	W_{SIW}	5.50
	Via	P	1.04
	Diameter	D_1	0.70
Resonator	Width	x_1	4.50
	Width	x_2	4.50
	Width	D_2	0.15
	Length	k	2.40

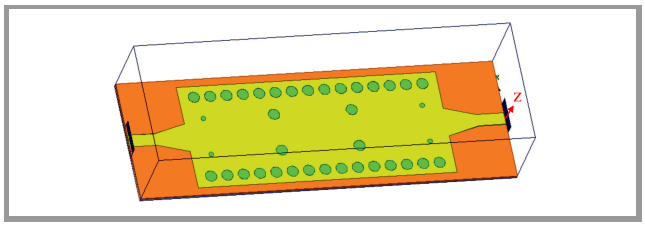


Fig. 6. 3D view of the proposed filter.

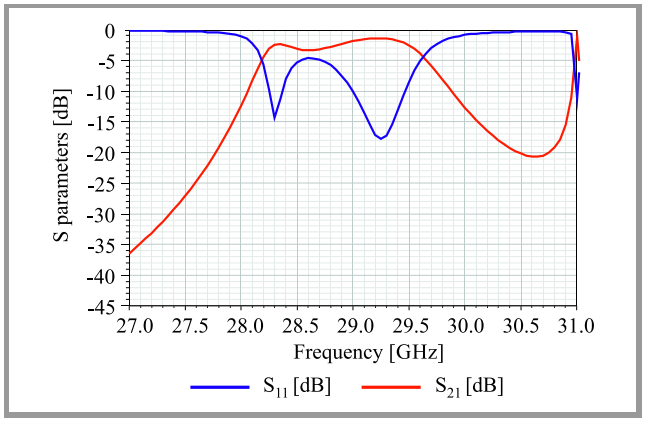


Fig. 7. S parameters of the proposed SIW band-pass filter with initial geometric parameters.

which provide a very poor reflection and transmission coefficients of 5 dB and 2 dB, respectively, for a passband greater than 1.45 GHz in the 28.20–29.65 GHz band. Accordingly, the structure is then optimized using HFSS software to improve the filter response.

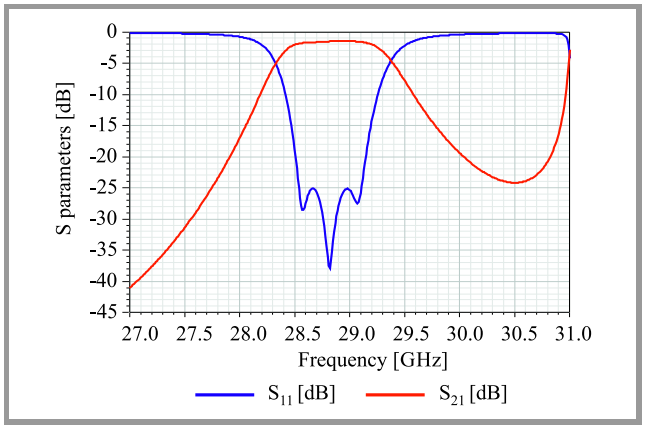


Fig. 8. S parameters of the optimized SIW band-pass filter.

Figure 8 shows a high frequency response for the optimized structure. The reflection coefficient is less than 25 dB and the transmission coefficient is approximately 1 dB for the resonance frequency of 28.82 GHz and a passband of 1.05 GHz. Table 3 presents, in detail, the different dimensional changes introduced based on optimization procedures. Optimized dimensions of the 3rd order SIW band-pass filter structure are presented in Table 3.

Table 3
Optimized dimensions of the filter

Parameter	Size [mm]	Parameter	Size [mm]
W_{50}	0.8	l_{50}	2.00
W_t	1.6	L_t	4.00
W_{SIW}	5.6	a_1	4.95
d_1	0.80	a_1	4.50
d_2	0.15	k	2.4

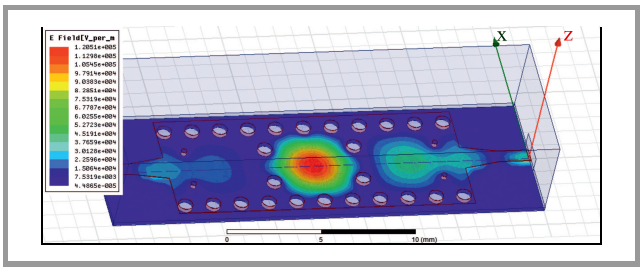


Fig. 9. Simulated E field distribution.

Figure 9 presents the simulated field layout of the proposed SIW band-pass filter for the TE101 propagation mode, at the center frequency with two circular propagating zones in a single SIW filter cavity.

6. SIW Structure Variation Effects

As shown in Fig. 10, after optimizing the iris inductive shunt filter with the use of the HFSS simulator, the frequency response becomes very close to the ideal specification with a passband of 1.045 GHz, center frequency of approx. 28.85 GHz, and reflection losses of less than 25 dB. The curves presented in Fig. 11 demonstrate that variations of the opening diameter influence the center frequency inversely to the WSIW width variation effect. In fact, an increase in the diameter causes right shifting, while a decrease in the diameter causes left shifting. As a result, the

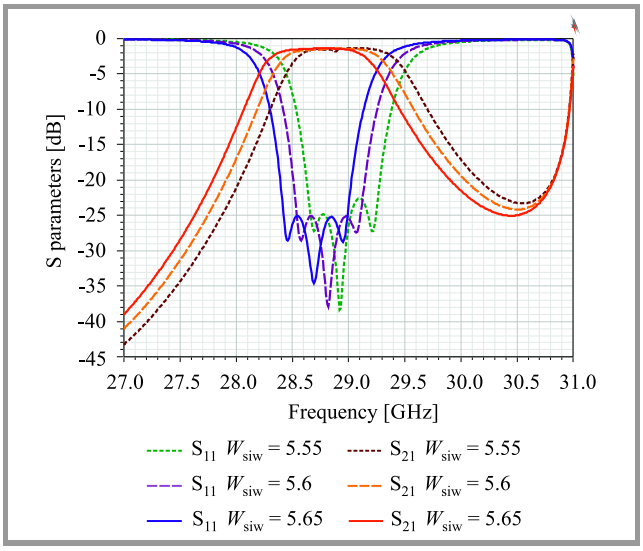


Fig. 10. W_{SIW} spacing variation effect.

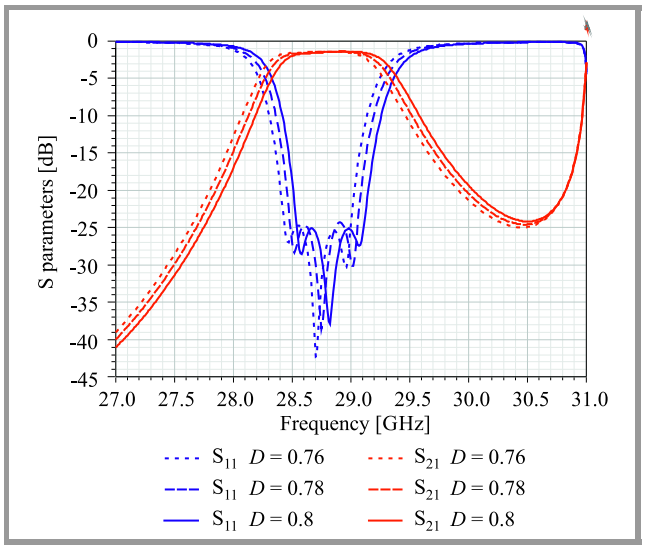


Fig. 11. Variation of the D diameter.

resonant frequency may be a function of both, with changes to the opening diameter and W_{SIW} width introduced.

7. SIW Band-pass Filter Based DGS Technique

In the next step, the proposed model has been modified by implementing the defected ground structure (DGS), as shown in Fig. 12. The DGS technique is implemented through three-slot forms at the ground's upper surface, in order to obtain a very low profile for a passband of 4 GHz and center frequency of 27 GHz. Table 4 presents the parameters of the implemented slots described in Fig. 12.

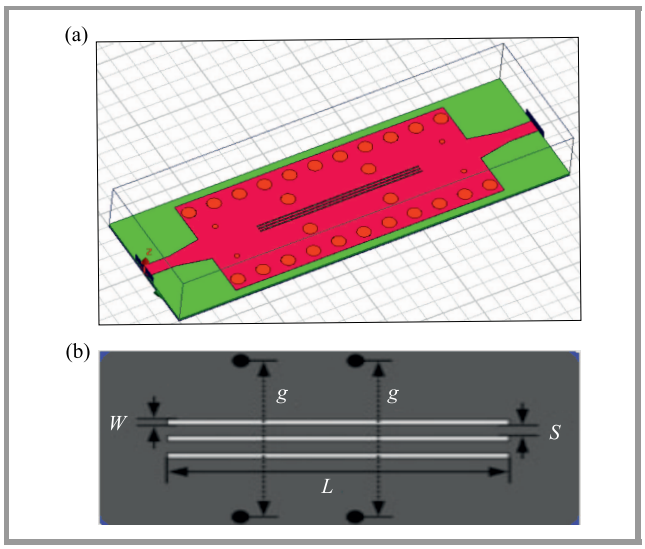


Fig. 12. Topology of the SIW band-pass filter using the DGS technique: (a) HFSS model, (b) slot geometry.

Table 4
Slot form dimensions

Parameter	Value [mm]	Parameter	Value [mm]
W	0.1	g	2.50
S	0.1	L	9.95

Simulation results presented in Fig. 13 show that the response of the new filter is improved, as the low return loss equals 27 dB and good transmission with 1 dB for a central frequency of approximately 29 GHz and a 4.12% fractional bandwidth.

Distribution of E field does not change when implementing slots at the ground Fig. 14. This demonstrates that the DGS technique allows to enhance the filter's response without affecting field distribution. In order to validate the two SIW filter models proposed above, a comparison with paper [11] using shunt inductive posts is performed

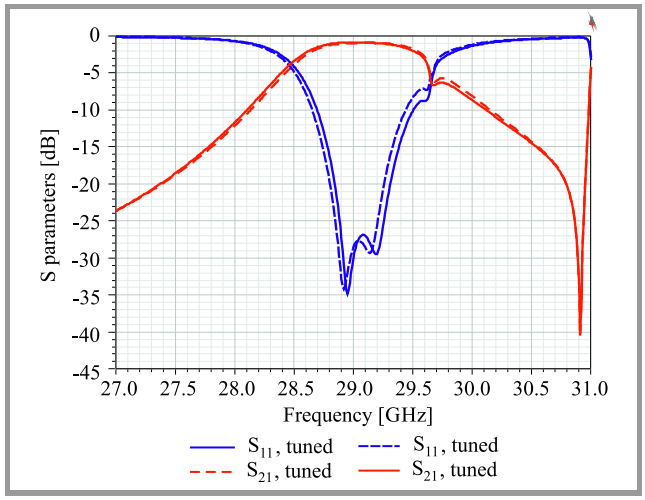


Fig. 13. S parameters of the SIW band-pass iris shunt inductive filter based on DGS.

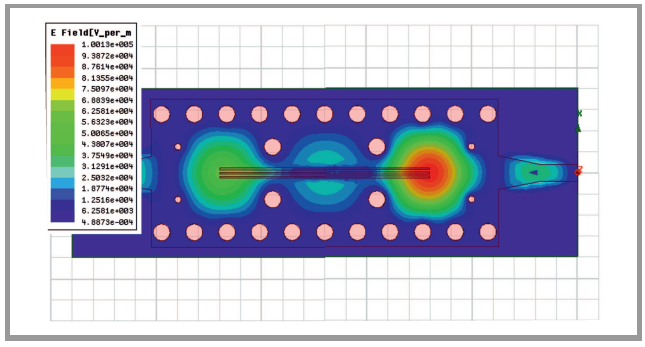


Fig. 14. Simulated E field distribution obtained by SIW DGS band-pass filter.

Table 5
EM frequency response comparison

Reference	FBW and F0	Insertion loss [dB]	Return loss [dB]
[11], filter using inductive shunt posts	3.87%, 29.18	1.5	15
First model – SIW iris inductive shunt filter	3.64%, 28.82	1.5	25
Second model – SIW based DGS filter	4.12%, 29.09	1.1	27

for the same filter order and passband of between 28.35 and 29.75 GHz, as presented in Table 5.
The proposed filter designs offer good electromagnetic performance in the Ka band. They are characterized by a low return loss and good transmission compared with the designs presented in other studies that validate both the accuracy of the modeling method relied upon and the efficiency of the selected EM simulator.

8. Conclusion

This paper presents two novel topologies of a third order band-pass iris shunt inductive filter using the SIW technology and the DGS technique. Both filters use the coupling matrix method which turns out to be a very useful tool for direct filter synthesis. Analytical formulas are used to calculate the filter's different geometric parameters and efficient optimization techniques are employed to reduce the effective dielectric permittivity on certain waveguide sections in order to achieve high performance demonstrated by a low reflection coefficient and a good transmission coefficient for the entire passband. Filters of this type are designed for use used in the Ka frequency band that is very suitable for satellite communications.

References

[1] I. Ohta, K. Toda, M. Kishihara, and T. Kawai, "Design of cruciform substrate-integrated waveguide hybrids based on H-plane planar circuit approach", in *Proc. of Asia-Pacific Microw. Conf.*, Bangkok, Thailand, 2007, pp. 683–686 (DOI: 10.1109/APMC.2007.4554871).

[2] M. Bozzi, L. Perregrini, K. Wu, and P. Arcioni, "Current and future research trends in substrate integrated waveguide technology", *Radio Engin.*, vol. 18, no. 2, 2009 [Online]. Available: https://www.radioeng.cz/fulltexts/2009/09_02_201_209.pdf

[3] A. Coves *et al.*, "A novel passband filter based on a periodically drilled SIW Structure", *Radio Sci.*, vol. 51, no. 4, pp. 328–336, 2016 (DOI: 10.1002/2015RS005874).

[4] L. Silvestri *et al.*, "Modeling and implementation of perforated SIW filters", in *Proc. IEEE MTT-S Int. Conf. on Numer. Electromag. and Multiphys. Model. and Optimiz. NEMO 2016*, Beijing, China, 2016, pp. 209–210 (DOI: 10.1109/NEMO.2016.7561668).

[5] D. Dealandes and K. Wu, "Single-substrate integration techniques for planar circuits and waveguide filters", *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 2, 593–596, 2003 (DOI: 10.1109/TMTT.2002.807820).

[6] S. Moscato, R. Moro, M. Pasian, M. Bozzi, and L. Perregrini, "Two-material ridge substrate integrated waveguide for ultra-wideband applications", *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 10, pp. 3175–3182, 2015 (DOI: 10.1109/TMTT.2015.2461612).

[7] R. J. Cameron, "Advanced coupling matrix synthesis techniques for microwave filters", *IEEE Trans. on Microw. Theory and Techn.*, vol. 51, no. 1, 2003 (DOI: 10.1109/TMTT.2002.806937).

[8] R. Bouhmidi, B. Bouras, and M. Chetioui, "Multi-ports extraction technique for microwave passband filter optimization", *Int. J. of Microw. and Opt. Technol.* (*IJMOT*), vol. 14, no. 6, pp. 431–439, 2019 [Online]. Available: https://ijmot.com/VOL-14-NO-6.aspx

[9] X.-P. Chen and K. Wu, "Substrate integrated waveguide filter: Basic design rules and fundamental structure features", *IEEE Microw. Mag.*, vol. 15, no. 5, pp. 108–116, 2014 (DOI: 10.1109/MMM.2014.2321263).

[10] F. Xu and K. Wu, "Guided-wave and leakage characteristics of substrate integrated waveguide", *IEEE Trans. on Microw. Theory and Techn.*, vol. 53, no. 1, pp. 66–73, 2005 (DOI: 10.1109/TMTT.2004.839303).

[11] F. Parment, A. Ghiotto, T. P. Vuong, J. M. Duchamp, and K. Wu, "Air-filled substrate integrated waveguide for low-loss and high powerhandling millimeter-wave substrate integrated circuits", *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 4, pp. 1228–1238, 2015 (DOI: 10.1109/TMTT.2015.2408593).

Mehdi Damou received his Ph.D. in Telecommunications from Abu Bakr Belkaid University of Tlemcen, Algeria, in 2018. He is a lecturer and the Head of the Electronics Department at Dr. Tahar Moulay University of Saida, Algeria. His research interests include microwave and RF devices and components. He is working on developing antennas designs and microwave filters based on SIW technologies and efficient EM modeling techniques.
https://orcid.org/0000-0003-4448-3318
E-mail: bouazzamehdi@yahoo.fr
Laboratory of Technologies of Communications
Dr. Tahar Moulay University of Saida
Saida, Algeria

Yassine Benallou received his Ph.D. in Electronics from Djillali Liabes University of Sidi Bel Abbes, Algeria, in 2014. He is a lecturer at the Electronics Department of Dr. Tahar Moulay University of Saida, Algeria and has been a senior member at the Laboratory of Technologies of communications since its establishment. His academic research focuses on different fields, including modeling and designing of electronic circuits and systems, optimizing passive EM components and characterizing new materials for biomedical and renewable applications.
E-mail: benallou06@yahoo.fr
Laboratory of Technologies of Communications
Dr. Tahar Moulay University of Saida
Saida, Algeria

Mohammed Chetioui received his Ph.D. in Telecommunications from Abu Bakr Belkaid University of Tlemcen, Algeria, in 2018. Since then, he has been a lecturer at the Electronics Department of Dr. Tahar Moulay University of Saida, Algeria. His research interests include digital communications, signal processing, microwave circuits

and RF systems. He is working on designing passive/active microwave filters based on the microstrip technology and accurate optimizations.
E-mail: chetioui.mohammed@yahoo.fr
Laboratory of Telecommunications
Abu Bakr Belkaid University of Tlemcen
Tlemcen, Algeria

Abdelhakim Boudkhil received his Ph.D. in in Electronics from Abu Bakr Belkaid University of Tlemcen, Algeria, in 2018. He is an Assistant Professor at the Electronics Department at Dr. Tahar Moulay University of Saida, Algeria. His research experience concerns several fields, including digital, optical, microwave, and RF communication systems. His research interests focus more on optimizing and developing antennas based on integrated technologies and advanced techniques.
E-mail: boudkhil.abdelhakim@yahoo.fr
Laboratory of Telecommunications
Abu Bakr Belkaid University of Tlemcen
Tlemcen, Algeria

Redouane Berber received his M.Sc. in Electronics from Dr. Tahar Moulay University of Saida, Algeria, in 2008. He is an Assistant Professor and a deputy head at the Electronics Department at Dr. Tahar Moulay University of Saida, Algeria. His research interests focus specifically on electronic components and systems, as well as on optical communications and networking. He is working on investigating different multiple access techniques.
E-mail: red1ber@gmail.com
Laboratory of Technologies of Communications
Dr. Tahar Moulay University of Saida
Saida, Algeria

Paper

Vlasov Launcher Diagrammatic Design Using the RT Method

Andrzej Francik, Grzegorz Jaworski, Maciej Nowak, and Kacper Nowak

Faculty of Electronics, Wrocław University of Technology, Poland

https://doi.org/10.26636/jtit.2021.150321

Abstract—In this paper, a simple and fast method relied upon for designing a Vlasov launcher with a helical cut is proposed. The method is based on graphic interpretation of analytical relationships that link wave parameters (EM field mode) to the launcher's geometrical dimensions. Using the ray tracing method, a simplified graphic analysis may be carried out. The results obtained are not significantly different from those of rigorous full-wave analyzes. The family of normalized curves that is created in the process greatly facilitates the stage of optimizing the geometrical parameters of the Vlasov launcher.

Keywords—gyrotron output, ray tracing, Vlasov launcher.

1. Introduction

Over the past decade, interest in the gyrotron technology has increased significantly due to its potential applications in many fields, such as plasma heating, material processing, radar and communications systems, and medical research. One of the most promising fields is that of fusion energy-based power generation. A mode converter is a key component enabling efficient RF propagation in a high power gyrotron. In recent years, several designs of a quasi-optical mode converter have been presented [1]–[4]. The quasi-optical mode converter with a Vlasov launcher is still incorporated in modern gyrotron designs.

Bian *et al*. presented a broadband quasi-optical mode converter that could function at three segment-continuous frequencies in three different waveguide modes [5]. Gao *et al*. described the design and fabrication of a frequency-agile gyrotron used for frequency-chirped MAS DNP [6]. Alaria *et al.* described the design of a helical cut, smooth-wall Vlasov launcher for converting the $TE_{22,6}$ mode to a Gaussian mode, used in a 120 GHz, 1 MW gyrotron [7]. Zhang *et al.* described a study on a gyrotron with a quasi-optical mode converter for terahertz imaging [8].

The Vlasov launcher still appears to be attractive for designing conventional high-power gyrotrons operating in the high terahertz range. At this frequency range, the complicated structure of a Denisov launcher makes precise design and manufacturing processes increasingly difficult. A properly optimized internal converter based on a Vlasov launcher is compatible with the gyrotron's electro-optical system and may generate a clear Gaussian beam with an efficiency of more than 80% in an extraordinarily broadband range [1]. A further correction with the use of mirrors allows to obtain the efficiency of up to 90% [9].

For the design of a Vlasov launcher, approaches based on the geometric optics theory and specialized software, such as electric field integral equation code (SURF3D LOT/SURF-3D) are frequently used [10]. Full-wave electromagnetic simulation software, such as CST, is used as well [7], [8]. This article presents an analysis of propagation of electromagnetic waves in the gyrotron's internal components. The results of such an analysis serve as a basis for developing an effective launcher design procedure. The theoretical foundations of the ray tracing (RT) method are relied upon in the analysis in a manner that is presented exhaustively in [11]. The RT method is very good for designing key elements of a gyrotron, such as the waveguide (forming the input section of the launcher), the launcher, and the mirror system. It relies on much simpler mathematical formulas to describe the propagation of electromagnetic waves – especially when compared with methods based on Maxwell's field equations. The proposed design procedure was also verified, with good results, by comparing it with calculations based on data published in the literature of the subject.

The analysis of propagation of electromagnetic waves in the launcher, being the gyrotron's internal component, is presented and serves as a basis for developing an effective launcher design procedure. A Vlasov antenna [12], designed using the oversized circular waveguide technique, is commonly used as a launcher in microwave gyrotron systems.

2. Design Procedure

The procedure uses a graphic representation of the known analytical relationships describing RT power transmission in a launcher system and allows the designer to make optimal decisions during the design process.

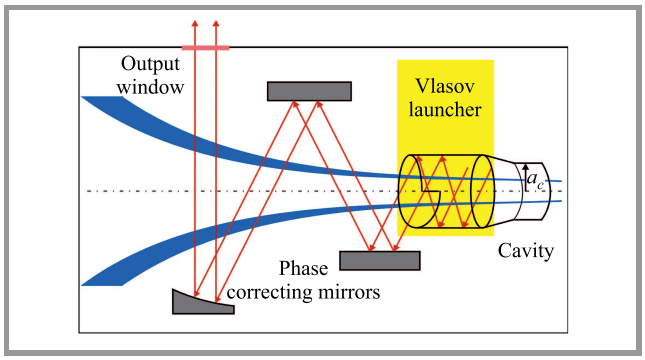


Fig. 1. Structure of the gyrotron's microwave output, with the Vlasov launcher highlighted in yellow. Blue arcs determine the spent electron trajectory, a_c – cavity radius.

An example of a typical internal structure of a gyrotron microwave system is shown in Fig. 1. The location of the launcher within the structure is highlighted in yellow. It is usually the gyrotron cavity that is the first designed part of an internal gyrotron microwave system. It is characterized by a resonant frequency that is equal to the desired gyro frequency in which the electron beam interacts with the cavity's resonant microwave field. This interaction results in the transfer of electron beam energy into the microwave field, with its energy reaching significant power levels (measured in megawatts). In the process of cavity design, its geometrical structure is selected, as is the microwave field mode and the dimensions that are calculated taking into account the power levels generated. After designing the cavity, the microwave field mode and the geometrical parameters of the cavity may be determined. These quantities usually serve as input data required to design the next element of the structure, namely the launcher. The field types, determined in such a way, are usually high-order TE_{mn} modes with m and n indexes whose values are much larger than 1. This, and the ability to work at high microwave power levels are the reasons behind using oversized microwave systems in situations in which the conditions allow to rely on the RT method. The launcher appearance and the ray presentation of the wave in the launcher are shown in Fig. 2.

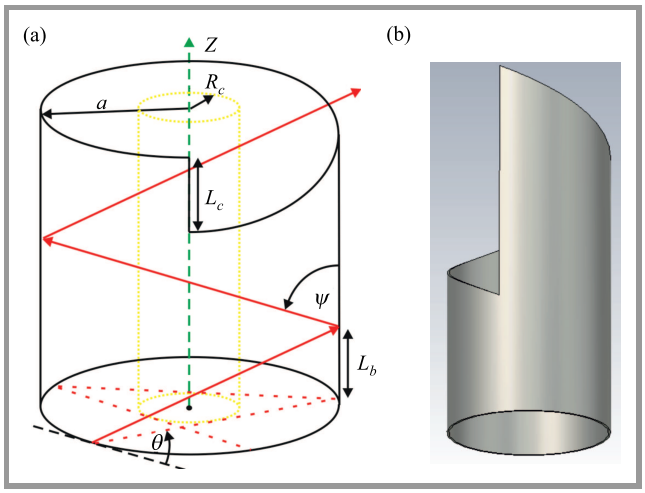


Fig. 2. Circular waveguide launcher having an L_c-length helical cut: a) ray presentation of the wave in the launcher, b) structural sketch.

In Fig. 2, most of the geometrical parameters of ray transmission (RT) are marked:

- a – waveguide radius,
- ψ – Brillouin angle,
- L_c – helical cut length (the smallest possible length of the launcher L),
- L_b – axial displacement (towards the waveguide axis) of two consecutive reflection points of the ray from the surface of the waveguide,
- R_c – caustic radius,
- θ – azimuth angle (angle 2θ is a central angle determined on the S plane, perpendicular to the waveguide axis and containing one ray reflection point, defining the length of arc $2\theta a$, where the chord is the projection of the ray on the S plane. The azimuth angle is clearly visible in Fig. 3, showing ray propagation of the wave beam in a circular waveguide.

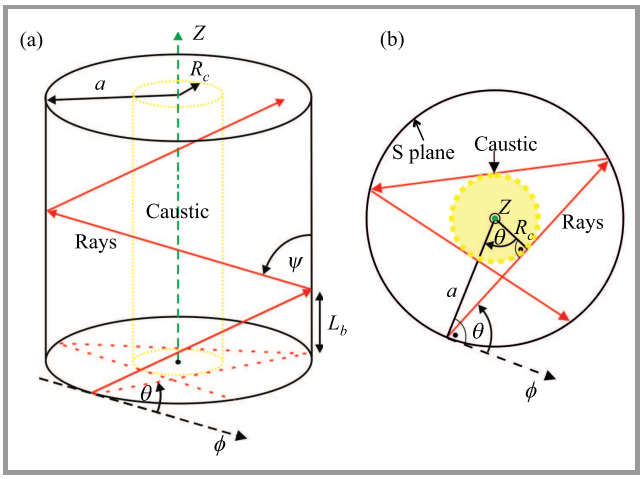


Fig. 3. Wave beam in a circular waveguide illustrated with the use of geometrical optics principle: a) side view, b) top view.

Another reflection point from the surface of the waveguide is located on the helix of the inclination angle τ, created on the waveguide's surface. The inclination angle τ, being another geometrical parameter of RT transmission, is visible on the unfolded surface of the launcher's waveguide, on the plane shown in Fig. 4. The straight lines, with their slopes measured in relation to the z axis, equal τ and represent the helix in the plane figure, while B0-B4 parallelograms shown are the Brillouin zones. Figure 4 offers an explanation to the strategy of selecting the cutting edges. The simplest helical cut of a circular waveguide is a straight section BD. If such a cut was to be made, the entire field

radiated from the Brillouin zone B1 would have encountered an obstacle. This would be a part of the waveguide wall designated as the Brillouin zone B2. This means that the field would obviously not be radiated into the free space. Therefore, this obstacle is removed by making a cut along the border between zones B3 and B4 and along the entire section of C'–C.

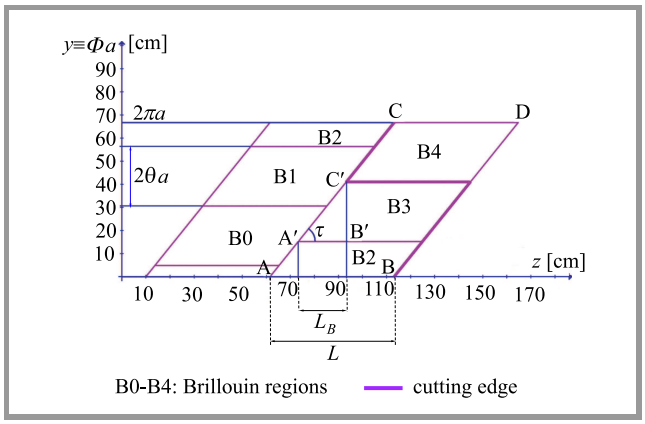


Fig. 4. The surface of a cylindrical waveguide launcher unfolded on a plane. The cutting edges of the launcher are represented using bold lines.

The RT representation shown above is satisfactorily accurate, ensuring the results of the analysis are not significantly different from the results of a rigorous full-wave analysis, when the ratio between the waveguide diameter ($D = 2a$) and the wavelength in free space λ_0 is greater than 10 [13], [14]. The field radiated from the Brillouin zone, marked in Fig. 4 as B3, may be considered as a set of rays running parallel, in the axial direction, and diverging spherically in the transverse direction. These rays, propagating – for instance – towards the focal toroidal mirror, are shown in Fig. 5 on the so-called Vlasov converter used in gyrotrons to convert launcher radiation into a Gaussian beam [15].

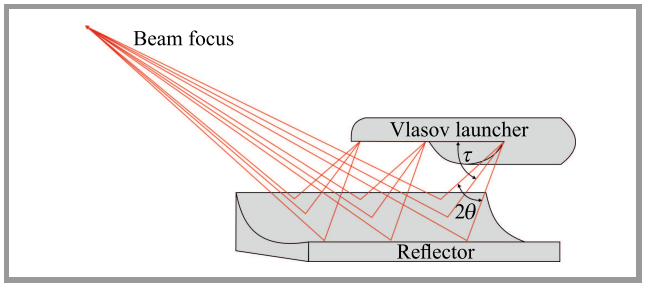


Fig. 5. Geometry of the Vlasov converter consisting of a circular waveguide launcher with helical cut and a double-curved reflector (toroid). τ is the helix inclination angle and 2θ is the radiation expansion angle transverse to the waveguide axis.

Angle 2θ marked in Fig. 5 is ν_{exp} and is a radiation expansion angle that is transverse to the waveguide axis. Expansion angle ν_{exp} is another important geometric parameter in RT transmission analysis. Mathematical dependencies describing the parameters introduced above have been formulated, inter alia, in [11]. They are based on the assumption that transmission in the launcher with the microwave structure considered here is expressed in the cylindrical coordinate system shown in Fig. 6.

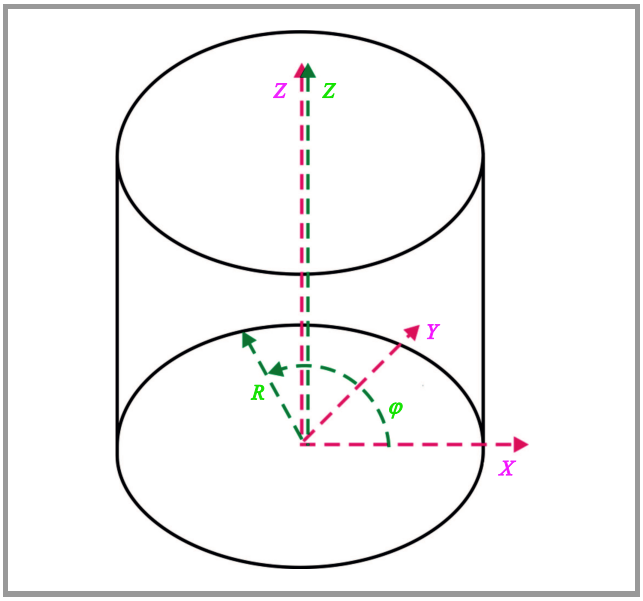


Fig. 6. The cylindrical coordinate system associated with the Vlasov launcher in such a way that cylinder axis is the axis of circular waveguide of launcher.

For the launcher and the TE_{mn} mode under consideration, the $u(r,\phi,z)$ field propagating in a cylindrical waveguide, which is the input section of the launcher, may be expressed, using the cylindrical coordinate system, in the following manner:

$$u(r,\phi,z) = A_0 J_m(k_r r)\mathrm{e}^{\pm \mathrm{j}m\phi}\mathrm{e}^{\pm \mathrm{j}k_z z}\ , \tag{1}$$

which:

$$k_r = \frac{\chi_{mn}}{a}\ , \tag{2}$$

$$J'(\chi_{mn}{}') = 0\ , \tag{3}$$

$$k_z = \sqrt{k_0{}^2 - k_r{}^2}\ , \tag{4}$$

where:

- A_0 – constant field amplitude,
- z and r – longitudinal and radial coordinate,
- ϕ – azimuth angle,
- J_m and J'_m m-th order Bessel function and its derivative,
- χ_{mn} and $\chi_{mn}{}'$ n-th zero of m-th order Bessel function and its derivative,
- k_r and k_z – radial and longitudinal wave number,
- k_0 – wave number in free space.

The geometrical parameters of the RT method for the considered launcher are described in the following analytical expressions:

$$\cos\psi = \frac{\vec{N}\vec{e}_z}{|\vec{N}||\vec{e}_z|} = \frac{k_z}{\sqrt{{k_r}^2 + {k_z}^2}} = \frac{k_z}{k_0}, \sin\psi = \frac{k_r}{k_0}, \quad (5)$$

where (for TE field):

$$k_r = \frac{\chi_{mn}{}'}{a} \ , \quad (6)$$

$$k_0 = \frac{2\pi}{\lambda_0} \ , \quad (7)$$

$$\cos\theta = \frac{\vec{N}\vec{e}_z}{|\vec{N}||\vec{e}_z|} = \frac{m}{\chi_{mn}{}'} \ , \quad (8)$$

$$R_c = a\cos\theta \ , \quad (9)$$

$$L_B = 2a\sin\theta \ \cot\psi \ , \quad (10)$$

$$\tau = \arctan\frac{\theta\tan\psi}{\sin\theta} \ , \quad (11)$$

$$L = 2\pi a\cot\tau \ , \quad (12)$$

$$\nu_{exp} = 2\theta \ . \quad (13)$$

Equations (6)–(13) allow one to design a launcher for the input data obtained from the gyrotron cavity design and from other gyrotron system requirements, such as:

- the volume of the vacuum space inside the gyrotron, that the launcher must fit in,
- trajectory of the spent electrons' motion with no launcher elements present,
- the so-called modal purity that needs to be sufficiently large, as it determines the share of Gaussian components in the wave beam propagated in the launcher.

In addition to those listed above, numerous additional requirements may be formulated as well, e.g. shock resistance, acceptable thermal expansion values, etc. A relatively easy assessment of the impact that the listed requirements exert on the geometrical parameters of the launcher is possible by transforming Eqs. (6)–(13) in such a way that the field mode parameters (n, m, $\chi_{mn}{}'$) appear in the equations and the Brillouin angle ψ is an independent variable. After applying relatively simple transformations, design Eqs. (6)–(13) take the following form:

$$d = \frac{D}{\lambda_0} = \frac{\frac{\chi_{mn}{}'}{\pi}}{\sin\psi} = \frac{\chi_{mn}{}'}{\pi\sin\psi} \ , \quad (14)$$

$$\cos\theta = \frac{m}{\chi_{mn}{}'} \ , \quad (15)$$

$$\sin\theta = \sqrt{1 - \left(\frac{m}{\chi_{mn}{}'}\right)^2} \ , \quad (16)$$

$$R_c = a\cdot\cos\theta = a\frac{m}{\chi_{mn}{}'} \ , \quad (17)$$

$$L_b = 2a\cdot\sin\theta\cot\psi = 2a\sqrt{1 - \left(\frac{m}{\chi_{mn}{}'}\right)^2}\cot\psi \ , \quad (18)$$

$$\tau = \arctan\frac{\tan\psi}{\frac{\sin\theta}{\theta}} \ , \quad (19)$$

$$L = 2\pi a\cdot\cot\tau = 2\pi a\cdot\frac{\sin\theta}{\theta}\cot\psi \ . \quad (20)$$

From Eqs. (18)–(20), one can eliminate waveguide radius a by entering normalized quantities of caustic radius r_c, axial shift L_b, and length of the launcher L:

$$r_c = \frac{R_c}{a} = \frac{m}{\chi_{mn}{}'} \ , \quad (21)$$

$$l_b = \frac{L_b}{D} = \sqrt{1 - \left(\frac{m}{\chi_{mn}{}'}\right)^2}\cot\psi \ , \quad (22)$$

$$l = \frac{L}{\pi D} = \frac{\sin\theta}{\theta}\cot\psi = \frac{\sqrt{1-\left(\frac{m}{\chi_{mn}{}'}\right)^2}}{\arccos\frac{m}{\chi_{mn}{}'}}\cot\psi \ . \quad (23)$$

In Eq. (23), length L of the launcher was normalized to circumference πD of the waveguide constituting the input section of the launcher. The analytical form of Eq. (23) can be simplified by entering the $\operatorname{sinc}\theta$ function:

$$\operatorname{sinc}\theta = \frac{\sin\theta}{\theta} \ . \quad (24)$$

The $\operatorname{sinc}\theta$ function, for relatively small values of the angle θ characterizing the high-order TE_{mn} fields, takes values slightly lower than one, as shown in Eq. (23), so it affects the launcher length to a relatively small extent. Therefore, it may be treated as a correction factor for the launcher length normalized to the πD circumference. The correcting factor of the launcher is expressed as:

$$l_k = \frac{l}{\operatorname{sinc}\theta} = \cot\psi \ . \quad (25)$$

Equations (14), (21), (22), and (25) allow for tracking the impact that the Brillouin angle ψ and field type (m, $\chi_{mn}{}'$) exert on normalized geometrical parameters of the launcher, such as waveguide diameter d, caustic radius r_c, axial shift l_b, and launcher length l_k.
In Appendices A and B, direct dependencies between the launcher length and the waveguide circumference are derived.

3. Launcher Design Supporting Graphs

Based on the analytical formulas provided in Section 2, specific the graphs and curves have been drawn to facilitate the design of the launcher.

3.1. Diagrams Presenting Dependence of Waveguide Diameter and Launcher Length on the Brillouin Angle

In this paper, graphs of the functions showing the relationship between $d(\psi)$ and $l_k(\psi)$ have been prepared. These graphs, just like analytical formulas, allow to trace the impact of the Brillouin angle (ψ) and field type ($m, \chi_{mn}{}'$) on the launcher's geometrical parameters d and l_k. This graphical presentation is, however, much more convenient to follow than analysis of Eqs. (14) and (25). The manner in which diagrams $d(\psi)$ and $l_k(\psi)$ are used in the launcher design process is presented below. Figure 7 shows

the parametric graph of function $d(\psi)$ for many realistically selected parameter values $p = \frac{\chi_{mn}'}{\pi}$. This graph clearly visualizes the impact of the field mode and of the Brillouin angle value on the normalized waveguide diameter d.

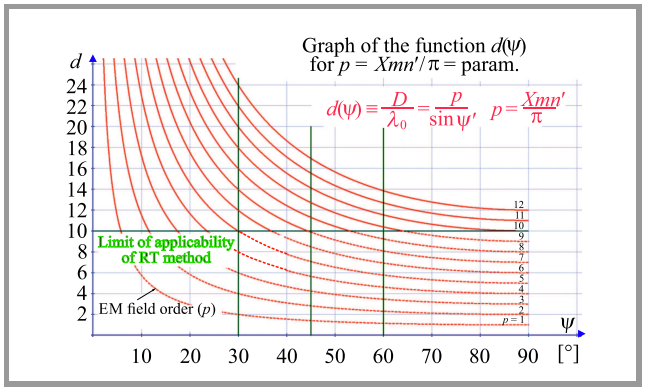


Fig. 7. Relationship between normalized waveguide diameter d and the Brillouin angle ψ.

The graph contains the following information:

- analytical formula describing function $d(\psi)$, $d(\psi) \equiv \frac{D}{\lambda_0} = \frac{p}{\sin\psi}$, where $p = \frac{\chi_{mn}'}{\pi} = parameter$,
- arbitrary limit of applicability of the RT method to the description of ray propagation in the launcher ($d_{min} = 10$). For, $d > 10$, errors in the analytical description of wave propagation taking place in the launcher have a satisfactorily low level of 1%. The error rate decreases with an increase in d [11],
- limits of the middle range of the Brillouin angle variation $\Delta(\psi)$ set arbitrarily at: $\psi_{min} = 30^\circ$ and $\psi_{max} = 60^\circ$. The Brillouin angle values adopted in projects are usually within that range,
- straight line marking the middle of the range $\Delta(\psi)$: $\psi_c = 45^\circ$,
- values of parameter p for all curves of function $d(\psi)$,
- trend indicator determining the direction of the shift in characteristic $d(\psi)$ with the increase of in parameter p (with the increase of the field mode order),
- parts of the $d(\psi)$ characteristics below the marked limit of applicability of the RT method (dashed line) cannot be used in this range because errors in the analytical description of wave propagation in the launcher are too large.

Variable ψ is independent and its range covers all potential values of angle $\psi \in [0, 90^\circ]$, while the range of the dependent variable d is arbitrarily set to $d \in [0, 26]$. The range of variation of parameter p ($p \in [1, 12]$) was adopted arbitrarily, based on a review of the value of the Bessel function derivative zeroes χ_{mn}' for various field modes [16].
The $l_k(\psi)$ function graph, shown in Fig. 8, presents the impact of the ψ angle value on the launcher length l_k.

Figure 8 contains the following information:

- analytical formula describing the $l_k(\psi)$ function $l_k(\psi) \equiv \frac{L}{\pi D \operatorname{sinc}\theta} = \cot\psi$,
- limits of the middle range of the Brillouin angle variation $\Delta(\psi)$, set arbitrarily at: $\psi_{min} = 30^\circ$ and $\psi_{max} = 60^\circ$,
- straight line marking the middle of the range $\Delta(\psi)$: $\psi_c = 45^\circ$,
- ranges of Brillouin angle ψ in which launcher shortening or elongation occurs in relation to its length at $\psi = 45^\circ$.

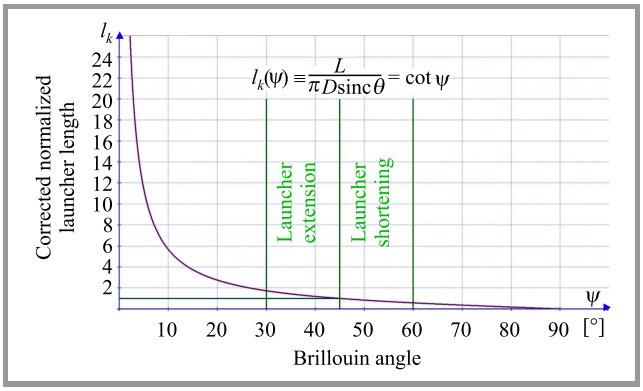


Fig. 8. Relationship between normalized launcher length l_k and the Brillouin angle ψ.

Similarly to Fig. 7, independent variable ψ covers all potential values of angle $\psi \in [0, 90^\circ]$, while the range of the dependent variable l_k is arbitrarily set to $l_k \in [0, 26]$. There is a relatively small difference between d and l_k values for a given angle ψ. This makes it possible to apply the same ordinate axis for both of these quantities when plotting functions $d(\psi)$ and $l_k(\psi)$ in one common coordinate system (Fig. 9). Additionally, this method of presentation facilitates observation of the impact that field mode and the Brillouin angle ψ have on the launcher's geometrical parameters d and l_k.

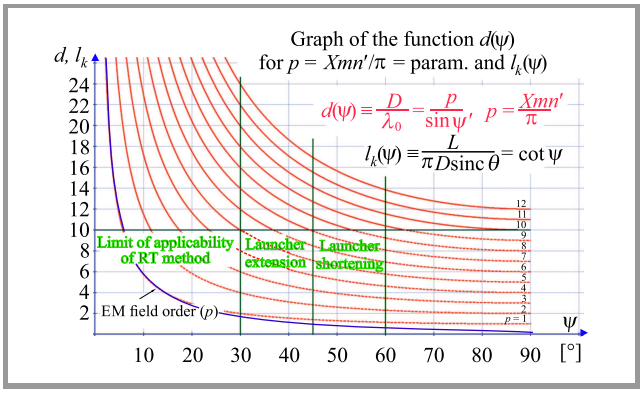


Fig. 9. Relationship between normalized launcher length l_k, normalized and corrected waveguide diameter d, and the value of Brillouin angle ψ.

3.2. Relationship Between sinc (θ) Function and Field Mode Parameters

Equation (24) was transformed into Eq. (26) using Eqs. (15) and (16), and was then used to create the chart shown in Fig. 10.

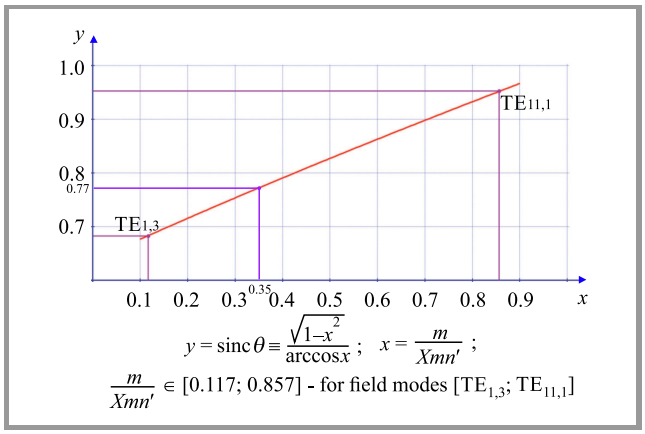


Fig. 10. Relationship between function $y = \operatorname{sinc}\theta$ and the field mode, represented by $x = \frac{m}{\chi_{mn}'}$.

$$y = \operatorname{sinc}\theta = \frac{\sin\theta}{\theta} = \frac{\sqrt{1-\left(\frac{m}{\chi_{mn}'}\right)^2}}{\arccos\frac{m}{\chi_{mn}'}}. \tag{26}$$

The graph shown in Fig. 10 is a supplement to the graphs shown in Fig. 9. It shows the values of function $\operatorname{sinc}\theta$, referred to as field parameters, allowing the designer to quickly evaluate the function value for the adopted field mode. Knowledge of the function value allows to correct launcher length l_k, previously determined graphically in Fig. 9. The length corrected in accordance with Eq. (25) is equal to $l = l_k \cdot \operatorname{sinc}\theta$.

The range of variable $x = \frac{m}{\chi_{mn}'}$ variation was determined after calculating its values for parameters m and χ_{mn}' given in the paper [16]. The results of these calculations are presented in Table 1.

The review of values m/χ_{mn}' from Table 1 was relied upon to adopt the range of variable $x \in [0.1, 0.9]$. The range of the dependent variable y was adopted as $y \in [0.6, 1]$, based on calculations of values for a given x. The two extreme points have been marked on the $\operatorname{sinc}\theta$ graph. They were described by field mode symbols corresponding to these values.

4. Graph-assisted Launcher Design Method

4.1. Design Strategies

As mentioned earlier, the operating frequency, the TE field mode, and the geometrical parameters of a gyrotron cavity are usually basic input data for the launcher design process. With waveguide radius a known, and with the knowledge of λ_0 and m, n, indexes of TE_{mn} mode are sufficient to calculate all other parameters. Unfortunately, as one can see in Fig. 1, radius a differs from the gyrotron cavity radius a_c, $(a > a_c)$. This requires that the cavity and the launcher be connected by a tapering transition section known as a "taper". As a result, there is some freedom in the selection of the waveguide radius in the launcher design. This allows for the creation of various design strategies that depend on additional requirements applicable to the launcher. One of the additional requirements consists in determining the proper ratio between the waveguide diameter and the wavelength, ensuring sufficiently high mode purity of the output Gaussian beam. The next requirement may consist, for instance, in determining the right length of the launcher selected, so that it does not appear on the trajectory of electrons moving to the collector (the so-called spent electrons).

4.2. Launcher Design for Given Field Mode Indexes

Initial input data:

- field mode: TE_{95},
- frequency, (wavelength): f_0, (λ_0).

Design steps:

Step 1. Selection of the $d(\psi)$ characteristics with parameter p closest to the setpoint.

For the assumed field mode TE_{95} as $\chi_{mn}' = \chi_{95}' = 25.8913$, parameter p is calculated as $p = \frac{\chi_{mn}'}{\pi} = 8.24145676$. Within the range of characteristics $d(\psi)$ (waveguide diameter), we look for characteristics in which parameter p is closest to the calculated value of 8.24145676. This characteristic is

Table 1
Values of variable $x = \frac{m}{\chi_{mn}'}$ for selected field modes TE_{mn}

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	$m=11$
$n=1$	3.832	1.841	3.054	4.201	5.317	6.415	7.501	8.578	9.647	10.71	11.77	12.83
m/χ_{mn}'	0	0.543	0.655	0.714	0.752	0.779	0.800	0.816	0.829	0.840	0.850	0.857
$n=2$	7.016	5.330	6.706	8.015	9.282	10.52	11.73	12.93	14.11	15.29	16.45	17.60
m/χ_{mn}'	0	0.188	0.298	0.374	0.431	0.475	0.511	0.541	0.567	0.588	0.608	0.625
$n=3$	10.17	8.536	9.969	11.35	12.68	13.99	15.27	16.53	17.77	19.00	20.22	21.43
m/χ_{mn}'	0	0.117	0.200	0.264	0.315	0.357	0.393	0.423	0.450	0.474	0.494	0.513

marked in the graph (Fig. 11) in magenta, with $p = 8$. At this design stage, it is possible to move the operating point P freely along the selected characteristic within the permitted range above the limit of applicability of the RT method. One may see that when point P is moved upwards, waveguide diameter ($d \equiv D/\lambda_0$) increases.

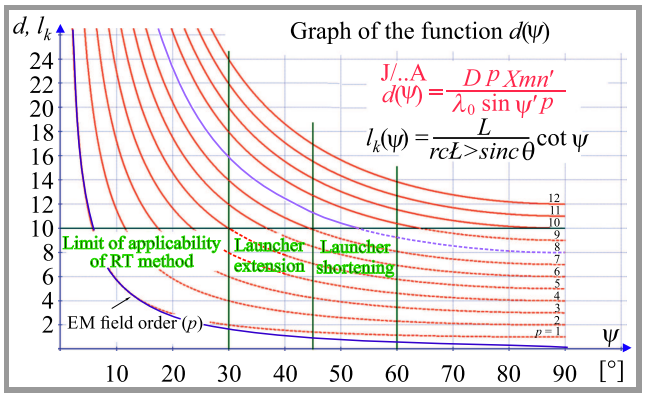


Fig. 11. Highlighted area shows the $d(\psi)$ characteristic with the p parameter closest to the calculated value.

Step 2. Determining the operating point on the selected $d(\psi)$ curve.

As mentioned previously, the additional requirement consists in the need to determine the proper ratio between waveguide diameter and wavelength D/λ_0 for high Gaussian mode purity of the output beam. In this example, the D/λ_0 value is determined by the selection of the Brillouin angle value $\psi = 45°$, as shown in Fig. 12. The value read is $D/\lambda_0 = 11.33$. If this value is greater than the D/λ_0 of the gyrotron cavity, it will be necessary to apply the taper transition between the cavity and the launcher.

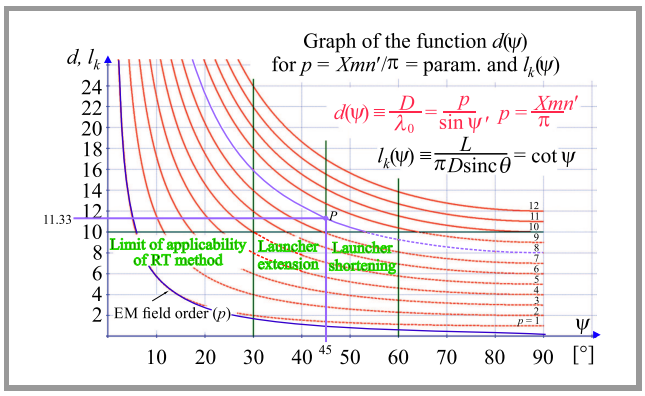


Fig. 12. Illustration of operating point P selection based on characteristic $d(\psi)$.

Step 3. Determining the operating point on $l_k(\psi)$ curve.

This curve illustrates the relationship between normalized launcher length l_k and the Brillouin angle ψ. The operating point, designated as $P1$, is positioned at the intersection between the vertical line $\psi = 45°$ and the characteristic $l_k(\psi)$, as shown in Fig. 13.

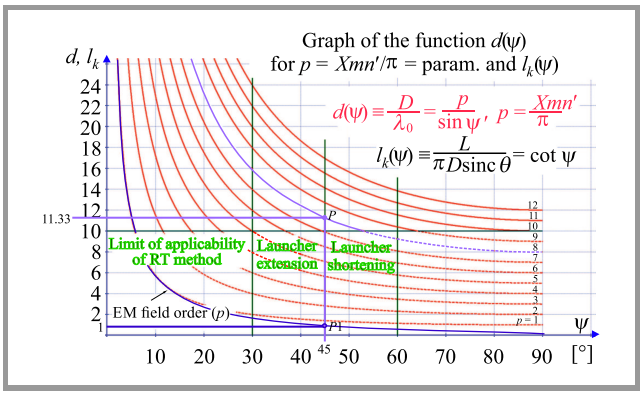


Fig. 13. Relationship between normalized launcher length l_k, normalized and corrected waveguide diameter d, and the value of Brillouin angle ψ.

Based on Eq. (25), the normalized value of corrected launcher length $l_k(\psi)$ is equal to 1 (the ordinate of point $P1$ for $\psi = 45°$). Therefore, in order to determine normalized length l of the launcher equal to Eq. (23), the ratio $L/(\pi D)$ and the value of the $\operatorname{sinc}\theta$ function should be calculated for the assumed field mode.

Step 4. Calculation and graphic visualization of the $\operatorname{sinc}\theta$ function value.

According to Eq. (8) and field parameters given in step 1:

$$\theta = \arccos\frac{m}{\chi_{mn'}} = \arccos\frac{9}{25.8913} = 1.21\ [rad]\ ,$$

therefore: $\operatorname{sinc}\theta = 0.771$.

To show the value of the sinc θ function, the plot of the function's value depending on the field mode, as presented in Fig. 10, will be used. In this plot, the value of the function calculated above is the ordinate of point P, as shown in Fig. 14.

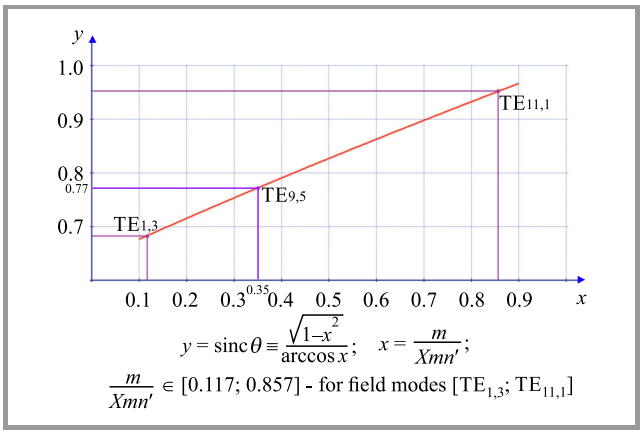


Fig. 14. Visualization of the value of the $y = \operatorname{sinc}\theta$ function for the selected field mode, represented by $x = \frac{m}{\chi_{mn'}}$.

Step 5. Calculation of the ratio between launcher length L and waveguide circumference πD.

This ratio is the standard length of the launcher $l = \frac{L}{\pi D}$, as introduced in Eq. (23). Based on Eq. (25) $l_k = \frac{l}{\operatorname{sinc}\theta}$,

hence, $l = l_k \ \mathrm{sinc}\,\theta$. The normalized, corrected launcher length l_k was determined in step 3 and is equal to 1, while $\mathrm{sinc}\,\theta$, determined in step 4, is 0.771. So, $l = l_k \mathrm{sinc}\,\theta = 0.771$.

Step 6. Calculation of waveguide diameter D for the assumed wavelength λ_0.

The value of $\frac{D}{\lambda_0}$ was determined in step 2 as the ordinate of point P. Hence, $D = y_p\lambda_0 = 11.33\lambda_0$. For example:

- for $\lambda_0 = 3$ cm ($f_0 = 10$ GHz):
  $D = 11.33 \cdot 3 = 33.99$ cm,
- for $\lambda_0 = 1.875$ cm ($f_0 = 16$ GHz):
  $D = 11.33 \cdot 1.875 = 21.24$ cm.

Step 7. Calculation of launcher length L for the assumed wavelength λ_0.

The normalized launcher length $l = \frac{L}{\pi D}$ has been calculated in step 5. Hence, $L = \frac{\pi}{D}$, where, based on step 6: $D = y_P \cdot \lambda_0$. So $L = \pi l D = \pi l y_P \cdot \lambda_0$. After substitution into this equation where according to step 5, $l = 0.771$, and step 2 and 6, $y_P = 11.33$, $L = \pi \cdot 0.771 \cdot 11.33 \cdot \lambda_0 = 27.45 \cdot \lambda_0$ is obtained. For example:

- for $\lambda_0 = 1.875$ cm, $f_0 = 16$ GHz:
  L$= 27.45 \cdot 1.875 = 51.47$ cm.

The results obtained in steps 6–7 conclude the main part of the design procedure and make it possible, as shown above, to conveniently and simply calculate the main parameters of the launcher: its length and waveguide diameter for a given operating frequency. These results can, therefore, be used to design launcher models operating at other frequencies and in other field modes.

Other launcher parameters, such as: R_c, L_B, τ, and ν_{exp}, can be calculated using analytical formulas (9)–(13).

4.3. Launcher Design Method for Assumed Field TE Mode and Standard Circular Waveguide Diameter

The task presented here can be treated as an extension of the previous task and aims to use a standard circular waveguide in the launcher in order to reduce the costs of implementing the launcher model operating at a different (usually lower) frequency. The following initial input data is used:

- field mode: TE_{95},
- frequency: $f_0 \cong 16$ GHz,
- waveguide diameter: $D \cong 21.15$ cm, (determined in step 6 of the previous example),
- normalized waveguide diameter:
  $d_p = \frac{D}{\lambda_0} = 11.33$ cm, (determined in step 2).

Design steps:

Step 1. Selection of the standard circular waveguide.

From the table of standard circular waveguide sizes, a C10 waveguide with a diameter of $D = 21.514$ cm and a nominal working band of 0.039–1.29 GHz has been selected [17]. Its diameter is as close as possible to value D assumed in the initial input data.

Step 2. Calculation of wavelength λ_0 and operating frequency f_0.

The normalized waveguide diameter assumed in initial input data is equal to $\frac{D}{\lambda_0} = 11.33$. Hence: $\lambda_0 = \frac{D}{11.33} = 1.90$ cm and $f_0 = 15.80$ GHz.

Step 3. Calculation of launcher length.

Similarly to step 7 of the previous example, $L = 27.45 \cdot \lambda_0 = 52.13$ cm.

Step 4. Calculation of caustic radius R_c.

The calculation will be performed for a standard waveguide with a diameter of $D = 21.514$ cm selected in step 1 and the value $\frac{m}{\chi_{mn'}} = 0.3476$ calculated in step 4 of the previous task.

The calculation is performed for a standard circular waveguide selected in step 1 with a diameter of $D = 21.514$ cm and by taking the values calculated in step 4: $\frac{m}{\chi_{mn'}} = 0.3476$.

$R_c = a \cdot \cos\theta = \frac{D}{2} \cdot \frac{m}{\chi_{mn'}} = \frac{21.514}{2} \cdot 0.3476 = 3.74$ cm.

Step 5. Calculation of distance L_B (traveled by the ray in the axial direction between two successive reflections from the waveguide wall).

The calculation will be performed for a standard circular waveguide selected in step 1 with the diameter of $D = 21.514$ cm. We then take the values calculated in step 4, $\frac{m}{\chi_{mn'}} = 0.3476$, for which $\theta = \arccos\frac{m}{\chi_{mn'}} = \arccos\frac{9}{25.8913} = 1.21$ [rad]. Hence, $\sin\theta = \sin 1.21 = 0.94$, and from Eq. (19), $L_b = D\sin\theta\cot\psi = 21.514 \cdot 0.94 \cdot \cot 45^\circ = 20.17$ cm.

Step 6. Calculation of helix inclination angle τ.

For $\psi = 45^\circ$ and $\mathrm{sinc}\,\theta = 0.771$:

$\tau = \arctan\frac{\tan\psi}{\mathrm{sinc}\,\theta} = 0.91$ [rad]$= 52.35^\circ$.

As expected, the helix inclination angle τ is slightly larger than the Brillouin angle ψ.

5. Conclusions

This paper presents an effective method for designing a Vlasov launcher that is a part of the gyrotron's microwave power transmission system. The method uses a graphical representation of known analytical relationships describing RT power transmission in such a system. An image of the design space makes it much easier for an engineer to take optimal decisions in the design process. In particular, for a given EM field mode $(m, \chi_{mn'})$, the engineer can efficiently, easily and quickly determine the optimal launcher duty points: point P (Fig. 13), Brillouin angle ψ, diameter d of the launcher, and point P1 specifying length l_k of the launcher. They can also assess the impact of the operating points' displacement on the launcher's geometrical parameters and determine the allowed displacement ranges. The method was presented based on the example of a Vlasov launcher with a helical cut, but its assumptions may be used for all other launcher geometries. At the final stage of the design process, the described method allows to calculate the launcher geometry taking into account the wavelength.

This is an additional advantage of the method concerned, resulting from the normalization of dimensions d and l_k depending on the wavelength.

The relatively large degree of design freedom offered by this method may greatly facilitate the design of the launcher, with additional requirements, such as those related to the limited installation space inside the gyrotron taken into consideration. Examples with detailed steps are presented, describing the design procedure relying on the method concerned. The data used in these examples, i.e. field mode (TE_{95}) and wavelength ($\lambda_0 = 3$ and 1.875 cm), were adopted arbitrarily. It was also shown how to use the presented graphs in order to easily implement a standard-size circular waveguide into the launcher design. The geometrical parameters of the launcher determined in the examples for the assumed values of λ_0 can be immediately recalculated for other values of λ_0, e.g. for typical terahertz wavelengths at which gyrotrons usually operate.

The correctness of the design procedures developed in this paper was verified by comparing the results of the Vlasov launcher design procedures described in previously published works with the results obtained, for the same input data, after implementing the design procedures this paper is concerned with. Such comparisons were made for projects presented in [7] and [15], containing also experimental verification of the radiators developed and confirming the usefulness of the design procedures devised.

Appendix A

Direct relationship between launcher length and waveguide circumference

Equations (27) and (28) describe the relationships between waveguide diameter and launcher length on the one hand, and the Brillouin angle on the other. It is possible to displace variable ψ from the system of these equations. In this way, we obtain a direct relationship between these two quantities without mediating the Brillouin angle ψ.

$$d = \frac{D}{\lambda_0} = \frac{\frac{\chi_{mn}'}{\pi}}{\sin\psi} = \frac{\chi_{mn}'}{\pi \sin\psi} , \tag{27}$$

$$l = \frac{L}{\pi D} = \frac{\sin\theta}{\theta}\cot\psi = \operatorname{sinc}\theta \cot\psi , \tag{28}$$

where:

$$\frac{d^2}{p^2} - \frac{l^2}{(\operatorname{sinc}\theta)^2} = 1 , \tag{29}$$

or

$$l = \operatorname{sinc}\theta \cdot \sqrt{\left(\frac{d}{p}\right)^2 - 1} . \tag{30}$$

Equation (30), after renormalization of d and L, takes the form of:

$$L = \operatorname{sinc}\theta \cdot \sqrt{\left(\frac{\frac{D}{\lambda_0}}{p}\right)^2 - 1} \cdot \pi D . \tag{31}$$

This relationship shows that for $L > 0$, the condition $D/\lambda_0 > p$ must be met. The smallest zero value of the derivative of the Bessel function is $\chi_{mn}' = 1.8412$ [16]. Thus, it follows that the lowest value of parameter $p = \frac{\chi_{mn}'}{\pi}$ is equal to $p_{min} = 0.58607216$ and, consequently, that the smallest value of D/λ_0 should be greater than 0.58607216. This is a milder condition than that which determines the use of the RT method (requiring that $D/\lambda_0 > 10$). This condition is always met when designing the launcher using the RT method.

Appendix B

Direct relationship between launcher length and waveguide diameter for $\psi = 45^\circ$

From the Eq. (27) for $\psi = 45^\circ$ we get:

$$\frac{d_{45}}{p} = \frac{2}{\sqrt{2}} , \tag{32}$$

but from Eq. (30):

$$l_{45} = \operatorname{sinc}\theta \cdot \sqrt{\left(\frac{d_{45}}{p}\right)^2 - 1} = \operatorname{sinc}\theta , \tag{33}$$

which can also be seen from the Eq. (28):

$$l = \operatorname{sinc}\theta \cot\psi .$$

Because, $\cot 45^\circ = 1$, so $l_{45} = \operatorname{sinc}\theta$. Therefore,

$$d_{45} = \frac{D_{45}}{\lambda_0} = \frac{\sqrt{2}}{\pi} \cdot \chi_{mn}' , \tag{34}$$

and

$$l_{45} = \frac{L_{45}}{\pi D_{45}} = \operatorname{sinc}\theta . \tag{35}$$

Hence, after renormalization:

$$D_{45} = \frac{\sqrt{2}}{\pi} \cdot \chi_{mn}' \cdot \lambda_0 , \tag{36}$$

and

$$L_{45} = \sqrt{2} \cdot \chi_{mn}' \cdot \operatorname{sinc}\theta \cdot \lambda_0 . \tag{37}$$

References

[1] C.-H. Du, X.-B. Qi, and P.-K. Liu, "Theoretical study of a broadband quasi-optical mode converter for pulse gyrotron devices", *IEEE Transac. on Plasma Sci.*, vol. 44, no. 10, pp. 2348–2355, 2016 [Online]. Available: http://ieeexplore.ieee.org/document/7572128/ (DOI: 10.1109/TPS.2016.2606497).

[2] Z. Li and J. Feng, "Design of a Vlasov mode converter of 263 GHz gyrotron oscillator for DNP-NMR", *The J. of Engineer.*, vol. 2018, no. 14, pp. 709–713, 2018 [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/joe.2018.0124

[3] M. K. Alaria, N. Singh, U. Singh, A. Bera, and A. K. Sinha, "Development of 170 GHz, 0.1 MW short pulse gyrotron", in *Proc. Fusion Engineer. and Design*, vol. 144, 2019, pp. 87–92 [Online]. Available: https://doi.org/10.1016/j.fusengdes.2019.04.073

[4] M. Pilossof and M. Einat, "High-average-power second harmonic w-band gyrotron with room-temperature solenoid", *IEEE Transac. on Electron Devices*, vol. 67, no. 4, pp. 1804–1807, 2020 (DOI: 10.1109/TED.2020.2971653).

[5] B. Hui-Qi *et al.*, "Analysis of a broadband quasi-optical mode converter for gyrotrons working in multi modes", in *Proc. IEEE Asia Pacific Microwave Conf. (APMC)*, Kuala Lumpur, Malaysia, 2017 (DOI: 10.1109/APMC.2017.8251547).

[6] C. Gao *et al.*, "Frequency-chirped dynamic nuclear polarization with magic angle spinning using a frequency-agile gyrotron", *J. of Magnetic Resonance*, vol. 308, 2019 (DOI: 10.1016/j.jmr.2019.106586).

[7] M. K. Alaria, A. K. Sinha, and H. Khatun, "Design and development of mode launcher for high frequency gyrotron", *Infrared Physics & Technol.*, vol. 75, pp. 187–192, 2016 (DOI:10.1016/j.infrared.2015.12.011).

[8] C. Zhang, W. Fu, and Y. Yan, "Study on a gyrotron quasi-optical mode converter for terahertz imaging", *J. of Electromagnetic Waves and Applications*, vol. 35, no. 2, pp. 176–184, 2020 (DOI: 10.1080/09205071.2020.1828186).

[9] X. Li, "Study of High-harmonic Gyro-devices in the THz Range", Ph.D. Thesis, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK, 2016 [Online]. Available: https://qmro.qmul.ac.uk/xmlui/bitstream/handle/123456789/23214/LI_Xiang_FinalPhD_171016.pdf

[10] T. Kariya *et al.*, "Development of over-MW gyrotrons for fusion at 14 GHz to sub-THz frequencies", *Nuclear Fusion*, vol. 57, no. 6, 2017 [Online]. Available: https://iopscience.iop.org/article/10.1088/1741-4326/aa6875

[11] J. Jin, "Quasi-Optical Mode Converter for a Coaxial Cavity Gyrotron", *Forschungszentrum Karlsruhe*, Karlsruhe, no. 7264, 2007 [Online]. Available: https://publikationen.bibliothek.kit.edu/270067871/3814966 [in German].

[12] S. N. Vlasov and I. M. Orlova, "Quasioptical transformer which transforms the waves in a waveguide having a circular cross section into a highly directional wave beam", *Radiophysics and Quantum Electronics*, vol. 17, no. 1, pp. 115–119, 1974 [Online]. Available: http://link.springer.com/10.1007/BF01037072

[13] L. A. Vainshtein, *Open Resonators and Open Waveguides*. Boulder, Colorado: Golem Press, 1969.

[14] V. M. Babic and V. S. Buldyrev, *Short-Wavelength Diffraction Theory*. Moscow: Springer-Verlag Berlin Heidelberg, 1972 (ISBN: 9783642834615).

[15] M. Blank, "High Efficiency Quasi-Optical Mode Converters for Overmoded Gyrotrons", Ph.D. Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Massachusetts, 1994 [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/34089/32053304-MIT.pdf

[16] C. A. Balanis, *Advanced Engineering Electromagnetics, 2nd Edition*. Wiley and Sons, 2012 (ISBN: 9780470589489).

[17] H. Nickel, "Cross reference for hollow metallic waveguides", *Spinner*, 2020 [Online]. Available: https://www.spinner-group.com/images/download/technical_documents/SPINNER_TD00036.pdf

Grzegorz Jaworski received his Ph.D. degree from Wrocław University of Science and Technology in 1999. Between 2003 and 2004, he was with the Ørsted-DTU/EMI Department of Denmark Technical University, working on next generation SAR systems. Between 2006 and 2007, he participated in developing antennas for the Columbus module of the International Space Station. Currently, he is an Assistant Professor at the Electronics and Telecommunication Department, Faculty of Electronics, at Wrocław University of Science and Technology. Currently, his interests focus on high frequency techniques and technologies for telecommunications, radars, industry and medical applications.

https://orcid.org/0000-0002-9172-7437
E-mail: grzegorz.jaworski@pwr.edu.pl
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland

Andrzej R. Francik received his M.Sc. Eng. in Electronics from Wrocław University of Technology, Poland, in 1969. He started working as a university lecturer in 1970, and then as a research worker employed by the Institute of Telecommunications and Acoustics. His publications focus primarily on theoretical and practical problems related to the microwave module of ESR spectrometers. In 1978, Francik earned his Ph.D.E.E. for analyzing systematic distortions originating from the microwave unit. His research interests include power combiners, combgenerators, mixers, detectors, MICs, microwave instruments and, recently, gyrotron technologies. In 1992, he received a D.Sc. degree for publishing a monography titled "Instrumental Effects in Homodyne Electron Paramagnetic Resonance Spectrometers" (Ellis Horwood & PWN, Chichester, Warszawa, 1989). In 2001, he was employed as an Associate Professor at his mother Institute. Currently, he is working, as a Professor Emeritus specializing in the terahertz technology.

https://orcid.org/0000-0001-6414-7106
E-mail: andrzej.francik@pwr.edu.pl
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland

Maciej Nowak received his Ph.D. from Wrocław University of Science and Technology in 2016. He is currently an Assistant Professor at the Telecommunications and Teleinformatics Department, Faculty of Electronics, Wrocław University of Science and Technology. He is a member of the Wrocław Terahertz Center. His research interests include: terahertz spectroscopy, spectral imaging techniques and machine learning.

https://orcid.org/0000-0002-7747-4867
E-mail: maciej.nowak@pwr.edu.pl
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland

Kacper Nowak received his Ph.D. from Wrocław University of Science and Technology in 2012. He is currently an Assistant Professor at the Electronics and Telecommunication Department, Faculty of Electronics, Wrocław University of Science and Technology. His research interests include: terahertz spectroscopy, gyrotron technology, industrial automation, networking and programming.

https://orcid.org/0000-0002-5980-8237
E-mail: kacper.nowak@pwr.edu.pl
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland

Paper

Analysis of the Discrete-time Multi-queue System with a Cycle-based Scheduler

Wojciech Burakowski and Maciej Sosnowski

National Institute of Telecommunications, Warsaw, Poland

https://doi.org/10.26636/jtit.2021.152121

Abstract—This paper presents an analysis of a discrete-time multi-queue system handling a number of packet streams. The analysis focuses on calculating system state distribution and packet sojourn time distribution. The method relied upon for determining system state distribution is based on creating a number of equations that are solved numerically. Next, based on the distribution calculated in such a manner, we derive relations for packet sojourn time distribution. The models studied may be useful for instance in a system supporting a number of virtual links (each of a constant bitrate) that share a common physical link. Isolation of performance of those virtual links needs to be assured. Finally, we present some exemplary numerical results showing the usefulness of the proposed analysis for supporting the system dimensioning process.

Keywords—discrete-time queueing system with vacations, system state distribution, packet sojourn time distribution, virtualized system.

1. Introduction

The paper presents an analysis of a FIFO-type discrete-time queueing system handling a number of packet streams in which service access of specific streams is governed by a cycle that has been assumed *a priori*. The cycle is repeated periodically and consists of a number of time slots, dedicated to handling packets assigned to a predefined stream. Therefore, from the point of view of a given stream, it is the system with vacations in which the so-called active state and vacation periods may be distinguished.

Packets belonging to a given stream may only be serviced in the active periods (time spent serving packets from this stream), while in the vacation periods, these packets cannot be serviced. The important feature here is that such an approach guarantees performance isolation between the packet streams handled. It means that performance-related parameters, i.e. delays and loss of packets belonging to a given stream, are not disturbed by servicing other streams.

The idea behind a system with vacations fed by one packet stream is illustrated in Fig. 1. The system may operate in one of two potential states, i.e.: active period A, when packet service is available, and vacation period V, when packet service is not available.

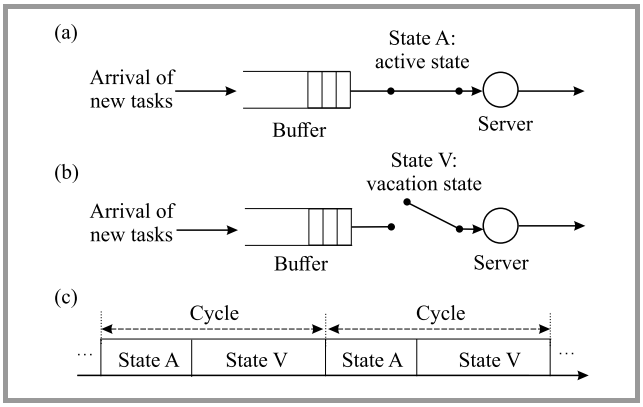


Fig. 1. Block diagram of a system with vacations fed by one packet stream: (a) in state A (active period), (b) in state V (vacation period), (c) states A and V alternate accordingly to a cycle assumed on an *a priori* basis.

The system under analysis is a model of a solution supporting a number of virtual links that share one physical link. Hence, performance isolation between packet streams within the specific virtual links is required.

Figure 2 shows an example of two virtual links established between virtualized servers, each with two virtual machines (VM). One virtual link is dedicated to transferring packets from VM1 to VM3, while the other is dedicated to transferring packets from VM2 to VM4. Such an approach to a virtualized system was successfully implemented and tested, for instance, in the IIP System [1] that was designed for creating a number of parallel Internets (with different protocol stacks) sharing the same physical resources, i.e. physical links and virtualization-enabling devices.

The system analyzed in this paper is of the FIFO discrete-time type, with a constant time of servicing packets from a given stream. These packets are serviced only in specific time slots within the active periods, as new packets from the considered stream may arrive into the system at the beginning of each slot only. For the vacation periods, we assume that they consist of a number of time slots (named

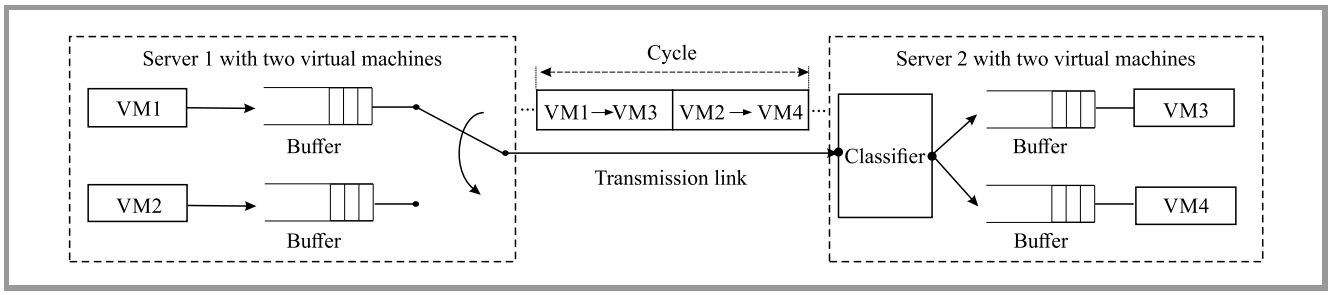


Fig. 2. Scenario with two virtual links established between two virtualized machines: virtual link 1 for data transfer between VM1 and VM3, and virtual link 2 for data transfer between VM2 and VM4.

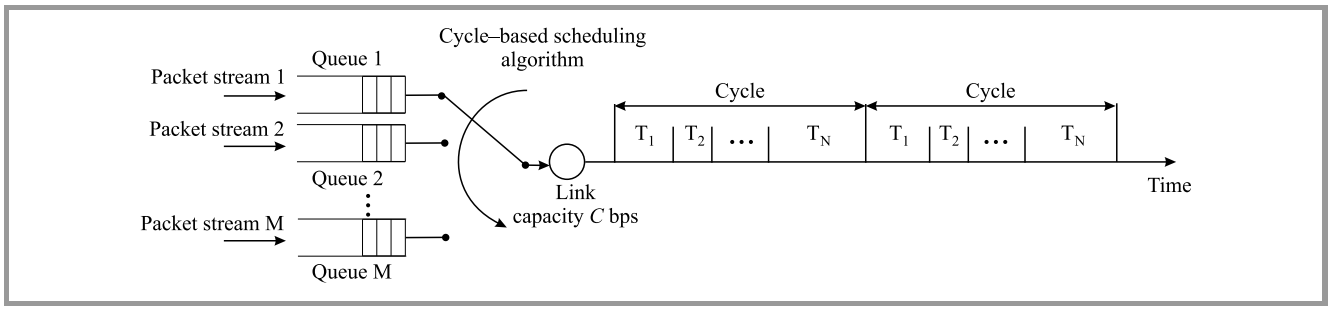


Fig. 3. The proposed discrete-time multi-queue system fed by M independent packet streams.

dummy slots) of the same length as the active periods. Again, new packets arriving during the vacation periods may arrive only at the beginning of dummy slots as well. For such a system with an infinite buffer size, we propose methods allowing to calculate state distribution by numerically solving a number of equations. Then, calculation of sojourn time distribution is proposed as well, using analytical formulas being an extension of the method described in [2], concerning a system without vacations.

Systems with vacations have been studied by many authors making different assumptions. A decent survey of these methods may be found e.g. in [3]–[8]. Unexpectedly, according to our best knowledge, no analysis is available of the system that is discussed in this paper.

Let us mention the papers that directly correspond to the research problem discussed in this paper. In [9], the authors consider a system with vacations but with continuous time and present formulas for mean waiting times. In [10], the authors extend the analysis for the system with general service times of packets and derive relations for system state distribution and packet loss ratio (with a finite buffer).

An approximated method for calculating mean waiting times in the considered system fed by a Poissonian stream was proposed in [11]. Finally, an extension of the cycle-based scheduler, capable of providing service with a lower priority for tasks during periods that were not dedicated to them was described in [12].

2. Details of the Studied System

The system under consideration belongs to the family of discrete-time queueing systems with vacations that are fed by M independent packet streams (as shown in Fig. 3), with their buffer being of the infinite size. Access that packets belonging to specific streams have to a commonly shared link of capacity C is governed by a cycle-based scheduler. For this purpose, the system allocates a buffer to each packet stream and assigns a period of time in consecutive cycles. More precisely, cycle time duration T is divided into M periods of T_m $(m = 1, \ldots, M)$, where $T = \sum_{m=1}^{M} T_m$. During the T_m period, only packets belonging to stream m may be transmitted.

The system studied may be analyzed from the point of view of each separate packet stream. This is possible thanks to the use of the cycle-based scheduler which ensures performance isolation between the packet streams serviced. The above means that packet transfer characteristics (defined, for instance, by delay and loss rate) concerning a given packet stream are not disturbed by servicing packets belonging to the remaining streams.

Figure 4 shows the discussed system from the point of view of packet stream m $(m = 1, \ldots, M)$. This packet stream identifies its packet queue and those periods during the consecutive cycles in which the packets from this stream may be served. So, stream m identifies its own cycle in which active period T_{mA} and vacation period T_{mV} may be observed. It needs to be noted that the cycle visible to each packet stream is of the same length as the length of the cycle in the scheduler. The further analysis assumes that the duration of the cycle, as well as the duration of the periods dedicated to serving packets from particular streams, are constant.

Moreover, we assume that packets belonging to the same stream have a constant length. For the sake of simplicity,

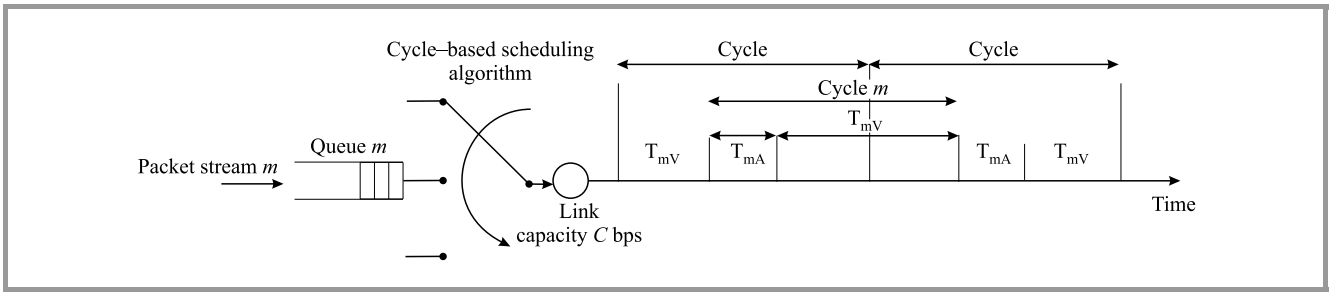


Fig. 4. System from the point of view of handling packets belonging to stream *m*.

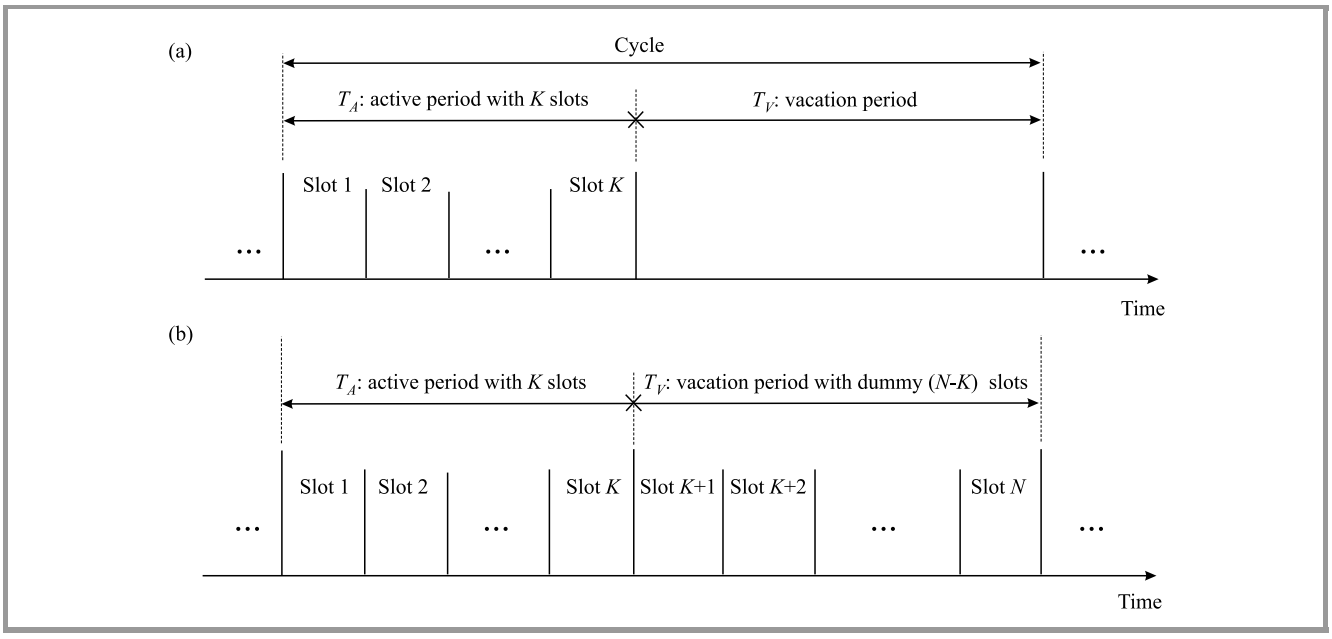


Fig. 5. Time slots within the cycle: (a) real system, *K* time slots in the active period, (b) model proposed for the analysis, with a cycle consisting of *N* slots, *K* slots in the active period, and *N*–*K* dummy slots in the vacation period.

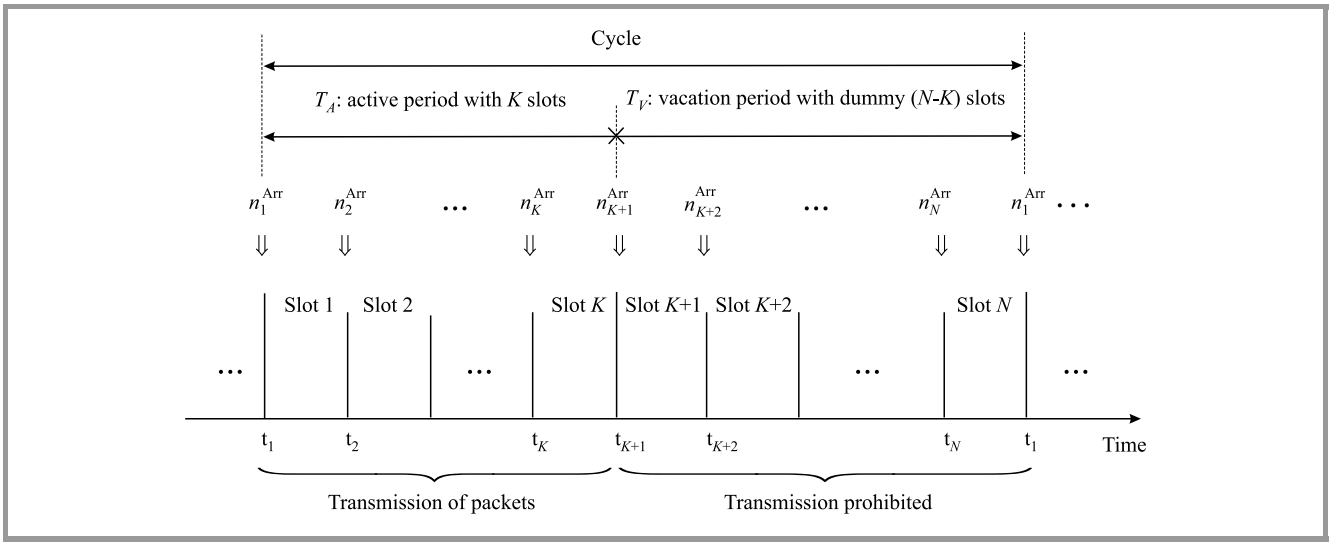


Fig. 6. Packet arrival and service processes.

the numbering of packet streams is omitted here, since the system under consideration is the same from the point of view of each stream. Thus, packets belonging to a given stream are transmitted within a time slot belonging to active periods only, as shown in Fig. 5a. On the other hand, new packets may arrive into the system any time. In order to unify the considerations, we introduce dummy time slots into the vacation periods. The lengths of these time slots are the same as the lengths of time slots in the active periods, as shown in Fig. 5b. However, no packet transmission is allowed in the dummy time slots. Such an assumption is made to count packet sojourn times in a number of slots. Finally, the system under consideration may be classified as a discrete-time system with vacations, where the cycle consists of K time slots in an active period and N time slots in total.

New packets arrive into the system in batches, at the beginning of each slots t_i $(i = 1, \ldots, ,N)$ only. These moments occur just after the previous slot has finished, and just before the first waiting packet is taken for transmission (for slots from 1 to K only). Therefore, at these moments, all packets in the system wait in a queue (for service). The arrival process of new packets and the service process is illustrated in Fig. 6, where n_i^{Arr} is the number of new packets arriving into the system at time t_i. We analyze the system assuming that the packet arrival process at specific t_i moments may differ for different i, $i = (1, \ldots, N)$.

3. Analysis

In our analysis, we will consider a system with an infinite buffer size. The first observation is that the system maintains constant properties with respect to time slots that have the same position within the cycle. The following relation takes place:

$$\Pr\{k \text{ packets in the system at time } t_i\} = \Pr\{k \text{ packets in the system at time } t_i + \tau\}, \quad \text{where } \tau = j(T_A + T_V),\ j = 1, 2, \ldots. \tag{1}$$

3.1. System State Distribution

The system state is described by the number of packets available in the system at a given time. Let us define the system state n_i $(n_i = 0, 1, 2, \ldots,)$ when the i-th time slot $(i = 1, \ldots, N)$ in the cycle starts. As only one packet may be served during one time slot belonging to the active period, and due to the fact that no packet service occurs during the vacation period, we can write the following relations for those periods (see Fig. 6):

$$\begin{cases} n_i = \max\left(n_{(i-1)} - 1, 0\right) + n_i^{Arr}, & \text{for } i = 2, \ldots, K+1 \\ n_i = n_{(i-1)} + n_i^{Arr}, & \text{for } i = K+2, \ldots, N, 1 \end{cases}, \tag{2}$$

where n_i^{Arr} denotes the number of new packets arriving into the system at t_i $(i = 1, \ldots, N)$ and $n_{(i-1)}$ is the number of packets being in the system at the time slot that is located before the i-th time slot (e.g. time slot N is before the time slot 1).

For the sake of simplicity, let us continue our analysis under the assumption that the number of packets arriving into the system at moments t_i $(i = 1, \ldots, N)$ is described by the same probability distribution function. Therefore, probabilities $\Pr\{n_i^{Arr} = j\}$, $j = 0, 1, 2, \ldots$, do not depend on the position of the slot in the cycle. The above assumption is not critical in our approach. In fact, the presented method can be easily adapted to a scenario in which packet arrival processes are not the same for different t_i. However, for a given i these distributions need to be identical. Thanks to Eq. (2), we can write the following set of equations for $i = 2, \ldots, K+1$:

$$\left[\Pr\{n_i = 0\}\ \Pr\{n_i = 1\}\ \Pr\{n_i = 2\}\ \ldots\right] = \left[\Pr\{n_{(i-1)} = 0\}\ \Pr\{n_{(i-1)} = 1\}\ \Pr\{n_{(i-1)} = 2\ \ldots\}\right] \cdot A, \quad \text{for } i = 2, \ldots, K+1, \tag{3}$$

where:

$$A = \begin{bmatrix} \Pr\{n_i = 0 | n_{(i-1)} = 0\} & \Pr\{n_i = 1 | n_{(i-1)} = 0\} & \ldots \\ \Pr\{n_i = 0 | n_{(i-1)} = 1\} & \Pr\{n_i = 1 | n_{(i-1)} = 1\} & \ldots \\ \Pr\{n_i = 0 | n_{(i-1)} = 2\} & \Pr\{n_i = 1 | n_{(i-1)} = 2\} & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}.$$

Assuming that the arrival process does not depend on the system state, we count the items of matrix A as:

$$\Pr\{n_i = m | n_{(i-1)} = k\} = \begin{cases} \Pr\{n^{Arr} = m\}, & \text{for } n = 0 \\ \Pr\{n^{Arr} = m - k + 1\}, & \text{for } k > 0 \text{ and } m \geq k - 1 \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Taking into account Eqs. (3) and (4), matrix A is:

$$A = \begin{bmatrix} \Pr\{n^{Arr} = 0\} & \Pr\{n^{Arr} = 1\} & \Pr\{n^{Arr} = 2\} & \ldots \\ \Pr\{n^{Arr} = 0\} & \Pr\{n^{Arr} = 1\} & \Pr\{n^{Arr} = 2\} & \ldots \\ 0 & \Pr\{n^{Arr} = 0\} & \Pr\{n^{Arr} = 1\} & \ldots \\ 0 & 0 & \Pr\{n^{Arr} = 0\} & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}.$$

Similarly to Eq. (3), using Eq. (2), we can write the following equations for $i = K+2 \ldots, N, 1$:

$$\left[\Pr\{n_i = 0\}\ \Pr\{n_i = 1\}\ \Pr\{n_i = 2\}\ \ldots\right] = \left[\Pr\{n_{(i-1)} = 0\}\ \Pr\{n_{(i-1)} = 1\}\ \Pr\{n_{(i-1)} = 2\ \ldots\}\right] \cdot B, \quad \text{for } i = K+2, \ldots, N, 1, \tag{5}$$

where:

$$B = \begin{bmatrix} \Pr\{n_i = 0 | n_{(i-1)} = 0\} & \Pr\{n_i = 1 | n_{(i-1)} = 0\} & \ldots \\ \Pr\{n_i = 0 | n_{(i-1)} = 1\} & \Pr\{n_i = 1 | n_{(i-1)} = 1\} & \ldots \\ \Pr\{n_i = 0 | n_{(i-1)} = 2\} & \Pr\{n_i = 1 | n_{(i-1)} = 2\} & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}.$$

Still, assuming that the arrival process does not depend on the system state, we count the items of matrix B in the following manner:

$$\Pr\{n_i = m | n_{(i-1)} = k\} = \begin{cases} \Pr\{n^{Arr} = m-k\}, & \text{for } m \geq k \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

Matrix B is:

$$B = \begin{bmatrix} \Pr\{n^{Arr}=0\} & \Pr\{n^{Arr}=1\} & \Pr\{n^{Arr}=2\} & \dots \\ 0 & \Pr\{n^{Arr}=0\} & \Pr\{n^{Arr}=1\} & \dots \\ 0 & 0 & \Pr\{n^{Arr}=0\} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}.$$

By combining Eqs. (3) and (5), we can write the formulas for the distribution of the number of packets in the system at time t_1:

$$\left[\Pr\{n_1=0\}\ \Pr\{n_1=1\}\ \Pr\{n_1=2\}\ \dots\right] = \left[\Pr\{n_1=0\}\ \Pr\{n_1=1\}\ \Pr\{n_1=2\}\ \dots\right] \cdot A^K \cdot B^{(N-K)}. \quad (7)$$

Finally, on the basis of Eq. (7) together with:

$$\sum_{k=0}^{\infty} \Pr\{n_N = k\} = 1, \quad (8)$$

we get the number of equations that can be used to numerically calculate system state distribution at time t_1 – see Eq. (2). On the basis of these values, we can calculate system state distributions for the remaining times t_i $(i = 2,\dots,N)$, using Eqs. (3) and (5).

Matrices A and B are of an infinite size, due to the unlimited buffer size. However, in practice, we can limit the size of these matrices assuming that we consider the probabilities of packet number arrivals that are greater than the assumed threshold ε, e.g. $\varepsilon \geq 0.0001$. The formula describing the probability that k packets are present in the system is:

$$\Pr\{n = k\} = \frac{1}{N} \sum_{i=1}^{N} \Pr\{n_i = k\}, \quad (9)$$

where $\Pr\{n_i = k\}$ is the probability that in the i-th slot $(i = 1,\dots,N)$ k packets are present in the system.

3.2. Packet Sojourn Time Distribution

The packet sojourn time is defined as the period between the arrival of a packet into the system and the completion of its service. Here, it is counted for a number of time slots. Packet sojourn time distribution is computed as a function of system state distribution and packet arrival distribution. The analysis begins by recalling the formula that was derived for the packet sojourn time in the case of the discrete-time queueing system without vacations and FIFO discipline fed by packets arriving to the system in each slot accordingly to the same probability distribution [2]:

$$\Pr\{D = k\} = \begin{cases} 0, & \text{for } n = 0 \\ \dfrac{\Pr\{X = k\}}{\rho}, & \text{for } k > 0 \end{cases}, \quad (10)$$

where: $\Pr\{D = k\}$ denotes the probability that packet sojourn time is equal to k $(k = 1,2,\dots)$ time slots and $\Pr\{X = k\}$ denotes the probability that at a time immediately following the end of a slot (and just before a packet is taken into service in the next slot), there are k packets in the system, and ρ is the load of each slot.

Unfortunately, Eq. (10) cannot be adapted to the presented system comprising both active and vacation periods, as the load varies between slots and depends also on the slot's position in the cycle. In particular, when packets arrive into the system in each slot based on the same probability distribution, then the first active slot after a vacation period has a greater load than other active slots remaining in the cycle. Equation (10) is nevertheless useful for checking the correctness of a more general formula in which we assume that no vacation periods are present in the system under consideration.

Our approach focuses on those packets that finish their services within given time slots that are a part of the active period. The sojourn time of such packets is equal to the number of time slots. We need to know the time slot during which this packet arrived into the system and how many packets had to be served earlier, i.e. between the moment it arrived into the system and the moment it is taken for service. When the FIFO queuing approach is relied upon, those packets that arrive into the system after the arrival of the packet in question exert no impact on its sojourn time. Thus, a strict dependency exists between the number of packets waiting in the queue at the arrival of the new packet into the system and its sojourn time. Notice that at the period between the arrival moment of a packet and the moment of ending its service the system is in the busy period, meaning that packets are served in all time slots within the active periods in the interval under consideration.

Let us now focus on a packet that is taken into the service in the i-th time slot $(i = 1,2\dots,K)$ and its sojourn time in the system is k time slots. For such a packet, we define two parameters (Fig. 7): position $d_n(i)$ of the time slot in the cycle at the time of arrival of the packet and the number of time slots in the active periods $d_{n_A}(i)$ located between the packet's arrival and the beginning of its service. Notice that the value $d_{n_A}(i)$ denotes the number of waiting packets being present in the system at the arrival of the packet in question. These packets are served before the packet under consideration, due to the fact that the FIFO order of precedence has been adopted. The values of parameters $d_n(i)$ and $d_{n_A}(i)$ are calculated in the following manner:

$$d_n(i) = N - \text{mod}[k - i - 1, N], \text{ for } i = 1,\dots,K, \quad (11)$$

where $\text{mod}[x,y]$ is the modulo function, and:

$$d_{n_A}(i) = \begin{cases} k, & \text{for } i \geq k \\ i + \left\lfloor \frac{k-i}{N} \right\rfloor K + \\ \max(\text{mod}(k-i,N) - (N-K), 0), & \text{for } i < k \end{cases}, \quad (12)$$

where $\lfloor x \rfloor$ is an integer part of x.

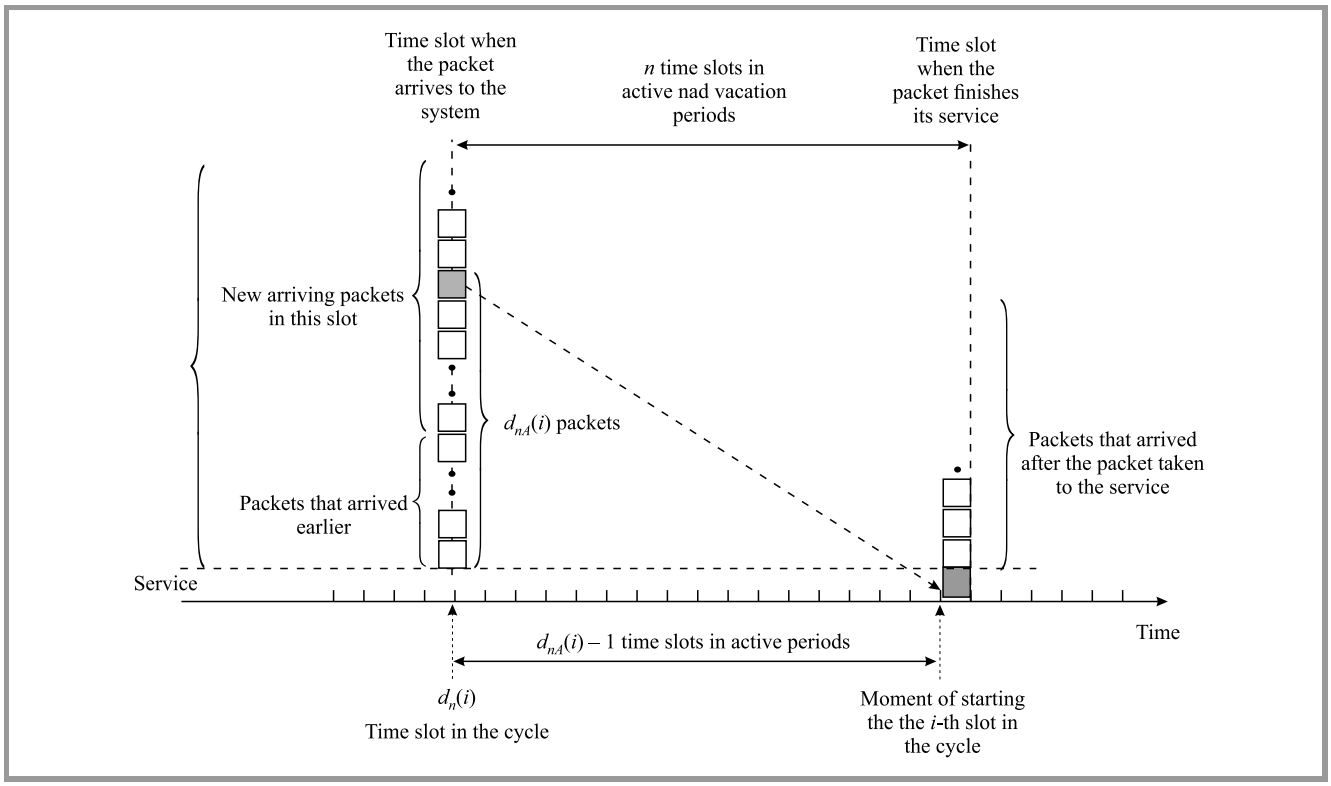


Fig. 7. Illustration of parameters $d_n(i)$ and $d_{n_A}(i)$.

To calculate the probability that the sojourn time of the packet which has finished its service in the i-th time slot $(i = 1,\ldots,K)$ and has spent k time slots in the system we use the formula:

For $k = 0$

$$\Pr\{D_i = k\} = 0 \quad . \tag{13}$$

For $k \geq 1$ and $d_n(i) = 2,\ldots,K+1$

$$\Pr\{D_i = k\} = \frac{\left(\begin{array}{l}\left[\Pr\{X_{d_n(i)-1}=0\}+\Pr\{X_{d_n(i)-1}=1\}\right]\Pr\{n_i^{Arr}\geq d_{n_A}(i)\}+ \\ \sum\limits_{m=1}^{d_n(i)-1}\Pr\{X_{d_n(i)-1}=m+1\}\Pr\{n_i^{Arr}\geq d_{n_A}(i)-m\}\end{array}\right)}{\rho_i} \quad . \tag{14}$$

For $k \geq 1$ and $d_n(i) = K+2,\ldots,N,1$

$$\Pr\{D_i = k\} = \frac{\left(\begin{array}{l}\Pr\{X_{d_n(i)-1}=0\}\Pr\{n_i^{Arr}\geq d_{n_A}(i)\}+ \\ \sum\limits_{m=1}^{d_n(i)-1}\left[\Pr\{X_{d_n(i)-1}=m\}\Pr\{n_i^{Arr}\geq d_{n_A}(i)-m\}\right]\end{array}\right)}{\rho_i} \quad , \tag{15}$$

where:

- $d_n(i)$ is calculated by Eq. (11),
- $d_{n_A}(i)$ comes from Eq. (12),
- $\Pr\{X_i = m\}$ $(m = 1,\ldots,N)$ denotes the probability that the number of packets in the i-th time slot $(i = 1,\ldots,N)$ equals m,
- $\Pr\{n_i^{Arr} \geq l\}$ $(l = 0,1,\ldots;\ i = 1,\ldots,N)$ denotes the probability that in the i-th time slot at least l new packets arrive,
- ρ_i is the load of the i-th time slot $(i = 1,\ldots,K)$ such as:

$$\rho_i = 1 - \Pr\{X_i = 0\} \, . \tag{16}$$

The index $d_n(i) - 1$ from Eqs. (13)–(15) refers to the time slot that precedes the time slot indexed by $d_n(i)$:

$$d_n(i) - 1 := \begin{cases} N, & \text{if } d_n(i) = 1 \\ d_n(i) - 1, & \text{otherwise} \, . \end{cases} \tag{17}$$

Finally, the probability that packet sojourn time in the system lasts k time slots is:

$$\Pr\{D = k\} = \begin{cases} 0, & \text{for } k = 0, \\ \sum\limits_{i=1}^{K} \frac{\rho_i}{\sum\limits_{j=1}^{K}\rho_j Pr}\{D_i = k\}, & \text{for } k > 0 \, . \end{cases} \tag{18}$$

Equations (13)–(15) should transform to Eq. (10) when there are no vacation periods in the analyzed system. In this case, we do not distinguish the positions of particular time slots. Moreover, the following relation takes place:

$$d_n = d_{n_A} = k \, . \tag{19}$$

From Eqs. (13)–(15) we get:

$$\Pr\{D=k\} = \begin{cases} 0, & \text{for } k=0, \\ \dfrac{\left([\Pr\{X=0\}+\Pr\{X=1\}]\Pr\{n^{Arr}\geq k\}+\sum_{m=1}^{k-1}\Pr\{X=m+1\}\Pr\{n^{Arr}\geq k-m\}\right)}{\rho}, & \text{for } k>0 \end{cases} . \quad (20)$$

For this system, we can write the following system state distribution equations:

$$\Pr\{X=k\} = (\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}=k\}+ \sum_{m=1}^{k}\Big[\Pr\{X=m+1\}\Pr\{n^{Arr}=k-m\}\Big], \quad \text{for } k=0,1,2,\ldots \quad (21)$$

By applying Eq. (21) to:

$$\Pr\{n^{Arr}\geq k\} = \Pr\{n^{Arr}\geq k-1\}-\Pr\{n^{Arr}=k-1\}, \quad \text{for } k=1,2,\ldots \quad (22)$$

the identity of Eqs. (20) and (10) for the system without vacations can be proved (see Appendix A).

3.3. Numerical Examples

Here, results corresponding to the system are presented with 15 time slots per cycle, with the first 5 slots constituting the active period ($T_A=5$) and the remaining 10 slots belonging to the vacation period ($T_V=10$). We will compare system state distributions and packet sojourn time distributions under the assumption that the utilization factor of the active periods equals 0.9 and that the packet arrival distribution is the same in each slot. Two following scenarios are verified:

- 1 – the packets arrive into the system according to binomial distribution with $p=0.3$, i.e. $\Pr\{n^{Arr}=1\}=0.3$ and $\Pr\{n^{Arr}=0\}=0.7$.
- 2 – the packets arrive into the system according to geometric distribution with $p=0.23$, $\Pr\{n^{Arr}=k\}=(1-p)p^k$ for $k=0,1,2,\ldots$.

For scenario 2, the variance of the arrival process distribution is greater than in scenario 1. The characteristics describing system state distributions for the two scenarios under consideration are presented in Figs. 8 and 9.

Figure 8 shows the mean $E[n_i]$ and the variance $var[n_i]$ of the random variable holding the number of packets being in the system at t_i $(i=1,2,\ldots,N)$. These values are greater in scenario 2 than in scenario 1. As expected, the lowest mean values of the number of packets in the system are observed at the beginning of vacation periods (slot 6), while their maximum values occur at the beginning of active periods (slot 1).

Figures 9 and 10 show the system state distribution calculated from Eq. (9) and the packet sojourn time distribution calculated from Eq. (18), respectively. These distributions may be characterized by long tails, since some packets may be kept waiting for their service even for a number of cycles.

The numerical results were confirmed by a discrete event simulation software written by the authors. The simulator validates both system state distribution at t_i $(i=1,\ldots,N)$ and packet sojourn time distribution. Results of the simulations are not presented, as they are indistinguishable from analytical results.

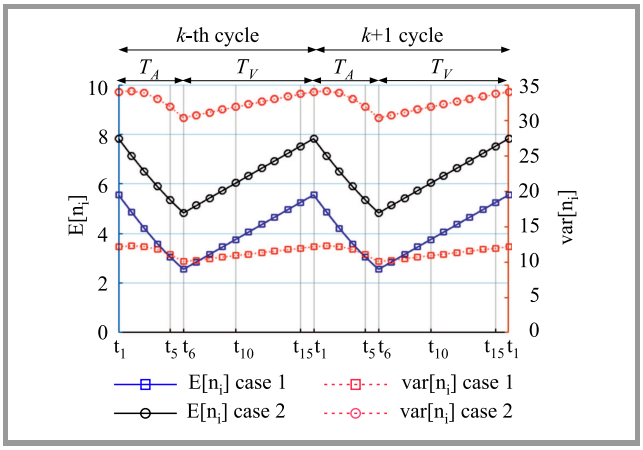


Fig. 8. Mean $E[n_i]$ and variance $var[n_i]$ of packets in the system at times t_i $(i=1,2,\ldots,N)$.

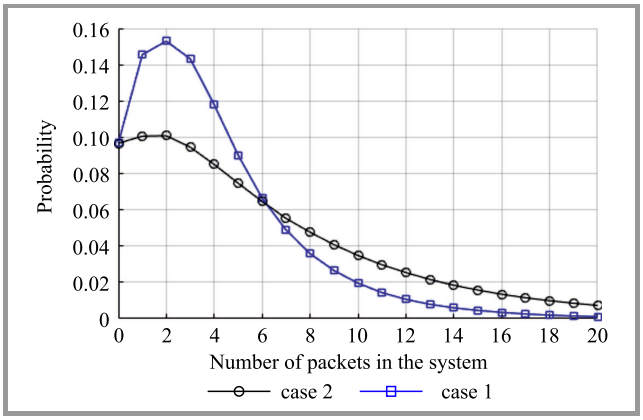


Fig. 9. System state distribution for $T_A=5$ and $T_V=10$) based on Eq. (9).

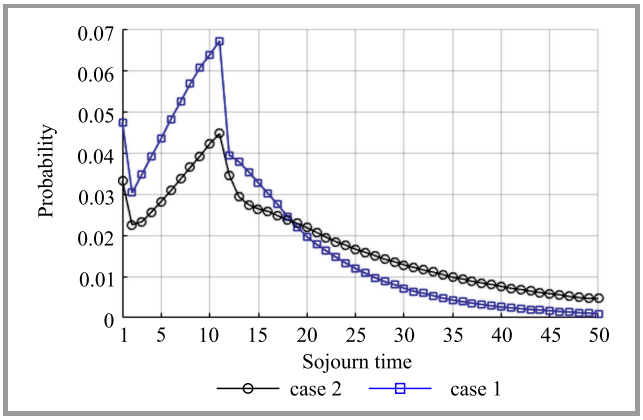


Fig. 10. Packet sojourn time distribution for $T_A=5$ and $T_V=10$ derived from Eq. (18).

4. Summary

In this paper, we have presented an analysis of a discrete–time multi-queue system with a cycle-based scheduler, in which the packet arrival process may be different for time slots with various positions within the cycle. The analysis of the system focused on calculating system state distribution and, based thereon, packet sojourn time distribution. The formulas presented provide exact solutions. The queueing system refers well to e.g. a system supporting a number of virtual links requiring performance isolation when they share a common physical link as it was implemented in the mentioned IIP System supporting a number of parallel internets.

Appendix A

Proof of Eqs. (10) and (20) identity for the system without vacations

We want to prove that for $n \geq 1$:

$$\Pr\{D=n\} = \frac{\Pr\{X=n\}}{\rho} = [\Pr\{X=0\}+\Pr\{X=1\}]\frac{\Pr\{n^{Arr}\geq n\}}{\rho} + \sum_{m=1}^{n-1}\Pr\{X=m+1\}\frac{\Pr\{n^{Arr}\geq n-m\}}{\rho}.$$

So, actually we need to prove that for $n \geq 1$:

$$\Pr\{X=n\} = [\Pr\{X=0\}+\Pr\{X=1\}]\Pr\{n^{Arr}\geq n\} + \sum_{m=1}^{n-1}\Pr\{X=m+1\}\Pr\{n^{Arr}\geq n-m\},$$

using a mathematical induction.

Base case

For $n=1$:

$$\Pr\{X=1\} = (\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}\geq 1\} = (\Pr\{X=0\}+\Pr\{X=1\})(\Pr\{n^{Arr}\geq 0\}-\Pr\{n^{Arr}=0\}) = \Pr\{X=0\}+\Pr\{X=1\}-(\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}=0\} = \Pr\{X=1\},$$

since $(\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}=0\} = \Pr\{X=0\}$ – see Eq. (21).

Inductive step

Let us assume that for an arbitrary $n=k$, $k \geq 1$:

$$\Pr\{X=k\} = (\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}\geq k\} + \sum_{m=1}^{k-1}[\Pr\{X=m+1\}\Pr\{n^{Arr}\geq k-m\}]. \quad (23)$$

For $n=k+1$:

$$\Pr\{X=k+1\} = \underbrace{(\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}\geq k+1\}}_{C1} + \underbrace{\sum_{m=1}^{k}[\Pr\{X=m+1\}\Pr\{n^{Arr}\geq k-m+1\}]}_{C2}$$

Let us rewrite particular components. The first one ($C1$):

$$C1 = (\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}\geq k+1\} = (\Pr\{X=0\}+\Pr\{X=1\})(\Pr\{n^{Arr}\geq k\}-\Pr\{n^{Arr}=k\}) = \underbrace{(\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}\geq k\}}_{C11} - \underbrace{(\Pr\{X=0\}+\Pr\{X=1\})\Pr\{n^{Arr}=k\}}_{C12}$$

The $C2$:

$$C2 = \sum_{m=1}^{k}[\Pr\{X=m+1\}\Pr\{n^{Arr}\geq k-m+1\}] = \sum_{m=1}^{k}[\Pr\{X=m+1\}\Pr\{n^{Arr}\geq k-m\}] - \underbrace{\sum_{m=1}^{k}[\Pr\{X=m+1\}\Pr\{n^{Arr}=k-m\}]}_{C22} = \underbrace{\sum_{m=1}^{k-1}[\Pr\{X=m+1\}\Pr\{n^{Arr}\geq k-m\}]}_{C21} + \Pr\{X=k+1\} - C22$$

Finally:

$$\Pr\{X=k+1\} = C1+C2 = C11-C12+C21+\Pr\{X=k+1\}-C22 = (C11+C21)-(C12+C22)+\Pr\{X=k+1\} = \Pr\{X=k+1\},$$

since, from Eq. (23): $C11+C21 = \Pr\{X=k\}$ and also from Eq. (21): $C12+C22 = \Pr\{X=k\}$.

Since both the base case and the inductive step have been proved as true, by mathematical induction the statement holds for all $n \geq 1$, Q.E.D.

References

[1] W. Burakowski, "Virtualized network infrastructure supporting co existence of Parallel Internets", in *Proc. 13th ACIS Int. Conf. on Software Engin., Artif. Intell., Network. and Parall./Distrib. Compu. SNPD 2012*, Kyoto, Japan, 2012 (DOI: 10.1109/SNPD.2012.67).

[2] B. Vinck and H. Brunnel, "Relationships between delay and buffer contents in ATM queues", *Electron. Lett.*, vol. 31, no. 12, 1995 (DOI: 10.1049/el:19950662).

[3] N. P. Dellaert, *Production to Order. Models and Rules for Production Planning*. Berlin Heidelberg: Springer, 1988 (ISBN: 9783540513094).

[4] M. J. A. Eenige, *Queueing Systems with Periodic Service*. TU Eindhoven, Netherlands, 1996 (ISBN: 9038603487).

[5] H. Takagi, *Queueing Analysis: Vacation and Priority Systems*, pt. 1. Amsterdam: North-Holland, 1991 (ISBN: 9780444817709).

[6] N. Tian and Z. G. Zhang, *Vacation Queueing Models – Theory and Applications*. New York: Springer, 2006 (ISBN: 9780387337210).

[7] B. T. Doshi, "Queueing systems with vacation: A survey", *Queueing Systems*, vol. 1, pp. 29–66, 1986 (DOI: 10.1007/BF01149327).

[8] J. C. Ke, C. H. Wu, and Z. G. Zhang, "Recent developments in vacations queueing models: A short survey", *Int. J. of Oper. Res.*, vol. 7, no. 4, pp. 3–8, 2010 [Online]. Available: http://www.orstw.org.tw/ijor/vol7no4/2-Vol.%207,%20No.4%20pp.3-8.pdf

[9] D. R. McNeil, "A solution to the fixed-cycle traffic light problem for compound Poisson arrivals", *J. of Appl. Probab.*, vol. 5, no. 3, 1968, pp. 624–635 (DOI: 10.2307/3211926).

[10] A. Chydzinski and B. Adamczyk: "Analysis of scheduler for virtualization of links with performance isolation", *Appl. Mathem. & Inform. Sci.*, vol. 8, no. 6, pp. 2653–2666, 2014 (DOI:10.12785/amis/080601).

[11] M. Sosnowski and W. Burakowski. "Analysis of the system with vacations under Poissonian input stream and constant service times", *J. of Telecommun. and Inform. Technol.*, no. 3, 2013 [Online]. Available: https://www.il-pib.pl/czasopisma/JTIT/2013/3/3.pdf

[12] W. Burakowski and M. Sosnowski, "On cycle based schedulers with time alternating priorities", in *Proc. of 27th Int. Telecommun. Netw. and Appl. Conf. ITNAC 2017*, Melbourne, Australia, 2017, pp. 1–6 (DOI: 10.1109/ATNAC.2017.8215377).

Wojciech Burakowski works at the Institute of Telecommunications, Warsaw University of Technology and at the National Institute of Telecommunications and is a member of the Architectures and Applications for the Internet Team. He has participated, since 1990, in several COST and EU Framework Projects. He is a member of the Telecommunications Section of the Polish Academy of Sciences. He was a chairman and a member of many technical program committees at national and international conferences. He is the author or co-author of approximately 250 papers published in books, international and national journals and conference proceedings. Burakowski is also the author of approximately 80 technical reports. His research areas include network techniques for the Internet (ATM, IP QoS, Future Internet), heterogeneous networks (fixed and wireless), network architectures, clouds (including MEC), traffic control, simulation techniques as well as network mechanisms and algorithms. His latest area of interest is in designing wide area research networks for 5G.

https://orcid.org/0000-0002-8486-8004
E-mail: W.Burakowski@il-pib.pl
National Institute of Telecommunications
Szachowa 1
04-894 Warsaw, Poland

Maciej Sosnowski received his M.Sc. degree in Telecommunications from Warsaw University of Technology in 2012. He is a member of the Internet Technologies and Applications Department at the National Institute of Telecommunications, Poland, and concurrently holds a research position within the Architectures and Applications for the Internet (AAI) Team, operating at the Institute of Telecommunications, Warsaw University of Technology. His research interest lies in the areas of queuing theory, virtualization techniques, Internet of Things, cloud computing, multi-access edge computing, and 5G networks.

https://orcid.org/0000-0003-0563-1111
E-mail: M.Sosnowski3@il-pib.pl
National Institute of Telecommunications
Szachowa 1
04-894 Warsaw, Poland

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Editors' Note on the Special Section

The dynamic, not to say the rapid, development of wireless communication technologies has had enormous impact on nearly all aspects of our everyday life and in a large number of cases has changed them significantly in the process. The same also applies to the technologies and services that are related to the introduction of the fifth generation of mobile networks. As this development represents an evolution in communication architectures which should ensure much faster speeds, it will not be long before its particular and continued significance will become even more apparent to us all. Accordingly, the mission statement of the annual convention of the Federation of Telecommunications Engineers of the European Community to be held in Cracow and planned for September 28–29, 2020, was "5G – Opportunities and Threats". Regrettably, the pandemic restrictions prevented the organization of the event in the traditional conference format with audience participation. Therefore, we invited our authors to submit their articles for publication in the special section. Ultimately, 5 articles by the authors from Italy, Greece, Finland, the UK and Poland have been selected for publication.

These articles have been divided into three sections that cover different areas of research. The first one, including two papers, presents the view of 5G networks from the operator's perspective. The first article in this part, authored by Peter McCarthy-Ward, Andy Valdar, Stuart Newstead, and Stuart Revell entitled "5G New Business Opportunities – New Business Models, Pricing and Use Cases", presents a profitability analysis and return of investment following the introduction of 5G networks from the operator's perspective, with three exemplary practical usage scenarios for 5G. Then, the article "5G Is Out There: How to Ride the Market Storm and Thrive" by Edward Smith and Mauro Ugolini presents the changes in the commercial model adopted by the mobile ICT industry effected by the introduction of the 5G technology. The article discusses the influence of the 5G technology on the service market and its significance as an element of the competitive strength of mobile network operators.

The second section is devoted to a presentation of a number of exemplary practical solutions that are based on the 5G technology. This section includes the article entitled "C-V2X Communications for the Support of a Green Light Optimized Speed Advisory (GLOSA) Use Case", prepared by an international team of twelve authors. The article focuses on the use of the 5G technology in vehicular communications, or more precisely within the framework of cooperative Intelligent Transportation Systems (C-ITS), in the context of cellular V2X (C-V2X) technologies. The article draws on the research results obtained within the framework of the 5G-DRIVE project promoting cooperation between the EU and China. The particular case of the Green Light Optimized Speed Advisory (GLOSA) is discussed in the article in more detail.

The last part of the special section contains two articles on different aspects of the research into the upcoming development of the fifth-generation mobile network technology. The article prepared by Piotr Remlein and Urszula Stachowiak and entitled "Security Verification in the Context of 5G Sensor Networks" focuses on the ways the safety of network solutions can be evaluated. With the example of a tool that performs automated symbolic analysis (Tamarin prover), the authors present how to verify the correctness of the operation of safety protocols in a formal way. The article also discusses an example of modeling the process of the DTLS 1.2 handshake protocol enriched with the TCP Syn Cookies mechanism which is dedicated to preventing DoS attacks. The authors have proved that the Tamarin software can be successfully used to evaluate safety protocols in 5G networks.

The last article in this section, authored by Małgorzata Wasilewska and Łukasz Kułacz, is entitled "Machine Learning-Based Small Cell Location Selection Process". The paper discusses an algorithm to determine the location in small cells in a dense metropolitan area network. The algorithm proposed in the article makes use of machine learning methods, such as k-means clustering and spectral clustering, while it uses ray tracing to model channels. The article considers two scenarios for the choice of the base station: the one that is based on an arbitrary choice and the other that is based on the level of signal strength. To evaluate both scenarios, the value of the average bitrate is used. The authors claim and prove that the machine learning method can be successfully used in the solved example.

We recommend reading these interesting articles.

Maciej Sobieraj and Piotr Zwierzykowski
Guest Editors

https://doi.org/10.26636/jtit.2021.preface2

Paper

5G New Business Opportunities – New Business Models, Pricing, and Use Cases

Peter McCarthy-Ward[1], Andy Valdar[2], Stuart Newstead[3], and Stuart Revell[4]

[1] *Freelance Lecturer & Consultant, British Telecom retired*
[2] *University College London*
[3] *Ellare*
[4] *RTA Communicating Systems Ltd*

https://doi.org/10.26636/jtit.2021.152221

Abstract—This paper addresses how network operators may gain a reasonable return on their investment into 5G infrastructure. It first considers the 5G mobile network costs structure then applies this to three typical use cases.

Keywords—5G services, business challenges, infrastructure.

1. Introduction

The drive to build and operate 5G networks continues to be a priority for policymakers across the world. The pressure on mobile network companies to transition rapidly to 5G platforms and services is immense. Less attention has been given to how those investing in 5G infrastructure will make a reasonable return on their investment. This paper considers some of the opportunities available to 5G investors and some of the constraints and limitations on how those opportunities may be exploited. We begin with an examination of pricing in relation to network services, then turn to the cost structure of 5G networks, and finally look at how 5G's new network features can support new revenue growth.

2. Pricing

The costs of producing a product are recovered through pricing. The microeconomic theory of pricing is charmingly straightforward. It states that the price for a good will settle at the point where supply matches demand. That point of equilibrium is reached when the price a customer is willing to pay matches the marginal cost incurred in producing the good. If the price is higher demand will drop. If the price is lower there is no incentive to supply. Marginal cost is the cost added by producing an additional unit of supply.

For the theory to hold, certain simplifying assumptions have to be made – such as effective competition, buyer rationality, perceived value, portfolio independence and cost recovery timescales. This said, Fig. 1 provides a good summary of the fundamentals, showing that supply equals demand at the intersection of P1 and V1, when the price (and marginal cost) will be P1.

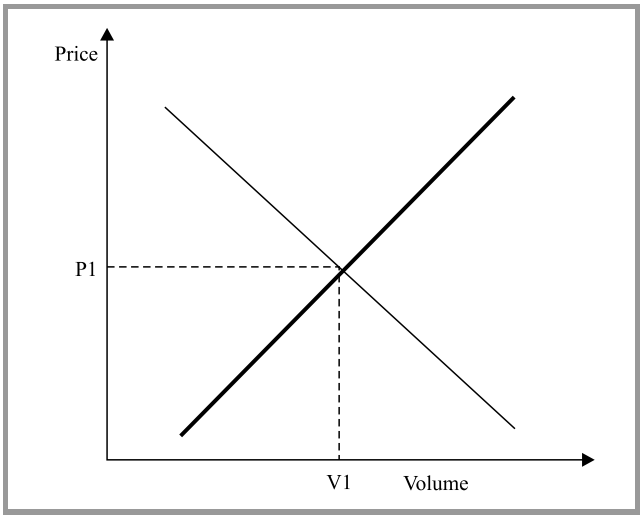


Fig. 1. Price-volume relationship.

The simplifying assumptions of microeconomic theory become particularly relevant when considering network pricing. Communication networks have very high fixed costs and very low variable costs. Network marginal costs rise as a stepped function. The cost of an incremental megabyte of traffic is close to zero, until network capacity is exhausted. The next megabyte requires network expansion and has a huge marginal cost. It is for this reason that regulators often look at long run incremental costs (the cost of providing the whole product or service) rather than marginal costs (the cost of providing a unit of that product or service) when considering regulated prices and interconnection.

A network can supply a portfolio of services, and network operators have options over how to recover fixed costs across that portfolio.

Network operators also have the choice (or a regulatory obligation) to offer wholesale as well as retail services. Here the consideration is the balance to be struck between the potential for better loading of network capacity, and the risk to retail volumes and prices of supporting a competing mobile virtual network operator (MVNO).

High capacity digital networks have led to new forms of competition. What were traditional network services such as text messaging and voice traffic are now supplied by over-the-top providers (OTTPs). New on-line services such as video content, broadcast and catch-up TV and so on, once seen as the future for operators, are increasingly dominated by more fleet-footed OTTPs. Here the risk to operators is not simply the foregone added value opportunity, but also the price volume P1-V1 possibility of network services becoming bundled with OTT packages with the choice of operator moving from the consumer to the OTTP.

These complications and refinements above and beyond basic microeconomic theory make network service pricing particularly challenging. We next consider how 5G cost structures differ from those of current networks, and how its features change the service and portfolio options available to network operators.

3. 5G Costs

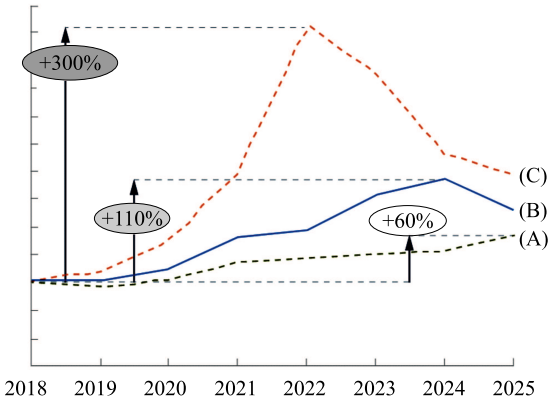
5G specifications relating to the air interface were agreed in 2017 and for the 5G architecture in 2018. Later work is addressing the specification of the 5G next generation core (NGC). Initial deployments of 5G networks thus precede the availability of NGC equipment and will rely on the cores of existing 4G networks. While eventually, the 5G network will become stand-alone and capable of providing an omnipotent facility covering fixed and mobile communications, there will be a period of parallel running of 4G and 5G.

5G's use of higher frequency bands (3.4, 3.8, and 24.25 to 27.5 GHz) gives greater user bandwidth, but at the expense of reduced cell sizes. However, new spectrally-efficient forms of multiplexing the data onto the radio carriers together with the use highly directional multiple input multiple output (MIMO) antenna technology, gives a major increase in bits-per-Hz. So, we can expect a more cost-effective way of carrying greatly increased user data rates.

The 5G NGC will exploit several new network technologies within an IP integrated architecture [1], [2]. An important innovation is network slicing, whereby the capacity is partitioned so that an appropriate part through the NGC is dedicated to a service type or even an individual customer. This enables the operator to guarantee network performance, something new for IP networks. It also enables better network utilization since capacity can be used optimally for the class of traffic carried – with consequent operational cost savings for the operator.

New technologies that promise to reduce 5G network equipment costs are network functions virtualization in which many of the functions within the NGC are realized in software run on standard processors. Further economies can be gained by hosting the functions in one or more clouds. Crucially, the functions and network capacity can be applied dynamically, enabling tracking of instantaneous traffic demand – giving operational cost savings and potential new revenue opportunities. Further possible service features and capital cost savings are expected by deploying edge computing and, possibly, content-distribution network technologies.

The capital cost of 5G network deployment per bit of user data carried will decrease as will operational costs in managing 5G network capacity. However, the 5G network deployment will not be contiguous for many years, the existing 4G networks being needed to provide full mobile coverage – so the operators will have the burden of running two networks. This tension was neatly captured by



Scenarios consider costs associated wit legacy network evolution, small-cell densification, and the addition of a 5G macro layer

Scenarios:
(A) 25% data growth
(B) 35% data growth
(C) 50% data growth

Note: Total cost of ownership includes capital expenditures and operational expenditures for radio access networks and transmission but not core networks. Data are based on 3 operators in a European country. Results are rounded.

Fig. 2. Total cost of ownership of mobile networks over time [3].

McKinsey in a 2018 report [3]. The humps in the chart of Fig. 2 show cost of ownership of the radio access network (excluding core) peaking and then reducing as 5G build matures.

In conclusion, the capital costs of access and core network build are high and near certain. Spectrum costs have already been incurred. Parallel running will increase costs.

These investments create opportunities for improved network efficiency, where gains are probable but not guaranteed. They enable much higher data rates but whether operators will be able to generate significantly higher charges is not assured, and costs and revenues are dependent on assumptions about data growth where reasonable projections span a broad range. They enable new services, and thus new sources of revenue, though those opportunities will be contested between fixed and mobile networks and by MVNOs and service providers.

4. New 5G Services

Three types of use case are used as umbrella terms in describing potential 5G services:

- enhanced mobile broadband (eMBB),
- massive machine-type communications (mMTC),
- ultra-reliable low-latency communications (URLLC).

4.1. eMBB

In 5G, mobile broadband will be "enhanced", especially at the radio layer, to provide:

- more extensive coverage,
- denser coverage in highly populated areas (e.g. stadiums, commuter train stations, shopping centers),
- higher capacity (more connections per area, more total data carriage per area),
- higher speeds and lower latency for individual users,
- mobility service while travelling at higher speeds,
- more overall reliability,
- content caching at the base station – "multi-access edge computing",
- seamless management of access method (mobile, public, and private Wi-Fi).

Table 1 shows the performance goals for eMBB.

Mobile operators will want to fill their new capacity quickly, and as efficiently as possible. Only in this way will the theoretical unit cost reductions (cost per bit per Hz) be realizable. If this can be achieved, then operators could see profit growth from eMBB, even without premium pricing (other than for early adopters) – but the downside risk looms large.

Table 1
Performance goals for eMBB [4]–[5]

Use case area	Category	Performance goal
eMCC	Speed and throughput	1–10 Gbit/s connections
		Cell aggregate throughput: 20 Gbit/s downlink (DL), 10 Gbit/s uplink (UL)
		Indoor throughput 10 Mbit/s per m^2
		User experience DL 100 Mbit/s/UL 50 Mbit/s
	Latency	4 ms user plane, 10 to 20 ms control plane
	Mobility	Stationary 0 km/h
		Pedestrian 10 km/h
		Vehicular 10–120 km/h
		High speed 120–500 km/h

Most mobile operators are also fixed network operators. They can achieve economies of scope by running 5G fibre backhaul themselves, and by managing the access method for each device more efficiently. It could well be the case that some operators reflect these economies by offering a single solution for a device, a family or a small business. BT's recent promotions show signs of thinking along these lines, currently in the form of the converged "Halo" portfolios. So eMBB is an extension of home/office broadband and public Wi-Fi.

In principle, retail pricing for eMBB could also incorporate added-value elements, such as better experience in densely-covered areas, or "boost-it" temporary quality increments to, for example, speed up file transfer by caching it at the network edge.

Where these quality elements might be more relevant is in wholesale pricing to a wide collection of potential new service providers. These could range from virtual reality game providers, to highways agencies managing motorways, to factories and warehouses controlling robots and humans. Given the need to fill up their networks efficiently, networks may find themselves being as creative with wholesale pricing as they have been to-date with retail pricing.

4.2. mMTC and URLLC

This section explores some of the new opportunities created by mMTC and URLLC. The performance goals for these new technologies are summarized in Table 2.

The performance goals for mMTC and URLLC are based on the requirements of industry vertical sector use cases previously not supported by mobile network technologies. Although the requirements and needs have been known for a long while, e.g. industrial control systems, automotive telematics, connected health, in some cases the network economics, liability, security and performance have not

Table 2
5G performance goals for mMTC and URLLC [4]–[5]

Use case area	Category	Performance goal
mMTC	Density	1,000,000 nodes per km^2
URLLC	Security and reliability	Highly secure/resilient
	Latency	Deterministic quality of service (jitter and latency)
		Low latency: 1 ms user plane, 10 to 20 ms control plane

matched requirements and hence services have not been deployed.

Addressing the new use cases will mean deployments into new areas, both virtually and physically, and may require the use of multiple networks to create the end-to-end connections required. These connections may need to be negotiated in real time to set up and tear down the required connection and agree the required service level agreements (SLAs) to ensure the QoS profile for the specific use case can be delivered. Once the connection is negotiated and agreed, slicing mechanisms can then be used to ensure the necessary QoS service parameters can be supported. Slicing is a network feature that enables the physical network infrastructure to be portioned to provide a QoS controlled end-to-end path for defined services.

There will be fierce competition for leadership in the delivery of new use cases. MNOs able to leverage existing assets such as spectrum, backhaul, radio, core, billing systems and sites together with new techniques and technologies like orchestration and slicing will be best placed to secure and consolidate this leading position. Orchestration is a policy-driven function to coordinate the hardware and software components of a network to automate the way network requests are managed and delivered.

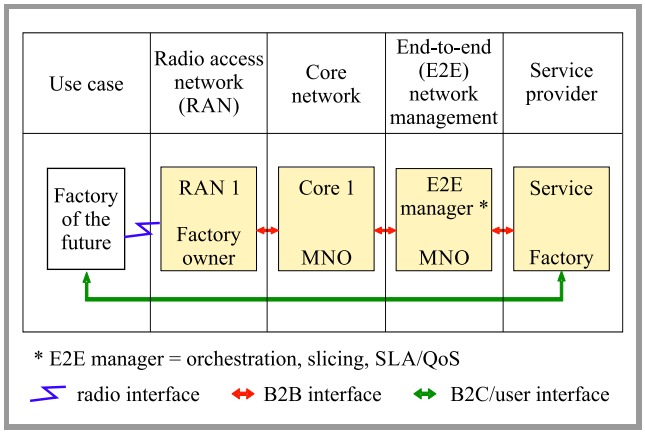


Fig. 3. Factory of the future, Industry 4.0 application.

Figure 3 shows a simple model to support an Industry 4.0 use case, two private entities collaborating, while Figs. 4–5 demonstrate some potential use cases.

In the example above, the factory location did not have good mobile coverage. In partnership with the MNO, the factory purchased and installed a 5G in-building radio access network (RAN), capable of eMBB, URRLC and mMTC functionality. The new network provided ability to run applications to operate critical control systems. Traditional mobile, telephony and broadband services can be enhanced through the indoor coverage, enabling one network investment to address multiple applications. The factory RAN is controlled by an MNO core and an E2E manager. The E2E manager could also be used for orchestrating other assets such as authenticated Wi-Fi or fixed communication technologies. The service provider is the factory which means that this could operate as a private network in the factory and, outside of the building, the employees would seamlessly connect to the MNO network.

Figure 4 shows a slightly more complex model where three entities are collaborating, two private and one public.

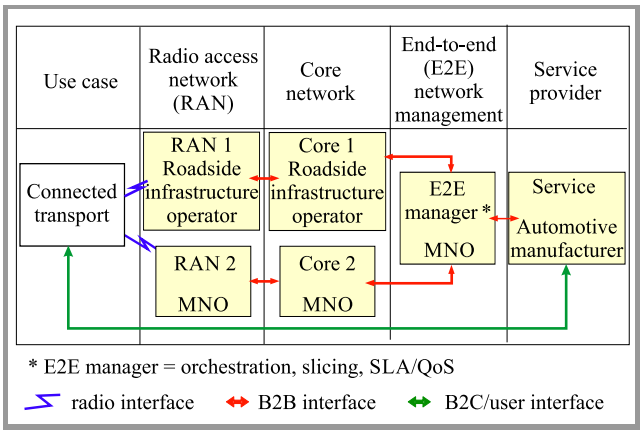


Fig. 4. Connected transport use case.

In this connected transport use case, two RANs are providing connectivity to the car, both with their own core network. Core 1 could be a separate physical network, or a virtual network operated by an MNO. The E2E management is being orchestrated by the MNO. The service provider is an automotive manufacturer. The approach on the RAN is based on a public roadside operator connecting 5G nodes to existing infrastructure to provide coverage to areas where traditionally the B2C MNO did not provide adequate mobile coverage.

As with the factories of the future example, the network is capable of eMBB, MMTC and URLLC thereby allowing the infrastructure to support multiple applications such as private radio, vehicle-to-everything (V2X) applications, mapping, mobile telephony, and broadband. In some areas, where multiple networks exist, the car can transparently switch between the two RANs, managed by the E2E orchestrator. The relationship with the consumer is through the car manufacturer and hence the other entities are not visible to the user.

The third example in Fig. 5 shows how such a modular business eco-system could address a local health and social care use case. In this case the complexity has increased again, reflecting the complexity and fragmentation of health and social care systems. The common approach enables applications such as telehealth, vital sign monitoring, secure large file transfer, robotic surgery, traditional mobile telephony, and broadband services to run over a common network infrastructure owned by multiple parties.

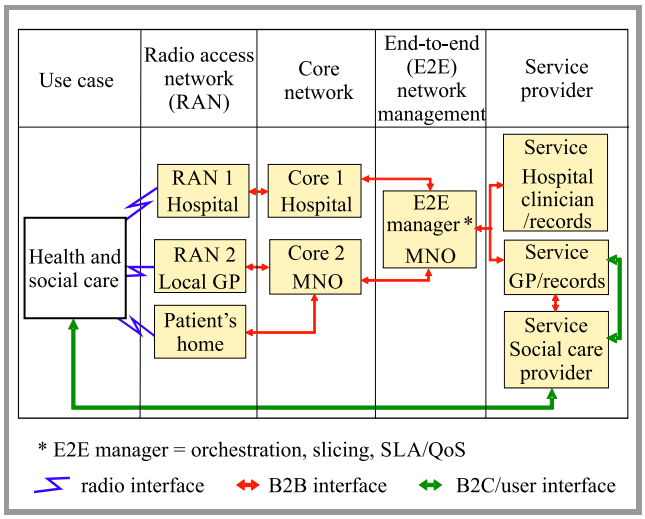


Fig. 5. Health and social care use case.

Figure 5 shows three RANs in the hospital, GP surgery and the patient's home. None of these needs be owned by the MNO, although the MNO can be the orchestrator of the system and the fourth RAN could be the MNO which means the patient can also be connected outside of the three main RAN areas, providing service based on context and location.

Implementing this sort of extended multiple player value chain is a complex process. The work of the third generation partnership project (3GPP) will define and help create a common technology base, however we are still far from standardizing the commercial common approach for multiple network/business owner interactions, and further work is required.

5. Conclusions

5G is being built in anticipation of a continuing and major growth in user data volumes. It will create a much greater capacity to connect and carry data traffic. But that comes at a cost – license fees, new access and core infrastructures, more cell sites and raised transitional costs during parallel running.

Capitalizing on the growth in data is not necessarily straightforward. Experience in fixed markets is that customers expect growing data capacity but not growing data prices. Competition will be fierce, between mobile network operators, fixed operators, MVNOs and service providers.

It is likely that 5G build will create capacity greater than demand, at least for an interim period. Price pressures on simple data packages will be acute. The challenge for operators will be to avoid commoditization of data services, through bundling with other services and terminals, differential levels of quality or establishing and building brand and reputational values.

The initial signs are that MNOs are trying to find means of differentiating their 5G data service from those of their competitors. EE began by charging a premium for 5G, aiming to capitalize on the enthusiasm of early adopters. Vodafone opted for innovation, dropping data limits and offering tiered pricing based on data speeds. Three positioned as the value for money player, offering 5G at no extra cost to existing customers. These initial positions will change as 5G build progresses and as market reaction to the different offers becomes apparent. But it is clear that 5G data prices will have to be innovative and find new sources of perceived user value to succeed.

5G also creates new opportunities, both for greater levels of network efficiency and for new sources of revenue. Investor, regulatory and competitive pressures are likely to ensure that operational efficiencies are delivered, or that failure to achieve them is punished.

Capitalizing on new sources of revenue is more complex. The sorts of use cases enabled by 5G require new competences to deliver and bring new sources of competition into play. Solutions will tend to be customer or sector specific, to require management across a range of networks, often with different owners, and to require ongoing management and oversight.

This combination of bespoke solutions and ongoing support of complex systems requires development and stewardship resources largely new to MNOs. It means new forms of partnership working. It means new requirements for B2B interfaces, business models and SLAs. The good news for MNOs is that part of the skill set required is expertise in the management of interconnected network infrastructures and application of technologies such as orchestration and slicing. This should play to existing core strengths.

The less good news is that, although necessary, network skills are not sufficient. MNOs will have to consider whether to develop, recruit, acquire or partner in order to get service development, solutions architecture, customer relationship management, contracting and contract management, billing and other skills necessary for success. In part, that choice will be driven by whether MNOs wish to contribute to a solution, by supplying an off-the-shelf capability and leaving leadership, management and ownership to others, or whether they wish to take the more costly but potentially more lucrative alternative of owning and leading the solution themselves.

It is likely that pragmatism will prevail, with MNOs choosing to lead on solutions which rest heavily on their core strengths and moving to a supplier-basis for other contracts. A further revenue opportunity lies in wholesale markets, supplying network services to MVNOs and others who may compete at the retail level. Network slicing holds the prospect of a richer and more varied wholesale portfolio.

The commercial opportunities of 5G are real but will not be

straightforward to seize and capitalize upon. Where once mobile licenses were considered both a permit to operate and a license to print money, now they are a commitment to spend against a significantly less confident possibility of a return.

6. Acknowledgement

This article first appeared in the Institute of Telecommunications Professionals Journal (vol. 13, no. 2, 2019) and is published here with the kind permission of the ITP.

References

[1] A. Sutton, "5G Network architecture", *J. of the Institute of Telecommun. Professionals*, vol. 13, no. 4, pp. 19–25, 2019.

[2] A. Valdar, "Developments in telecommunications network architecture", *J. of the Institute of Telecommun. Professionals*, vol. 12, no. 4, pp. 27–33, 2018.

[3] "The road to 5G: the inevitable growth of infrastructure cost", McKinsey & Co., 2018 [Online]. Available: www.mckinsey.com/industries/technology-media-andtelecommunications/our-insights/the-road-to-5g-the-inevitableinfrastructure-growth-of-infrastructure-cost

[4] Recommendation ITU-R M.2410-0, 2017 [Online]. Available: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2410-2017-PDF-E.pdf

[5] UK 5G Innovation Network, "Technical report on 5G network Architecture and Security", 2018 [Online]. Available: https://uk5g.org/5g-updates/research/technical-report-5g-network-architecture-and-secur/

Peter McCarthy-Ward retired from BT in 2008 after 24 years in regulatory, marketing and strategic roles. From 2005 he was BT's Project Director for Ofcom's Strategic Review of the industry and for the development and implementation of the resulting undertakings. Peter runs the strategy, product management and marketing module of University College London's MSC, and sits on the ITP Journal's Editorial Board.
E-mail: peter.mccarthyward@btinternet.com
Freelance Lecturer & Consultant

Andy Valdar has had a wide-ranging career in telecoms, covering network strategy, planning, standards, regulation, product management, as well as teaching and writing. After 30 years working for BT including three years on secondment to the UN in India, he joined University College London in 1999 directing MSc programmes. Now retired, he retains his Visiting Professor role at UCL and lectures on a part-time basis. Andy has authored four text books on telecoms, is a Director on the ITP board, former chairman of the ITPJ Editorial Board, and was President of FITCE Europe for three years. He is a fellow of the IET and ITP and is a Chartered Engineer.
E-mail: a.valdar@ucl.ac.uk
Visiting Professor
Department of Electronic & Electrical Engineering
Faculty of Engineering Science
University College London
London, United Kingdom

Stuart Newstead is a digital expert. He's been an expert for so long that he remembers when digital was called telecoms. He founded independent consultancy Ellare in 2002 and provides strategic and analytical advice to clients in over 20 countries. Clients include BBC, gigaclear, Squirro, Nominet, mi-Pay, Blackberry, BT, Oxfam, Zzoomm, Esprit Capital, Elliot advisers, Albion Ventures, plus a number of startups. Stuart was a Vice President at O2 until 2002 and, prior to that, a General Manager at BT.
E-mail: stuart.newstead@ellare.net
Ellare
Oxford, United Kingdom

Stuart Revell is Managing Director of RTACS Ltd. An industry expert, covering both commercial and technical disciplines with over 35 years experience in the technology industry working for leading companies including Freescale Semiconductors, Motorola and Compare Industrial. Starting his career as an electronics design engineer, Stuart has experience in the semiconductors, electronics, hardware and software, ranging from industrial control systems through to complex ICT solutions, mobile and consumer solutions and has managed international teams covering both commercial and technical disciplines.
E-mail: stuart.revell@rtacs.com
RTA Communication Systems Ltd
London, United Kingdom

Paper

5G Is Out There: How to Ride the Market Storm and Thrive

Edward Smith[1] and Mauro Ugolini[2]

[1] *Wokingham U3A, Wokingham, United Kingdom*
[2] *Department of Engineering, Roma Tre University, Rome, Italy*

https://doi.org/10.26636/jtit.2021.151521

Abstract—We examine the changes in the commercial model adopted by the mobile ICT industry, due to the advent of 5G technology. This includes consideration of the challenges involved in rolling out a new infrastructure, which new markets this is likely to open up and how this affects partnering decisions. We show that as technology horizons expand, their degree of overlap increases and previously complimentary technologies may compete with each other, impacting the size of the addressable market. It is expected that 5G, whilst supporting its existing markets, will offer additional machine to machine, low latency and highly reliable services. We consider the synergies and the drivers for adoption for the wider 5G propositions and consider the impetus for more bandwidth and services and how the new technology impacts selling approaches. We identify the risks and uncertainties for the network providers and the likely requirements for a sustainable 5G business model and will describe our view of the steps necessary for a 5G successful outcome.

Keywords—5G, eMBB, mMTC, URLLC.

1. More than an Evolution

In this paper we will cover the changes in the commercial model adopted by the mobile ICT industry, due to the advent of 5G technology and the impacts it is likely to engender.

We will examine the challenges offered in rolling out a new infrastructure, which new markets this is likely to open up and how this affects partnering decisions. We will also show that as technology capabilities expand, their degree of overlap increases and previously complimentary technologies can become competition for each other. Competition is likely to impact the size of the market.

We suggest that 5G will not be just an evolution of 4G services, supporting its existing markets. The synergies between 5G and the application space will be examined and the drivers for adoption identified. We will ask where the impetus for more bandwidth might come from and how this will change the way mobile capacity is sold. We will investigate each of the following main propositions, the last two of which require strong partner collaboration, and evaluate the prognosis for their commercial success:

- eMBB, extended massive broadband, representing the evolutionary path from the existing proposition,
- mMTC, massive machine type communications, delivering high density radio services for machine to machine (M2M) applications,
- URLLC, ultrareliable low latency communications, fostering improvements in latency and reliability to allow mission critical applications.

We will examine where the risks and uncertainties for the networks providers are and the likely requirements for a sustainable 5G business model. We will conclude by presenting our view of the steps necessary for a 5G successful outcome.

2. The Impact

5G is expected to provide a user data rate of up to 100 Mbps in the downlink direction, an increase in spectral efficiency, a latency of 1–4 ms, a connection density of up to 1,000,000 devices per km^2 and improvements in availability, reliability and energy utilization [1]. An increase in capacity of 5G over LTE by a factor of between 1000 and 5000 fold is anticipated [2], requiring, as a consequence, an increase in backbone capacity [3].

4G remains the fastest growing network, reaching 5.6 billion users worldwide by 2022 and consuming 79% of all mobile data traffic by 2021. The corresponding figures for 5G are 400 million subscribers and 1.5% of traffic [4], [5]. 5G is expected to generate opportunities, with a global economic impact of 12.3 trillion USD [6], even if the economics and ability to monetize the investment remain unclear [2].

A low to mid-band 5G network has similar form and cost to an equally dense 4G implementation. The high bandwidth deployments delivering high performance rely on networks that have considerably more node density than LTE [7], [8]. Unless costs fall dramatically, this will challenge most business cases [7].

Investments in 5G can be deferred by building on existing LTE infrastructures. For example, network operator

EE's initial UK implementation[1] uses 4G to provide the WAN, signaling and control plane, and the 5G user plane for data [9]. Deployment of 5G could double network costs, with a 60% increase in capital expenditures, for standalone deployments. Simulation of a 5G build-out demonstrated that around 20% more macro-sites would be needed and an increase in new small cells equivalent to 100–150% of existing macro-cells may be required [10].

Network sharing can reduce total cost of ownership by 30% and the cost of small cell deployments by 50%, as shown in Fig. 1, while improving network quality, minimizing the impact of urban works and reducing the level of street furniture [10].

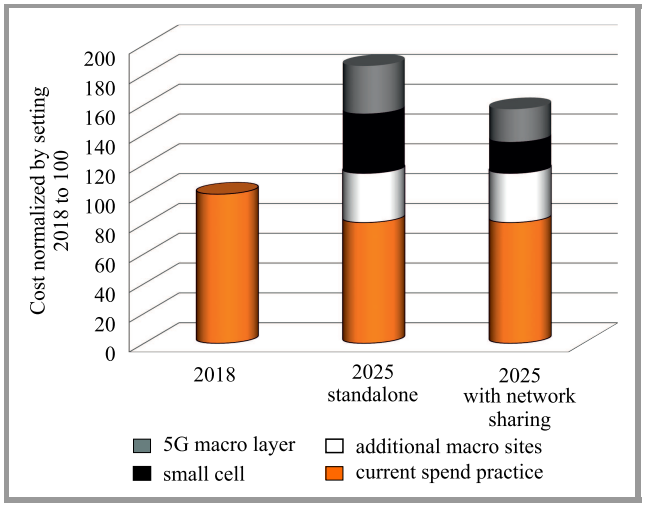


Fig. 1. The impact of network sharing on deployment costs [10].

Software defined networks, which should reduce operational costs, and network function virtualization, impacting capital costs, are often cited as a means of reducing deployment costs. In the short term, the contribution from these is likely to be significantly less than network sharing.

Network slicing is a virtualization technique offering significant advantages for mobile network operators (MNOs), allowing them to operate separate virtual infrastructures, with their own service characteristics across a common physical network. This allows say the operation of a dedicated network slice for the emergency services or to allow a mobile virtual network operator (MVNO) to occupy its own network slice. However, there seems to be little quantitative information showing how much network slicing will contribute to the overall 5G business case.

5G will focus initially on eMBB and as the service matures, URLLC and mMTC will be added [9]. So far, we have considered 5G in terms of its capabilities and managing the cost of its introduction. We now progress to examine the role of competition between technologies.

[1]In January 2020, following a security review, the UK government limited the involvement of Huawei to 35% of non-core 5G infrastructure. In May, the US prevented Huawei from incorporating American technology in their solutions, undermining Huawei's supply chain and creating delivery uncertainty. A subsequent UK government review has prohibited UK telecommunications providers from buying new infrastructure from Huawei and requiring the removal of their equipment from MNO's 5G infrastructures by 2027.

3. Competition Amongst Communication Technologies

Technology introduction can be commercially hazardous. For example, ISDN (Integrated Services Digital Network) – a network supporting digital voice and data services, enjoyed limited market success, before being displaced by broadband services for data and more recently SIP (Session Initiation Protocol) – a method of setting up associations across IP, mainly used for voice services. Similarly SMDS (Switched Multimegabit Data Service) and ATM (Asynchronous Transfer Mode) – a high speed connection orientated service, had short commercial lives, yielding the high speed data space to MPLS (Multiprotocol Label Switching) – a high speed service optimized for IP.

Even where technologies have been successful, longevity brings its own issues and many mobile applications still widely exploit GPRS (General Packet Radio Services) [11], making it difficult to eliminate the costs associated with the old technology, to maximize the revenues accruing the new technology base.

A number of use cases posited for 5G demand high bandwidth, but not necessarily mobility, making ultra broadband services a viable option [12]. Further, wireless LAN capabilities have supported the growth in use of mobile devices, but can divert traffic away from the mobile network [13]. New Wi-Fi standards (802.11ax, also known as Wi-Fi 6) support dense M2M environments, improving the average throughput by a factor of at least four [14]. In the mMTC area, wireless networks using low power consumption and unlicensed spectrum also provide competition.

Having discussed a number of cases where the success of a technology has been impacted by the emergence of alternatives and the capabilities of existing technologies, which are more established, we move onto consider the area of the market for massive broadband capabilities.

4. The Market For Radio Broadband (EMBB)

eMBB represents an evolution of the existing mobile market place, where we begin our analysis of its commercial position. Figure 2 shows the existing usage for mobile services, with the growth in traffic being driven by video.

How much of the new value generated by 5G will flow to mobile network operators (MNOs) is uncertain. As LTE evolved, MNOs experienced rapid traffic growth from over-the-top providers video and declining average revenue per user due to intense price competition. Ultimately, wireless carriers have focused on lower cost per bit to fend off the risk of falling behind competitors [15].

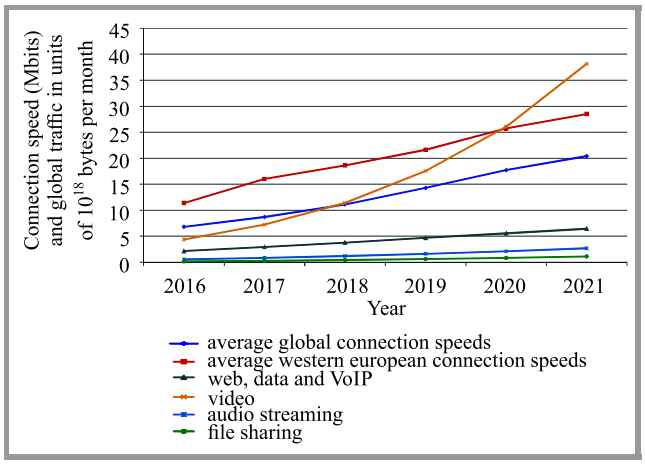


Fig. 2. Expected connection speeds and usage for mobile networks going forward [5].

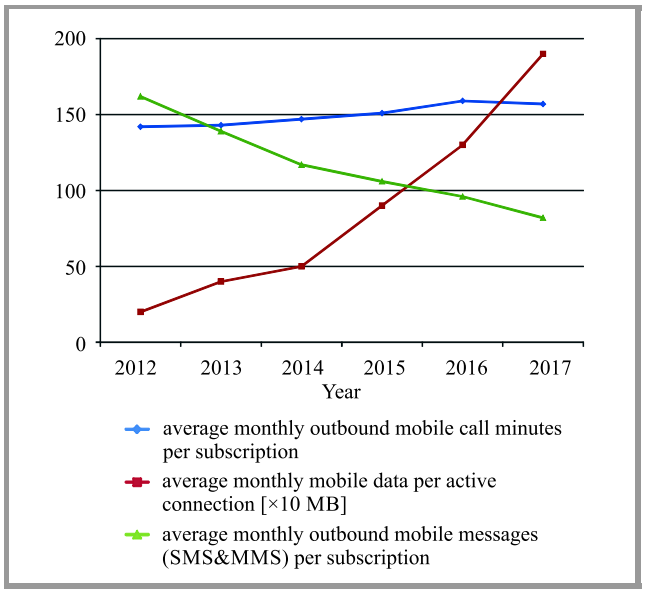


Fig. 3. UK mobile network traffic volumes [16].

As an example, Ofcom report that in the UK the average monthly data use per residential fixed broadband connection and per mobile SIM increased in 2017 by over 40%, to 190 GB and 1.9 GB respectively [16]. The Ofcom derived figures for mobile network traffic volumes as a function of time are shown in Fig. 3. Figure 4 clearly shows that whilst data growth is expected to support revenue increase, revenue accumulation lags traffic growth, and this situation, together with changes in tariffing strategies from prepaid approaches to pay-monthly services, are characteristic of a mature market, where costs are imperative and much differentiation comes from creative tariffing. It is expected that this pattern of revenue growth is typical of most EU countries despite the rapid growth in subscribers and higher data volumes seen across the world [15].

In addition, we see that in the UK 69% people connected to Wi-Fi rather than cellular networks from their mobile devices [17]. During the peak commuter time of 5 and 6 pm

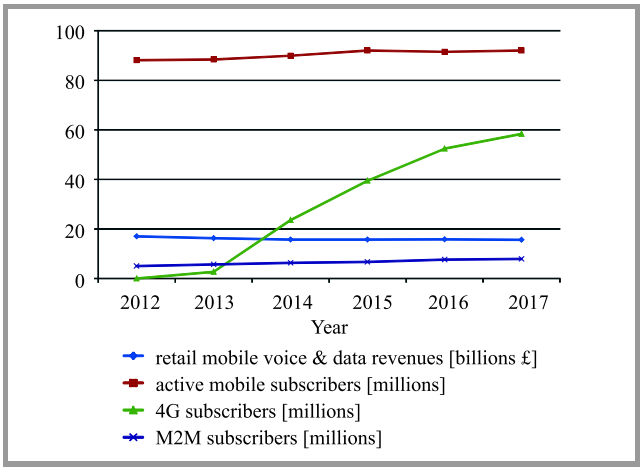


Fig. 4. Mobile subscriber and revenue details from the UK [16].

data volumes on cellular networks peaked, while Wi-Fi use was higher between 6 and 10 pm, transferring more data over Wi-Fi than on mobile networks [17]. Actually, this circumstance is not just typical of the UK, because these observations are consistent with global trends reported elsewhere [5].

After this exam of the proposition which forms the base case for 5G, we move on to consider the segment that serves machine to machine communications.

5. The Market for Machine Communications (MMTC)

mMTC offers density of the order of 1000000 nodes per km^2 [1], with an evolution path from 4G services. Cisco predicts a rapid rise in the number of M2M connections, but only a third of them connected to a mobile network. Numbers of mobile network connected and low power wide area (LPWA) M2M devices appear to be growing rapidly. Cisco believes that connected home applications are expected to have the largest share by 2023, but the connected car will be the fastest growing application [14]. Gartner, on the other hand, believes that initially outdoor surveillance cameras will be the largest M2M market, but will be surpassed by connected cars in 2023 [18].

Figure 4 shows the take-up of M2M subscriptions in the UK, as an example, and Table 1 gives the picture on a worldwide basis.

Worldwide, smart meters provide an example of an M2M network where instrumentation, networks and systems blend to provide a solution. A system integrator leads on the contract and coordinates delivery, but MNOs provide only some communications elements of the solution. They take less of the risk, but the system integrator orchestrates the opportunity to add value in terms of additional services. Most smart meter solutions use mixed technologies, including 2G mobile, multipoint radio and power line technology, with systems integration being performed by

a power distribution company in Italy and an independent integrator in the UK. MNOs therefore need to modify their view of the market and their commercial approach, needing different commercial skills and approaches to partnership [11].

Table 1
M2M projected connections [14]

Technology (total values shown)	Billions connections in 2018	Billions connections in 2023	CAGR
Networked M2M devices	5.90	14.70	19%
M2M devices connected to a mobile network	1.14	4.45	31.0%
Smartphones connected to a mobile network	4.05	5.37	5.8%
LWPA connections	0.22	1.88	530%
Mobile connections	5.10	5.70	2.0%

Low power wide area networks (LPWA) supports M2M applications requiring low bandwidth, wide geographic coverage and low power consumption, module and connectivity costs, addressing needs that cellular networks cannot meet by themselves. LPWA will grow from 2.5% of M2M connections in 2018 to 13% by 2023, from 223 million to 1.9 billion globally [14] and is expected to compete fiercely with 5G.

LoRaWAN is a proprietary LPWA solution, uses unlicensed spectrum (868 MHz) and is limited to a transmission power of 25 mW or less. It permits the installation of a gateway at arbitrary points not governed by installation and regulations issues. LoRaWAN solutions have been used to monitor traffic flows, view pedestrian crossings and, as an example, Thames Water in the UK has considered them for their smart meters. Such networks are suitable only for small data packets, unsuitable for mission critical applications and support devices whose batteries need to last many years.

5G technologies can support a range of smart city scenarios as demonstrated by the 2019 FITCE congress [19], which included papers on autonomous vehicles and virtual reality (VR) enhanced tourism, fiber to the antenna, fiber based and 5G costs models for smart city applications and the costs and benefits of deploying 5G enabled light poles. Work has also been performed to investigate the potential, economically sustainable, business models for 5G and network slicing. Although smart cities are a key application, local authorities are unsure of the requirements and their supporting business cases, creating a barrier to deployment [20].

Whilst the machine to machine segment is an evolution of the 4G standards, ultra-reliable low latency communication (URLLC) is not, and represents a number of challenges which are covered in the next section of the paper.

6. The Market for Low Latency Communications (URLLC)

The expected latency delay for eMBB is 10 ms [3], although the target is set at 4 ms [1]. To meet the sub 1 ms requirement [1] for URLLC, performance targets of 0.3 ms device processing, 0.1 ms delay each way across the radio network and 0.5 ms within the core network are needed [21], as illustrated in Fig. 5.

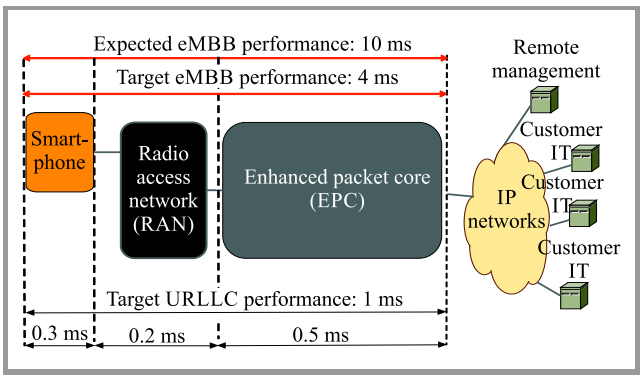


Fig. 5. URLLC latency requirements.

This requirement for coverage of a small area and the use of a high frequency radio interface is commercially challenging and a private network or hybrid solution may be more effective than the MNO network. For example, surgical applications are unlikely to be remote, but take the form of microsurgery within a hospital, using radio to replace delicate cable connections [22]. Industrial use is another use case, where private 5G networks connecting to the public networks could be used. Such use cases may use multiple providers [14].

URLLC will support augmented reality (AR), virtual reality (VR), edge computing and autonomous vehicles [23], but these technologies have not so far achieved significant market penetration. It is believed that immersive multimedia environments have high market potential, with gaming being a key focus. Edge computing has yet to be defined in detail [22] and in some cases MVNO may add operational complexity by requesting that devices under their control are included in an MNO's infrastructure.

Connected cars may not need 5G, but will be able to use its enhancements and autonomous vehicles may require it [24].

We now move onto make an overview assessment of the 5G market place.

7. Market Analysis

The Braudel rule states that "freedom becomes value when it changes the limits of the possible in the structures of everyday life" [25]. We now look at the three key application areas (eMBB, mMTC, URLLC) in turn and identify

which factors allow that application to meet the requirements of Braudel's rule.
The growth in mobile traffic appears to have been driven, historically, by a killer application, as shown in Table 2.

Table 2
The mobile generations and their associated killer applications

Generation	Killer application	Year launched
Analog	Mobile business voice	1986
2G	SMS and consumer mobility	1992
3G	Smartphone and data	2003
4G	Video and OTT applications	2012
5G	?	2019

Figure 6 shows the growth in smartphone sales which begins when 3G provided the highest data speeds and continued to accelerate with the introduction of 4G, levelling off before 4G became the dominant mobile technology. This Cisco estimate to be between 2019 and 2020 [5].

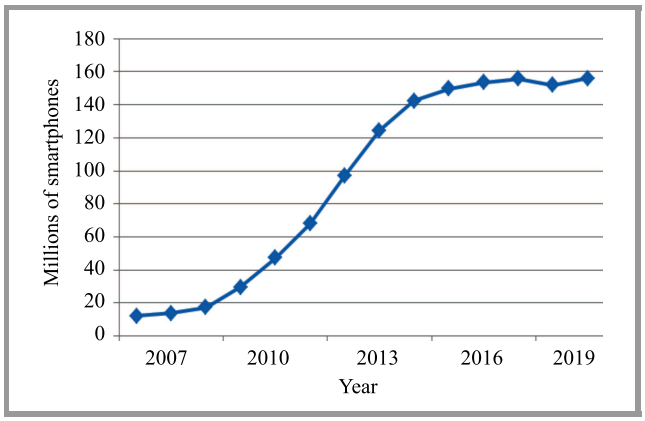


Fig. 6. The saturating smartphone market [26].

As a further example of a situation that is common in other European countries, plotting Ofcom produced figures for data volumes in the UK. This is expected to be typical for the EU) [16] against the revenues of Facebook [27] and Netflix [28]. We see, as shown in Fig. 7, a linear relationship consistent with a link between growth in mobile data and the development of video based, over-the-top provider (OTT), services.
In a recent survey McKinsey found that the majority of chief technology officers (CTO) see enhanced mobile broadband (eMBB) and the Internet of Things (IoT) as the most significant applications for 5G. The uncertain economics of 5G are encouraging the evaluation of alternative business models, which will need investment in operational support systems (OSS) and business support systems (BSS), to provide the required service wrap [29].
The development of business cases, operations and maintenance strategies and commercial strategies, significantly lags strategies for pilots and technology. The business case

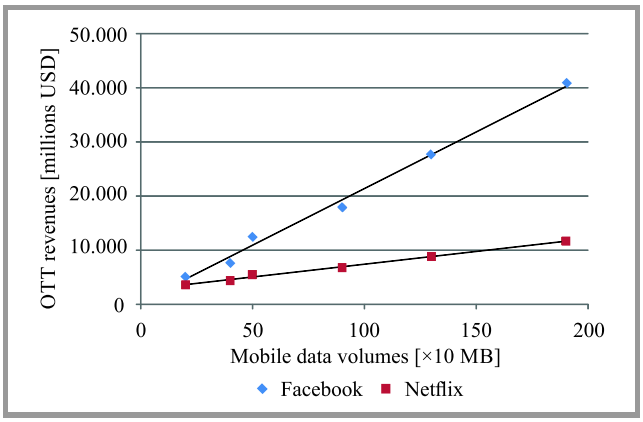


Fig. 7. The relationship between mobile data volumes and new applications media revenues [16], [27], [28].

was the top concern of all European CTOs, contrasting with 11% in North America.
Operators primarily see 5G as an opportunity to establish network leadership. M2M appears initially not to be a core objective for 5G, suggesting that the existing M2M capability is considered sufficient for most use cases [29]. Although URLLC has significant potential, it is not yet a major area of investment and its applicability is open to question. European feedback is skeptical about the new use cases.
Delivering the new use cases is critical dependent on partnership, but whilst MNOs are proficient collaborating on standards, they have been less effective in collaborating with third parties. Investment in OSS and BSS is also a concern. However, much of the value of M2M will be generated from advanced network capabilities, which is the MNO's core competence [29].
Exploiting Keen and Mackintosh's interpretation[2] of Braudel's view of product adoption yields the concepts presented in Table 3.

Table 3
Examining the maturity of 5G technology in terms of Keen and Mackintosh's matrix

	Promise	Chaos and clutter	Business mainstream
Freedom	Monitor closely	Experiment	Make strategic
Convenience	Ignore	Avoid	Incorporate selectively
Feature	Distrust	Retreat	Buy at right price
No clear target	Do not invest	Short sell stock	Wish you had bought it

eMBB is a mainstream use case, but delivers much more bandwidth than appears to be required. An average mobile broadband speed of 43.9 Mbps is envisioned for 2023 [14],

[2]Keen and Mackintosh matrix is a formal, space-time conceptualization that is used to delineate the impact of the analyzed technology, and a taxonomy of the technology-based applications.

which is sufficient for ultra high-definition (UHD) video. Yet 5G speeds will be 13 times higher than this by 2023, with an average bandwidth of 575 Mbps [5]. The challenge for the product management function within an MNO is to identify what will take this from "feature" to "freedom" as social media/video did for 4G.

The smart meter use case, described earlier, illustrates some of the pitfalls of the mMTC proposition, which splits between high bandwidth requirements, e.g. closed circuit television (CCTV) and connected cars and low bandwidth and low power applications (e.g. remote sensors). Some of the low power applications generate freedoms but have attracted competition from other technologies. Connected cars appear to be an interesting use-case, but it is unclear why this need cannot be satisfied by eMBB. To take advantage of this, as a revenue opportunity, MNOs need to work closely with the motor industry.

On present evidence it would seem that the high bandwidth case is at the convenience level and low bandwidth at low power is at the freedom level, but with well-established and attractive competition. Many applications require integration into a wider solution and therefore need bringing into line with a solution's business.

URLLC is the proposition at the earliest stage of development, with some clear technical challenges. This coupled with the role of third parties and private networks in its delivery make it likely that this will be a complex and costly proposition to deliver. Most use cases look to be convenience or feature stage. Automated vehicles look the nearest to a freedom, but existing developments of in-vehicle sensors and intelligence, coupled with information from other vehicles and roadside units, should be capable of being satisfied with a response time of 10 ms as opposed to 1 ms.

The games industry is an innovative market place, which is said to need low latency. The migration is towards mobile and app-based gaming platforms and growth of access using mobile technology make it demanding of a mobile platform. Multiplatform games are the most likely to use VR, which will demand high bandwidth, particularly in its high definition form. Multiplayer games have strict latency requirements, with any latency over 75 ms or so causing players and actions to fall out of synchronization. However, a 75 ms response time is within current 4G capabilities and those of eMBB's. It may be a freedom for 5G but it does not seem to require URLCC, which is the most uncertain business area and needs the most development.

8. Conclusions

We have examined the challenges offered in rolling out a new infrastructure, which new markets are likely to open and how this affects partnering decisions, showing that as technology capabilities expand their degree of overlap increases. However, previously complimentary technologies can become competition for each other, impacting the size of the market.

Developing 5G successor technologies is likely to be an expensive business, particularly given the increased base station densities that the new technology demands. We reassert the view made by others that network infrastructure sharing provides a way of mitigating some of the costs, but other avenues such as automation need to be explored.

We believe that the demand for any product is fueled by customer needs and wants and that there are several technologies that meet many of the needs addressed by 5G. The hazardous nature of the route to commercial success is outlined with descriptions of several casualties of the invisible hand of the market.

The massive broadband market is expected to be the provider of the lion's share of mobile revenues, but this market has become price sensitive and much traffic is offloaded onto broadband Wi-Fi connections. Whilst the detail of this is expressed in terms of the UK market, we expect the principles to extend to the rest of Europe.

The evolution of low latency and machine to machine services was examined, in particular considering the low power segment of the latter. This is an area where mobile operators face existing, established competition, with a strong record of having the necessary systems integration skills. The market for low latency services appears to be harder to define and refers to supporting technologies which in themselves are not well established and at present niche applications.

We attempted to pull the analysis together by distilling the needs and wants of the market place into Braudel's rule, focusing on the impact of the technology on the limits of everyday life. The applications, spurring the growth across the generations of mobile technology, were unclear at the time of technology launch: SMS for 2G, the smartphone and data for 3G and video and OTT applications for 4G.

The pressure is on mobile operators to embrace the new opportunity and not be left behind technologically, but what will drive the adoption by end customers? As before this is not obvious and whilst the new applications identified by mMTC and URLLC may make a difference, there is no clear indication that this will be sufficient. As McKinsey have shown, in Europe at least, business cases are incomplete and the biggest worry for operators. A better understanding of the likely development of the market is vital and something it is in the interest of academia, regulators, adjacent sectors and operators to acquire. Whilst it's true that a technologically driven white knight has ridden to the rescue in the past, it has always squeezed the profits of the operators in return.

In short, MNO need to understand the priority applications to focus on. Approaches involving the building of utility models, partnership analysis and optimization may be helpful in analyzing through the complexity of the market. To summarize, we believe that 5G increases the capability overlap with terrestrial services and requires further non-consumer markets to be addressed to generate a sus-

tainable business model. Both factors require a change in thinking from the MNO industry, if a successful outcome is to be achieved.

References

[1] P. McCarthy-Ward, A. Valdar, S. Newstead, and S. Revell, "5G – new business opportunities", *J. of the Institute of Telecommun. Profess.*, vol. 13, no. 4, pp. 19–25, 2019 [Online]. Available: http://www.fitce2020.pl/5G-New%20Business%20Opportunities%20FITCE%202020%20pdf.pdf

[2] A. Sutton and R. Tafazolli, "5G The future of mobile communications, *J. of the Institute of Telecommun. Profess.*, vol. 9, no. 1, pp. 10–16, 2015.

[3] A. Sutton, "Deployment of the EE 5G network", *J. of the Institute of Telecommun. Profess.*, vol. 13, no. 4, pp. 27–31. 2019 [Online]. Available: https://www.academia.edu/41625209/Design_and_Deployment_of_the_EE_5G_Network

[4] "5G forecast report from Ovum. Global 5G Subscription Forecast 2019–2022", Telecom Statistics, Telecomlead, 23 March 2018. [Online]. Available: http://www.telecomlead.com/telecomstatistics/5g-forecast-report-from-ovum-83098

[5] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, White Paper", Czech Marketplace, 2017. [Online]. Available: http://www.czechmarketplace.cz/news/cisco-visual-networking-index-global-mobile-data-traffic-forecast-update-2016-2021-white-paper

[6] A. Andonian, A. Karlsson, A. Axel, and K. Nonaka, "Japan at a Crossroads: The 4G to 5G (R)evolution", McKinsey, Jan. 2018. [Online] Available at: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/japan-at-a-crossroads-the-4g-to-5g-revolution

[7] M. Collins, A. Das, A. Menard, and D. Patel, "Are you ready for 5G?", McKinsey, 2018 [Online]. Available: https://www.mckinsey.com/industries/telecommunications/our-insights/are-you-ready-for-5g

[8] S. Temple, "5G and demand alternative network principles", *J. of the Institute of Telecommun. Profess.*, vol. 9, no. 1, pp. 18–21, 2015.

[9] A. Sutton, "5G rollout, performance and next steps", ITP Seminar – The Reality of 5G, Nov. 2019.

[10] F. Gripink, A. Ménard, H. Siguardsson, and N. Vucevic, "Network sharing and 5G: A turning point for lone riders", McKinsey, Feb. 2018 [Online]. Available: https://www.mckinsey.com/industries/telecommunications/our-insights/network-sharing-and-5g-a-turning-point-for-lone-riders

[11] E. A. Smith and M. Ugolini, "Rolling out smart meters in Europe", *J. of the Institute of Telecommun. Profess.*, vol. 11, no. 1, pp. 19–24, 2017.

[12] M. Ugolini and E. Smith, "Il mercato della banda ultralarga in Europa", *Rivista AEIT*, vol. 104, no. 3/4, pp. 12–21, 2019 [Online]. Available: https://www.aeit.it/aeit/edicola/aeit/aeit2019/aeit2019_02_cisa/aeit2019_02_riv.pdf [in Italian]

[13] M. Ugolini and E. Smith, "The market potential of 5G: An Anglo-Italian view", *Rivista AEIT*, vol. 104, no. 1/2, pp. 48–55, 2019 [Online]. Available: https://www.aeit.it/aeit/edicola/aeit/aeit2019/aeit2019_01_cisa/aeit2019_01_riv.pdf [in Italian]

[14] Cisco Annual Internet Report (2018–2023) [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[15] D. Littmann, P. Wilson, C. Wigginton, B. Haan, and J. Fritz, "5G: The chance to lead for a decade", Report, Deloitte Consulting LLP, 2018 [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-tmt-5g-deployment-imperative.pdf

[16] CMR Communications Market Report, Ofcom, 2 Aug. 2018 [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0022/117256/CMR-2018-narrative-report.pdf

[17] Mobile matters, Researching people's experience of using Android mobile services, Ofcom, 10 October 2019 [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0038/169769/mobile-matters-report.pdf

[18] A. Weissberger, "Gartner: 5G IoT endpoints to triple between 2020 and 2021; Surveillance cameras to be largest market over next 3 years", IEEE ComSoc Technology Blog, 17 Oct. 2019 [Online]. Available: https://techblog.comsoc.org/2019/10/17/gartner-5g-iot-endpoints-to-triple-between-2020-and-2021-surveillance-cameras-to-be-largest-market-over-next-3-years/

[19] Collection of congress papers "2019 CTTE-FITCE: Smart Cities & Information and Communication Technology (CTTE-FITCE)" [Online]. Available: https://ieeexplore.ieee.org/xpl/conhome/8890754/proceeding

[20] M. Ugolini and E. A. Smith, "Closer look to the future of smart cities", in *Proc. of 2019 CTTE-FITCE: Smart Cities & Inform. and Commun. Technol.*, Ghent, Belgium, 2019 (DOI: 10.1109/CTTE-FITCE.2019.8894827).

[21] "The Tactile Internet", ITU-T Technology Watch Report, August 2014 [Online]. Available: https://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000230001PDFE.pdf

[22] ITP Seminar – The Reality of 5G, ITP Seminar Round Table Debate, London, November 2019.

[23] S. Newstead, "5G or not 5G? That is the question", *J. of the Institute of Telecommun. Profess.*, vol. 13, no. 4, pp. 9–16, 2019.

[24] T. Degrande, F. Vannieuwenborg, S. Verbrugge, and D. Colle, "Adoption of cooperative intelligent transport systems in Flemish passenger cars: A review of European policy options", in *CTTE-FITCE: Smart Cities & Information and Communication Technology*, Ghent, Belgium, 2019 (DOI: 10.1109/CTTE-FITCE.2019.8894817).

[25] P. G. W. Keen and R. Mackintosh, *The Freedom Economy*. Berkeley: McGraw-Hill, 2001, pp. 30–33 (ISBN: 9780072133677).

[26] S. O'Dea, Global smartphones sales to end-users 2007–2020, 28 Feb. 2020 [Online]. Available: https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007

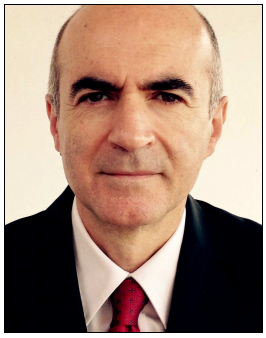
[27] Facebook Revenue 2009–2020 | FB, Macrotrends [Online]. Available: https://www.macrotrends.net/stocks/charts/FB/facebook/revenue

[28] Netflix Revenue 2006–2020 | NFLX, Macrotrends [Online]. Available: https://www.macrotrends.net/stocks/charts/NFLX/netflix/revenue

[29] F. Grijpink, T. Härlin, H. Lung, and A. Ménard, "Cutting through the 5G hype: Survey shows telcos' nuanced views", McKinsey Insights, Feb. 2019 [Online]. Available: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/cutting-through-the-5g-hype-survey-shows-telcos-nuanced-views

Edward Smith retired from BT in May 2016, and now focuses on research in network modeling, telecommunications history and the development of new ICT propositions. He partners with researchers in the UK and Italy on a range of projects. In his professional life he has worked with major clients across a wide range of technologies, having first developed IT skills with United Biscuits handling real time process control systems and Elida Gibbs, developing transaction processing middleware and data communications solutions. He holds B.Sc. and Ph.D. degrees from the University of Leicester and a Post Graduate Certificate in Commercial Management from De

Montfort University. He is a Fellow of the British Computer Society, a Chartered Information Technology Practitioner, a Chartered Engineer, ISEB accredited consultant and a TOGAF accredited Enterprise Architect.
https://orcid.org/0000-0003-0896-0455
E-mail: edward.a.smith@btinternet.com
Wokingham U3A
Wokingham, United Kingdom



Mauro Ugolini is an expert in business innovation and in process digitalization, who has held several executive positions in global ICT companies. A pioneer of the Mobile Internet development, he has launched several hi-tech startups. He graduated in Physics (cum laude) at Sapienza University of Rome, gained an Executive MBA from Alma Graduate School at University of Bologna and earned a Ph.D. in Telecommunication Engineering at Roma Tre University. He is a consultant for business development and innovation lecturer for Italy's Ministry of Education, University and Research (MIUR). He is also a chartered Master in Economics, General Management and Market Management and a recognized author of works in the fields of Information Technology, Management and Communication. As a courtesy professor of the Engineering Department at Roma Tre University, he is focused on the research for the development of the Networked Society, pursued with the aim of increasing the responsibility and sustainability levels in the current global economy, through the analysis and the study of the properties of the so called networked digital ecosystems.
E-mail: mauro.ugolini@uniroma3.it
Department of Engineering
Roma Tre University
Rome, Italy

Paper

C-V2X Communications for the Support of a Green Light Optimized Speed Advisory (GLOSA) Use Case

Ioannis P. Chochliouros[1], Anastasia S. Spiliopoulou[1], Alexandros Kostopoulos[1], George Agapiou[1], Pavlos Lazaridis[2], Zaharias Zaharis[3], Tao Chen[4], Athanassios Dardamanis[5], Michail-Alexandros Kourtis[6], Marinos Agapiou[7], Uwe Herzog[8], and Latif Ladid[9]

[1]Hellenic Telecommunications Organization, Athens, Greece
[2]University of Huddersfield, Huddersfield, United Kingdom [3]Aristotle University of Thessaloniki, Thessaloniki, Greece
[4]VTT Technical Research Centre of Finland, Espoo, Finland [5]SmartNet, Athens, Greece
[6]ORION Innovations Private Company, Athens, Greece [7]Independent Consultant, Athens, Greece
[8]Eurescom, Germany [9]University of Luxembourg/SnT, Luxembourg

https://doi.org/10.26636/jtit.2021.152321

Abstract—Rapid expansion of 5G affects a number of sectors, including vehicular communications relying on cooperative intelligent transportation systems (C-ITS). More specifically, in the context of the Internet of Vehicles (IoV), a particular emphasis is placed on modern cellular V2X (C-V2X) technologies aiming to further improve road safety. This work originates from the detailed scope of the ongoing 5G-DRIVE research project promoting cooperation between the EU and China, with the aim of demonstrating IoV services that rely on vehicle-to-infrastructure (V2I) communications. With the C-V2X approach serving as a point of departure, we analyze and describe a specific green light-optimized speed advisory (GLOSA) use case, for which we provide a detailed descriptive framework, a proposed architectural framework for trials, as well as specific KPIs for the joint assessment of trials between the EU and China. We also discuss the context for performance test procedures to be conducted as part of the intended trials. GLOSA provides end-users with short-term information on upcoming traffic light status to optimize traffic flows, help prevent speed limit violations, improve fuel efficiency, and reduce pollution.

Keywords—5G, cellular V2X, cooperative awareness messages, cooperative intersection, cooperative ITS, green light-optimized speed advisory, intelligent transportation systems, traffic efficiency, vehicle-to-everything.

1. Introduction

1.1. 5G as an Enabler for Modern ITS

The 5G technology used in mobile and wireless communications fosters the evolution of modern societies and economies. This strongly affects the way in which mobile communications and enormous amounts of data generated by various applications may transform our modern way of living. 5G supports connectivity requirements of a number of heterogeneous infrastructures, thus facilitating the use of convenient solutions designed by various market actors. This means that 5G contributes to the creation and operation of a new form of underlying network infrastructures offering increased agility and adaptability levels, and effectively supporting a variety of technologies. This, in turn, fosters development and economic growth, as 5G becomes capable of promoting innovative solutions that may be beneficial for the participating end-users (i.e. residents, corporations, state authorities, and many others) and, therefore, of improving our living standards.

In particular, as 5G promotes also the inclusion of cognitive and automation features, as well as the latest advances in artificial intelligence (AI), platforms are established and operated that support the design and the offering of new services, such as mission-critical (MC) communications and services based on the Internet of Things (IoT). Such facilities offer enhanced capacity and performance for all communication methods, serving various sectors [1]. With that taken into consideration and in view of the wide-scale and/or global digital transformation taking place, unique opportunities may be offered by offering new processes and unprecedented opportunities for the benefit of the industrial sector [2]. 5G is not a simple transition to a next generation of mobile communications. Instead, it becomes a true enabler of progress, increasing the level of performance, reducing latency and boosting network/service reliability, even for MC applications. Furthermore, 5G is also a promoter of market growth, as it supports a great number of the so-called vertical industries that will play an important part in the new reality and will become active participants of the new digital era [3].

Introduction and inclusion of cooperative intelligent transportation systems (ITS) [4] is expected to play an important role in this global evolutionary process. ITS are perceived

to be a sector that has been evoking the largest amount of interest recently, attracting significant national and international investments aiming to promote novel solutions. As ITS are explicitly correlated with every-day activities, policy makers, the automotive industry, telecom providers, transportation experts and other "actors" have shown significant interest in that particular area. The aim of their actions has been to offer a portfolio of new services and, most importantly, to deal with existing obstacles, such as traffic congestion, road safety, transport efficiency and environmental conservation. All this means that further growth has been thwarted. ITS may be considered as effectively including all types of transportation infrastructures, such as streets, highways, bridges, tunnels, railways, ports and airports, as well as vehicles (not only cars, but also buses, lorries, trains, aircraft and waterborne vessels).

The fact that ITS affect the majority of various modes of transport, simultaneously fostering multi-modality, is an important factor too. The range of the services offered on the market is immense and depends on the enabling technologies applicable that are relevant for the specific cases. ITS may include, for example, management enhancing services, such as those related to navigation, traffic signal control systems, container management systems, variable message signs and enforcement systems for monitoring applications (e.g. CCTV systems). However, the notion may also be extended to cover more advanced applications that integrate live data from other sources (such as parking guidance and information systems or weather information systems), working in more complex environments and being able to correlate data from various sources.

1.2. Challenges for Cooperative ITS in Vehicular Applications

Applications for connected vehicles, in a broad sense of the word, are an area that is attracting strong interest due to the potential progress, especially when the support offered by 5G-based innovative features is taken into consideration. This calls for a new strategic approach that will address previous concerns and/or measures, such as those applied by ETSI [5], as well as those developed occasionally in Europe and at other locations around the world. ITS support effective, combined and inclusive use of communication, computer and control tools to meet stricter system performance-related and operational requirements. This approach also takes into account additional requirements established in order to satisfy other priorities, such as environmental concerns, reduced energy consumption, better performance and level of service compliant with applicable criteria, enhanced security even in complex operational environments, and mobility, just to mention a few.

"G5" – not to be confused with 5G, is such a reliable standard for car-to-car communications. It is based on the evolution of the Wi-Fi standard (i.e. IEEE 802.11.p), being one of the iterations of the original IEEE 802.11 solution, aiming to provide wireless access in vehicular environments. G5 relies on data exchanged between vehicles moving at high speeds, as well as between vehicles and roadside infrastructure. The process is known as V2X communication and relies on the ITS-licensed band of 5.9 GHz (5.85–5.925 GHz). With the benefits offered by 5G taken into consideration and in order to ensure enhanced functionalities, the development of new services, occasionally based on the hybrid communication approach, should be expected.

Modern communications technologies and the related in-vehicle and on-the-road infrastructures have all brought about several advantages for drivers, the automotive industry as such and for other stakeholders, especially in the transportation and the emergency services ecosystems, providing better driving conditions and promoting higher security levels. Recently, research efforts have been focusing on cooperative ITS (C-ITS), where vehicles communicate with each other and/or with the infrastructure [6]. Thus, C-ITS goes a step further by intending to improve the quality and to increase the reliability of information available, such as data about the vehicles, their location and the environment in which they move. This results not only in improving the existing applications relied upon by road users, but also in developing new ones, thus leading to better efficiency of transport and to increased safety [7]. As a critical policy challenge for modern societies is to reduce the number of road accidents and to advance road safety and security, it is essential for vehicles to detect what is happening around them, predict what may potentially occur next and take both protective and proactive measures, accordingly.

1.3. Cellular V2X Communications Supporting C-ITS

In a broader context, we could consider V2X communication as being a variety of a wireless sensor-based system, allowing vehicles to share and exchange information with each other and with other parts of the infrastructure via dedicated communication channels. When attempting to compare this solution with "standard" sensors (such as radar, LIDAR, lasers, ultrasonic detectors, etc.) one needs to mention that the utilization of a V2X system offers extra benefits, as such an approach enable to obtain out-of-sight information, to test hidden threats and to expanding the driver's perception. In consequence, driving safety and efficiency are improved, and driving comfort is boosted by various driving automation systems [8].

The vehicle-to-everything (V2X) technology is the next big feature to evolve further the automotive and transportation industry [9]. Through various communication technologies, V2X allows a vehicle linking to other vehicles, pedestrians, road infrastructure, the Internet, and other entities in the transportation ecosystem. V2X also provides the infotainment experiences, and eventually supports transition towards thing in terms of autonomous driving. In [10], the authors forecast that, by the year 2022, there will be more than 125 million vehicles connected by means of various V2X technologies, offering an unprecedented level of safety, expanding the range of known transportation services and offering other novel advantages. Thus, it becomes

obvious that it is now the critical phase to test and deploy V2X technologies and infrastructures, especially by combining those processes with the deployment and testing of modern 5G infrastructures.

In recent years, different regions in the world have conducted intense V2X trials. Two distinct V2X technical paths are followed by the automotive industry. ETSI ITS-G5, an approach based on 802.11p technologies [11], is one of them. 3GPP LTE-V2X [12] that is rooted in 3GPP standards is the other. Different regions have their own preferences concerning the technologies. China opts for cellular V2X (C-V2X or LTE-V2X) as its national standard. In Europe, the debate is ongoing on how to adopt the technologies. The 5G Automotive Association (5GAA), the international association with the mission to promote C-V2X technologies, expects that the first commercial deployments of V2X will occur in China and Europe, while deployments in the US and other parts of Asia will follow closely. Considering the fact that the life cycle of road infrastructure equals usually 30 years, and that the life cycle of a car is 10–15 years, selection of V2X will be critical for the future evolution of technologies.

Due to compatibility reasons, it is crucial that the various regions cooperate to ensure harmonization of technologies. With this kept in mind, Europe and China have established cooperation on validation of the C-V2X technology through joint research and trials. Car manufacturers, road authorities, telecom vendors, and mobile operators team up to test key C-V2X technologies and use cases. The 5G-DRIVE EU-funded project [13] is one of the undertakings concerned with working on C-V2X trials with China. The aim is to compare the performance in joint V2X use cases, and to identify any potential interoperability problems.

The work is organized as follows. Section 1 serves as an introduction. Firstly, it identifies the important role of 5G for the promotion of modern ITS, then it focuses on the challenges for cooperative ITS in vehicular applications. Lastly, it discusses modern cellular V2X communications supporting the automotive industry in the 5G era. Section 2 presents a wider scope of the 5G-DRIVE project, aiming to promote trials and tests between the EU and China for two scenarios, one of which is relevant to V2X communications, especially those relying on C-V2X technologies. Then, Section 3 presents a more detailed framework for C-V2X communications, initially by discussing recent evolutionary options stemming from 5G growth, and then by introducing the selected green light optimized speed advisory (GLOSA) use case. More specifically, we assess the innovative features of this important use case which offers major benefits for the market and we propose a dedicated novel architecture within the 5G-DRIVE-specific context, based on C-V2X, being one of the novelties introduced by this work to international literature. Then, we propose several KPIs for the assessment of trials and the scope of the intended performance test procedures. Section 4 completes the work by offering general overview of the scheduled trials intended to focus upon compliance tests performed in accordance with the ETSI EN 302 571 standard. Based on some early findings, we also assess trials performed in the EU and China.

2. 5G-DRIVE Project Framework

2.1. Trial and Experiment Sites in the EU

The 5G-DRIVE project [13] is an active element of the well-established H2020 ICT-22-2018 Call (EU China 5G Collaboration) scheme, aiming to ensure collaboration between the EU and China in order to work out solutions guaranteeing effective use of 5G technologies and to handle spectrum-related challenges before the expected wide scale roll-out of 5G. Figure 1 shows the fundamental 5G-DRIVE concept that focuses on the three illustrated core streams. Figure 1 depicts also the expected conceptual flow: from research, to adaptation to existing test-beds, to commercial test-bed deployments, to real-world trials of 5G radio access networks (RAN) and of the target 5G networks. 5G-DRIVE supports collaboration between solid research competence, commercial grade test-beds and some of the stakeholders who will eventually become major customers of the 5G systems and the proposed applications.

The 5G-DRIVE project incorporates extensive 5G test-bed installations provided by three research organizations – members of the consortium: the University of Surrey (UoS), the VTT Technical Research Centre of Finland (VTT), and the Joint Research Centre of the European Commission (JRC) [14]–[15]. While all three test-beds are equipped with commercial grade hardware to ensure reliability of the obtained results, each one has been arranged to serve a dedicated purpose. The Surrey test-bed supports capacity provision in very dense deployments over a 4 km^2 area. The VTT's Espoo test-bed validates the use of slicing and V2X. The JRC facility in Ispra, Italy, allows the testing of new V2X technologies in any part of the network in a fully-controlled environment. All test-beds allow for gradual introduction and testing of new equipment, as well as for the inclusion of new mechanisms, algorithms and protocols. In the research stream, the project investigates network and RAN slicing, mobile edge computing (MEC), massive multiple-input multiple output (MIMO) for 5G New Radio (NR) [16], as well as software-defined network (SDN) and network function virtualization (NFV) techniques applied to different traffic and load scenarios. The techniques and mechanisms relied upon in the research stream are currently under development and will be used in the most appropriate test-bed, with the final aim of deploying them in all three test-beds, if possible.

The project's experimental test-beds/facilities intend to satisfy several exact prerequisites [17], including the following:

- they need to offer an adequate level of flexibility, thus allowing for the inclusion – and use – of various characteristics of the involved technologies (distinct physical layers, frequency bands, etc.) at the differ-

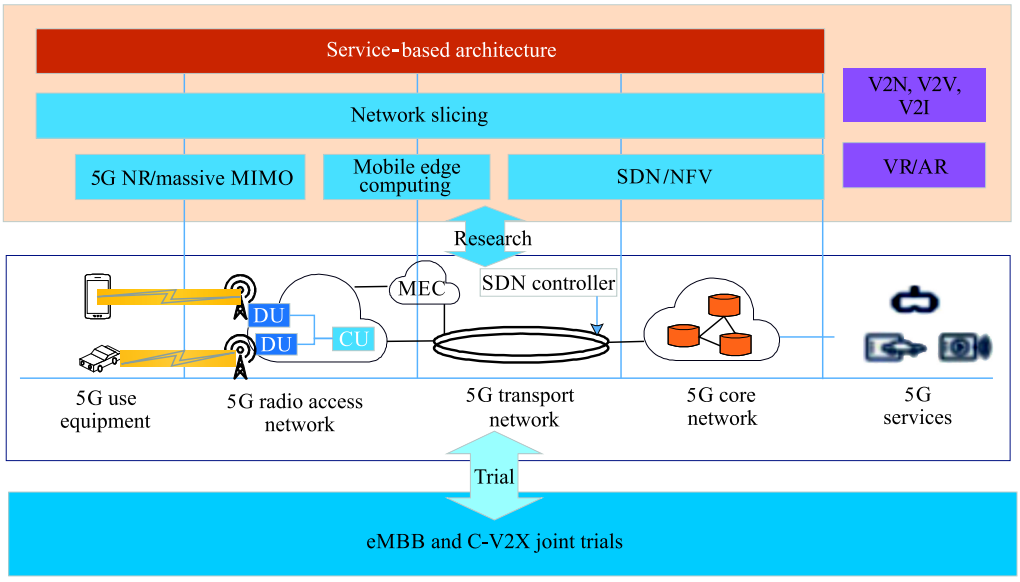


Fig. 1. Overall concept of the 5G-DRIVE project.

ent layers and in the related network components, on a case-specific basis,

- they need to be easily reconfigurable, thus allowing for dynamic adaptation to any specific testing requests made by the experiment participants,
- they need to support open-source solutions, so that their potential may be expanded by relying on the competence of the broadest possible scientific community,
- they need to offer testable and repeatable results, where necessary, to guarantee fair testing and reasonable comparison of the distinct technologies,
- they need to be complete in order to enable incorporation of all potential elements of a modern 5G ecosystem, such as mobile network operators (MNOs), virtual operators, end-users, machine-to-machine (M2M) applications, and the IoT,
- they need to support heterogeneity of the radio and optical interfaces tested, but also of specific contexts, including body-centric communications, vehicular networks, advanced robotics, etc.,
- they need to be site-agnostic, to the extent possible, with the purpose of testing technologies and the related facilities in dissimilar backgrounds,
- they need to be topology-agnostic, with the aim of supporting all potential wireless solutions, including cellular and satellite technologies, and topologies ranging from small cells to macro-cells.

2.2. Aims and Objectives of the 5G-DRIVE Effort

The main goal of 5G-DRIVE is to conduct 5G trials focusing on two dedicated scenarios:

- the first one is related to enhanced mobile broadband (eMBB) (3.5 GHz) – a priority band in the two regions for early introduction of very high-rate services. It is used for rendering typical mobile broadband services, as well as augmented reality (AR) services and is supported by the Surrey and Espoo trial sites,
- the other scenario is related to LTE-V2X-based Internet of Vehicles (IoV) using the 5.9 GHz band for vehicle-to-vehicle and the 3.5 GHz band for vehicle-to-network applications. The demonstrations will rely on real-life setups. Corresponding trials are conducted at Espoo and JRC sites.

The overarching aim of the project is to establish reliable collaboration and interactivity between recent 5G developments in the European Union and China, through joint trials and research activities in order to facilitate technology convergence, spectrum harmonization and business innovation, before any probable large-scale market efforts related to 5G deployment are undertaken. The expectation is to define requirements for the joint provision of corresponding services by the consortium members, especially with the participation of mobile operators and other stakeholders from the automotive and intelligent transports markets. The collaborative assessment allows to select the use cases, define trial requirements and to perform subsequent implementation and analysis. The inclusion of variable stakeholders is important to ensure that the trials and solutions meet the requirements of different vertical domains.
The 5G-DRIVE effort is based on existing 5G norms and on those that are currently under development, namely on

releases 13-14 of the 3GPP standard. Some issues covered by future releases have been taken into account as well, where possible. The project relies on testing opportunities offered by the three existing 5G test-beds provided by the participating partners, all aiming to test and validate innovative applications and related services. For this purpose, the project has developed key 5G technologies and pre-commercial test-beds for eMBB and V2X services in collaboration with a twinned Chinese project led by China Mobile, intending to compare the implementation methodologies and results. Trials for testing and validating key 5G functionalities, services and network planning solutions are actually taking place in eight cities across the EU and China. The main targets of this close collaboration scheme are illustrated in Fig. 2.

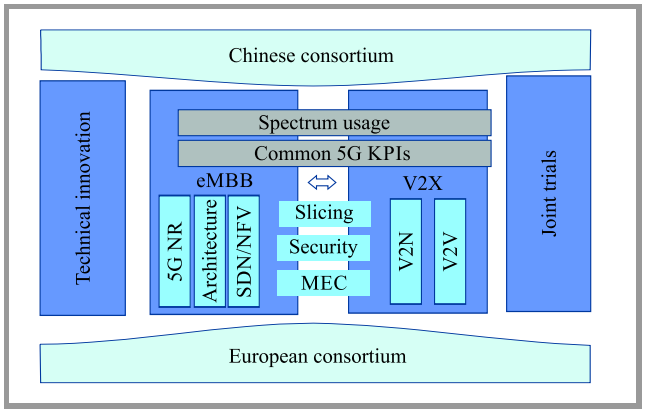


Fig. 2. EU-China 5G-DRIVE collaboration.

3. Scope for Collaborative V2X Communication

3.1. General Framework

Vehicle-to-Everything (V2X) communications were conceived as a means of transferring information from a vehicle to any entity that may have an impact on the vehicle, and *vice versa*. It is a vehicular communication system that integrates other more specific types of communications, such as vehicle-to-infrastructure (V2I), vehicle-to-network (V2N), vehicle-to-vehicle (V2V), vehicle-to-pedestrian (V2P), vehicle-to-device (V2D) and vehicle-to-grid (V2G) solutions. Thus, V2X may be perceived as a wireless technology enabling the exchange of data between a vehicle and its surroundings, whatever the latter may be. The core motivations behind V2X are to increase road safety, boost traffic efficiency and generate energy savings. The process of sharing, speed and position data between a given vehicle and the surrounding vehicles and/or infrastructures may enhance the driver's awareness of potential dangers and may significantly increase their ability to avoid collisions, resulting in a reduced number fatalities and in lower severity of injuries [18]. Fluidity of traffic may be increased as well by providing warnings about imminent traffic congestions and by suggesting alternative routes. Eco-friendly driving styles may be encouraged as well, reducing CO2 emissions by adaptive cruise control solutions and by relying on smarter transportation management systems.

Two types of existing V2X communication systems may be distinguished, according to their underlying technology: they may be WLAN-based or cellular-based.

IEEE first published the specification of WLAN-based V2X (IEEE 802.11p) in 2012 [19]. 802.11p is an extension of the 802.11a (Wi-Fi) standard that was defined by IEEE in 2009. In 2012, 802.11p was included in the overall IEEE 802.11 standard, but the informal term of 802.11p still continues to be used. The 802.11p-based multiple access mechanism (carrier sense multiple access protocol with collision avoidance, CSMA-CA) is a statistical protocol for establishing direct communications, as well as for V2V and V2I connections [20]. Primary V2X communication relies on the WLAN technology and is established directly between vehicles that make up the so-called vehicular ad-hoc network, providing that V2X senders are located within each other's range. Therefore, no infrastructure is required for the vehicles to communicate, which is of fundamental importance to guarantee safety in remote areas. WLAN is suitable for V2X communications as it is a low latency solution. It transmits several messages that are known as cooperative awareness messages (CAM), decentralized environmental notification messages (DENM) or basic safety message (BSM) [21]. The amount of data contained in these messages is very low. The radio technology used relies on the WLAN IEEE 802.11 family of standards and is known in the US as wireless access in vehicular environments (WAVE), and in Europe as ITS-G5.

3GPP initiated early standardization of cellular V2X (C-V2X) as part of release 14 in 2014 [22]. The applicable specifications were published in 2016. Because C-V2X relies on LTE as its underlying technology, it is often referred to as C-V2X to differentiate it from the 802.11p-based V2X technology. The functionalities it supports include direct communication (V2V, V2I) and wide area cellular network communication (V2N). In release 15, 3GPP continued the process of standardizing C-V2X based on 5G [23]. Therefore, to indicate the underlying technology, the term of new radio V2X (NR-V2X) is often used in contrast to LTE-based V2X (LTE-V2X). In either case, C-V2X is a generic term used to refer to cellular-based V2X communications, irrespective of the specific underlying type of mobile communication technologies used (4G or 5G). In release 16, 3GPP has further enhanced C-V2X functionality [24], and work in this regard is still in progress. 3GPP's main goal is to ensure enhanced data rates and to offer advanced ITS services. In other words, it strives to achieve objectives considered to be the key part of the future vehicular telecommunications landscape.

3.2. GLOSA Use Case – Conceptual Description

As far as the 5G-DRIVE V2X use cases are concerned, GLOSA and intelligent intersection are two solutions that are currently in their deployment and continuous assess-

ment phase. European trials use the lab and the stable communication environment available in Ispra, Italy and the dedicated ITS back-office services tailored for the needs of the Espoo, Finland trial site. Mobile-edge computing and virtual traffic light technologies are implemented as well for checking LTE-based communication channels and C-V2X user equipment (UE) nodes.

The architecture of the system follows the European mobile edge computing (MEC) guidelines [25]–[26] and is aligned with the C-ITS DEMN message format [27]. Even if the devices are different between Europe and China, the results are intended to be comparable, as the trials rely on a common set of KPIs.

Extreme fuel consumption experience in urban settings is closely associated with driving in heavy traffic that is also characterized by larger speed fluctuations and frequent stops at intersections. An improvement in the flow of city traffic allows to lower fuel consumption and CO2 emissions by reducing the number of stops, shortening delays and maintaining moderate speeds throughout the entire journey. One way to eliminate the unnecessary stop-and-go driving in cities is to optimize traffic signal timing in a dynamic manner, by adapting it to the current traffic situation.

Historically, signal timing optimization tools were developed to reduce delays and stops experienced by urban drivers. In recent years, new methods in traffic signal optimization have integrated the changes in the drivers' behavior in order to achieve the highest level of performance at intersections with traffic lights [28]. Today, vehicles can be equipped with a variety of sensors, driver assistance tools and safety related systems. Integration of cellular communication systems with cars/vehicles has resulted in an improvement of both safety and comfort levels. The solutions are now used in millions of cars and a trend showing that continued rapid growth may be expected is evident. In fact, many of the use cases, previously described in the respective ETSI ITS specifications and/or other documents [5]–[6], are now a reality thanks to existing cellular network connections. Cellular networks allow a variety of warnings to be sent to cars. These include, for instance, warnings about slow or stationary vehicles, road work warnings, weather condition warnings, hazard warnings, in-vehicle signage and speed-limits. Connected vehicles are capable of establishing two-way wireless communications (of the V2V and V2I variety) that may be effectively used for a variety of mobility and safety-related applications. One such application is known as GLOSA [29]. The system uses accurate information about traffic signal timing and locations to guide drivers, relying on I2V communications to offer speed recommendations that make the commute smoother, reducing stopping and allowing to smoothly pass through traffic signals [30]. A GLOSA implementation may be evaluated for two types of traffic signal timing schemes: predictable fixed-time signal timing and unpredictable actuated-coordinated signal timing.

GLOSA systems have been shown to be able to reduce both CO2 emissions and fuel consumption [31] by providing drivers with speed recommendations when approaching a traffic light. For the system to reach its maximum potential, it is essential to appropriately forecast all different types of traffic light configurations, including also adaptive systems in which signals may change with lead times as short as 1 s. The GLOSA application offers the benefit of timely and accurate information about traffic light cycles and locations via I2V communication. In consequence, drivers are provided with speed advice, ensuring that more constant speeds may be maintained and that less time is spent while the vehicle is stopped at traffic lights [32]. A reliable solution requires that traffic intensity, communications between traffic lights and vehicles, as well as driver behaviors be modeled in an effective manner. Research has been performed in each one of the aforementioned areas, but comprehensive simulations taking into account the dynamics of all parameters are scarce in the international literature.

The main goal of the GLOSA service is to predict the green phases of traffic lights and to provide drivers with reliable information enabling them to pass traffic lights during the green light phase. The green-wave assistant and the deceleration assistant are the two key applications of the system. The C-Roads Platform [33], being a joint initiative of the European Union Member States and road operators, established for testing and implementing C-ITS services in light of cross-border harmonization and interoperability, offers two methods of integrating the GLOSA service with vehicles: via ETSI G5, directly with the on-board computer of the vehicle, and via the mobile network, with a dedicated smartphone application. The green wave assistant shows information allowing the driver to reach the green phase at the next signal-controlled junction. Consequently, unnecessary stopping and acceleration may be prevented. The deceleration assistant works by informing the driver that he/she cannot reach the green phase at the next signal-controlled junction. With the help of these two applications, drivers may adapt their driving behaviors based on the information sent and received by the vehicle, therefore boosting efficiency and comfort levels. The GLOSA application offers the advantage of providing timely and accurate information about traffic light cycles and positions through I2V communication. In consequence, drivers are provided with speed advice, ensuring that more constant speeds may be maintained and that less time is spent while the vehicle is stopped at traffic lights.

3.3. Proposed Architecture

GLOSA is a Day-1 signage C-ITS service [34] aiming to inform end-users (i.e. drivers) about the speed that needs to be sustained, within the applicable legal limits, to approach an upcoming traffic light during the green phase. Although the individual user interface (UI) features are manufacturer specific, GLOSA notifications provided to end-users usually follow the structure below:

- the upcoming traffic light phase will change in x seconds (countdown, with x usually varying between 10 and 20 s),

- maintain 35 km/h – an optimal value to reach the upcoming traffic light during the green phase, in line with the applicable legal speed limits prevailing in a given,
- the driver will reach the upcoming traffic light during the red phase.

GLOSA provides drivers of approaching vehicles with short-term information on the status upcoming traffic lights, thus optimizing traffic flows and helping prevent speed limit violations, improving fuel efficiency and reducing pollution [35]. The system requires that intersection topology be determined, including geographical coordinates of entry and exit lanes that are transmitted in the form of a V2X I2V MAP message. This information is used by the receiving vehicles to calculate the relevance of the received notifications, taking into account their position relative to a given traffic light. The MAP message [36] is the basic requirement in many use cases, including GLOSA, as it provides a geographic reference for other messages. Not all vehicles have the same map installed on board and some of them may even have no map at all. Therefore, the MAP message is a platform that serves as an independent point of reference for all stations being a part of the ITS.
In Fig. 3, the architecture proposed for the test phase is presented. The key components of this particular use case are:

- A *physical/virtual traffic light* along with its controller to orchestrate the transitions between the red, amber and green phases. For the purpose of the experiment evaluating this particular use case, the traffic light may be either of the physical (i.e. a commercial, end-user product) or a virtual variety (software running on / communicating with the roadside unit (RSU) and performing the transitions ordered by the controller);
- An *LTE-V2X RSU* installed at the traffic light location (if physical) or running on / communicating with a finite state machine (FSM) (if virtual). The RSU will periodically broadcast signal phase and timing messages (SPAT) (e.g. every 100 ms) to all vehicles nearby;
- An *ITS-G5 RSU* installed at the traffic light location (if physical) or running on / communicating with the FSM (if virtual). The RSU will periodically broadcast signal phase and timing messages (SPAT) (e.g. every 100 ms) to all vehicles nearby;
- *Two On-Board Units (OBUs)* (one ITS-G5, one LTE-V2X) deployed in the test vehicles. The OBUs will receive and process the SPAT messages locally to compute the relevant GLOSA information in various forms, e.g. time remaining until transition to the next traffic light phase, optimum speed to reach the traffic light in the green phase, etc. Once computed, the GLOSA information will be relayed to an on-board laptop (or UI device), where it will be communicated, both visually and audibly, to the driver;
- The *JRC internal communications network* will provide connectivity between the RSUs and various supporting services running in the JRC data center, e.g. the experiment management console, the traffic light controller, the log server, etc.;
- *Physical/virtual* servers in the JRC data center running the above-mentioned supporting services. For the purpose of implementing and experimentally evaluating this use case, these servers may be provisioned either physically or with the use of virtual machines (VMs).

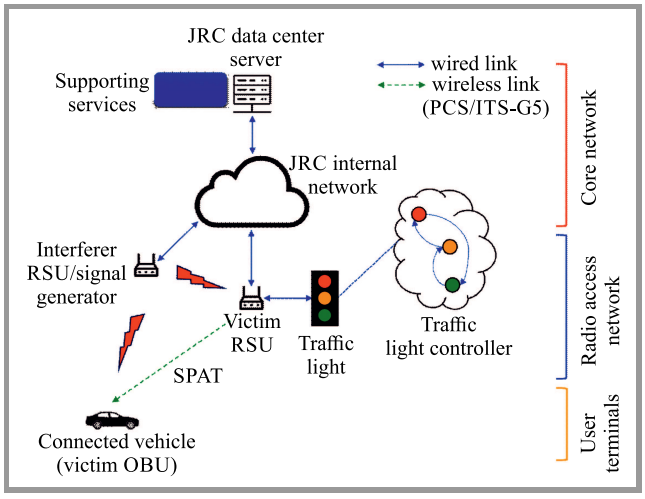


Fig. 3. Architecture of the GLOSA performance test.

During the course of the performance tests, both ITS-G5 and LTE-V2X units are exposed to external harmful interference. This interference will be generated either by ITS-G5 or LTE-V2X units themselves (depending on the specific test being run) or, alternatively, by an external signal generator.
It needs to be mentioned that the RSU also broadcasts MAP messages. The vehicles present nearby may receive these messages on their OBUs and process them locally along, inter alia, with their own positioning, speed and direction data. By doing so, on-board V2X modules can notify drivers of the optimum speed allowing to reach the upcoming traffic light during the green phase or, alternatively, can notify them that the traffic light will nevertheless transition to red imminently. The dynamic information is disseminated using the SPAT V2X I2V message and contains the time remaining until the traffic light changes and information about the recommended speed, applicable to a group of entry lanes. MAP and SPAT messages are already standardized [36] and profiled [37]. However, the interpretation of their content at the receiving side (cooperative vehicles), and the relation between this content and the actual current status of the traffic light controller may still lead to confusion.

Knowledge on how to interpret the SPAT content at the receiving side is particularly critical in the case of connected automated vehicles (CAVs) [38]. A CAV is assumed to be an ETSI ITS G5 [11] equipped vehicle with automated driving capabilities. In this sense, it is an improved version of a cooperative vehicle. It uses vehicle sensors and V2X transmissions to monitor its surroundings, as well as V2X communications to interact with other cooperative vehicles, CAVs and with cooperative intersections (CIs) [39]. Such a vehicle relies on the flow of information obtained for trajectory and maneuver planning, platoon organization and for implementation of improved, advanced driver assistance system (ADAS) functionalities. In a broader sense, a CAV may be a vehicle of any type. Multiple CAVs can form a vehicle platoon. In fact, the automated behavior of CAVs approaching a cooperative intersection (CI) will depend strongly on the information provided via the SPAT message. It is assumed that a CI has the form of an ETSI ITS G5 [11] equipped traffic light controller. In addition to such infrastructure sensors as inductive loops and cameras relied upon to detect all types of road users, the CI uses V2I communication to exchange data with cooperative vehicles and CAVs. The CI uses the flow of information for adaptive optimization of traffic light timing, simultaneously taking into account additional policy parameters, such as specific traffic priorities and demands pertaining to vulnerable road users [40]. In addition, it uses V2I communications to interact with cooperative vehicles, for example to provide recommendations concerning speed values and lanes to be chosen. Multiple CIs may be connected to enable traffic light coordination along a specific traffic corridor.

Based on the correct interpretation of the SPAT content, and together with other environmental information obtained via the onboard sensors, CAVs will decide whether to adjust the speed to the suggested value or to prepare for stopping. As far as the 5G-DRIVE context is concerned, GLOSA is an attractive V2X use case allowing to improve traffic flow in urban areas. It provides drivers with optimum speed recommendation as they approach an intersection with traffic lights. The recommendation, usually displayed on the car's dashboard, will instruct the driver to maintain the current speed, slow down or speed up. GLOSA may also provide time-to-green information when the vehicle is stationary at a traffic light. GLOSA takes advantage of real-time traffic sensing and infrastructure information that may then be relayed to vehicles aiming to reduce fuel consumption. The GLOSA system may operate in the manner presented in Fig. 4.

Fig. 4. LTE-V2X GLOSA demonstration by China Mobile, Huawei and ASTRI in the context of the 5G-DRIVE project.

3.4. KPIs and Performance Test Procedures

Since 5G-DRIVE aims to compare the benefits of 5G in Europe and China, some common test scenarios have been planned for implementation both in China and at the European trial sites. Common KPIs are defined to measure the same parameters and to compare the results. The trials aim to capture, analyze and discuss the following KPIs that can be either Quality of Service (QoS) or Quality of Experience (QoE) metrics. The former group includes: transmission range, average data rate, total message/s per channel and latency, i.e. the contribution of the radio access network to the total elapsed time, measured from the instant the RSU sends a packet to the moment when the OBU receives it. The packet error rate (PER) is measured as well, i.e. the ratio of packets unsuccessfully received by the OBU vs. the total number of packets sent by the RSU (percentage-wise). Other parameters include spectral efficiency and channel busy ratio (CBR, defined as the ratio between the time the channel is sensed as busy and the total observation time). It is a measure of channel load perceived by a vehicle, and depends on the number of vehicles within its transmission range and on their individual message generation rates, which makes it a metric suitable for increasing packet delivery performance.

Other related KPIs belonging to the same category include the following: co-channel and adjacent-channel interference and duty cycle. QoE metrics, in turn, include the following: deployment complexity, as well as ease of configuration and setup. Chinese KPIs are more exhaustive, since a greater emphasis is attached to network optimization, whereas 5G-DRIVE is application-driven. Because of the low volume of data/information transmitted in GLOSA (traffic light state machine timing data), KPIs such as peak/user-perceived data rate are not a critical concern in the context of this specific use case. Instead, the following most relevant service-level performance indicators have been defined for GLOSA [41]: PER and latency, i.e. the contribution of the radio access network to the total elapsed time, measured from the instant the RSU sends a packet to the moment when the OBU receives it.

For the purpose of the GLOSA trial, a commercial ITS-G5 roadside unit has been deployed at the JRC Ispra trial site/campus, covering a section of its internal road. The RSU sits at the junction of two suburban-type roads that are 420 and 220 m long, respectively, at a height of approximately 10 m. In addition, the RSU is connected to the trial-site's internal network infrastructure to allow remote configuration, management and traffic monitoring. The com-

mercial RSU runs a Linux-based operating system, thus allowing the execution of custom user space applications (such as a virtual traffic light for the GLOSA service). This setup is presented in Fig. 5.

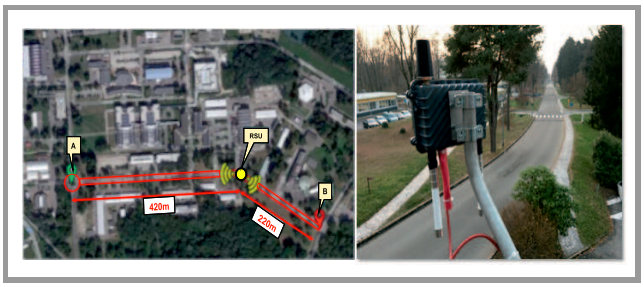


Fig. 5. ITS G5 RSU deployed in the Ispra campus.

The GLOSA setup described above is to be extended by deploying an LTE-V2X unit co-located with the ITS-G5 RSU. The LTE-V2X unit will also be connected to the campus' internal network for remote configuration, management and traffic monitoring purposes. In the context of the GLOSA trial, both technologies will be tested in a mutually-exclusive fashion to avoid harmful interference between them. The C-ITS road infrastructure described above is complemented by two OBUs (one ITS-G5, one LTE-V2X) co-located in a test vehicle for the purpose of the GLOSA trials. The OBUs are connected to a laptop for experiment configuration and traffic monitoring purposes.
As far as the specific test setup is concerned, coexistence tests will feature two C-ITS systems: an *interferer* and a *victim*. Each system will comprise two C-ITS units (ITS-G5 or LTE-V2X, depending on the particular test scenario). The victim system units will actively exchange C-ITS messages (e.g. standard CAM broadcasts) used in regular C-ITS operation. By contrast, the interferer system units will initially be set to operate in the idle mode (i.e. without broadcasting any C-ITS messages via the 5.9 GHz channel). Later, they will start sending C-ITS application messages of a user-controlled length (len_{msg}, in bytes) and packet rate (R_{pkt}, in packets/s) with the aim of causing harmful interference affecting the victim system. The following KPIs will be continuously monitored and recorded in the victim system.
The proposed GLOSA performance test procedure required that the following steps be taken:

- the vehicle starts driving towards the RSU at 30–40 km/h;
- the vehicle enters RSU's transmission range and the OBU starts receiving SPAT messages;
- the GLOSA client app successfully processes traffic light phase information encapsulated in SPAT messages. The in-vehicle UI displays the traffic light change-of-phase timer;
- traffic light phase and SPAT phase information are synchronized. The success criterion implicates that the traffic light phase changes when the in-vehicle timer expires;
- previous steps are repeated by setting $\text{len}_{\text{msg}} = (50, 100, 200, 300, 400, 500)$ bytes, and by setting $\text{R}_{\text{pkt}} = (300, 600, 650, 700, 750, 800)$ packets/s, in the interferer RSU or in an external signal generator. Now the success criterion implicates that for higher values of len_{msg} and R_{pkt}, the interferer system blocks communication between victim units.

As the procedure described above is a performance test, no actual pass/fail success criteria are set. The goal of the test is to evaluate resilience of ITS-G5 and LTE-V2X units against external harmful interference. The previously notified KPIs will be evaluated.

4. Discussion

It is quite interesting to mention that the actual GLOSA use case has also been examined in parallel with the C-V2X intelligent intersection use case, although the latter has more strict requirements. The purpose of such an approach was to ensure joint applicability and to extend visibility and market impact by using the same infrastructure and/or equipment. Offering a greater variety of services to the users and/or other market actors involved was another of the objectives pursued.
The tests will comprise two main activities. Firstly, there will be a subset of compliance tests specified in the ETSI Harmonized Standard for radio-communications equipment operating in the 5855–5925 MHz frequency band (ETSI EN 302 571) using standalone ITS-G5 and LTE-V2X devices [42]. The purpose of these tests is to verify the conformance of commercial/pre-commercial V2X devices with the European standard for radio transmissions in the 5.9 GHz band – a market-entry condition as laid down in the EU Radio Equipment Directive. Additionally, they will also help acquire the hands-on skills that are necessary to operate V2X equipment in a test environment. In the second phase, the aim will be to evaluate co-channel and adjacent-channel interference of ITS-G5 over LTE-V2X devices and *vice-versa*, both in conducted mode (laboratory) and in the anechoic chambers of the JRC Ispra site.
The subset of the conformance tests defined in ETSI EN 302 571 allows to verify the following:
Frequency stability: The goal of this test is to ensure that V2X equipment can operate on the specific carrier center frequencies (f_c) that correspond to the nominal carrier frequencies covering the 5860–5920 MHz band in 10 MHz steps. The actual carrier center frequency for any of the above 10 MHz channels shall be maintained within the range $f_c \pm 20$ ppm.
RF output power: The purpose of this test is to ensure that the mean equivalent isotropically radiated power (EIRP) of V2X equipment during transmission bursts does not exceed 33 dBm.

Power spectral density: The aim of this test is to ensure that the power spectral density (PSD) of the V2X device during transmission bursts does not exceed 23 dBm/MHz EIRP.

Transmit power control (TPC): TPC is a scheme supporting coexistence with the European Committee for Standardization (CEN) dedicated short-range communication (DSRC) at toll plazas and is used as a single mechanism by decentralized congestion control (DCC) solutions to reduce communication channel congestion. The goal of this test is to ensure that TPC operates as expected in the presence of CEN DSRC devices, as well as during DCC.

Transmitter unwanted emissions: The purpose of this test is to ensure that V2X devices respect the radio frequency (RF) power limits for radio frequency emissions outside the 5 GHz ITS frequency band (outside of 5855–5925 MHz).

Decentralized Congestion Control (DCC): DCC is a mandatory ITS-G5 mechanism to ensure that the radio channel is not congested by too many transmissions within a certain geographical range. The mechanism is such that the equipment adapts its transmission behavior dynamically, based on how occupied the channel is at a given moment. The aim of this test is to ensure that DCC operates as expected in the presence of interferer signals in the 5.9 GHz band.

As mentioned above, several ITS-G5/LTE-V2X coexistence tests will be performed as well to evaluate the degree of harmful interference between commercial/pre-commercial off-the-shelf ITS-G5 and LTE-V2X equipment in the absence of any active coexistence mechanisms in their radio protocol stacks. The outcome of these experiments will provide a baseline scenario for further tests once specific coexistence mechanisms have been specified, implemented and deployed in the ITS-G5 and LTE-V2X radio protocol stacks.

The main conceptual differences consist in the fact that in China the trials are LTE network-based, whereas the European ones have been performed with the use of PC5 devices, while also relying on an LTE network. The other distinction is that European V2X applications use the decentralized environmental notification messages (DEMN) message format, whereas in China, BSM type messages are used. The KPIs have been selected to offer complementary measurements between the 5G large-scale trial in China and the 5G-DRIVE project in Europe. However, some application-specific measures related to ETSI- and SAE-standardized message deliveries between the digital infrastructure and the vehicle need to be taken into consideration as well. The aims of the Chinese and the European V2X tests are slightly different. Chinese trials are led by a telecom operator and, therefore, they focus more on network side operation and on latencies in different communication planes. As expected, in Europe the tests focus on pure communication and cover also communication delays between cloud services and vehicle components (OBUs, message parser, etc.), since the main objective is to understand the benefits of deploying intelligent vehicle components. The packet error rate target in effect in Europe is less than 5%, whereas the Chinese target is less than 10%. Latency targets are equivalent, as they equal 100 ms. In China, latency is divided between control and data planes (CP and DP) only, whereas European V2X trials do distinguish different latency sources.

5. Acknowledgement

This paper has been drawn up based on the research performed under the H2020 5G-DRIVE (Harmonized Research and Trials for Service Evolution between EU and China) project, funded by the European Commission under Grant Agreement no. 814956.

References

[1] 5G-PPP, "5G strategic deployment agenda for connected and automated mobility in Europe" [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2020/10/20201002_5G_SDA_for_CAM_Final.pdf

[2] "Report ITU-R M.2243 (00/2011): Assessment of the global mobile broadband deployments and forecasts for International Mobile Telecommunications", *Int. Telecommunication Union* [Online]. Available: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2243-2011-PDF-E.pdf

[3] J. G. Andrews *et al.*, "What will 5G be?", *IEEE JSAC, Special issue on 5G Wireless Commun. Systems*, vol. 32, no. 6, pp. 1065–1082, 2014 [Online]. Available: https://arxiv.org/pdf/1405.2957

[4] L. Figueiredo *et al.*, "Towards the development of intelligent transportation systems", in *Proc. IEEE Intelligent Transportat. Systems*, Oakland, CA, USA, 2001, pp. 1206–1211 (DOI: 10.1109/ITSC.2001.948835).

[5] ETSI, "TS 103 723 V1.2.1 (2020-11): intelligent transport systems (ITS); profile for LTE-V2X Direct Communication", 2020 [Online]. Available: http://www.etsi.org/deliver/etsi_ts/103700_103799/103723/01.02.01_60/ts_103723v010201p.pdf

[6] A. Festag, "Cooperative intelligent transport systems in Europe", *IEEE Vehicular Technol. Mag.*, vol. 12, no. 2, pp. 89–97, 2017 (DOI: 10.1109/MVT.2017.2670018).

[7] ISO/CEN, "Cooperative intelligent transport systems (C-ITS). Guidelines on the usage of standards", 2020 [Online]. Available: https://www.itsstandards.eu/app/uploads/sites/14/2020/10/C-ITS-Brochure-2020-FINAL.pdf

[8] NGMN, "V2X White Paper", 2018 [Online]. Available: https://www.ngmn.org/wp-content/uploads/V2X_white_paper_v1_0-1.pdf

[9] GSMA, "Connecting vehicles – today and in the 5G era with C-V2X" [Online]. Available: https://www.gsma.com/iot/wp-content/uploads/2019/08/Connecting-Vehicles-Today-and-in-the-5G-Era-with-C-V2X.pdf

[10] H. Bhatia, "125 Million+ connected cars shipments by 2022; 5G cars by 2020", *Counterpoint*, 2018 [Online]. Available: https://www.counterpointresearch.com/125-million-connected-cars-shipments-2022-5g-cars-2020/

[11] ETSI, "EN 302 663 V1.3.1 (2019-10): intelligent transport systems (ITS); ITS-G5 access layer specification for intelligent transport systems operating in the 5 GHz frequency band" [Online]. Available: https://www.etsi.org/deliver/etsi_en/302600_302699/302663/01.03.01_30/en_302663v010301v.pdf

[12] 3GPP, "TS 22.185 v16.0.0 (2020-07): service requirements for V2X services" [Online]. Available: https://www.3gpp.org/ftp//Specs/archive/22_series/22.185/

[13] 5G-DRIVE H2020 5G-PPP Project [Online]. Available: https://5g-drive.eu

[14] I. P. Chochliouros *et al.*, "Testbeds for the implementation of 5G in the European Union: the innovative case of the 5G-DRIVE project", in *Proc. Artificial Intelligence Applications and Innovations*, 2019, pp. 78–92, (DOI: 10.1007/978-3-030-19909-8_7).

[15] A. Kostopoulos *et al.*, "5G trial cooperation between EU and China", in *Proc. IEEE Int. Conf. on Commun. Workshops (ICC 2019 Workshops)*, 2019, Shanghai, China, pp. 1–6 (DOI: 10.1109/ICCW.2019.8756985).

[16] ETSI, "TS 138 300 V15.3.1 (2018-10): 5G; NR; overall description" [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138300_138399/138300/15.03.01_60/ts_138300v150301p.pdf

[17] R. Verdone and A. Manzalini, "5G experimental facilities in Europe", *NetWorld 2020 ETP* [Online]. Available: https://www.networld2020.eu/wp-content/uploads/2016/03/5G-experimentation-Whitepaper-v11.pdf

[18] J. Wang, Y. Shao, Y. Ge, and R. Yu, "A survey of vehicle to everything (V2X) testing", *Sensors*, vol. 19, no. 2, 2019 (DOI: 10.3390/s19020334).

[19] IEEE, "IEEE 802.11p-2010: IEEE Standard for Information Technology (IT) – Local and metropolitan area networks – Specific requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments" [Online]. Available: https://standards.ieee.org/standard/802_11p-2010.html

[20] 5GAA, "An assessment of LTE-V2X (PC5) and 802.11p direct communications technologies for improved road safety in the EU", 2017 [Online]. Available: http://5gaa.org/wp-content/uploads/2017/12/5GAA-road-safety-FINAL2017-12.05.pdf

[21] ETSI, "EN 302 665 V1.1.1 (2010-09): intelligent transport systems (ITS); communication architecture" [Online]. Available: https://www.etsi.org/deliver/etsi_en/302600_302699/302665/01.01.01_60/en_302665v010101p.pdf

[22] 3GPP, "TR 21.914 v14.0.0: release 15 description; summary of Rel-14 work items", 2018 [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/21_series/21.914/

[23] 3GPP, "TR 21.915 v0.4.0: release 15 description; summary of Rel-15 work items," 2018 [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/21_series/21.915/

[24] 3GPP, "TR 21.916 v1.0.0: release 16 description; summary of Rel-16 work items," 2020 [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/21_series/21.916/

[25] S. Kekki *et al.*, "MEC in 5G Networks, ETSI White paper", *European Telecommunications Standards Institute (ETSI)*, no. 28, 2018 [Online]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf

[26] ETSI, "ETSI GS MEC 003 V1.1.1: mobile edge computing (MEC); framework and reference architecture" [Online]. Available: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/01.01.01_60/gs_MEC003v010101p.pdf

[27] J. Santa, F. Pereniguez-Garcia, A. Moragón, and A. Skarmeta, "Experimental evaluation of CAM and DENM messaging services in vehicular communications", *Transportat. Res. Part C: Emerging Technol.*, vol. 46, pp. 98–120, 2014 (DOI: 10.1016/j.trc.2014.05.006).

[28] S. I. Guler, M. Menendez, and L. Meier, "Using connected vehicle technology to improve the efficiency of intersections", *Transportat. Res. Part C: Emerging Technol.*, vol. 46, pp. 121–131, 2014 (DOI: 10.1016/j.trc.2014.05.008).

[29] A. Stevanovic, J. Stevanovic, and C. Kergaye, "Green light optimized speed advisory systems: impact of signal phasing information accuracy", *J. of the Transportat. Res. Board*, vol. 2390, no. 1, pp. 53–59, 2013 (DOI: 10.3141/2390-06).

[30] K. Katsaros, R. Kernchen, M. Dianati, and D. Rieck, "Performance study of a green light optimized speed advisory (GLOSA) application using an integrated cooperative ITS simulation platform", in *Proc. Int. Wireless Commun. and Mobile Comput. Conf. (IWCMC)*, Istanbul, Turkey, 2011, pp. 918–923 (DOI: 10.1109/IWCMC.2011.5982524).

[31] T. Tielert *et al.*, "The impact of traffic-light-to-vehicle communication on fuel consumption and emissions", in *Proc. Internet of Things 2010 Conf. (IoT2010)*, Tokyo, Japan, 2010, pp. 1–8 (DOI: 10.1109/IOT.2010.5678454).

[32] M. Chao-Qun, H. Hai-Jun, and T. Tie-Qia, "Improving urban traffic by velocity guidance", in *Proc. Int. Conf. on Intelligent Comput. and Automation (ICICTA-2008)*, vol. 2, pp. 383–387 (DOI: 10.1109/ICICTA.2008.288).

[33] C-ROADS [Online]. Available: https://www.c-roads.eu/platform.html

[34] M. Lu *et al.*, "C-ITS (cooperative intelligent transport systems) deployment in Europe – challenges and key findings", in *Proc. ITS World Congress*, pp. 1–10, 2018.

[35] B. Asadi and A. Vahidi, "Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time", *IEEE Transac. on Control Systems Technol.*, vol. 19, no. 3, pp. 707–714, 2010 (DOI: 10.1109/TCST.2010.2047860).

[36] SAE, "SAE J2735_201603, dedicated short range communications (DSRC) message set dictionary", 2016 [Online]. Available: https://www.sae.org/standards/content/j2735_201603/

[37] Talking Traffic consortium, Dutch profiles and ITF, "Smart mobility community for standards and practices", 2019 [Online]. Available: http://www.smartmobilitycommunity.eu/talking-traffic-dutch-profiles-and-itf.

[38] M. Galvani, "History and future of driver assistance", *IEEE Instrumentation Measurement Mag.*, vol. 22, no. 1, pp. 11–16, 2019 (DOI: 10.1109/MIM.2019.8633345).

[39] L. Chen and C. Englund, "Cooperative intersection management: A survey", *IEEE Transac. on Intelligent Transport. Systems*, vol. 17, no. 2, pp. 570–586, 2015 (DOI: 10.1109/TITS.2015.2471812).

[40] M. Hafner, D. Cunningham, L. Caminiti, and D. Del Vecchio, "Cooperative collision avoidance at intersections: Algorithms and experiments", *Trans. Intel. Transportation Systems*, vol. 14, no. 3, pp. 1162–1175, 2013 (DOI: 10.1109/TITS.2013.2252901).

[41] 5G-DRIVE Project, "Deliverable D4.2: Joint specifications for V2X trials", 2018.

[42] ETSI, "EN 302 571v2.1.1 (2017-02): Intelligent Transport Systems (ITS); Radiocommunications equipment operating in the 5855 MHz to 5925 MHz frequency band; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU" [Online]. Available: https://www.etsi.org/deliver/etsi_en/302500_302599/302571/02.01.01_60/en_302571v020101p.pdf

Ioannis P. Chochliouros graduated from the Dept. of Electrical Engineering at the Polytechnic School of the Aristotle University of Thessaloniki, Greece, holding also an M.Sc. (D.E.A.) and a Ph.D. from Université Pierre et Marie Curie (Paris VI), France. His practical experience as an engineer has been mainly in telecommunications, as well as in various construction projects in Greece and in the wider Balkan area. Since 1997, he has been working at the Competition Department and as an engineer-consultant of the Chief Technical Officer of the Hellenic Telecommunications Organization (OTE). He was closely involved in OTE's major national and international business activities, as

a specialist-consultant for technical and regulatory affairs, especially for the evaluation and the adoption of innovative e-infrastructures and e-services in Greece and abroad. He has also served as the Head of Technical Regulations Department of OTE's Division for Standardization and Technical Regulations, representing OTE in international standardization bodies, and has been involved in a great variety of projects regarding European and international standardization, with an emphasis placed on modern technologies. In addition, he has also worked as an independent consultant, participating in several European and/or international research and business studies. Since 2005, he has been the Head of OTE's Fixed Network R&D Programs Section and has been involved in different national, European and international R&D projects and market-oriented activities, many of which have received international awards. During his professional career, he has participated – either as a co-ordinator or a scientist-researcher – in more than 62 European and national research programs (IST, STREP, EURESCOM, FP6, FP7, ESA, H2020 and the 5G-PPP phases I-III). He is the author/co-author of three international books and has published more than 250 distinct scientific or business papers/reports in international papers.

https://orcid.org/0000-0002-4208-1676
E-mail: ichochliouros@oteresearch.gr
Hellenic Telecommunications Organization
Athens, Greece

Anastasia S. Spiliopoulou is a lawyer and a Member of the Athens Bar Association. She also holds a post-graduate diploma from the Law School of the National and Kapodistrian University of Athens, Greece. She has many years of professional experience in telecommunications and IT-related issues and has been involved in many legal cases related to regulatory issues, as well as to the deployment and operation of modern electronic communication networks, and to the provision of communications services. She is OTE's (Hellenic Telecommunications Organization) expert for a great variety of regulatory issues affecting both European and national policies. Apart from her expertise in regulatory matters concerned with networking and service offering, she specializes also in privacy- and data protection-related issues. She is the author/coauthor of more than 110 papers published in international papers and has participated in numerous conferences, acting in the capacity of the invited speaker on several occasions.

E-mail: aspiliopoul@ote.gr
Hellenic Telecommunications Organization
Athens, Greece

Alexandros Kostopoulos holds two M.Sc. degrees in Telecommunications (University of Athens) and in Computer Science (University of Piraeus). He received his Ph.D. from Athens University of Economics and Business (AUEB) in 2013. He was a visiting lecturer at AUEB, as well as a postdoctoral researcher at the Institute of Computer Science, in Forth. He is currently working at the Research and Development Department of OTE (Hellenic Telecommunications Organization), focusing on European and national research projects. He has been involved, inter alia, in various 5G-PPP projects (phases I-III). He has published several papers in scientific journals and conferences and has joined numerous events in the capacity of a speaker.

E-mail: alexkosto@oteresearch.gr
Hellenic Telecommunications Organization
Athens, Greece

George Agapiou received his Diploma in Electrical Engineering from the University of Louisville, Kentucky, in 1985, as well as M.Sc. and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology, in 1987 and 1991, respectively. From 1984 to 1985, he worked at Philip Morris at Louisville, Kentucky as a Maintenance Engineer. Since 1996 he has been a telecom engineer at OTE and COSMOTE, where he worked on various European projects in the area of Mobile and Optical Communications. Currently, he holds the position of the Head of Core Lab Testing. He has participated in various IST, STREP, EURESCOM, FP6, FP7, ESA, H2020 and 5G-PPP (phases I-III) projects and has published more than 100 papers in scientific journals and proceedings and has been a co-author of three technical books.

E-mail: gagapiou@oteresearch.gr
Hellenic Telecommunications Organization
Athens, Greece

Pavlos Lazaridis is a Professor of Electronic and Electrical Engineering at the University of Huddersfield, UK. He received his Electrical Engineering degree from the Aristotle University of Thessaloniki, Greece, in 1990, the M.Sc. degree in Electronics from Université Pierre et Marie Curie, Paris 6, France, in 1992, and the Ph.D.

degree in Electronics and Telecommunications from Ecole Nationale Supérieure des Télécommunications (ENST) and Paris 6, Paris, in 1996. From 1991 to 1996, he researched semiconductor lasers, wave propagation, and nonlinear phenomena in optical fibers for the Centre National d'Etudes des Télécommunications (CNET) and was teaching at ENST. In 1997, he became the Head of the Antennas and Propagation Laboratory, TDF-C2R Metz (Télédiffusion de France/France Télécom Research Center), where he conducted research on antennas and radio-coverage for cellular mobile systems (GSM), digital audio broadcasting (DAB), and digital video broadcasting-terrestrial (DVB-T). From 1998 to 2002, he was with the European Patent Office, Rijswijk, the Netherlands, as a Senior Examiner in the field of Electronics and Telecommunications. He is leading the EU Horizon 2020 projects titled ITNMOTOR5G and RISE-RECOMBINE, for the University of Huddersfield, UK.

E-mail: P.Lazaridis@hud.ac.uk
Department of Engineering and Technology
University of Huddersfield
Huddersfield, United Kingdom

Zaharias D. Zaharis received his B.Sc. in physics, M.Sc. in electronics, Ph.D. and a Diploma degree in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1987, 1994, 2000, and 2011, respectively. From 2002 to 2013, he was with the administration of the telecommunications network, Aristotle University of Thessaloniki, and since 2013 he has been with the Department of Electrical and Computer Engineering of the same university. His current research interests include design and optimization of antennas and microwave circuits, signal processing on smart antennas, development of evolutionary optimization algorithms, and neural networks.

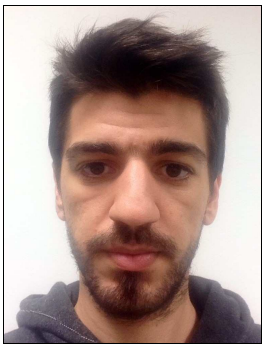
E-mail: zaharis@auth.gr
Aristotle University of Thessaloniki
Thessaloniki, Greece

Tao Chen received his Ph.D. degree in telecommunications engineering from the University of Trento, Italy in 2007. Then, he joined the VTT Technical Research Centre of Finland and currently he is a senior researcher focusing on the area of connectivity. He is a docent (adjunct professor) at the University of Jyväskylä, and a project coordinator for 5G PPP COHERENT. His current research interests include software-defined networking for 5G mobile networks, massive IoT in 5G, dynamic spectrum access, energy efficiency and resource management in heterogeneous wireless networks, and social-aware mobile networks.

E-mail: Tao.Chen@vtt.fi
VTT Technical Research Centre of Finland
Espoo, Finland

Athanassios Dardamanis is the CEO of Smartnet S.A. and has graduated from the IT Dept. of the National and Kapodistrian University of Athens, Greece. He has extensive experience in a variety of technical implementation and design activities and has worked with operators, enterprises, organizations and public bodies in Greece. He has successfully implemented numerous technical projects for various customers and has supported the company's strive towards widening the scope of its presence in the area of modern technological applications. He has extensive administrative and technical experience obtained while working in the capacity of a top executive of several very well-known IT companies in the Hellenic market (such as Altec and Alcatel).

E-mail: adardamanis@smartnet.gr
SmartNet
Athens, Greece



Michail-Alexandros Kourtis received his Ph.D. from UPV/EHU in 2018 and his Diploma and M.Sc. in Computer Science from the Athens University of Economics and Business, in 2011 and 2013, respectively. Since 2015, he has worked on cybersecurity applications for Network Function Virtualization, as well as on the definition of privacy and security risk metrics for virtualized infrastructures. His research interests include QoE, QoS, video processing, video quality assessment, image processing, LTE, 5G, network function virtualization, and software-defined networks. He is a contributor to the OPNFV open-source project known as Yardstick, and an active member, participant and contributor at the IETF NFVRG. He is also a TPC member and a reviewer at various conferences and journals. He has joined several EU-funded projects, with some of them implemented within the framework of 5G-PPP (phases I–III).

E-mail: akourtis@orioninnovations.gr
ORION Innovations Private Company
Athens, Greece

Marinos Agapiou has completed his B.S. at the National Kapodistrian University of Athens and is currently complementing his MBA at the Athens University of Economics and Business in Greece. He is the author of several papers dealing with telecommunications and market analysis.

E-mail: minosagap@gmail.com
Independent Consultant
Athens, Greece

Uwe Herzog is a Program Manager at Eurescom. He has more than 20 years of professional experience. He has been involved in a number of EC research projects as project coordinator, work-package leader and in other roles, including projects such as WINNER I and II, XIFI, CONCORD, ENVIROFI, eMobility NetWorld, UbiSec&Sens, WSAN4CIP, EURO-5G, SPEED-5G, ORPHEUS, Euro-5G and To-Euro-5G. Uwe has also been an evaluator and reporter in the FP7 and H2020 Call evaluation process in the area of Future Networks and SME-Instrument on Open Disruptive Innovation. He holds an M.Sc. degree in electrical engineering from the University of Chemnitz, Germany, and an MBA from the University of Mannheim and ESSEC Business School Paris.

E-mail: herzog@eurescom.eu
Eurescom
Heidelberg, Germany

Latif Ladid is the Founder and current President of the IPv6 Forum and has been a Board Member of 3GPP since 1999. He is also the Chairman of the European IPv6 Task Force, the Chairman of the 5G World Alliance, and the Chairman of the following technical Committees: IEEE COMSOC IoT subCommittee, IEEE COMSOC 5G subCommit, and ETSI IPv6 ISG (including 5G). He is the Co-Chair of IEEE COMSOC SDN-NFV subCommittee. He is a member of the UN Strategy Council, a Member of the present EU Future Internet Forum, representing the Luxembourg government. He sits on the IPv6 Ready Logo Program Board. He is a Program Coordinator at the University of Luxembourg, contributing to several EU-funded projects on next generation technologies.

E-mail: latif.ladid@uni.lu
Faculty of Science, Technology and Medicine
Department of Computer Science
Université du Luxembourg
Luxembourg

Paper

Security Verification in the Context of 5G Sensor Networks

Piotr Remlein and Urszula Stachowiak

Poznań University of Technology, Poznań, Poland

https://doi.org/10.26636/jtit.2021.153221

Abstract—In order to develop reliable safety standards for 5G sensor networks (SN) and the Internet of Things, appropriate verification tools are needed, including those offering the ability to perform automated symbolic analysis process. The Tamarin prover is one of such software-based solutions. It allows to formally prove security protocols. This paper shows the modus operandi of the tool in question. Its application has been illustrated using an example of an exchange of messages between two agents, with asynchronous encryption. The scheme may be implemented, for instance, in the TLS/DTLS protocol to create a secure cryptographic key exchange mechanism. The aim of the publication is to demonstrate that automated symbolic analysis may be relied upon to model 5G sensor networks security protocols. Also, a use case in which the process of modeling the DTLS 1.2 handshake protocol enriched with the TCP SYN Cookies mechanism, used to preventing DoS attacks, is presented.

Keywords—5G, automated symbolic analysis, Internet of Things, security protocols, sensor networks.

1. Introduction

New communication technologies, such as 5G, are vulnerable to various type of attacks. Devices in such networks deal with huge volumes of seemingly non-valuable data. This means that security rules in such systems are often disregarded by equipment manufacturers, and that users usually do not care about maintaining proper protection mechanisms. The advent of 5G shows that negligence in the implementation of security measures is an enabler of attacks aimed at harming private and public property. Thus, security of 5G-based sensor networks (SN) or IoT systems is an important issue exerting a significant impact on the development of these technologies [1].

To achieve adequate level of security, newly developed systems rely on cryptographic protections that were known previously. However, they are often quite clumsy in terms of design, and the level of security of new solutions is not properly verified. Therefore, it is important to achieve a formal proof of security at the design stage. For this purpose, tools for automatic security verification relying on symbolic analysis are often used [2], such as Tamarin [3].

This paper describes how Tamarin may be used in testing and validation of security protocols. The paper addresses issues related to the security level of 5G sensor networks and the Internet of Things with respect to the characteristics of 5G SN devices.

This paper is structured as follows. In Section 2, a brief description of selected security standards for 5G SN and IoT is presented. Section 3 contains an introduction to Tamarin software. The manner in which this software may be used for protocol security analysis is described, with a simple message exchange between two agents in an IoT or a sensor network, with asynchronous encryption in the transport layer security (TLS) protocol, supporting secure exchange of cryptographic keys, used as an example. Section 4 shows how to search for errors in the model based on the simple example of a Diffie-Hellmann algorithm model. Sections 5 and 6 describe how Tamarin software may be used to validate security protocols based on the DTLS 1.2 handshake. Reference to TLS version 1.2 is given and a model representing migration from TLS-over-TCP to TLS-over-UDP is examined. The Jun Kim model with a modification of the DTLS protocol aimed at adding a collateral TCP SYN-Cookies mechanism [4], [5] is described as well. Datagram TLS seems to be another promising protocol in 5G SN applications, where the broadly understood system resources, such as computational power, operating memory, the number of round-trips necessary to commence the sending of application messages, and the energy used for computing and data transmission, are usually very limited. The summary and conclusions are presented in Section 7.

2. Selected Security Mechanisms for 5G SN and IoT

The limited hardware resources of 5G SN devices often do not allow for the full implementation of typical security mechanisms and advanced cryptographic algorithms. Thus, special security protocols should be developed, such as TLS and datagram transport layer security (DTLS). They are designed to offer specific security rules relied upon while communicating. TLS is a better version of the secure socket layer (SSL) and the terms are often used interchangeably.

TLS utilizes the transmission control protocol (TCP). DTLS is designed for applications whose communications use the user datagram protocol (UDP) and is designed to be no different from TLS. It should also provide a degree of

security that is similar to the one offered by TLS. Because DTLS is based on UDP, communications are unreliable. It is less resource intensive, making it better suited for 5G SN than TLS. The DTLS protocol solves two problems that TLS can experience with datagram transport. In the case of DTLS, it is not possible to use stream ciphers. Therefore, it is possible to decode records from datagrams individually, because the order of delivery of datagrams may differ from the original order. Unlike TLS, it utilizes explicit transport layer messages [6].

These protocols support confidentiality, integrity and authentication at the transport layer. Both of them use asymmetric encryption and the X.509 certification protocol. They use traditional security mechanisms but can also be modified to fit the limited hardware resources of IoT and 5G SN devices [7]. The cryptographic system parameters used during a single session are agreed upon using the TLS handshake protocol. Communication participants can choose the protocol version and the cryptographic algorithms. Optionally, both sides can authenticate each other and define a shared secret key. TLS supports three modes of key agreement: elliptic curve Diffie-Hellman (ECDHE), pre-shared key (PSK), and PSK with ECDHE [8].

The TLS handshake consists of three main phases. The first of them is the key exchange phase, where the keys and cryptographic parameters are established to allow encryption of the transmission. In this phase, the client transmits a *ClientHello* message that includes a random, one-time identifier named "once". The response has the form of a *ServerHello* message that specifies the negotiated cryptographic parameters of the connection. The participants negotiate a shared communication key. Then, in the server parameter exchange phase, other security parameters that are required for authentication are specified. In order to do this, the server sends two types of messages: *EncryptedExtensions* and *CertificateRequest*.

The third and final phase is concerned with server authentication and, optionally, client authentication. The same set of commands is exchanged: *Certificate*, *CertificateVerify*, and *Finished* [8]. The TLS protocol is a hardware resource intensive solution and may lead to overloading the system. It requires additional steps to be taken when setting up communication and calls for significant amounts of memory for storing the certificates. Version 1.3 from 2018 is the most recent iteration of TLS/DTLS. This specification skips many of the outdated cipher algorithms. Version 1.3 is optimized for efficiency and is intended for IoT applications [8], [9].

3. Introduction to Tamarin

To verify the security level of a specific protocol, computational or symbolic analyses are relied upon. Symbolic verification is a relatively quick method that is less detailed than the calculation-based approach. During the process of creating new protocol models, large assumptions are used. Hence, the idea of creating software solutions to increase the level of automation while using this method was conceived. This type of automated analysis has been intensively researched over the past years. Tools of this type use preprogrammed functions with cryptographic primitives and have a predefined model of intruder/attacker behaviors. Their usefulness was proven during the analysis of TLS protocol's version 1.3. Automated symbolic analysis was used at the pre-implementation verification stage. Although this type of analysis is used on a frequent basis, the solutions obtained are characterized by a high level of discrepancy between what can be demonstrated by traditional computational methods and what is offered by fully automatic tools [6], [10].

Tamarin is an open-source software-based solution designed for symbolic analysis and verification of security protocols. It consists of a collection of tools capable of solving many more problems than alternative tools, such as ProVerif, Scyther, and Maude-NPA [4], [11]. Tamarin was designed by the Information Security Group at ETH Zurich. Research is currently underway on new security protocol verification techniques potentially applicable to this tool. Actual security protocols for 5G systems or the IEC 9798 standard are being modeled as well [10], [12], [13].

Tamarin's operation is based on a dedicated language, in which the analyzed protocols are modeled and rules are created for agents – both those involved in information exchange and attackers. The models created may be used for automated generation of proofs related to the analyzed protocols. Proofs are always performed according to similar principles, i.e. a logically justified number of messages exchanged between agents is created for the protocol under investigation. Then, the degree to which these messages are susceptible to a specific attack method is verified. If all known attacks taken into account in the analysis prove to be ineffective, then a symbolic proof of security for the examined instances is obtained. Otherwise, Tamarin provides a counterexample describing the attack.

This type of software performs proofs while working in two modes. The first is fully automated, uses heuristics to find counter-arguments for the performed proofs. This mode, however, does not always allow to obtain an unambiguous answer in the form of a security proof. This is due to the complex nature and numerous message relationships of the investigated model under. The second, interactive mode allows a counterexample to be generated using a graph that includes the agents involved in a given type of attack (Fig. 1). This makes it possible to understand how a specific attack can be carried out. Tamarin was created using the Haskell programming language [10], [12], [13].

During the phase focusing on the protocol's security level, an intruder/attacker model and a communication channel model are created in addition to the symbolic protocol model. The intruder model is based on a set of theorems and algorithms that formalize the knowledge and the capabilities of the attacker. One of the most common models used in Internet security research to describe intruder behavior is the Dolev-Yao model. It is also used in Tamarin. Here,

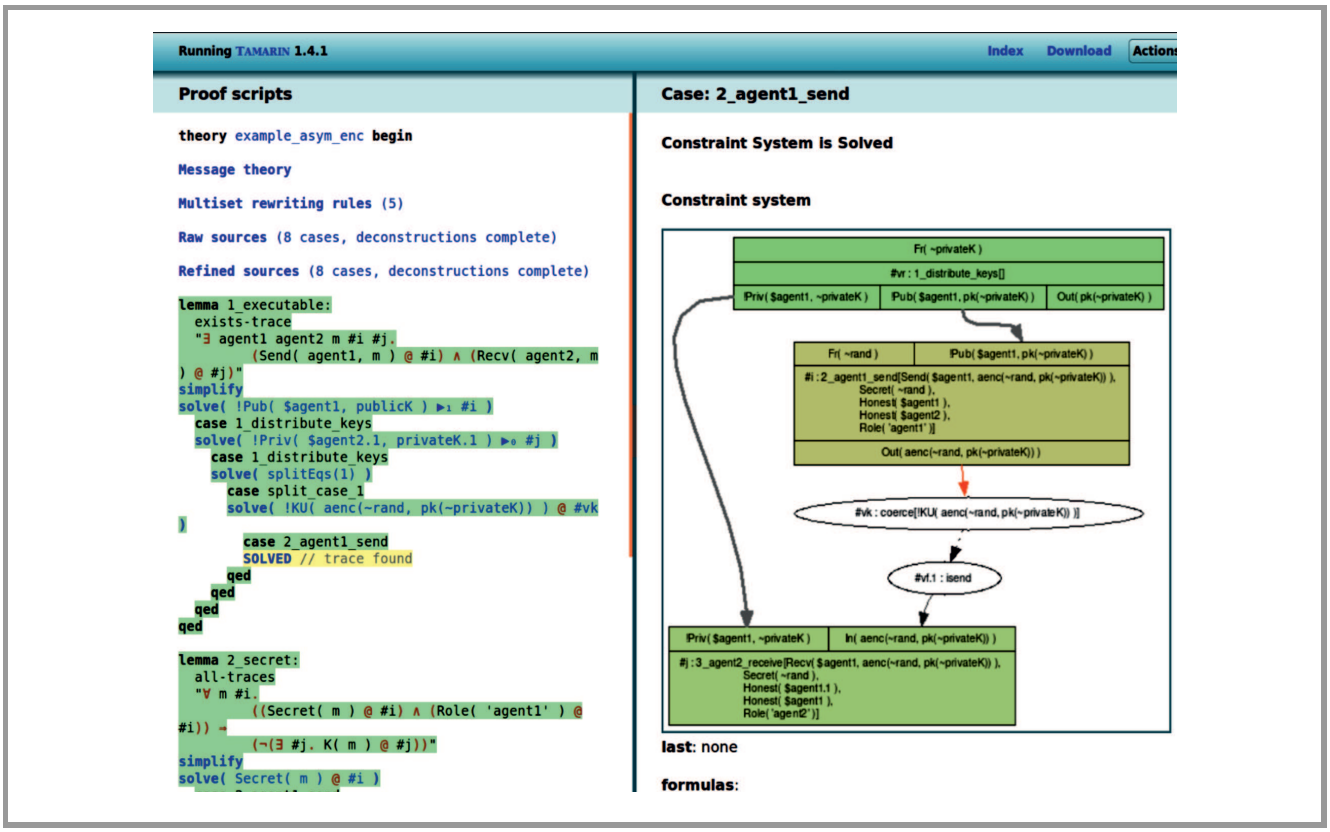


Fig. 1. Tamarin tool interactive mode.

the attacker can eavesdrop on the transmission, adding information from the message to the fact set. The foe can use knowledge about functions, terms and equations that are not marked as secret. The model also allows the attacker to rely on deduction to derive new facts, and to prepare messages with false facts it is familiar with. One may conclude that the Dolev-Yao model in which the intruder can eavesdrop on information and send modified data to legitimate agents participating in the communication process is a man-in-the-middle type of attack [12]–[14].

Recently, it has been observed that the Dolev-Yao model is not accurate enough for IoT applications. For example, it does not take into account physical attacks which are a common cause of security breaches in 5G SN and IoT networks [14]. Difficulty with modeling rather complex attacker behaviors observed in 5G SN and IoT networks is a major disadvantage as well. Here, the activities of the intruder are not always obvious. The intruder often collects sensitive information and passes it to another agent for exploitation. This makes it necessary to model the intruder's behavior related to specific incidents within sensor networks [14].

Protocols studied with the use of Tamarin are modeled by a set of messages exchanged between agents participating in the communication process. Tamarin does not take into account a scenario in which messages may be lost. It is also necessary to define constraints related to the order in which messages are transmitted using the protocol model.

In the Tamarin model, a specific rule represents the sending of a message and generates an Out(message) fact. At the same time, the message is added to the intruder's knowledge set through the embedded rules. The sending of a particular message is considered complete when the fact Out(message) is transformed into In(message). The occurrence of an In(message) fact models the arrival of the message at the destination [12]. Tamarin's operation is based on a set of translation rules that describe the behavior of the model at the inference stage. The rules specify transitions between states defined in the model of the system or the protocol under analysis. At the inference stage, all potential combinations of the defined rules are analyzed.
A simple model of an exchange of messages between two agents using asymmetric encryption is described below. The code snippet begins with the name of the model that is being proven, in this case the name is "example". The next line is the command that starts the main "begin" block, where custom and built-in functions are defined at the beginning.
The example shows only the built-in asymmetric-encryption mechanisms that are necessary for the model to work:

- p(K) – a function that generates a public key from the private key K,
- aenc(m, $pubK$) – asymmetric encryption – a function that encrypts messages m with $pubK$ key, using an asymmetric encryption algorithm,

- adec(m, K) – an asymmetric decryption function that decrypts an "asymmetrically encrypted" message m with key K.

It is worth noting that K and *pubK* are chained using the p() function, so that *pubK* = p(K). Also, if used in the way described above, i.e., with a public key for encryption and private for decryption, the model represents the confidentiality property. There is, although, no reason not to swap *pubK* and K to encrypt with a private key, thus modeling integrity/undeniability.
In Tamarin, the rules are defined based on the following notation: **[Conditions/facts preceding] −− [Action Facts]-> [Conditions/facts following]**
The main parts of the defined rules are presented in square brackets. The first one refers to the definition of conditions that must be fulfilled to initiate a given rule. The next part is used to define facts representing the given condition, used to prove lemmas. The last part defines the requirements (consecutive facts) which trigger subsequent rules. There may be an optional "let in" fragment in the syntax. This fragment primarily defines variables the use of which should improve code clarity.
In the presented example, the first key distribution rule (*1_distribute_keys*) generates a private key and a public key. The public key so created is then distributed to the two parties involved in the message exchange.

Example 1

```
1. Theory example
2. Begin
3. Builtins: asymmetric-encryption
4. Rule 1_distribute_keys:
5.  [Fr(~privateK)]-->[!Priv($X,~privateK), !Pub($X,
    pk(~privateK)), Out(pk(~privateK))]
6. Rule 2_agent1_sent:
7.  [Fr(~rand), !Pub($agent1, publicK)]--
   [Send($agent1, aenc(~rand, publicK)),
8. Secret(~rand), Role('agent1')]->[Out(aenc(~rand,
   publicK))]
9. Rule 3_agent2_receive:
10.  let received_rand = dec(encrypted_msg,privateK) in
11.  [!Priv($agent2, privateK), In(encrypted_msg)]--[
     Recv($agent2, encrypted_msg),
     Secret(received_rand), Role('agent2')]->[ ]
12. Lemma 1_executable: exists-trace
13.  "Ex agent1 agent2 m #i #j.
   Send(agent1,m)@i & Recv(agent2,m) @j"
14. Lemma 2_secret: all-traces
15.   "All m #i. Secret(m) @i & Role('agent1') @i
      ==> (not (Ex #j. K(m)@j))"
16. End
```

In the description of the rule mentioned above, there is a definition of the Fr() fact that is the random value generated. The next section with facts is omitted because it is empty (square brackets omitted). The last part refers to !Priv() and !Pub() facts, assigning public and private keys to agents 1 and 2, and defines the Out() fact. The Out() fact models the sending of a message through a public channel that is also accessible to attacking agents. In this case, the Out() fact serves the purpose of revealing the public key to the attacker. In order to indicate that it is a permanent fact, it is preceded by an exclamation mark. Such a fact is read, but not consumed, i.e. it does not disappear from the set of available facts when the rule containing it on the left is triggered.
In the next code snippet, the second rule (2_agent1_sent) is defined. This rule accepts the public key, generates a random number and sends it, in an encrypted form, along with the public key. Using the Fr() fact, a random number is generated by the Pub() fact, the public key is accepted, and using the Out() fact, the encrypted random number is sent. The next section of this rule defines the following facts: Send() – the fact reports that *agent1* has sent a message, Secret() – the fact specifies that the value should be secret, Role() – the fact that specifies an identity.
In the (3_agent2_receive) rule, the following facts are defined: In() to receive the message, Priv() to retrieve the private key and adec() function to decode the message. The decrypted message is substituted by the received_rand parameter in the let-in block. The next part of this rule is similar to the previous one, except that the Recev() fact is used to inform that the message was received by *agent2*.
Although this is not shown in the example, it is possible to specify constraints when the rule definitions are presented. The constraints are triggered by the use of action facts. Their purpose is to refine the set of further analyzed protocol usage scenarios, for instance ensuring that *agent1* and *agent2* are always two distinct agents. Then the lemmas that Tamarin will require to be met to prove protocol security are defined. Two lemmas are used to check the feasibility of the model defined in the example and to check whether confidentiality of the transmitted messages is ensured. The first one is used to check if there exists *agent1* that sends a message to *agent2* (whether the model allows, at the very least, communication between agents) and the second one checks if value n, being the secret of *agent1*, has not been disclosed to the attacker, represented by the absence of fact K(n) at any given time j in the public channel. In Tamarin, two modes of lemmas can be identified, namely "exist-trace" and "all-trace". The first requires one condition to be satisfied. The other lemma requires that all conditions be satisfied.
In Tamarin, variables are defined by specific prefixes: ∼ denotes a new variable, defined by the Fr() fact, $ denotes a variable that is public and thus does not need to be entered by Fr(), # denotes a temporary variable. The lack of a prefix before a variable name indicates a variable to which the message being sent is rewritten [12].

4. Simple Example of Error Detection

In this section, the Diffie-Hellman (DH) algorithm will be demonstrated, allowing two agents to establish a shared symmetric encryption key without the agents having to transmit any secret data over an untrusted network. A detailed description of the algorithm and its properties can be found in [15]. The presented model is a modification of the solution provided by the Tamarin authors [10], [12].

The aim of the modification was to prove that when adding the second half of the Diffie-Hellman algorithm, the two sides of communication will create the same session key. However, Tamarin responded to such a model by means of a counter example, so this modification demonstrates how the software can be used to identify initially unnoticeable reasoning errors while modeling new protocols. The following section also explains the exact meaning of symbolic analysis in the context of modeling security properties.

At the beginning of the model, information is provided on the types if built-in functions and cryptographic primitives used. For the DH algorithm it will be: "diffie-hellman built-in". Adding this element gives the model access to the cryptographic basis of the DH algorithm – the power operator, "^". As the analysis performed by Tamarin is of the symbolic variety, it only means that:

- the operator symbol "^" has appeared in the model. It has no assigned function,
- symbol "^" satisfies the following relationship: $(a \hat{} b) \hat{} c = (a \hat{} c) \hat{} b$.

Due to the fact that Tamarin performs symbolic analysis in the protocol model, it is impossible to determine the values of variables or the implementation of functions. Only their mutual dependencies are defined for functions. The variable values are not considered in Tamarin either. In other programming languages, no different types of variables (terms) exist. In Tamarin, a term can be understood as a symbolic name assigned to a value. Variables in Tamarin, however, have their sorts specifying their properties, mainly in the context of security.

The next step is to define function symbols, the number of arguments for each of them, and information who can use such functions. Again, since Tamarin carries out a symbolic analysis, only a name needs to be defined. There is no way to determine, nor a need to specify what the implementation of the function is. We know as much about the function as follows from the mathematical definition of this concept. The model uses: mac/2, g/0, macKey/0 [private]. Note that if the function is [private], the attacker cannot deduce its value for given arguments, even if he knows all of them. As a result, a 0-argument (i.e. constant) macKey function will remain unknown to the attacker, unless sent unsecured in a network message.

Labeled translation rules constitute another peculiar element. The model proposed in [3] takes into account a half of the process of determining the shared session key, i.e. encrypting the private key by one of the parties, sending it over the network to the other party, and combining the received value with its private key through the other party in order to obtain the session key. In order to enable the parties to authenticate each other, the transmitted message will have a MAC code built on the basis of a secret shared MAC key (macKey constant) known only to the parties, but not to the attacker. The model consists of the following rules shown in Example 2.

Example 2

```
1. // Model of the first step of the DH algorithm. A sends
   // its ''encrypted'' private key to B
2. Rule Step1_A_sends_encrypted_private_key_to_B:
3.  [Fr(threadId:fresh), Fr(privateKeyA:fresh)] --[ ]->
    [Out(<g^(privateKeyA:fresh),mac(macKey,
    <g^(privateKeyA:fresh), A:pub, B:pub>)>),
    SendingAnEncryptedPrivateKey(threadId:fresh,
    A:pub, B:pub, privateKeyA:fresh)]
1. // The second step model of the DH algorithm. B takes
   // A's ''encrypted'' private key and ''encrypts'' it with
   // his private key to obtain the session key.
2. Rule Step2_B_accepts_encrypted_private_key_A:
3.  [SendingAnEncryptedPrivateKey(threadId, A, B,
      privateKeyB:fresh),
4.  In(<encryptedPrivateKeyA, mac(macKey,
      <encryptedPrivateKeyA, B, A>)>)]
5.  --SessionKeyApproval(threadId,
      encryptedPrivateKeyA^(privateKeyB:fresh))]->[]
```

The first rule allows an initiator of communication with a publicly known name A to start the process of negotiating the session key with a recipient with a publicly known name B. Both identifiers are stored in public variables which do not guarantee the uniqueness of the values at the time of their introduction. To start the negotiation process, initiator A obtains a fresh or a secret unique value for the protocol thread identifier (variable *tid*) and a fresh value for variable x, representing the private key of A. In the same step, A "encrypts" its private key by raising constant g to the power of x, and then places the obtained value in the message.

Occurrence of the *SendingAnEncryptedPrivateKey* fact in a multiset together with a message containing $g \hat{} privateKeyA$ triggers the second rule. This rule means that if the system state elements are: the *Sending AnEncryptedPrivateKey* fact and a message containing this "encrypted" key, the recipient indicated by the sender can receive the message and process the $g \hat{} privateKeyA = encryptedPrivateKeyA$ value using his private *keyPrivateKeyB* (which is a fresh variable) as follows: $g \hat{} privateKeyA \hat{} privateKeyB$.

The recipient assumes that this value is the session key, taking into account that if it swapped the roles with the other side, the other side would get the same key.

In the second rule, the verification of the MAC code message is implicitly written. The requirement for the existence of the In() fact conforms to the fact produced in the first step only if this fact contains a message with the MAC signature based on the constant macKey. This means that the binding of the parameters of the Out() fact generated by the rule: *Step1_A_sends_encrypted_private_key_to_B* and the In() fact required by the rule: *Step2_B_accepts_encrypted_private_key_A* can be understood as rewriting the value from the source variable – a parameter of Out() to the target variable – a parameter of In(). The macKey is the exception here, as it is not a variable, but a constant. Therefore, its value must be the same in the fact consumed by the *Step2_B_accepts_encrypted_private_key_A* rule and in

the fact present in the multiset. This technique is known in computational logic as pattern matching.
The next step is to write down the lemma to specify the security-related protocol properties. The lemma shown as Example 3 states that the attacker has no way of knowing the value negotiated as the session key at any time, even before it is actually agreed upon.

Example 3
```
1. Lemma Session_key_is_never_revealed: all-traces
2. "All #t1 #t2 threadId sessionKey.
     SessionKeyApproval(threadId, SessionKey)
3.   @ #t1 & K(sessionKey) @ #t2 ==> F"
```

The following lemma can be read as such. In every possible trace produced by Tamarin, the condition should be preserved for any moment in time *#t1* and *#t2* and for any *thread id* of *threadId* and any session key *sessionKey*, if the recipient has accepted the session key at any time *#t1* and the opponent knows this key at any time *#t2*, then this is an impossible situation (empty predicate set, logically always equal to untruth).
Since the right side of the implication is always false, the only possibility for the entire implication to be true is if the left side of the implication is false as well. So, each trace produced by Tamarin should conform to Example 4.

Example 4
```
1. "All #t1 #t2 threadId sessionKey. Not
     (SessionKeyApproval (threadId,
2.   sessionKey) @ #t1 & K(sessionKey) @ #t2)"
```

It may be formulated in a much simpler way. In none of the produced scenarios (traces) there is such a value that would be accepted as the session key and would be known to the attacker.
As an experiment, we will try to corrupt the protocol by introducing a rule that will expose a macKey constant over the unsecured network at any time Tamarin finds suitable (i.e., with no time nor causal restrictions), giving the opponent access thereto. The interpretation of this fact may be as follows: knowing the macKey constant, the adversary can spoof the MAC code in any message and, consequently, impersonate the initiator. The recipient, thinking that it is negotiating the session key with the initiator, will accept it after the negotiation. However, this will be the session key established not between the initiator and the recipient, but between the attacker and the recipient. The attacker, as one part to the communication process, knows the session key negotiated, which leads to the falsification of lemma Session_key_is_never_revealed.
The third rule that models such a leak is:

Rule mac_key_reveal: [] -- [macKeyReveal()] -> [Out(macKey)]

It has a blank left side, so it can be executed at any time.
As expected, as a result of the analysis of such a protocol model by Tamarin, we obtain an example of a trace for which the lemma Session_key_is_never_revealed is not true. Tamarin has thus shown a relationship between the secrecy of two seemingly unrelated protocol elements: the MAC key and the negotiated session key.
Further, a less security-stringent version of the correctness condition can be used by replacing the lemma: Session_key_is_never_revealed by: If_the_mac_key_does_not_leak_before (Example 5).

Example 5
```
1. Lemma If_the_mac_key_does_not_leak_before: all-traces
2. "All #t1 #t2 threadId seassionKey.
     SessionKeyApproval(threadId, sessionKey)
3.   @ #t1 & K(sessionKey) @ #t2 ==> Ex #t3.
     macKeyReveal() @ #t3 & #t3 < #t1"
```

This lemma is constructed so that it is acceptable for the attacker to know the session key, but only if before accepting this key as the session key, the adversary gets to know the macKey (revealing the macKey is equivalent to the macKeyReveal() label in the trace). In this case, Tamarin shows the correctness of the protocol – the fulfillment of the lemma for all possible traces. In practice, this proof can be interpreted, as the macKey must remain secret, but only until the parties determine the session key. Afterwards, its secrecy does not matter.
Here, an attempt will be made to address the problem of implementing only "half" of the DH algorithm. It will be shown that if B accepts session key x in communication with A and A accepts session key y in communication with B, then keys x and y will be identical.
For this purpose, the first step is to add the *SessionKeyApproval* label informing who accepted the key and in communication with whom. Therefore, it is necessary to modify the *Step2_B_accepts_encrypted_private_key_A* rule so that is assumes the form given as Example 6.

Example 6
```
1. Rule Step2_B_accepts_encrypted_private_key_A:
2.  [SendingAnEncryptedPrivateKey(threadId, A, B,
    privateKeyB:fresh),
3.  In(<encryptedPrivateKeyA, mac(macKey,
       <encryptedPrivateKeyA, B, A>)>)]
4. --[SessionKeyApproval(threadId,
       encryptedPrivateKeyA^(privateKeyB:fresh),
       A, B)]-> []
```

This change does not affect the operation of the protocol, but only the set of information available for the purpose of inference. Therefore, verification of the previously used lemmas is not performed and a new one is introduced – see Example 7.

Example 7
```
1. Lemma Both_sides_get_the_same_session_key: all-traces
2.  "All #t1 #t2 threadId1 threadId2 sessionKey1
    sessionKey2 A B.
3.  (SessionKeyApproval(threadId1, sessionKey1, A, B)
       @ #t1 &
4.  SessionKeyApproval(threadId2, sessionKey2, B, A)
       @ #t2) ==> (sessionKey1 = sessionKey2)"
```

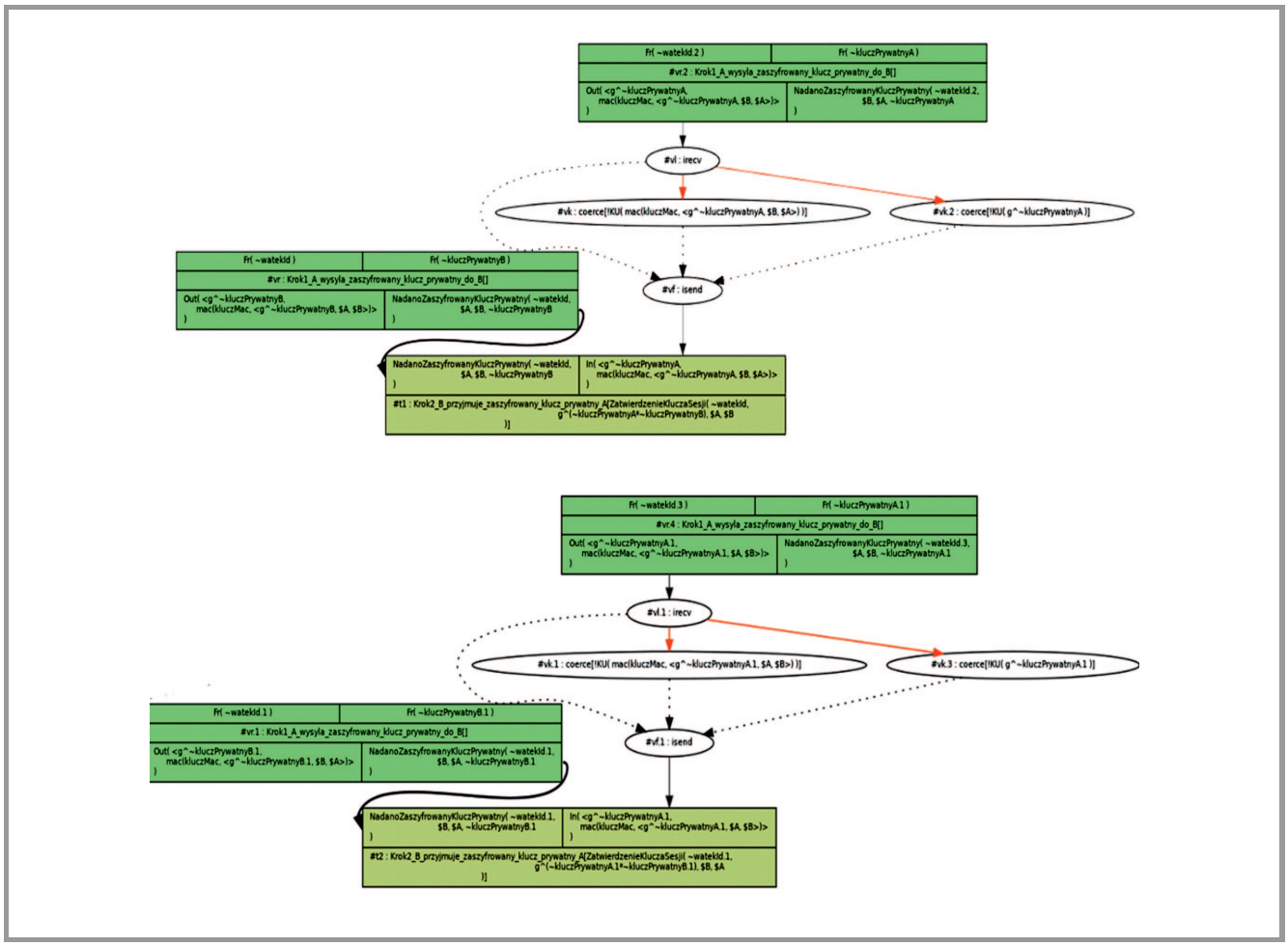


Fig. 2. A trace that does not meet the lemma – visualization of the attack.

It should be read as follows. If any party A has accepted the *sessionKey1* session key negotiated with B (at any time *#t1*) and party B has accepted the *sessionKey2* session key negotiated with A (at any time *#t2*), then both keys are equal. Tamarin is able to construct a counterexample for such a lemma, which means that assembling the "entire use case" of the DH algorithm from "two halves" does not ensure the correct operation of the protocol. In consequence, a situation is feasible where the parties using the same algorithm can generate different session keys.

To read such an attack scenario, it is necessary to examine the trace constructed by Tamarin (Fig. 2).

An explanation of the graphic notation used by Tamarin is required here. In rectangular frames, Tamarin shows the rules used. The notation $x.1$, $x.2$, $x.3$, $x.4$ refers to those variables that appear, in the rules, under the same name x, but have different values in specific rule instances (one rule can be used many times). The ovals show the opponent's actions based on the rules built into the opponent's model in Tamarin. In the trace of the attack scenario, we see four instances of the *Step1_A_sends_encrypted_private_key_to_B* rule: two resulting from the interaction of side A with side B, and two resulting from the interaction of side B with side A. It is worth noting that each time such a rule is applied, the private key of the initiating party is drawn anew (fresh variable). So, we have two sides A and B and four private keys: x, $x.1$, y and $y.1$. Either because of the attacker or because of a network property that does not keep the order of messages, sessions start to intertwine. As a result, party B (peer of the first session) accepts the session key from its first private key (y) and the second private key $A(x.1)$, while party A (peer of the second session) accepts the session key generated from its first private key ($x.1$) (x) and the second private key $B(y.1)$. On the A side, the $g^{\wedge}y^{\wedge}x.1$ session key is created and on the B side – $g^{\wedge}x^{\wedge}y.1$. Even if the ,$\wedge$ operator is used, two session keys will exist: $g^{\wedge}x^{\wedge}y.1$ and $g^{\wedge}x.1^{\wedge}y$, and since x and y are fresh, this guarantees that values x, $x.1$, y, and $y.1$ are different. Consequently, keys $g^{\wedge}x^{\wedge}y.1$ and $g^{\wedge}x.1^{\wedge}y$ accepted as session keys by B and A, respectively, are different. This is in contradiction to the lemma.

This problem can be solved by limiting the application of the *Step1_A_sends_encrypted_private_key_to_B* rule to the same initiator A and communication partner B at any given time. In other words, A is allowed to initiate negotiations with B only once in the entire trace. To do this, a label needs to be added first to the rule shown in Example 8.

Then, an auxiliary mechanism is relied upon to determine which traces generated by Tamarin can be considered worthy of further analysis. This mechanism is called a restric-

Example 8

```
1. Rule Step1_A_sends_encrypted_private_key_to_B:
2.   [Fr(threadId:fresh), Fr(privateKeyA:fresh)]--
3.   [StartEstablishingSessionKey(A:pub, B:pub)]->
     [Out(<g^(privateKeyA:fresh),
        mac(macKey, <g^(privateKeyA:fresh), A:pub,
        B:pub>)>),
4. SendingAnEncryptedPrivateKey(threadId:fresh,
        A:pub, B:pub, privateKeyA:fresh)]
```

tion. In this case, the restriction should allow only one label named *StartEstablishingSessionKey* to be present in the trail for a given initiator *A* and peer *B* (Example 9).

Example 9

```
1. Restriction Only_one_session_between_A_and_B:
2.  "All A B #t1 #t2. (StartEstablishingSessionKey(A, B)
      @ #t1 &
3.    StartEstablishingSessionKey(A, B)  #t2) ==> #t1
       = #t2"
```

Constraints are constructed similarly to lemmas. The above limitation can be understood as follows: for any initiator *A* and its peer *B* and for any moments in time *#t1* and *#t2*, if at time *#t1* and at time *#t2* there is a label *StartEstablishingSessionKey*(*A*, *B*) in the trace, *#t1* and *#t2* are always the same moment. For such a supplemented protocol model, Tamarin is no longer able to produce an attack scenario. Examples of the implementation of common restrictions can be found in [12].

5. Model of DTLS 1.2 Handshake

In this section, the use of the Tamarin prover to model security protocols and to verify their correctness is presented. The analysis concerns the DTLS 1.2 handshake protocol with the optional TCP SYNCookies mechanism modeled by the authors on the basis of documentation [19] and models provided by the Tamarin tool developers and Jun Kim [3]–[5].

In TLS/DTLS, a handshake is a step that is meant to negotiate symmetric encryption-decryption keys by the parties to the connection, one for each of them, without any confidential information being transmitted over an unsecured connection. The use of symmetric cryptography as a means of securing the connection and asymmetric cryptography only as a means of securely establishing symmetric keys stems from significant differences in the speed of operation and, hence, the device load. Asymmetric algorithms are, in general, more resource-demanding, but they guarantee a higher level of security. Such a mixed approach is especially appealing to applications in resource-constrained IoT environments.

The TLS handshake, in both version 1.2 and 1.3, has a certain disadvantage that is impossible to leave out when discussing IoT applications. These mechanisms are security measures for TCP connections established at the transport layer and both rely on the properties of this protocol. The use of stream ciphers for security-related data is only possible if the lower layer guarantees the delivery of packets in the exact order in which they were sent. Effective execution of the handshake in a short time frame will only be possible if the lower layer ensures the retransmission of those messages for which no confirmation has been provided in over an extended period of time. On the other hand, as far as the transport layer is concerned, protocols within the TCP/IP stack, UDP may be particularly interesting for IoT devices. While TCP requires the IoT device to store the connection and application state in its memory, UDP requires connection the state-related information only. At the same time, preventing the datagrams from being lost and reordered does not seem too complicated, provided that it is necessary at all. Therefore, the use of UDP reduces the amount of system resources used, being a big advantage of a protocol designed to work on IoT devices. It should also be emphasized that saving one round trip time per handshake when switching from TLS 1.2 to TLS 1.3 is a relatively poor gain in a scenario where it is always necessary to establish the TCP connection first when performing another 3-way TCP handshake.

The DTLS protocol is a solution that combines the security-related advantages of TLS with the simplicity of UDP. This protocol was created by extending TLS, i.e. allowing it to function properly on a transport layer that uses datagrams instead of connections. The essence of such a solution is to provide a mechanism that is as similar to TLS as possible, but copes with the characteristics of datagrams, i.e. the potential of datagrams getting lost in the network and changing the order of their delivery. DTLS compensates for the deficiencies of the UDP protocol only when performing a handshake. In DTLS handshake messages, there is a field for a sequence number, similar to that in TCP packet headers. DTLS also introduces timers responsible for measuring the time provided for the response of the other party.

When a predetermined value is reached, the last sent message is retransmitted, since either the message, or the response to it has been lost. Additionally, due to the fact that a single authentication certificate for one or both communication sides does not fit into a single datagram, DTLS supports protocol message fragmentation and reassembly at the receiver. In addition to establishing a DTLS "connection", i.e. after negotiating the session keys, the protocol maintains the properties of datagrams while handling application traffic.

As a result, the data that the applications send through the now-secured communication channel is encrypted in datagrams and is sent similarly to normal unsecured traffic. Encrypted datagrams can also get lost and can change their order during a transmission. The task of dealing with these anomalies is handled, however, by the application. DTLS only needs to ensure that the receiver can perform the decoding operation correctly, regardless of any missing and reordered datagrams. This is done by completely abandoning the use of stream ciphers in favor of stateless ciphers.

In addition, in order not to raise any alarms in an overzealous manner, thus unnecessarily breaking a secure session in the event of receiving a message that is not successfully decoded, DTLS adds an epoch number to each message that encapsulates the application data. The epoch is the period over which the same session keys are valid. The re-negotiation of these keys serves as a boundary of such an epoch. So, if a message sent by the client ending the handshake is overtaken, for example, by a message with the application data sent later, the server-side protocol may consider such a situation safe, because it has the ability to decrypt the application data (the handshake has already ended), and the epoch number in the decrypted datagram is greater, by one, than during the handshake. A similar situation occurs with the assumption that session keys are changed periodically and that a limited number of packets or bytes can be sent with a single key.

In the presented DTLS 1.2 handshake model, aspects related to mechanisms used for compensating for loss and reordering of datagrams may be considered as working correctly and securely because they are equivalent to the mechanisms known from TCP. Epoch numbers, in turn, are a mechanism used, in particular, during the exchange of application data. It follows that the DTLS handshake model prepared for Tamarin should be very similar to the TLS handshake model if the same protocol versions are compared.

The DTLS handshake model proposed in [3] is, in fact, similar to the TLS handshake model [5]. The first difference is that each handshake message is labeled with a HMAC signature. We model this signature by introducing a 1-argument HMAC function symbol unbound by any equality features. The attacker can compute the hash knowing all the required arguments, but cannot deduce arguments knowing the hash only. The TLS model equivalents of C_1 and S_1 rules will therefore initially look as shown in Example 10.

Example 10

```
1. Rule C_1:
2.  [Fr(~nc), Fr(~sid)]--[]->[Out(<$C, ~nc, ~sid, $pc,
      HMAC(<$C, ~nc, ~sid
3.      $pc>)>), St_C_1($C, ~nc, ~sid, $pc)]
1. Rule S_1:
2.    [In(<$C, nc, sid, pc, HMAC(<$C, nc, sid, pc>)>),
        Fr(~ns)]--[]->[Out(<$S, ~ns, sid, $ps>),
3.       St_S_1($S, $C, sid, nc, pc, ~ns, $ps)]
```

However, this model does not take into account one key aspect: the mechanism of preventing DoS attacks on IoT devices. A ready-made mechanism of this type, TCP SYN Cookies [18], can be used when the security protocol is based on the TCP transport layer. In the case of DTLS, a similar mechanism must be built into the security protocol. The specification states that implementation of this mechanism is optional. On the other hand, defense against DoS-type attacks in the case of an IoT network is a problem of such great importance that the mechanism in question should definitely be taken into account in the model. It is so due to the fact that a much greater number of "small" devices is expected to be present in an IoT network than in a typical computer network.

The solution proposed in [19] introduces a preliminary step into the protocol, preceding the sending of the *ClientHello* message by the client. Its task is to inform the server about the intention of establishing a secure session, without the server having to consume any resources. The server authorizes such an attempt to start a conversation by placing a cookie – a hashed value that contains connection parameters and a secret key known to the server only. Then, the client which acts, after receiving the answer, according to the standard procedure, confirms its authorization by attaching the previously received cookie to the *ClientHello* message. Only after receiving such a message and after verifying the correctness of the cookie, does the server allocate resources for the purpose of creating a secure session.

If it is only the server that allocates resources as a result of a complex client interaction, the attacker must maintain a sufficiently large number of fully functional malicious clients in order to consume all server resources. It may turn out to be unprofitable, unlike in a situation in which such protection is not used. Then the attacker would only need to craft an appropriate number of malicious *ClientHello* messages.

Adding a preliminary step to the model [5] requires defining two additional rules: C_0 and S_0.

The client sends the *ClientHello* message without a cookie. The server does not need to support the DoS protection mechanism, so it can respond immediately with the *ServerHello* message, and it can also request additional verification in the form of *ClientHelloVerify* – see Example 11.

Example 11

```
1. Rule C_0:
2.  [Fr(~nc), Fr(~sid)]--[]->
    [Out(<'client_hello', $C, $S,
    ~nc, ~sid, $pc,
3.    HMAC(<'client_hello', $C, $S, ~nc, ~sid, $pc>)>),
       St_C_0($C, $S, ~nc, ~sid, $pc),
4. // The client in this state can receive the ServerHello
   // directly St_C_1($C, $S, ~nc, ~sid, $pc),
5.    CookieFreeSession(~sid)
6. /* No match in the protocol; the server will decide
   /* whether or not to use cookies in this session by
   /* deleting or not deleting this fact when receiving the
   /* client_hello message
7. */]
```

The server can respond to *ClientHello* without a cookie by sending *ClientHelloVerify* with a freshly generated cookie. This decision is to be made by the server once per handshake (Example 12).

There are some further improvements introduced to the rules presented above. As a result of constant evaluation of the evolving DTLS 1.2 handshake model and based on the assessment of the produced counter-examples, the decision was made to:

- include both sender information and receiver information in each message to be sent,

Example 12

```
1. Rule S_0:
2.  Let cookie = h(<C, S, sid , ServerSecret(S)>) in
3.    [In(<'client_hello', C, S, nc, sid, pc,
        HMAC(<'client_hello', C, S, nc, sid, pc>)>),
4.    CookieFreeSession(sid)]--[ AssignsCookie(sid, S,
        C, cookie)]->
5.    [Out(< 'client_hello_verify', S, C, sid, cookie,
        HMAC(<'client_hello_verify', S, C,
6.          sid, cookie>)>)]
```

- include, in the argument, client and server state facts, not only session identifiers and sids, but also information about the other party,
- introduce the *CookieFreeSession*(*sid*) fact to ensure a constant decision on whether or not to use the cookie throughout the entire handshake. It is pointless for the server to first require the cookie and then resign from its verification.

Furthermore, due to the fact that the messages from the first stage of the handshake (*ClientHello* in the variant with or without cookies and optional *ClientHelloVerify*) require the type of message to be specified (the first field in the message containing a descriptive constant), the rules responsible for receiving them also need to be modified. The same applies to messages exchanged later.
The client sends *ClientHello* again, this time with a cookie. The condition for this rule to work is that the client has previously sent *ClientHello* without the cookie. So, there is a state from which it can recover the connection parameters, see Example 13.

Example 13

```
1. Rule C_1:
2.  [In(<'client_hello_verify', S, C, sid, cookie,
        HMAC(<'client_hello_verify',
          S, C, sid, cookie>)>),
3.  St_C_0(C, S, nc, sid, pc)]--[ResendsCookie(sid, C, S,
        cookie)]->
4.  [Out(<'client_hello_with_cookie', C, S, nc, sid, pc,
       cookie,
5.      HMAC(<'client_hello_with_cookie', C, S, nc,
          sid, pc, cookie>)>),
6.  St_C_1(C, S, nc, sid, pc)]
```

A rule that models the server's response to *ClientHello* with a cookie (Example 14).

Example 14

```
1. Rule S_1:
2.  Let cookie = h(<C, S, sid, ServerSecret(S)>) in
3.  [In(<'client_hello_with_cookie',;C, S, nc, sid, pc,
      cookie,
4.     HMAC(<'client_hello_with_cookie', C, S, nc,
         sid, pc, cookie>)>), Fr(~ns)]--
5.  [VerifiesCookie(S, C)]->
       [Out(<'server_hello', S, C, ~ns, sid, $ps
6.        HMAC(<'server_hello', S, C, ~ns,
            sid, $ps>) >),
7.      St_S_1(S, C, sid, nc, pc, ~ns, $ps)]
```

The *ServerSecret*/1 function symbol annotated as [private] has been introduced to generate the cookie. It returns, for each server, a secret value, and since it is a private symbol, the opponent cannot know this value for any of the servers appearing in the trace. That is, of course, if the value for the server is not sent directly over an unsecured network at any time. The second and third steps of the handshake remain generally unchanged, except for adding descriptive constants to messages, as well as sender and receiver identifiers to both messages and state-facts. It is also crucial that the rules have been supplemented with additional labels that will allow a conclusion on the use and verification of the cookie to be made later: Assigns Cookie(who, to_whom), ResendsCookie(who, to_whom) and VerifiesCookie(who, from_whom). The labels specified above are related to generating and assigning a cookie to the connection, sending the received cookie back to the server and verifying it by this server, respectively.
In order to best adapt the model described here to the requirements of the DTLS 1.2 specification, the possibility of the server deciding not to use the *ClientHelloVerify* mechanism was introduced. The server then responds to the *ClientHello* message without a cookie directly with the *ServerHello* message. To model this, an alternative version of the *S*_1 rule was introduced, known as *S*_1_no_cookie (Example 15).

Example 15

```
1. Rule S_1_no_cookie:
2.  [In(<'client_hello', C, S, nc, sid, pc,
       HMAC(<'client_hello', C, S, nc, sid,
3.    pc>)>), Fr(~ns), CookieFreeSession(sid)]--[]->
     [Out(<'server_hello', S, C,
4.     ~ns, sid, $ps, HMAC(<'server_hello', S, C, ~ns,
         sid, $ps>)>),
5.  St_S_1(S, C, sid, nc, pc, ~ns, $ps)]
```

6. Security of DTLS 1.2 Handshake

The model designed in the manner described above, based on the proof performed by Tamarin, is equivalent in terms of security to the TLS 1.2 handshake and TLS 1.3 handshake protocols, if the same lemmas are used for comparison.
The Meier's model [3] defines three lemmas that check the correctness-related properties of the TLS 1.2 handshake. The first lemma requires that both *keyS* and *keyC* session keys be secret, both from the client's and the server's point of view. In the Tamarin syntax, this is stated in the manner presented in Example 16.

Example 16

```
1. Lemma DTLS_session_key_secrecy:
2.  "not(Ex S C keyS keyC #k.SessionKeys(S, C, keyS,
       keyC) @ k & ( (Ex #i.
3.     K(keyS) @ i) | (Ex #i. K(keyC) @ i)) & not (Ex
       #r. RevLtk(S) @ r) & not (Ex #r. RevLtk(C) @ r)
4.     )"
```

The above notation can be interpreted that it is impermissible for both communication parties *S* and *C* to generate a pair of session keys that would be known at any stage of the task execution process by an attacking intruder without leaking the private key of *S* and leaking the private key of *C*. This property allows us to maintain four common high-level security features: confidentiality, integrity, availability and non-repudiation.

The second lemma (Example 17) is used to model the behavior of injective consensus ownership. This condition allows for the unconditional possibility of establishing a secure connection even when the adversary is able to prepare malicious messages and send them over the network.

Example 17

```
1. Lemma injective_agree:  all-traces
2.  "All sid actor peer params #i. Commit(sid, actor,
       peer, params) @ i ==>
3.  (Ex #j. Running(sid, actor, peer, params) @ j & j <
       i & not(Ex actor2 peer2 #i2.
4.       Commit(sid, actor2, peer2, params) @ i2 &
          not(#i = #i2))) (Ex #r. RevLtk(actor) @ r) |
5.       (Ex #r. RevLtk(peer) @ r)"
```

The lemma from Example 17 can be interpreted as follows: if an actor claims that it can send application messages via a secure channel to the peer, then such a peer had to be seen before as working (started a handshake) with identical parameters and, in addition, there is no other pair of *actor2* and *peer2* that would use the same connection parameters (session keys). The process of ensuring that *actor != actor2* and *peer != peer2* is performed by searching for Commit(*actor2*, *peer2*, ...) at a different moment than the one at which Commit(*actor*, *peer*, ...) appeared. The "!=" operator means "not equal". All of the above conditions must be fulfilled, unless there has been a leak of an actor's private key or a peer's private key.

The last lemma (Example 18) was introduced for the purpose of assessing the feasibility of the protocol, and therefore its correctness, not necessarily in terms of security. It requires the ability to establish a secure connection while satisfying all conditions defined at the rule level, without revealing keys.

Example 18

```
1. Lemma DTLS_session_key_setup_possible:  exists-trace
2.  "(All x y #i. Eq(x,y) @ i ==> x = y) & (Ex S C keyS
       keyC #k. SessionKeys(S, C, keyS, keyC) @ k &
3.       not (Ex #r. RevLtk(S) @ r) & not (Ex #r. RevLtk(C)
       @ r)
4.       )"
```

This lemma can be interpreted as follows. There is at least one possibility that any communication party *C* can negotiate session keys *keyS* and *keyC* by communicating with party *S* without revealing the private key of party *S* or the private key of party *C*. Additionally, the combination of two elements labeled Eq() must always imply equality between them. Tamarin makes it possible to show that all three lemmas are satisfied for the Meier model. In addition, it is necessary to introduce one more lemma to test the correctness of the *ClientHelloVerify* mechanism (Example 19).

Example 19

```
1. Lemma
 server_accepts_connections_only_from_clients_with_valid
 _cookie: all-traces
3.  "All sid S C Cparams Sparams #t0 #t2 #t3 cookie.
4.     (Commit(sid, S, C, Sparams) @ #t2 &
          Commit(sid, C, S, Cparams) @ #t3 &
          AssignsCookie(sid, S, C, cookie) @ #t0)
5.     ==> Ex #t1. (VerifiesCookie(sid, S, C, cookie)
            AssignsCookie(sid, S, C, cookie) @ #t0)
                (@ #t1 & (#t1 < #t2) & (#t1 < #t3)
6.              & (#t0 < #t1))"
```

It can be read as follows. Each server in communication with any client can confirm the successful establishment of a secure session, which is also confirmed by the client, but only on condition that it previously received a valid cookie from that client.

The fulfillment of this lemma could not be proven at first. Although no evidence of Tamarin looping while in operation was observed, Tamarin always exhausts its entire allocated RAM while producing a proof.

Hypothetically, this lemma is non-provable under the researched conditions due to increased model complexity which stems from the server being allowed to choose whether or not to use the cookie. To limit the degree of complexity, one of the choices may be forced, with the use of the cookie being the preferred solutions. Such an extortion could be introduced into the model by means of a imposing a constraint on the "correct" trace, i.e. the trace that is subject to further analysis, requiring that each server uses a cookie at least once in every conversation with the client – see Example 20.

Example 20

```
1. Restriction server_required_to_use_cookies:
2.  "All S C MS Skey Ckey #t2. Running(S, C,
       <'server', MS, Skey, Ckey>) @ #t2
3.     ==> Ex #t1. VerifiesCookie(S, C) @ #t1"
```

Additionally, a counter-example encountered during the model development phase shows that a constraint is needed in which the operations of sending and verifying cookies between server *S* and client *C* are possible only when *C* != *S*. In other words, the execution trace is subject to further analysis, provided that no client is present in it as a server in the same session – see Example 21.

Example 21

```
1. Restriction no_self_session_when_running:
2.  "All sid S C params #t. Running(sid, S, C, params)
3.   @ #t ==> not(S = C)"
```

However, we decided to take a more radical path and divide the presented model into two branches. One with the server never using a cookie, and the other unconditionally requiring the server to do so. Such an approach is practically justified, since we do not see any reason for the server

accepting the DTLS "connection" to allow a no-cookie and cookie handshake simultaneously.
The cookie-free solution is created by commenting the *S_0* rule, which allowed the server to send the *Client HelloVerify* message. This solution meets the first three lemmas. It is clear that the fourth lemma, server_accepts_connections_only_from_clients_with_valid_cookie, is always fulfilled for the no-cookie model, as the left side of the implication can never be true.
Instead, a cookie-enabled solution is created by commenting the *S_1_no_cookie* server rule, allowing it to respond with a *ServerHello* message to a *ClientHello* message without a cookie. In addition, the client's ability to move from the "*ClientHello* sent" state to the "*ServerHello* received" state was blocked. This solution meets the first three lemmas. In the case of the fourth lemma, it produces a counterexample which can be understood as follows:

- Client *C* starts a handshake with server *S*, thus initiating session *sid*. During that handshake, the *ServerHello* message is spoofed. The client receives a *ServerHello* message from an attacker with maliciously planted cryptographic parameters. This leads the client to believe that the handshake with the impostor may be continued and, as a result, the client claims the establishment of DTLS session keys with the host it considered to be *S*.
- Much later, server *S* receives a valid *ClientHelloWithCookie* message and claims it has verified the cookie in session sid with client *C*. This is a violation of the fourth lemma's time constraints. In parallel, there is another handshake in progress between server *S* and client *C*, using session identifier *sid.1*. Note that the attacker has enough information sniffed during the aforementioned handshake to spoof the *ClientHandshakeFinished* message of the current handshake in order to make it look like part of session *sid*. Also, server *S* is still waiting to finish the handshake with client *C* as part of session sid. As a result, the attacker can lead the server to believe it has negotiated a secure session sid with *C*, which is a prerequisite of the fourth lemma. And because client *C* claims establishing a secure session sid before the server verifies the cookie, lemma four is not fulfilled.

However, the attacker's actions are enough for the client to establish a safe session with the attacker, with the latter thinking he is talking to the server. The attacker can now position himself between the client and the server using the classic man-in-the-middle attack. This is in line with the proposed protocol model, as it examines the possibility of leakage of session keys and not the possibility of the attacker planting their own. In practice, protocols such as DTLS are secured against MITM with the help of PKI – requiring the server at least to have a certificate verifiable by an external, trusted oracle (a certificate authority). This is beyond the scope of the presented model.
It is worth noting that we are defending the server from DoS attacks, and the attacker's (malicious client's) actions have nothing to do with simplicity. Especially they cannot be executed if the client is to be stateless. As this is the only counterexample (attack scenario), the protocol analysis ends with the conclusion that with the exception of the MITM scenario (not considered in terms of security and producing the only counterexample in terms of cookies), the modeled DTLS protocol retains its security properties and the cookie mechanism works correctly. Based on the evidence, the DTLS 1.2 handshake with the additional *ClientHelloVerify* mechanism is a security equivalent of the TLS 1.2 handshake and the TLS 1.3 handshake.

7. Conclusions

There is no doubt that guaranteeing the appropriate level of security in 5G SN or IoT networks is an important requirement for the reliability of these networks.
Research focusing on improving the security of existing solutions, as well as on ensuring new and secure types of connections between devices operating within 5G SN or IoT networks, and on ensuring fully secure services rendered with the use of such networks, is ongoing. The deployment of new solutions involves the creation of new security protocols. In order to automate the process of checking correctness of the security protocols proposed, relevant software tools are created, such as Tamarin.
This paper presents how an automatic symbolic analysis tool can be used at the design stage to perform the security analysis and to verify the correctness of operation of newly proposed protocols used in 5G SN or IoT environments, as well as in other modern sensor networks.

References

[1] M. Nadimpalli, "Internet of Things – future outlook", *Int. J. of Innov. Res. in Comp. and Commun. Engin.*, vol. 5, no. 6, 2017 [Online]. Available: https://www.rroij.com/peer-reviewed/internet-of-things-future-outlook-85898.html

[2] S. Helme, "Perfect forward secrecy – an introduction", 2014 [Online]. Available: https://scotthelme.co.uk/perfect-forward-secrecy

[3] Tamarin prover [Online]. Available: https://github.com/tamarin-prover/tamarin-prover/blob/develop/examples/classic/TLS_Handshake.spthy

[4] J. Y. Kim, R. Holz, W. Hu, and S. Jha, "Automated analysis of secure Internet of Things protocols", in *Proc. of the 33rd Ann. Comp. Secur. Appl. Conf. ACSAC 2017*, Orlando, FL, USA, 2017, pp. 238–249 (DOI: 10.1145/3134600.3134624).

[5] J. Y. Kim, Automated-security-verification-of-IoT-protocols [Online]. Available: https://github.com/jun-kim/Automated-security-verification-of-IoT-protocols/blob/master/CoAP_DTLShandshake.spthy

[6] T. Cole, "Interview with Kevin Ashton – inventor of IoT: Is driven by the users", Smart Industry the IoT Business Magazin, 2018 [Online]. Available: https://www.smart-industry.net/interview-with-iot-inventor-kevin-ashton-iot-is-driven-by-the-users/

[7] T. Salman and R. Jain, "A survey of protocols and standards for Internet of Things", *Adv. Comput. and Commun.*, vol. 1, no. 1, 2017 (DOI: 10.34048/2017.1.f3).

[8] E. Rescorla, The Transport Layer Security (TLS) Protocol Version 1.3, draft-ietf-tls-tls13-28 - 20, 2018 [Online]. Available: https://tools.ietf.org/html/draft-ietf-tls-tls13-28

[9] E. Rescorla, H. Tschofenig, and N. Modadugu, The Datagram Transport Layer Security (DTLS) Protocol Version 1.3, draft-rescorla-tls-dtls13-01-13, 2017 [Online]. Available: https://datatracker.ietf.org/doc/html/draft-rescorla-tls-dtls13-01

[10] S. Meier, B. Schmidt, C. Cremers, and D. Basin, "The Tamarin prover for the symbolic analysis of security protocols", in *Computer Aided Verification 25th International Conference, CAV 2013, Saint Petersburg, Russia, July 13-19, 2013. Proceedings*, N. Sharygina and H. Veith, Eds. *LNCS*, vol. 8044, pp. 696–701. Springer, 2013 (ISBN: 9783642397981).

[11] J. Thakkar TLS 1.3 Handshake: Taking a Closer Look, 2018 [Online]. Available: https://www.thesslstore.com/blog/tls-1-3-handshake-tls-1-2/

[12] D. Basin, C. Cremers, J. Dreier, S. Meier, R. Sasse, and B. Schmidt, Tamarin-Prover Manual Security Protocol Analysis in the Symbolic Model, 2019 [Online]. Available: https://tamarin-prover.github.io/manual/tex/tamarin-manual.pdf

[13] D. Basin, C. Cremers, J. Dreier, R. Sasse, "Symbolically analyzing security protocols using Tamarin", *ACM SIGLOG News*, vol. 4, no. 4, 2017, pp. 19–30 [Online]. Available: https://hal.archives-ouvertes.fr/hal-01622110/file/tamarin-tool.pdf

[14] Q. Do, B. Martini, and K. R. Choo, "The role of the adversary model in applied security research", *Comp. & Secur.*, vol. 81, pp. 156–181, 2019 (DOI: 10.1016/j.cose.2018.12.002).

[15] W. Diffie and M. Hellman, "New directions in cryptography", *IEEE Trans. on Inform. Theory*, vol. 22, no. 6, pp. 644–654, 1976 (DOI: 10.1109/TIT.1976.1055638).

[16] The Illustrated TLS Connection [Online]. Available: https://tls.ulfheim.net/

[17] L. C. Paulson, "Inductive analysis of the Internet protocol TLS", *ACM Trans. on Inform. and Syst. Secur.*, vol. 2, no. 3, pp. 332–351, 1999 (DOI: 10.1145/322510.322530).

[18] G. Ferro, TCP SYN Cookies – DDoS defence, 2008 [Online]. Available: https://etherealmind.com/tcpsyn-cookies-ddos-defence/

[19] E. Rescorla and N. Modadugu, Datagram Transport Layer Security Version 1.2, 2012 [Online]. Available: https://tools.ietf.org/html/rfc6347

Piotr Remlein received his M.Sc. and Ph.D. degrees from Poznań University of Technology (PUT), Poznań, Poland in 1991 and 2002, respectively. In 2018, he received a D.Sc. degree from PUT. He has been employed at PUT since 1992, currently as an Associate Professor at the Institute of Radiocommunications, Faculty of Computing and Telecommunications. His scientific interests cover wireless networks, communication theory, error control coding, cryptography, digital modulation, continuous phase modulation, mobile communications, and digital circuit design. Dr. Remlein is the author of more than 120 papers, presented at national and international conferences and published in communications journals. He also acts as a reviewer for international and national conference and journal papers. He is a Senior Member of IEEE Communications Society and IEEE Information Theory Society.

https://orcid.org/0000-0002-7593-839X
E-mail: piotr.remlein@put.poznan.pl
Institute of Radiocommunications
Poznań University of Technology
Pl. M. Skłodowskiej-Curie 5
60-965 Poznań, Poland

Urszula Stachowiak received her B.Eng. degree in Information and Communication Technologies from the Faculty of Computing and Telecommunications, Poznań University of Technology, in February 2020. Since March 2020 she has been a master's student majoring in Computing and specializing in the Internet of Things. From July 2019 to November 2020, she was employed as a Telecommunications Analyst at Comarch, and has been working as a Software Engineer at Intel Corporation since December 2020. Her interests include topics related to the security of IoT and formal proving of communication protocol security.

https://orcid.org/0000-0002-6892-6876
E-mail: urszula.stachowiak@student.put.poznan.pl
Faculty of Computing and Telecommunications
Poznań University of Technology
Pl. M. Skłodowskiej-Curie 5
60-965 Poznań, Poland

Paper

Machine Learning-Based Small Cell Location Selection Process

Małgorzata Wasilewska and Łukasz Kułacz

Institute of Radiocommunications, Faculty of Computing and Telecommunications, Poznań University of Technology, Poznań, Poland

https://doi.org/10.26636/jtit.2021.151021

Abstract—In this paper, the authors present an algorithm for determining the location of wireless network small cells in a dense urban environment. This algorithm uses machine learning, such as k-means clustering and spectral clustering, as well as a very accurate propagation channel created using the ray tracing method. The authors compared two approaches to the small cell location selection process – one based on the assumption that end terminals may be arbitrarily assigned to stations, and the other assuming that the assignment is based on the received signal power. The mean bitrate values are derived for comparing different scenarios. The results show an improvement compared with the baseline results. This paper concludes that machine learning algorithms may be useful in terms of small cell location selection and also for allocating users to small cell base stations.

Keywords—base station selection, k-means clustering, spectral clustering, user equipment allocation.

1. Introduction

With the advent of 5G networks, one may notice increasing interest in the concept of small cells. Additional small cells positioned at locations where services are already available may significantly improve network performance and may boost the quality of service, depending on user needs. For example, deterioration in the quality of network access may often be observed in large gatherings, as most of people present in such scenarios use wireless devices. Such a group of devices connects to the base station and, consequently, neither this group nor other users of this particular base station are capable of obtaining satisfactory bitrates or service quality levels. An additional base station with a small coverage area (known as a small cell or a pico cell) positioned at the location where such a large group of devices is present may significantly improve the quality of service enjoyed by all users. While the use of small cells is justified in the aforementioned scenario, it is not quite obvious where exactly such cells should be located.

Sometimes, it is quite easy to determine where and for how long increased traffic rates may be expected. For example, a group of people actively using their mobile devices may be presented at a given location only for random periods of time only, or may be expected there periodically (bus stations, airports, etc). The above-mentioned scenarios are directly related to the location at which the increase in traffic takes place. For instance, if increased network traffic is observed at a bus stop – we know the location of the potential small cell base station. However, increased network traffic is not always closely related to a fixed location. Therefore, the authors have designed an algorithm that determines small cell installation locations with a given period of time, to match the highest demand levels. The results obtained with the use of this algorithm may be relied upon in many ways. It is possible to average the results (or to select critical, worst case scenarios), thus selecting a location for a stationary small cell. Such an approach may be used in network coverage planning or improvement processes. Another approach consists in using drones (UAVs) with a small cell base station hovering overhead. In this scenario, the position of such stations may be changed dynamically. The algorithm presented in this paper works regardless of the way the results are used.

In other publications concerning the application of machine learning techniques for handling small cell traffic two main aspects seem to prevail, namely assignment of user equipment (UE) to a given set of base stations (BSs) and positioning of BSs for best coverage. The articles dealing with the former of those aspects include [1]–[4]. Balapuwaduge *et al.* [1] focus on smarter assignment of UE to BS by employing an ML algorithm based on the hidden Markov model. The algorithm focuses on reliability and availability of network resources in order to select the best BS for a given piece of UE.

Yang *et al.* [2] employ reinforcement deep learning (DL) to position small cells in indoor scenarios, with a particular emphasis placed on company small cells. The problem presented may be generalized to the allocation of users whose behavior is predictable and those who behave in a more dynamic manner. The ML algorithm works based on data consisting only of allocation information for each piece of UE. Qi *et al.* [3] and Xu *et al.* [4] focus on the k-means clustering and the reinforcement k-means clustering algorithm, respectively. Both of those papers use ML for clustering UEs in order to achieve good load balance.

In the second group of papers which focus on BS positioning in order to achieve the best coverage, the use of drones is a popular solution [5], [6]. In [5], drones are to replace BSs in the event of an emergency. The main problem is

how to ensure the best possible coverage. The reinforcement learning approach, namely the Q-learning algorithm, is employed to determine the drones' positions based on whether a connection has been established between UE and the drone or not. Wang *et al.* [6] focus on problems that drones face while ensuring connectivity, namely co-channel interference, limited battery capacity and fast topology changes. In this case, ML is supposed to control not only the placement of drones, but also their transmission power, as it affects the level of interference and battery lifetime.

To recapitulate, this paper offers the following contributions:

- it combines the problem of allocating UE to BSs with the problem of choosing the stations' locations,
- it uses two simple unsupervised ML algorithms, namely k-means clustering and spectral clustering, in order to group UE on the basis of path loss data,
- it chooses the best BS location for each of the groups, in order to improve QoS and mean bitrate of the connections.

In the chapters below, the following are described: the proposed ML-based small cell deployment algorithm (Section 2), the system in which the simulations were performed (Section 3), detailed simulation results with conclusions (Section 4), and summary of the work performed.

2. ML-based Small Cell Deployment

The first thing one needs to do in order to successfully deploy small (pico) cells is to choose their optimized locations. In this paper, we propose the use of machine learning algorithms for this purpose. Such an algorithm will decide which small cells to use and which pieces of UE to assign to them. The main problem is what algorithm to use, considering the limitations of training data. In this section, different approaches to artificial learning are discussed and the best solutions are presented. We also describe how we employ the chosen ML algorithms for selecting BSs and assigning UE. Additionally, a detailed description of the input data that the presented algorithms rely on is given as well.

When it comes to selecting the right ML algorithm, one has to consider what types of data are available. In most cases, it is hard to obtain a labeled set of training data. Labeled data is a term used to describe data that consists of input features (usually referred as X labels), but also of their corresponding categories, or desired outputs, known as y labels. In order to obtain such a data set, it is usually necessary to manually label each input feature set. In the case of a computer simulation, it is much easier to generate training data along with their corresponding labels, but this is not always true.

In the system considered in this paper, labeled data would be generated for a set of many different combinations of user positions within the considered space. The input feature data could consist of the users' coordinates and other additional features, while output labels would indicate to which BS they are connected. It would be necessary to calculate, for all of the user locations considered, all bitrates to all of the possible BSs, while taking into account interference from all other BSs in order to determine how to allocate users to BSs. It is easy to imagine how computationally-intensive and time-consuming it would be to generate such a dataset. In order to address these issues, the authors propose to use ML algorithms that are not supervised and are able to learn based on data without any specified output labels. The algorithms that are explored in this paper are: k-means clustering and spectral clustering.

2.1. K-means Clustering

As the system under consideration consists of scattered users in who are in need of being allocated to a BS, clustering algorithms come to mind first. Clustering algorithms groups similar feature data points together. The resulting groups are called clusters. In this paper, the k-means algorithm has been proposed as a grouping method, as it is simple, yet effective.

The grouping process is performed in the following manner: initially, a random placement of centroids is picked (points around which clusters are centered). Then, all input instances are assigned to the closest centroid [7]. Then, the centroids are updated by minimizing the *inertia* criterion IC, given by:

$$IC = \sum_{n=0}^{N} \min_{c_i \in C}(||x_n - c_i||) \ , \tag{1}$$

where x_n is an instance from input dataset X, and c_n is the n-th centroid from the chosen centroid set C consisting of N centroids. The process of categorizing input data and assigning such data to clusters is repeated until the centroids stop moving. In the k-means algorithm, it is initially necessary to specify the number of clusters.

2.2. Spectral Clustering

Spectral clustering is another unsupervised grouping ML algorithm used in the experiments. Compared to the k-means algorithm, spectral clustering is capable of performing better on non-convex data, which is quite helpful in solving the problem presented in the paper. Spectral clustering creates a similarity matrix between the input data and then reduces the dimensionality of this matrix. After that, another clustering algorithm is used on the obtained matrix [8]. In the algorithm implemented for the experiments, spectral clustering performs k-means after dimensionality reduction. As it is the case with the k-means algorithm, spectral clustering requires that the number of clusters be specified before data grouping.

2.3. Clustering Algorithm Input Data

The algorithms outlined above require correctly defined input data. The input dataset consists of the pathloss values between each user and each of the potential BSs. In the analyzed simulation scenario, the authors had to limit the list of small cell locations to 28 potential sites. For the sake of simplicity, path attenuation was analyzed, calculated as the average attenuation for all resource blocks. Hence, there are 28 potential locations of pico-type BSs, and one feature instance representing a features dataset for one user has 29 values – 1 pathloss between the user and the macro BS, and 28 pathlosses between the user and each of the pico BSs. To sum up, the i-th input instance may be presented as the following vector: $[PL_{macroBSi}, PL_{picoBS1i}, PL_{picoBS2i}, \ldots, PL_{picoBS28i}]$, where $PL_{macroBSi}$ is a pathloss value between i-th user and the macro BS, $PL_{picoBSni}$ is a pathloss value between i-th user and the n-th pico BS. The data has been pre-processed before being used as input data. All of the pathloss values have been normalized and scaled to the 0–1 range, except for $PL_{macroBSi}$ that has been scaled to the 0–2 range in order to place a greater emphasis on this particular feature. Thanks to such alterations, algorithm should prefer to connect users to the macro BS, connecting them to pico BSs only in those cases in which such a step is required.

2.4. ML-based Algorithm

As explained in the previous section, both clustering algorithms group the input data into k groups based on their pathloss values concerning all BSs. The next step is to determine which BSs should be assigned to the created groups. The performance of the small cell location selection algorithm is considered in two scenarios, namely Scenario 1 and Scenario 2.

In Scenario 1, pieces of UE are directly associated with BSs indicated by the ML-based algorithm. After the piece of UE have been grouped, a comparative algorithm is implemented that searches for the best BS for a given UE group by checking which BS has the best (lowest) mean pathloss for the assigned users. One BS is assigned to each of the created clusters. The chosen BSs are dedicated to one cluster only, so if the number of clusters is k, the number of BSs used in the network is k as well.

In Scenario 2, ML algorithms perform clustering on the pieces of UE as well. Then, just as it was the case in Scenario 1, BSs are picked for each of the groups in the same manner. The main difference is that after the BSs have been chosen, the pieces of UE are associated with BSs based on the best received signal strength, just as in a typical LTE network.

3. System Description

In the model of their system, the authors analyzed a typical fragment of an urban environment in Madrid. It consists of several buildings of different heights, a grid of streets, a wide pavement typical of shopping districts, and a park. The method that was used for generating the radio environment relied on the ray tracing method which enabled to obtain a very precise fragment of the channel coefficients. This allowed for a good representation of the actual wave propagation conditions observed in a typical radio environment (in a dense urban area). At the same time, due to the high computational complexity of this method, the authors were forced to significantly reduce the potential locations of pico base stations to 28 points. In Fig. 1, the area of the analyzed network, with individual buildings marked, is presented. The macro BS covering a large part of this area, and the potential locations for pico BSs for which the channel was generated, are marked as well. Additionally, the locations of UE have been marked in the same figure.

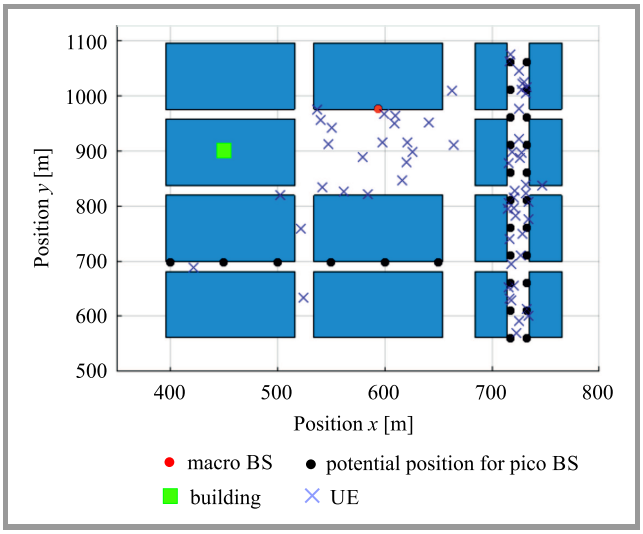


Fig. 1. Network topology subjected to analysis.

The user position generation method assumes that 30% of the pieces of UE are located in a park (close to the macro BS), 60% of them are on the pavement, and the remaining 10% of the pieces of UE are positioned elsewhere. UE positions are constant for all analyzed situations due to the complicated channel and the long lead time required for generating channel coefficients. The results presented in this paper should be understood as aiming to identify the best method for selecting the locations of small cells for a specific UE arrangement. The analyzed system is an LTE solution, with UE being assigned to the BS, by default, based on signal strength. During the simulations, this method assigning UE to BSs may be replaced with the ML-based algorithm that directly indicated UEs to BSs association. The scheduler used in the system is of the round robin variety. The system bandwidth is 20 MHz, which translates into 108 resource blocks. Downlink transmission was considered only. Since the entire system operates in exactly the same band, the authors did not take into account turning off the resource edge blocks. The macro BS has 16 antennas, and the pico BSs have 4 antennas. The TX

power of the macro BS is 46 dBm, and the transmit power of each of the pico BSs is 30 dBm.

Algorithm 1 Network simulation scheme

```
Input: small cell location set
Result: KPI set for all devices
 Generate channel coefficient between all (BS, UE) pair
 Associate each UE with BS
 for time slot t_i to simulation duration do
    if mod (t_i, t_assoc) = 0 then
       Associate each UE with BS
    end
     Allocate RBs to UEs
     Calculate interference
     Calculate SINR
     Calculate throughput
     Save KPIs
end
```

A detailed description of the simulator's operation is presented as Algorithm 1. The positions of small cells derived from the ML-based position selection algorithm from Scenario 2 are fed to the simulator as input data. In Scenario 1, information about the pattern of direct association of UE to BSs is an additional source of input data. At the initial phase of the simulator's operation, channel coefficients are generated between each piece of UE and a BS, separately for each RB. Then, depending on the scenario, pieces of UE are connected to their respective BSs on the basis of the received signal power or based on a direct indication from the proposed algorithm. Within the main simulation loop, where the simulation duration is set to 100 ms, the following operations are performed in sequence. Every t_i (in the simulation t_i equals 20 ms), the procedure of assigning pieces of UE to BSs is commenced. For each BS separately, the RBs are allocated, using the round robin algorithm, to all pieces of UE attached to a given BS. Then, interference is calculated separately for each UE and RB, and SINR for the allocated RBs is determined. Using the Shannon formula, throughput is calculated separately for each UE and RB and is then added up for all allocated RBs. The relevant metrics – key performance indicators (KPIs) – are saved for each time slot.

Once the simulation has been completed, the average throughput, as well as the first and the third quartiles of throughput are compared to evaluate the performance of the proposed solutions. The last two values allow to evaluate transmission performance for worst case and best case scenarios, respectively.

4. Experiment Setup and Results

Here, the results obtained for each of the proposed algorithms are presented. Transmission bitrate is the key value that is compared.

For both ML algorithms and for a number of k clusters, two results are compared for Scenario 1 and Scenario 2. Bitrates related to Scenario 1 are marked blue, while Scenario 2 results are presented with the use of yellow bars. The results of both scenarios are compared with the bitrate for $k = 1$, where there is only macro BS in use. All pieces of UE are assigned to this macro BS and the UE assignment method does not have any impact on the resulting bitrate. The dashed line presented in the graphs shows the bitrate level for $k = 1$ and is considered to be a benchmark value. In the following sections, the results for both ML algorithms and both scenarios are presented. First, results pertaining to the k-means algorithm are presented.

4.1. K-means Clustering Algorithm

First, *IC* values were calculated for each of the k values in order to see when the best value of k may be expected. Figure 2 shows the inertia values for different numbers of clusters k. One may observe that the best results should be obtained for $k = 2$, since for that value of parameter k, a peculiar, sudden change in the course of the inertia line is visible.

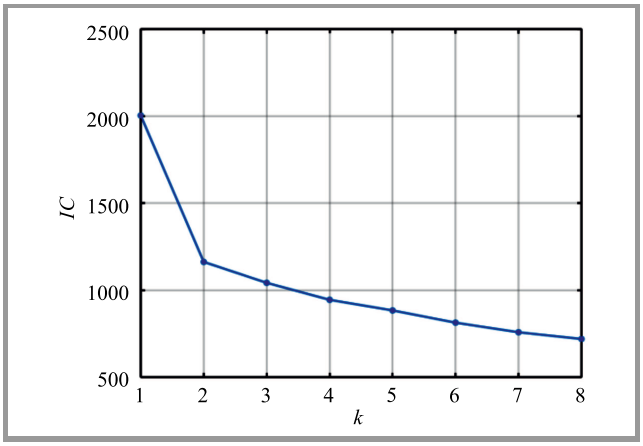


Fig. 2. Inertia of the k-means algorithm for different numbers of clusters. The graph the number of clusters for which the clustering results should be the best.

Figure 3 shows the mean bitrate for different numbers of clusters and for both Scenarios. For $k > 1$, there is a significant improvement in the mean bitrate. The best results of the k-means clustering algorithm (Scenario 1 results) have been obtained for $k = 5$. For that number of clusters, the macro BS and four pico BSs have been assigned to five clusters, and the mean bitrate improved 4.2 times compared to the mean bitrate benchmark value (results for $k = 1$ are presented). The best results in terms of the assignment of users to the same BSs without the k-means-based user grouping algorithm (Scenario 2) have been achieved also for $k = 5$, and the mean bitrate has improved 4.7 times. One can see, that the mean bitrate is better for Scenario 1-based allocations for $k = 2$ only. This means that only for a network with one macro BS and one pico BS the bitrate with k-means is better than the bitrate for Scenario 2 with the assignment to the same two stations.

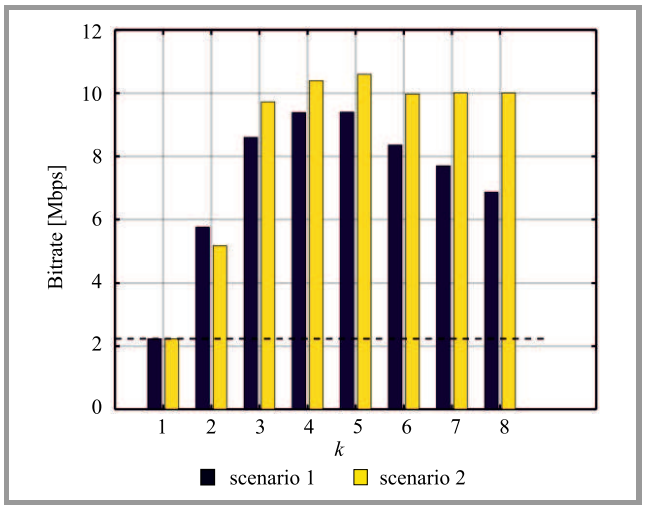


Fig. 3. Mean bitrate for station assignment based on k-means and without k-means (the same BSs are selected in both cases).

From Fig. 4, presenting results for the worst 25% of the connections, it is quite clear that Scenario 2 performs better for all BS numbers (all k values). The best results for Scenario 1 have been achieved for 2 BSs ($k = 2$), with bitrate improving 52.7 times. With the growing number of clusters, the results tend to get worse, although bitrate still remains better than for one cluster only. The assignment to the same BSs in Scenario 2 renders much better results, and the best outcomes have been achieved for $k = 3$ groups, with the bitrate improving 111.9 times.

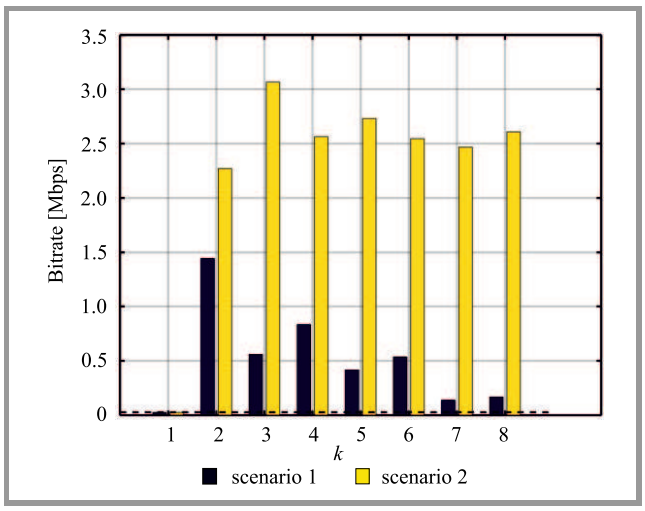


Fig. 4. First quartile (25th percentile) of user bitrate for station assignment based on k-means and without k-means (the same BSs are selected in both cases).

Although Scenario 2 performs, in most cases, in terms of bitrate for all users and in terms of transmission parameters for the weakest 25% of connections, Fig. 5 shows greater improvement for Scenario 1. For the best 25% of connections, the advantage caused by using more BSs is the greatest, and bitrate may be improved by up to 54.3 times for the best case of k-means-based grouping for $k = 4$. The best results for Scenario 2 have been achieved for $k = 5$, and bitrate has been improved 35.6 times.

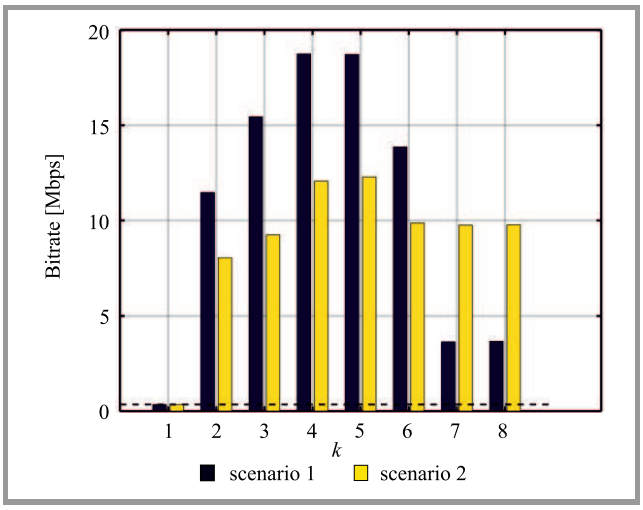


Fig. 5. Third quartile (75th percentile) of user bitrate for station assignment based on k-means and without k-means (the same BSs are selected in both cases).

4.2. Spectral Clustering Algorithm

The second set of results has been obtained using the spectral clustering algorithm. Similarly to the k-means algorithm, for the mean bit rate of connections (Fig. 6) only for two groups or two BSs ($k = 2$) the Scenario 1 grouping is better than the Scenario 2, where Scenario 1 achieved 2.6 times better bitrate and Scenario 2 achieved 2.3 better bitrate comparing with reference bit rate for $k = 1$.

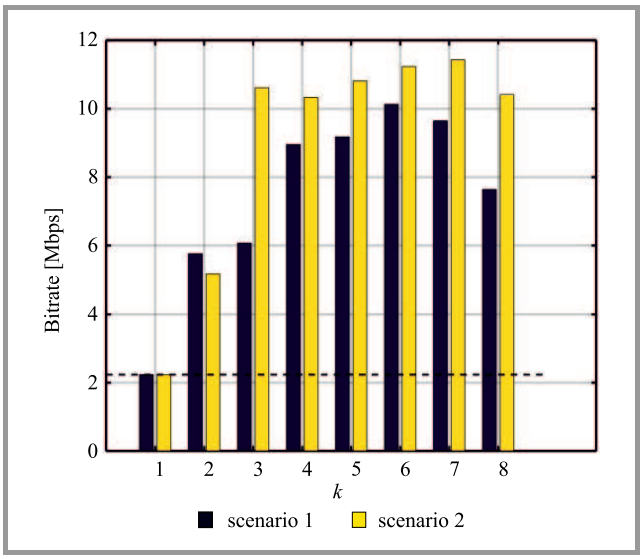


Fig. 6. Mean bitrate for station assignment based on spectrum clustering and without spectrum clustering (the same BSs are used).

In terms of the remaining results, those for Scenario 2-based grouping show a noticeable improvement compared

to the results from Scenario 1. For $k = 6$ and $k = 7$, Scenario 2 results for spectral clustering achieve a bitrate that is over 5 times better, while in Scenario 1, results achieved with the k-means method peak at $k = 6$ and reach a bitrate that is 4.5 times better.

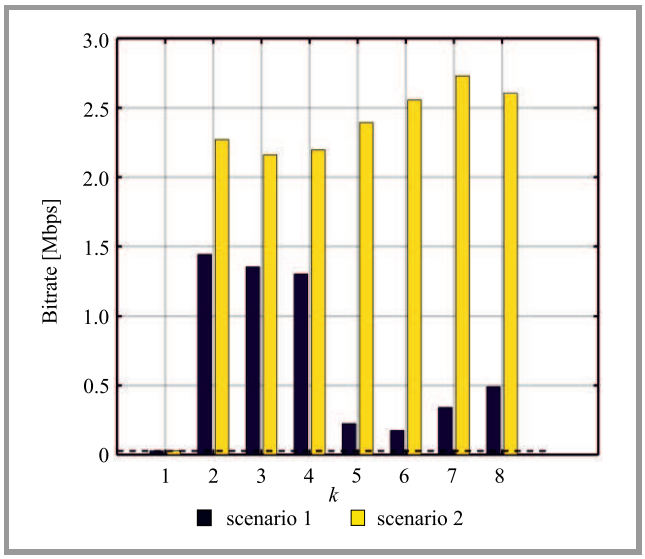


Fig. 7. First quartile of user bitrate for station assignment based on spectral clustering and without spectral clustering (the same BSs are used in both cases).

The results for the worst 25% of connections are presented in Fig. 7. In this case, spectral clustering based on Scenario 1 performs better than the k-means solution, especially for $k = 3$, and $k = 4$ for which bitrate improved 49 times and 47 times, respectively. Meanwhile, bitrate results achieved in the k-means Scenario 1 are only 20 and 30 times better, respectively. The grouping method used in Scenario 2 is also better than in each instance of Scenario 1.

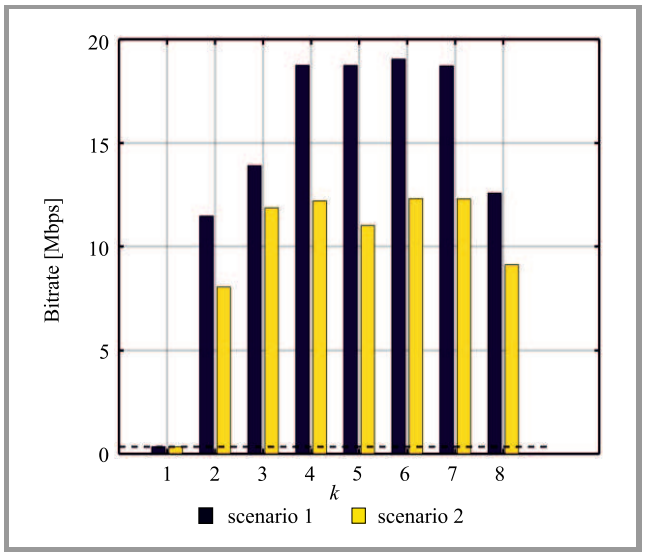


Fig. 8. Third quartile of user bitrate for station assignment based on spectral clustering and without spectral clustering (the same BSs are selected in both cases).

The last set of results concerns the best 75% of connections – see Fig. 8. Here, the results are also better when compared to those obtained using the k-means method. Scenario 1-based grouping achieved better results than Scenario 2 for each k value, and there are four k values for which Scenario 1 improved the bitrate 31 times (for $k = 4$, 6 and 7 – even 35 times).

5. Conclusion

The paper presents an algorithm for selecting the location of small cells using the ML technique. The presented simulation results showed that the choice of BS locations is performed with best users (75th percentile of throughput) preferred. However, average and the weakest (25th percentile of throughput) network users achieve lower bitrates in such a scenario. The presented algorithm is not universal and is effective in specific cases only, but it offers a promising point of departure for further studies. As an extension of the algorithm, the usage of the CRE parameter related to small cells may be considered. The application of other ML methods, such as reinforcement ML, could be taken into consideration as well.

6. Acknowledgements

This work was supported by the DAINA project no. 2017/27/L/ST7/03166 "Cognitive engine for radio environment awareness in networks of the future" (CERTAIN) funded by the National Science Centre, Poland.

References

[1] I. A. M. Balapuwaduge and F. Y. Li, "Hidden Markov model based machine learning for mMTC device cell association in 5G networks", in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Shanghai, China, 2019, pp. 1–6 (DOI: 10.1109/ICC.2019.8761913).

[2] J. Yang, C. Wang, X. Wang, and C. Shen, "A machine learning approach to user association in enterprise small cell networks", in *Proc. Int. Conference on Communications in China (ICCC)*, Beijing, China, 2018, pp. 850–854 (DOI: 10.1109/ICCChina.2018.8641148).

[3] W. Qi, B. Zhang, B. Chen, and J. Zhang, "A user-based K-means clustering offloading algorithm for heterogeneous network", in *Proc. 8th Annual Comput. and Commun. Workshop and Conf. (CCWC)*, Las Vegas, NV, 2018, pp. 307–312 (DOI: 10.1109/CCWC.2018.8301769).

[4] Y. Xu, W. Xu, Z. Wang, J. Lin, and S. Cui, "Load balancing for ultradense networks: a deep reinforcement learning-based approach", *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9399–9412, 2019 (DOI: 10.1109/JIOT.2019.2935010).

[5] R. de Paula Parisotto *et al.*, "Drone base station positioning and power allocation using reinforcement learning", in *Proc. 16th Int. Symp. on Wireless Commun. Systems (ISWCS)*, Oulu, Finland, 2019, pp. 213–217 (DOI: 10.1109/ISWCS.2019.8877247).

[6] L. Wang, Y. Chao, S. Cheng, and Z. Han, "An integrated affinity propagation and machine learning approach for interference management in drone base stations", *IEEE Trans. on Cognitive Commun. and Networking*, vol. 6, no. 1, pp. 83–94, 2020 (DOI: 10.1109/TCCN.2019.2946864).

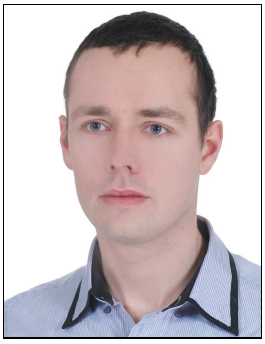
[7] J. MacQueen, "Some methods for classification and analysis of multivariate observations", in *Proc. of the fifth Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1, no. 14, 1967, pp. 281–297 [Online]. Available: http://digitalassets.lib.berkeley.edu/math/ucb/text/math_s5_v1_article-17.pdf

[8] U. Von Luxburg, "A tutorial on spectral clustering", *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007 (DOI: 10.1007/s11222-007-9033-z).

Małgorzata Wasilewska received her M.Sc. degree in Telecommunications from Poznań University of Technology, Poland, in 2017. She is currently pursuing her Ph.D., working at the Institute of Wireless Communications, PUT. Her main field of interest are wireless communications and machine learning.

https://orcid.org/0000-0002-3471-0516

E-mail: malgorzata.wasilewska@put.poznan.pl
Institute of Radiocommunications
Faculty of Computing and Telecommunications
Poznań University of Technology
Polanka 3
60-995 Poznań, Poland



Łukasz Kułacz received his M.Sc. degree in Telecommunications from Poznań University of Technology, Poland, in 2018. He is currently pursuing his Ph.D., working at the Institute of Wireless Communications, PUT. His main fields of interest are programming, wireless communications, and algorithm design.

https://orcid.org/0000-0002-3434-1917

E-mail: lukasz.kulacz@put.poznan.pl
Institute of Radiocommunications
Faculty of Computing and Telecommunications
Poznań University of Technology
Polanka 3
60-995 Poznań, Poland

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The authors JTIT LATEX style file is available:

https://www.itl.waw.pl/en/jtit-for-authors

Papers published should contain up to 10 printed pages in LATEX authors style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres, *Electron. Lett*., vol. 30, no. 3, pp. 262–264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, Informational representability: Abstract models versus concrete models, in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314

Biographies and photographs of authors. A brief professional authors biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the worlds most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishings databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(*Contents Continued from Front Cover*)

National Institute of Telecommunications

Editorial Office

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

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.nit.eu

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY 2/2021