CONTENT
Vol. 6. No. 2-2020

Published bi-monthly since 2015.

ISSN 2664-066X (Print)
ISSN 2664-0678 (Online)

Publisher
Institute of Radio and
Information Systems (IRIS),
Vienna, Austria

Address:
1010 Wien, Austria,
Ebendorferstrasse 10/6b
media-publisher.eu/
synchroinfo-journal

IMPROVING THE SYNCHRONIZATION OF OFDM SIGNALS IN THE SYSTEM DVB-T2

S. M. Al-Merahi,
Sana's University, Sana, Yemen

DOI: 10.36724/2664-066X-2020-6-2-2-6

ABSTRACT

Describe the principles of synchronization in systems of digital television broadcasting DVB-T2 and TDS-OFDM (DTTB), which provide all forms of synchronization with the signal analysis in time and frequency domains. In the system TDS-OFDM guard interval to obtain the pseudorandom sequence – m-sequence, which is used for synchronization and estimation of channel state. It allows to provide high interference immunity of the system and the acceleration of the synchronization process. It is proposed to use m-sequence in the system of DVB-T2 with placing it in the protective interval. The results of the modeling process correlation of m-sequence, transmitted together with the information of the cyclic prefix.

KEYWORDS: *sinchronization, OFDM signals, DVB-T2, TDS-OFDM, DTTB.*

I. INTRODUCTION

In the process of receiving OFDM signals, accurate synchronization of the receiver and transmitter in time and frequency is important. High-quality synchronization can significantly increase the receiver's noise immunity. The DVB-T2 synchronization system provides all forms of synchronization with signal analysis in the time and frequency domain – OFDM frames, symbols, time position of sampling intervals and frequency synchronization of subcarriers.

Synchronization in the time domain requires large hardware costs and allows one to achieve a lower accuracy of estimates, therefore it is used at the stage of rough estimation of time and frequency mismatch, and for an accurate estimate, algorithms in the frequency domain are used to a large extent.

Synchronization at the receiver is performed prior to demodulating the symbol subcarriers. The receiver must obtain information about the exact time position of the symbol to minimize the effects of intersymbol interference (ISI) and inter-subcarrier interference (ICI).

The proposed method of using PSP (m-sequences) in the DVB-T2 system will retain the possibility of using QAM-256 modulation while improving the quality of synchronization and its acceleration.

Synchronization principles in DVB-T2 system

For synchronization in the DVB-T2 system, P1 and P2 symbols are used (Fig. 1), as well as pilot signals located in the OFDM frame. P1 marks the start of a frame, and P2 follows P1 in every frame and provides access to physical layer PLPs in every frame. Several transport streams (TV, radio, HDTV, SD, mobile communication) can be transmitted in the DVB-T2 channel simultaneously, each of them being placed in its own PLP channel.

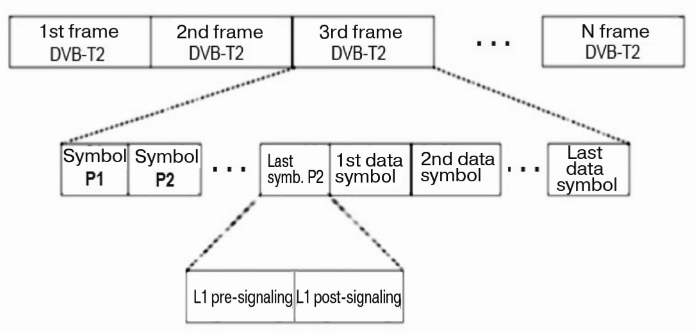


Figure 1. Placing P1, P2 symbols and L1 signals in a DVB-T2 frame [1]

Sampling clock synchronization is also performed, which is based on the information contained in the P1 symbol.

P1 symbol is used for synchronization and identification of DVB-T2 stream. It also contains information about the DVB-T2 frame - the number of subcarriers in OFDM (1k-32k) and the transmission format of the P2 symbol following P1 (MISO or SISO modes). Element P1 is an OFDM symbol with DBPSK modulation, two guard intervals on both sides (in the sum of 1/2 symbol duration).

Frequency synchronization is necessary to protect symbols from inter-subcarrier interference (ICI) and to maintain orthogonality in OFDM symbols. Orthogonality is violated when the carrier frequency shift appears, which appears in the absence of synchronization of the frequency generators on the transmitting and receiving sides. In addition, the frequency shift occurs due to the Doppler effect [1].

Sampling clocks are equally important. The offset of the sampling interval leads to severe distortion of the reconstructed OFDM signal. These distortions occur in multipath reception with fading, noise, and timing errors in symbol position. The sampling interval offset is divided into a sampling phase offset and a sampling frequency offset [2].

Let's consider the sequence of synchronization in the DVB-T2 receiver. The synchronization process begins with the detection of the P1 symbol. After the P1 symbol is detected, information about the symbol timing and frequency and the size of the fast Fourier transform (FFT) OFDM is obtained from it. After determining the size of the Fourier transform, the presence of the guard interval (GI) is determined. To detect the pilot signals, the basic information signal (OFDM symbols) is used and the L1 signal in P2 symbols is decoded. Since the receiver already knows the FFT length at this time, the receiver detects the pilot signals and can decode the data located at P2 and L1.

By decoding the signaling information in L1, the receiver determines the layout of the pilots and GIs in the OFDM data symbol stream. P1 marks the beginning of each frame. Detection of the P1 symbol is sufficient to start processing the DVB-T2 RF carriers. In addition, the P1 symbol is used to correct time and frequency synchronization.

The received radio signal, transmitted from the transmitting to the receiving antenna, may have a certain frequency shift Δf, when the transmitter and receiver move relative to each other - the Doppler effect. The frequency shift of the OFDM signal results in a loss of orthogonality between the subcarriers of the OFDM symbols, which causes a decrease in the quality of the demodulation of the received signal.

Receiving signals in urban environments with multi-beam signal propagation and Rayleigh fading is especially difficult in solving synchronization problems. Such signal propagation leads to significant distortion of its spectrum and complicates the task of synchronization and estimation of channel parameters in the frequency domain.

The task of frequency synchronization is to estimate and correct the frequency shift Δf with the required accuracy. Since a pair of Fourier transforms is used to generate and allocate orthogonal subcarriers in OFDM systems, signals are generated and transmitted as OFDM symbols. To minimize the effects of intersymbol interference in multipath reception, guard intervals are introduced between individual symbols.

Eliminating the loss of orthogonality of subcarriers during demodulation requires accurate phase and frequency matching of the transmitter and receiver over the entire frequency band of received signals. Phase and frequency mismatch is due to some spread and instability in the frequencies of the transmitter and receiver reference oscillators and Doppler shift in mobile communications.

The effect of synchronization errors increases with the number of subcarriers. Therefore, in order to minimize the phase distortion of the signal on the receiving side, and, consequently, the probability of error during decoding in DVD-T2, high requirements are imposed on time and frequency synchronization.

The use of such complex synchronization with pilot processing requires a significant amount of decision time and comprehensive synchronization of the DVD-T2 receiver. Synchronization time reduction was achieved in the TDS-OFDM system.

Description of the system receiving part TDS-OFDM (DTTB)

In China, a terrestrial digital TV and multimedia broadcasting (DTMB) system has been developed, which has shown good results and is accepted for use in some countries. In this system, no cyclic prefix is transmitted in the guard interval. It is replaced by a pseudo-random sequence (PRS) that is used for synchronization.

In the TDS-OFDM receiver, the radio frequency signal is fed to the tuner (Fig. 2), converted into an intermediate frequency signal, fed to the ADC and passed through the synchronization blocks (18), the channel equalizer (20), demodulation (22), the precorrection decoder (24). The basic configuration of the receiver includes the following types of subcarrier modulation: QPSK, 16QAM, 64QAM.

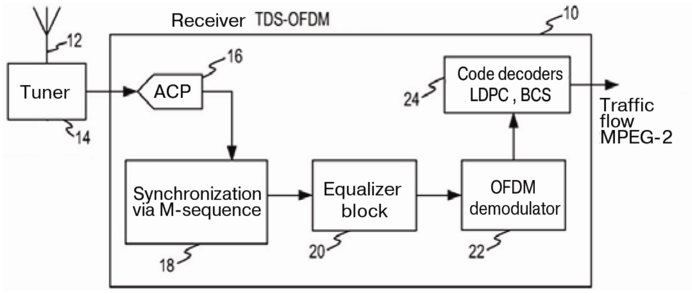


Figure 2. Generalized scheme of the TDS-OFDM receiver with blocks of synchronization, equalizer, OFDM demodulator and an LDPC decoder. The output of the receiver generates transport stream MPEG-2 or MPEG-4

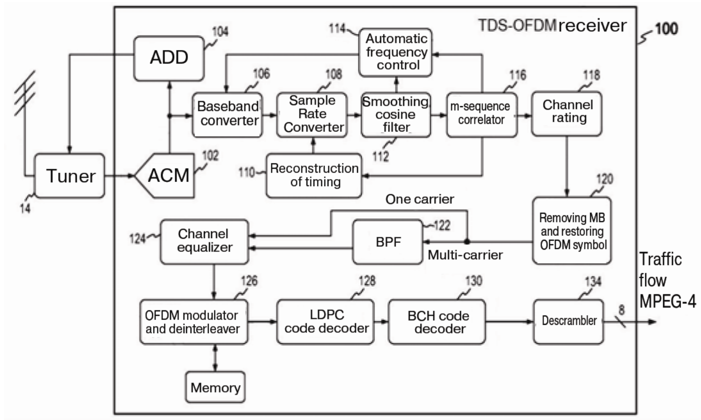


Figure 3. TDS-OFDM receiver circuit

In the TDS-OFDM system, a pseudo-random sequence (PRS), an m-sequence, which is used for synchronization and channel estimation, is shuffled into the guard interval.

Time synchronization of TDS-OFDM symbols is used to isolate the starting point of each symbol, which is necessary for accurate signal sampling. Symbol time offset (STO) estimation is performed in either the time domain or the frequency domain.

To synchronize symbols in the receiver, a training m-sequence is used, which allows you to quickly restore the frame and temporal relationships in the received digital signal, and to make a more reliable estimate of the channel state. Since no pilot signals are transmitted in TDS-OFDM (DTTB), all subcarriers are informational. This increases the system information capacity.

The DTTB TDS-OFDM modulation scheme contains 3780 symbols representing the result of the fast Fourier transform of size 3780. To ensure high noise immunity, the system uses precorrection with LDPC and BCS codes. Three LDPC codes are available at different rates.

Time estimation (timing) is performed by the correlation signal (block 116). The frequency response of the channel is obtained by taking the FFT from the impulse response of the channel (Fig. 3).

After correlator block 116 and channel estimator 118, OFDM symbols are detected and recovered. During modulation (at the transmitter), the PRS is placed in the guard interval, replacing the usual cyclic prefix. At the receiver, block 120 removes the PRP from the guard interval and recovers the OFDM symbol sequence. The reconstructed OFDM symbols are further processed in the frequency domain by a one-tap equalizer (block 124) with the elimination of the frequency response channel irregularities resulting from fading and multipath reception.

In the multi-subcarrier mode, the reconstructed symbols are time domain data and are then fed to an FFT (122) where a Fourier transform converts these symbols into corresponding frequency domain signals. Block 122 performs a 3780-point FFT.

Before synchronization starts, block 106 receives and processes the IF signal and generates a baseband signal. Then the signals are processed by the PRS correlator (116), which outputs a cross-correlation signal by comparing the incoming PRP and the same sequence generated by the local generator.

The correlator signal (block 116) is used to reconstruct the carrier, digital signal timing, and parameters used to estimate the channel state. The peak of the function generated by the correlator (Fig. 6, 7) marks the beginning of each frame of the OFDM signal, the frequency shift under the Doppler effect, and corrects the temporal relations.

In equalizer 124, a frequency domain response (AFC) is obtained by taking an FFT of a time domain response (channel impulse response). The equalizer estimates the parameters and equalizes the spectral components of the signal in the frequency domain of the channel.

The OFDM demodulator unit (126) reconstructs the OFDM symbols in the signal frame and converts the OFDM symbols to the frequency domain. The error precorrection decoder units (128 and 130) process the LDPC and BCH codes and generate a signal with erroneously received bits corrected.

The M-sequence located in the GI provides faster synchronization than in DVB-T2, fast frame and timing recovery, and reliable channel estimation. This sequence is transmitted to DTTB using spread spectrum technology and is used as a training sequence that increases the efficiency of channel condition estimation and for its correction on reception and accelerates receiver synchronization.

Figure 4 shows the data of a TDS-OFDM frame with a bandwidth of 420 (255, 165), expressed in the number of chips. In addition, variants of the PRP with the length of 511 (84) and 511 (434) are possible with the same frame length 3744 (information symbols) + 36 (TSP symbols). Thus, the PRS and postambulum are placed on the guard interval. In general, a preamble and a postamble can be placed on the guard interval. The PRS themselves (m-sequences) have a length of 255, 511, and the second numbers (165, 84, and 434) represent the auxiliary parts of these noise sequences.

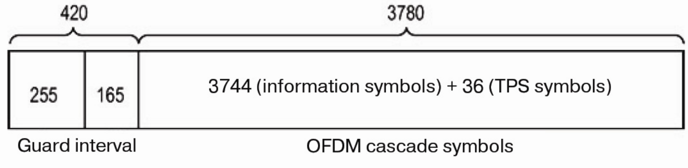


Figure 4. Schemes of the OFDM frame with the TDS-OFDM SRP located in the guard interval

The 3780 frame symbols consist of 3744 basic information symbols and 36 symbols for transmission parameter signals (TPS), which provide information to the receiver demodulator about the rate and length of the interleaving code.

Despite the high performance, the TDS-OFDM (DTTB) system does not allow the transmission of information with QAM-256 modulation, sensitive to interference, due to the appearance of mutual interference between the PSP and information symbols [3]. To reduce the appearance of interference in the TDS-OFDM (DTTB) system, in particular, the use of not one, but two PSS located in the GI is provided. Research is underway to completely eliminate the interference between the PRS and symbols.

Using guard interval to improve timing accuracy in DVB-T2

Synchronization improvements in DVB-T2 can be obtained by placing a pseudo-random sequence in the guard interval (possibly together with a cyclic prefix), as is done in the DTMB system. Unlike the DTMB system, the DVB-T2 system retains all pilot signals during the OFDM symbol and retains the initial and final portions of the cyclic prefix during the RFI. The rest of the prefix is required to terminate the transient at the beginning and at the end of each information character. The pseudo-random sequence is mixed in with the CPU or inserted in place of the cyclic prefix. The use of the PSP provides faster and more accurate symbol synchronization.

Since the weak point of the DTMB system is the interference between the PSP and a block of OFDM symbols [4], it is possible to apply the method of eliminating it proposed in [5, 6]. It consists in the use of additional pilot signals transmitted as part of information signals, and the method of "compressed sensor" (compressive sensing) for channel estimation in the time-frequency domain (Fig. 5). After the introduction of this method, the DTMB system can use QAM-256 modulation.

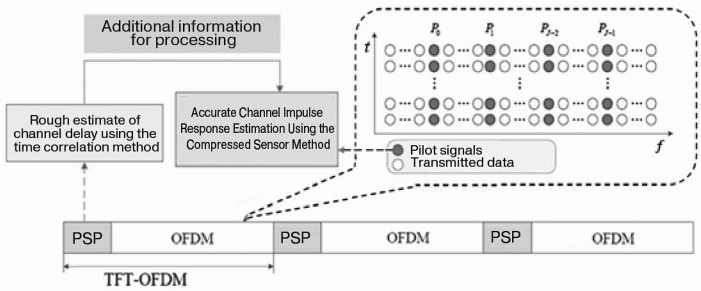
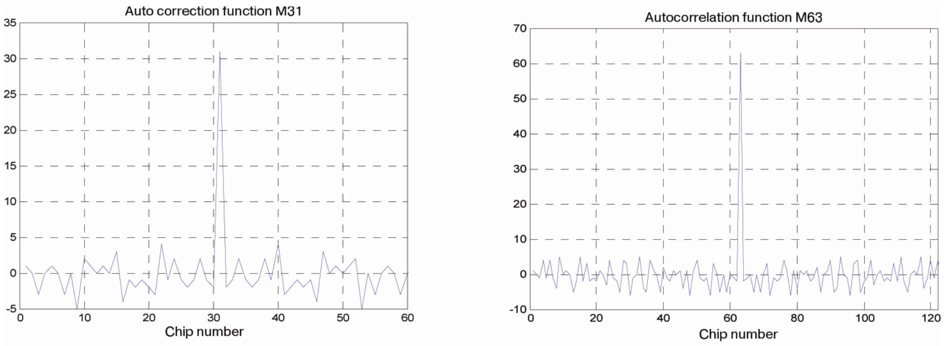


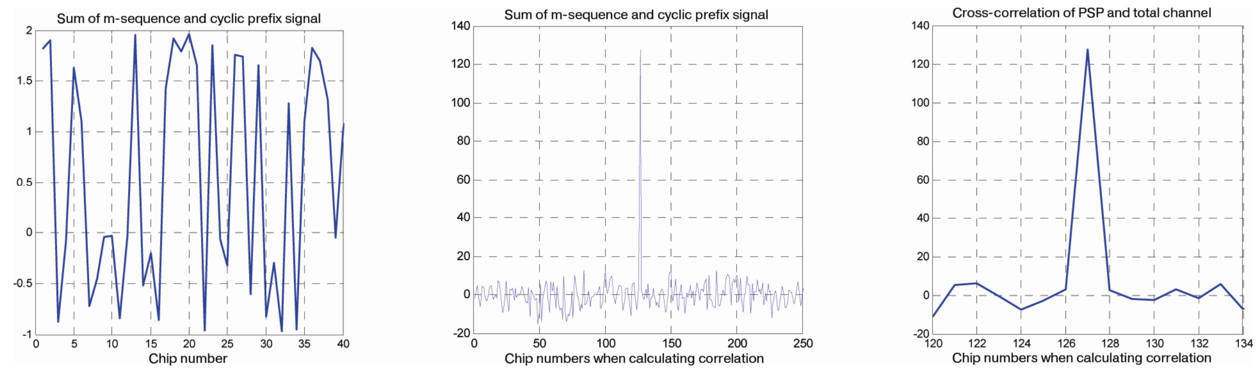
Figure 5. DTMB OFDM frame structure for channel estimation in time-frequency domains and for eliminating interference using PRP

Since in the proposed DVB-T2 structure, after mixing the PRS into the guard interval, the pilot signals are not excluded from the OFDM frame, no additional pilot signals are required. In DVB-T2, after the introduction of PRS and the above-mentioned method of eliminating interference, a high-quality QAM-256 transmission is possible. Shorter m-sequences (31 or 63 in length) can be used to reduce crosstalk between the PRS and the OFDM symbol block. They have a fairly high level of the correlation function burst relative to the side lobes (Fig. 6).



a b

Figure 6. Autocorrelation functions for m-sequences of length 31 chips (a) and 63 chips (b) obtained by simulation in Matlab environment



a b c

Figure 7. Signal after mixing the m-sequence (PSP with a length of 127 chips) to the CPU in the guard interval of the OFDM symbol (a) and signals at the correlator output (b, c) with different scales along the x axis

These PRS can be placed not in the center of the guard interval, but in its end. This will provide better protection against mutual interference between the PRS and OFDM symbol at high multipath delays.

The duration of the sequence is chosen slightly less than the duration of the guard interval in order to ensure attenuation of transients at the end of the first symbol and at the beginning of the next (Fig. 9).

It is possible to eliminate the effect of mutual interference between the PRS and the informational part of the OFDM symbol in the place of their joining by choosing the sequence duration somewhat less than the GI duration in order to ensure the decay of the transient processes at the end of the first symbol and at the beginning of the next one (Fig. 6). Offsets of the PRS to the end of the guard interval will help protect symbols for long delays in multipath environments.

Figure 7 shows the oscillograms after mixing the m-sequence (127 chips long) into the guard interval of the OFDM symbol while maintaining the cyclic prefix signal with subsequent processing of the total signal by the correlator on the receiving side. Modeling was performed in Matlab environment.

An m-sequence 127 chips long was used as a pseudo-random sequence, and another random sequence was taken as a CPU. It is obvious that the CPU signal, which is chosen as complex, does not affect the quality of the formation of the peak of the correlation function, which significantly exceeds the level of the side lobes. In order to reduce the level of the total signal, the amplitude of the m-sequence can be reduced by up to five times. In this case, the correlation function has a sufficient amplitude.

The cyclic prefix (CPU - signal copied into the guard interval) improves the transient at the beginning of each OFDM symbol. However, from this point of view, there is no need to copy the entire CPU, since the memory bandwidth and the information part of the symbol are spaced in time. Transients in a channel with a bandwidth of more than 7 MHz decay with a duration of less than 1 μs. Therefore, you can not transfer the entire CPU, but insert the start and end of the CPU at the end of the first character and at the beginning of the next character, respectively.

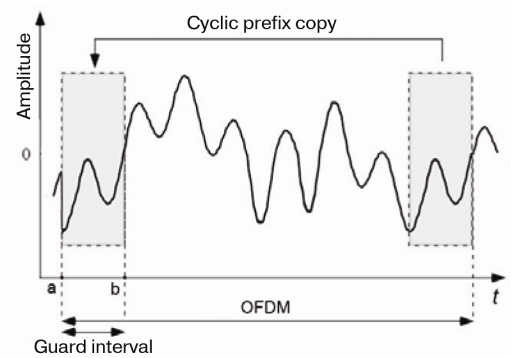


Figure 8. OFDM symbol with a guard interval and with CPU copying from the end of the symbol to this guard interval

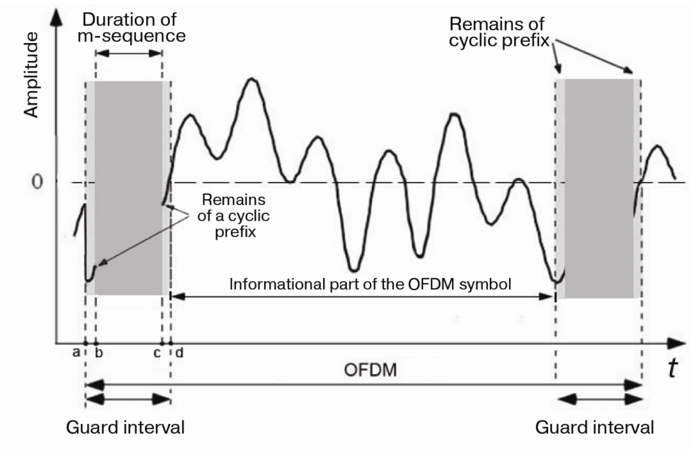


Figure 9. OFDM symbol with guard interval and partial CPU copying and freeing up space for pseudo-random sequence placement

Figures 8 and 9 show one OFDM symbol with a guard interval and with CPU copying from the end of the symbol to this guard interval. Figure 9 shows an OFDM symbol with a guard interval and partial copying of the CPU and freeing up space for placing an additional signal.

Conclusion

The proposed method of using PSP (m-sequences) in the DVB-T2 system will retain the possibility of using QAM-256 modulation while improving the quality of synchronization and its acceleration. Preservation of the main elements of the system without fundamental changes, in particular, preservation of all pilot signals transmitted in the OFDM frame, will ensure the compatibility of the modified DVB-T2 system with the existing one and will allow receiving signals using the bandwidth to old DVB-T2 receivers with a gradual increase release of new receivers with better synchronization.

References

1. *Morshed Md. S.* Synchronization Performance in DVB-T2 System. Tampere university of technology. Department of communications engineering, 2009.

2. *Yang L., Venkatachalam D.* Receiver For An LDPC based TDS-OFDM Communication System. Патент № US 20080025424 A1. 2008.

3. *Dai L., Wang J., Wang Z.* Time Domain Synchronous OFDM Based on Simultaneous Multi-Channel Reconstruction. Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, China, 2013.

4. *Bagade A. B., Deshmukh A.* Comparison on CP-OFDM and TDS-OFDM Using Compressive Sensing Theory in Wireless Systems. Journal of Innovation in Electronics and Communication Engineering, Vol. 5(2), July – Dec 2015.

5. *Yang F., Ding W., Dai L.* Joint Time-Frequency Channel Estimation Method for OFDM Systems Based on Compressive Sensing. IEEE. Research Institute of Information Technology. China. 2014.

6. *Jain S.* Time-Frequency Training OFDM using Matlab for high speed environments. International Journal of Scientific and Research Publications, Vol. 3, Issue 7, July 2013.

MODIFIED METHODS OF CIRCUIT SIMULATION OF RADIO ENGINEERING DEVICES IN THE TIME DOMAIN

Georgy A. Dolin, Anastasiya Y. Kudryashova,
Moscow Technical University of Communications and Informatics, Moscow, Russia,
dolin1974@gmail.com, asykka@bk.ru

DOI: 10.36724/2664-066X-2020-6-2-7-11

ABSTRACT

Today, different modeling methods are used for computer analysis of circuits of radio engineering devices (RED) in the time and frequency domains. The article provides a comparison and highlights the features of using the methods of nodal potentials and variable states. Developed methods of optimization of electrical circuits and discusses the possibility of calculating the margin of stability when changing the parameters of the circuit elements and the search of critical parameter values; theoretically and experimentally confirmed the advantages of using MEAs in the analysis of RED; proposed and implemented ways to eliminate the major disadvantages of the IPU; expanded and improved methods for obtaining the mathematical model of the circuit; the mathematical method allows to obtain the characteristic polynomial of a circuit without calculating its transfer function; the developed block for processing parameters of electrical circuit elements using scaling coefficients can significantly improve the accuracy of calculations; the use of speed-optimized algorithms makes it possible to analyze fairly complex circuits on a medium-performance PC. Developed software allows to analyze a wide class of linear, linearized, and nonlinear circuits for the RED, containing the active elements. The analysis of real electrical circuits proves the validity of all the proposed methods.

KEYWORDS: *circuit modeling, nodal potential method, variable state method, CAD, design automation*

INTRODUCTION

Several methods are used for radio electronic designs (RED) modeling: the method of contour currents, no-load and short-circuit, superposition, nodal potentials and state variables. For computer modeling, it is advisable to use the last two. Method of nodal potentials. The advantage of this method is the comparative simplicity of the process of representing the mathematical model of the scheme when forming a system of equations of the scheme [4]. In the method of nodal potentials, the mathematical model has the form.where Y is the matrix of nodal conductivities; X is the vector of nodal potentials; J is the vector of right-hand sides (the setting vector).

Another advantage of the nodal potential method is that you can immediately get the potential values of all the nodes of the RED, which greatly facilitates further analysis, i.e. to determine the current in the branch or the voltage between two nodes, you do not need to solve a system of equations every time. However, the disadvantage of this method is the complexity of constructing frequency characteristics and evaluating the stability of the RED. To build frequency characteristics, it is necessary to calculate the required parameter (current or voltage) for each frequency value. This is because the method of nodal potentials does not have the ability to obtain the characteristics of the chain in the form of an algebraic expression.

Stability estimation in the nodal potential method can only be performed indirectly since this method does not allow obtaining the characteristic polynomial of the transfer function. To assess stability using the nodal potential method, various stability criteria can be used that do not require knowledge of the characteristic polynomial. However, these criteria in the case of an unstable system do not provide information on how to change it to make it stable.

I. Circuit Simulation

Modern circuit modeling programs include five main stages [4]: entering a graphical description of the circuit, entering a text description of the circuit, an analysis task, modeling, and output of the analysis results.

At the first stage, the user creates a schematic or equivalent diagram of the device in a graphical format, i.e., as it is shown in the drawing. All elements of the scheme are conventional graphic designations in accordance with Spice Standard.

At the second stage, the graphic image of the diagram is converted to a text format. This procedure is usually hidden from the user and is performed by the computer itself. However, if necessary, the user can change or add some additional information to the text description (parameters of active element models, electric power sources, etc.). A text description is the main format that contains all the necessary information about the circuit: its topology (the relative position of elements), the ratings of passive elements, the parameters of active element models, and the parameters of electric power source models. It is based on this format that a mathematical model of the scheme is created.

At the third stage, after converting the schema to a text format, the user sets the type and parameters of the analysis. There are generally three main types of analysis: transient analysis (in the time domain), frequency response analysis, and DC transfer function analysis.

At the fourth stage, the circuit is modeled. It includes creating a mathematical model of the scheme and calculating its characteristics in accordance with the type of analysis and task data.

Finally, at the last stage, the results of the analysis are displayed. The results of the analysis are either graphs or tables. At the request of the user, they can be output to a printing device or to a magnetic data carrier.

A simplified diagram of the algorithm of the RED circuit modeling system is shown in Fig. 1.

In circuit design automation programs, solving the system of equations (1) is a frequently repeated task. The problems of solving this system using a computer are as follows. When calculating electronic circuits, the y matrix is usually very sparse, i.e. it contains many zero elements. In the process of solving (1), the matrix Y is transformed. At the same time, non-zero elements appear in place of zero elements, which leads to additional costs for the solution time and memory. Therefore, the first problem is to maintain a high degree of sparsity of the matrix Y in the solution process.

Another problem is the problem of maintaining the accuracy of results at the level of rounding errors in the source data.

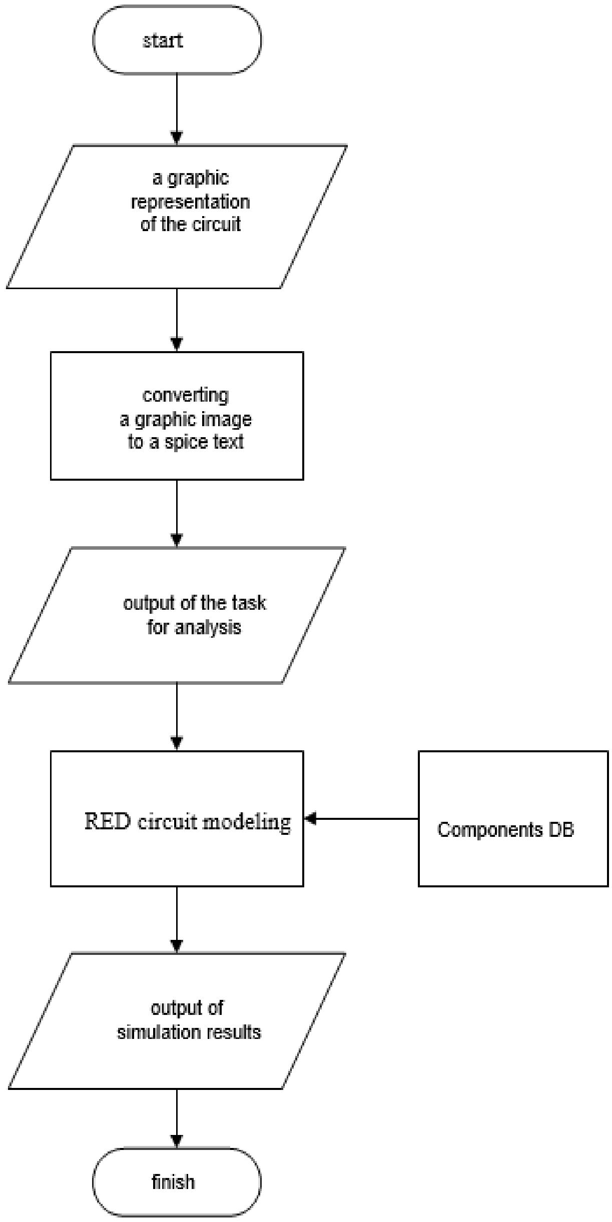


Figure 1. Simplified scheme of the algorithm of the circuit modeling system operation

Method of state variables. A distinctive feature and undoubted advantage of this method is the possibility of obtaining a mathematical model of the designed device in the normal Cauchy form, i.e., allowed with respect to derivatives [1]. Thus, you can use integration programs that are regularly available in the mathematical software of a computer.

This method is effective in designing radio engineering devices for which information is enclosed in the envelope of the carrier signal (high - quality AM and FM filters, amplifiers, correctors, special amplifying devices of multi-channel telecommunications covered by deep feedbacks, etc.), as well as modeling devices, systems and communication networks.

The method of state variables allows you to get the transfer function as a ratio of two polynomials, which is more informative than the complex number that must be found for each frequency value in the nodal potential method. Stability assessment can be performed without resorting to indirect methods, but directly by finding the roots of the characteristic polynomial.

The main disadvantage of the state variables method is the complexity of forming a mathematical model of the scheme and the orientation to explicit methods for solving systems of differential equations, which are practically not suitable for calculating schemes with a large spread of time constants due to restrictions on the calculation step. Implementation of implicit methods is difficult due to the complexity of calculating the Jacobi matrix.

Currently, most modern industrial RED circuit modeling programs use the nodal potential method as the basic one. However, when solving some problems (for example, determining the stability of RED nodes covered by feedback and calculating the stability margin), it is more appropriate to use the method of state variables.

In recent decades, there has been a noticeable trend in the development of electronics to move from the design of individual devices for narrow purposes to the design of complex hardware complexes designed to solve a wide range of tasks in changing external conditions. In this regard, the radio engineering devices (RED) design engineer has several new problems. The main one can be formulated as follows: how to improve the quality of the initial design of the RED, so that it can be partially or completely excluded from working out on a laboratory (material) layout? This is due to the fact that usually the initial design of the RED is very far from meeting the requirements put forward by the technical task for a separate device, system or complex that this device should be part of. Its subsequent improvement on the layout involves significant forces of designers and production workers in the design process, whose work is ineffective due to numerous and unavoidable alterations. As a result of all this, the design process is unacceptably stretched and becomes excessively expensive [2].

In addition to the above, trends in the development of modern radio technology are characterized by a significant increase in the complexity of requirements for electrical and structural parameters of RED, the widespread introduction of semi-conductors and microelectronics, and increased requirements for reliability and serial production of products. Increasing requirements for electrical parameters led to the need to use more complex circuit solutions and optimal selection of parameters of elements and circuit diagrams. Along with the above, the requirements for the size and weight of the RED, as well as for energy consumption, have increased, which necessitated the use of modern element base in these devices. The equivalent schemes used in synthesis and analysis have also become more complex. Particularly noteworthy is the increase in the volume of equivalent circuits due to the use of semiconductor devices and microchips (for example, a linear equivalent circuit of a high-frequency transistor contains at least 15-20 active and reactive elements).

In addition to defining equivalent circuits, the developer of radio equipment needed a more thorough and accurate analysis to optimize the characteristics, as well as deterministic or probabilistic calculation of variations in the characteristics of circuits when analyzing the product's serial capability. Conducting such an analysis using traditional methods is difficult, and in some cases impossible. Known analytical relations are usually derived with many assumptions and simplifications, which significantly narrows the scope of their application and can lead to unacceptable errors. Therefore, the method of full-scale modeling for determining and refining the parameters of RED is widely used. However, the possibilities of the experiment are limited due to high material costs and the fundamental impossibility of solving some design issues. These questions include:

- – determination of average values of device parameters without conducting statistical tests, since it is difficult to select active elements with average values of parameters for the experiment;
- – investigation of the influence on the measured characteristics of par-zit devices structural elements and measuring equipment;
- – investigation of the behavior of the circuit in difficult-to-reproduce external conditions;
- – requirements for reducing development time, etc.

II. Modified Methods

A few methods are used to model radio engineering devices (RED): the method of contour currents, no-load and short-circuit, overlap, nodal potentials and state variables. To calculate the steady-state periodic modes, we will use two representations of the mathematical model scheme.

The first type of model is set by the joint system

$$f\left(x,\ x,\ t\right)\ =\ 0 \tag{1a}$$

$$x\left(0\right)\ -\ x\left(T\right)\ =\ 0 \tag{1b}$$

Here f is a system of differential equations; x(t) is the desired vector function of variables that determine the state of the circuit; T is the period.

Thus, a nonlinear boundary value problem is defined, which can be solved by known methods. The second type of model reflects the properties of an electrical circuit to a greater extent:

$$f(v(t),t) = \frac{d}{dt}q(v(t)) + i(v(t)) + u(t) = 0 \quad (2a)$$

$$v(0) - v(T) = 0$$

(2b)

where u(t) is the vector of input signals; v(t) is the vector of nodal voltages; i(v(t)) and q(v(t)) are the vectors of nodal currents and nodal charges (fluxes), respectively.

This type of model avoids the difficulties associated with the problem of charge conservativeness (1).

The use of finite-difference methods for solving the boundary value problem (1) and (2) leads to a system of nonlinear equations, the number of unknowns of which is determined by the number of sampling points. For example, if the inverse Euler method is chosen to approximate the derivative in (1b), we have a system of equations of the form:

$$f_i\left(\frac{x(t_j) - x(t_{j-1})}{h_j}, x(t_j), t_j\right) = 0, j-1,\ldots,M, \quad (3)$$

where $h_j = t_j - t_j$-1, M is the number of sampling points.

Combining this system with the periodicity conditions (1b), we obtain the system of nonlinear equations necessary for calculating *x(t)* and *x(0).* The application of the Newtonian iterative process leads to a solution at the iteration step of a linear system of the form:

$$\left(L + B\right)\left(xk+1 - xk\right) = -f\left(xk\right) \quad (4)$$

where *k* is the index of the Newtonian iteration.

The L and B matrices have the structure:

$$L = \begin{bmatrix} \frac{C_1}{h_1} + G_1 & & & \\ -\frac{C_1}{h_2} & \frac{C_2}{h_2} + G_2 & & \\ \bullet & \bullet & \bullet & \bullet \\ & & -\frac{C_{m-1}}{h_m} & \frac{C_m}{h_m} + G_m \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & \bullet & 0 & \frac{C_m}{h_m} \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & 0 & 0 \end{bmatrix},$$

where Ci is the capacitance and Gi is the conductivity of the circuit [1].

Another common method for solving problem (1) or (2) is the method of reducing the boundary value problem to the Cauchy problem (shooting method). This method leads to the solution of a system of nonlinear equations (1b) with respect to the vector x(0), and at each step of the iterative process, the vector x(T) is the solution of the Cauchy problem with the corresponding initial conditions. This approach is in good agreement with circuit modeling programs, since it allows using standard integration methods that are well-developed for electrical modeling tasks.

It can be noted that if the system of nonlinear equations

$$x(0) - x(T, x(0)) = 0 \quad (5)$$

solve by the method of simple iteration

$$x(0)k+1 = x(T, x(0)k) \quad (6)$$

until the condition for the end of the iterative process is met for the specified error e:

$$\left\| x(0)k+1 - x(0)k \right\| < e \quad (7)$$

what we get is the application of standard approaches for finding periodic solutions by sequential integration over a large time interval. The main goal, of course, is to reduce the initial cost of modeling [3]. For this purpose, it is advisable to consider the possibility of applying the Newton method for solving system (5), which requires fewer steps. Using the Newtonian process leads, at the iteration step, to a system of linear equations of the form:

$$\left[\frac{\partial x(T)}{\partial x(0)} - I\right]\Delta x(0)^k = x^k(0) - x(T)^k, \quad (8)$$

where I is the unit matrix [2].

The main problem that arises in this case concerns the cost of forming the Jacobian for (8) or the sensitivity matrix $\frac{\partial x(T)}{\partial x(0)}$ under the initial conditions. Representing this matrix as differences requires multiple solutions to the Cauchy problem and is practically inefficient. For this reason, an economical method for calculating derivatives $\frac{\partial x(T)}{\partial x(0)}$ is proposed by finding solutions to systems of equations in variations:

$$\frac{\partial f}{\partial \dot{x}} \bullet \frac{d}{dt}\left(\frac{\partial x(t)}{\partial x(0)}\right) + \frac{\partial f}{\partial x} \bullet \left(\frac{\partial x(t)}{\partial x(0)}\right) = 0 \quad (9)$$

The main idea of this proposal is to combine operations performed during system integration (1a) and (9) with implicit methods. This leads to the solution of a system of linear equations with the same coefficient matrix at each step of integration, which significantly reduces costs. Algorithms of this type were for a long time basic for steady-state oscillation modeling programs [5].

However, despite saving operations, methods for reducing to the Cauchy problem with Newtonian iteration do not fully meet the increasing requirements for complexity characteristics. Their main limitations are: the cost of calculating the matrix of derivatives; the cost of Gaussian elimination for solving the system (8) with a dense matrix, proportional to N3, where N is the number of integrable variables.

The main areas of improvement of algorithms are related to attempts to avoid calculating the matrix $\frac{\partial x(T)}{\partial x(0)}$, as well as reducing the cost of solving a system with dense matrices. The development of these computational procedures based on the application of Krylov subspace methods is discussed below [4].

We also indicate the previously popular Skelbo extrapolation algorithm due to its simplicity of implementation. His idea is to use an extrapolation scheme to speed up the convergence of the sequence obtained by simple iteration (6): x(0) 0, x(0) 1, ...

The computational scheme can be represented as: x(0) = yn,

$$\begin{aligned} x(0)r+1 &= X\left(x(0)r,\ T\right),\ r = 1,\ 2,\ ...,\ m, \\ yn+1 &= EXTRAP\ \left(x(0)0,\ x(0)1,\ ...,\ x(0)m\right) \end{aligned} \qquad (10)$$

III. Conclusion

The advantages of the algorithm are that it allows you to avoid calculating derivatives $\frac{\partial x(T)}{\partial x(0)}$ when solving nonlinear problems, allows you to achieve quadratic convergence, and provides ease of implementation based on the standard integration methods used.

References

[1] 1. G. A. Dolin, "Object-Oriented Representation of Mixed Models Knowledge in the Design of Electronic Devices in CAD Electra," *2020 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russia, 2020, pp. 1-5, doi: 10.1109/IEEECONF48371.2020.9078546.

[2] 2. G. A. Dolin, "Using a distrib ed database of parameters of electronic components in course design," Methodological issues of teaching infocommunications in higher school, 2018. No 2. pp. 10-14.

[3] 3. G. A. Dolin, "Formation of the knowledge base of red expert design for conducting coursework and lectures," Methodological issues of teaching infocommunications in higher school, 2018. No 2. pp. 46-53.

[4] 4. G.A. Dolin, "Circuit Analysis, Synthesis and Simulation of Radio Devices in Electra CAD," *2019 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russia, 2019, pp. 1-6, doi: 10.1109/SOSG.2019.8706814.

[5] 5. G. A. Dolin, "Automation of synthesis of structural and basic electrical circuits of RED in production and object-oriented expert systems," 2018 Actual problems of radio and film technologies. II all-Russian scientific and technical conference. Saint Petersburg, Saint Petersburg state Institute of film and television, October 24-27, 2018. Collection of works. pp. 40-45

[6] 6. K. V. Boychenko, J. Rivory, I. V. Boychenko, A.Y. Kudryashova, "Interactive Space as a Signal Processing Device," 2020 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Svetlogorsk, Russia, 2020, pp. 1-6, doi: 10.1109/SYNCHROINFO49631.2020.9166033.

[7] 7. K. V. Boychenko, I. V. Boychenko, A.Y. Kudryashova, "Interactive built environment in shaping users orientation and navigation in space ," 2020 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 2020, pp. 1-4, doi: 10.1109/IEEECONF48371.2020.9078658

[8] 8. K. V. Boychenko, I. V. Boychenko, A. Y. Kudryashova, "Interactive Built Space as the New Means of Information Communication," 2019 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO), Russia, 2019, 10.1109/SYNCHROINFO.2019.8813912.

THE INVESTIGATION AND EVALUATION MULTISERVICE NETWORK NGN/IMS FOR MULTIMEDIA TRAFFIC

Bayram G. Ibrahimov
Azerbaijan Technical University, Baku, Azerbaijan, i.bayram@mail.ru

Arif H. Hasanov
Military Academy of the Republic of Azerbaijan, Baku, Azerbaijan

DOI: 10.36724/2664-066X-2020-6-2-12-15

ABSTRACT

The functional architecture of the subsystem picture messaging-IMS (Internet Protocol Multimedia Subsystem) for multimedia session management for multiservice NGN/IMS networks that determine the interaction systems and NGN protocols are investigated. The basic element of the IMS core network architecture are session control function CSCF (Call/Session Control Function) which is implemented on the SIP-server (Session Initiation Protocol) using the protocol.
A block diagram of the functioning of the traffic service model of multimedia services in the NGN/IMS network when establishing sessions that use the network elements of the user data server HSS (Home Subscriber Server), CSCF control system core, signaling and media gateways (SGW & MGW, Signaling Gateway & Media Gateway). The characteristics of the effectiveness of the IMS and found that in the NGN / IMS multimedia services network in real time significantly alter the traffic characteristics as a service, and efficiency, which requires new models and approaches to assess the main indicators of quality of the service Triple Play.
Based on the analysis of the quality NGN/IMS networks in the provision multimedia services, the mathematical model for evaluating the quality of services. This model takes into account the properties of self-similar random process with Hurst index and describes the quality of functioning of NGN/IMS architectural concept. Based on the model studied when performing voice services with the ability to activate multimedia applications, video telephony, multimedia traffic, Triple Play services.
According to the study of mathematical models of NGN / IMS networks using SIP-servers analytical expressions to assess the indicators of a common set of operating characteristics of IP-based networks, which take into account the recommendations of ITU-T Y.1540 and Y.2000 and determined the average waiting time in the queue IMS system in the provision of multimedia services with the required parameters, providing a guaranteed quality of service QoS (Quality of Service). According to the latest recommendations of the ITU-T and based on SMO with queues can be defined by five network characteristics that are indicative of the effectiveness of multi-service NGN/IMS networks and an important indicator of QoS.

KEYWORDS: *multimedia traffic, SIP protocol, IMS subsystem, HSS server, quality of service, Diameter protocol, SIP-server.*

Information about authors

Ibrahimov Bayram Ganimat oglu, *professor of Azerbaijan Technical University, Bak, Azerbaijan*
Hasanov Arif Hasan oglu, *an associate of the Military Academy of the Republic of Azerbaijan, Baku, Azerbaijan*

Introduction

At present, in the field of telecommunications, as the number of multimedia services and the volume of useful and service traffic increases, it is necessary to create effective multi-service communication networks of the next generation based on the multimedia IMS platform, which provide the necessary quality of services provided to subscribers.

Studies have shown [1, 2] that one of the most important tasks of multiservice NGN / IMS networks using SIP servers is to support the quality of service of multimedia traffic, primarily to minimize the time characteristics of service access delays. It should be noted that so far this issue has not been studied well enough and remains poorly studied [3, 4].

The efficiency of the IMS multimedia communication subsystem largely depends on the capacity of NGN / IMS networks using SIP servers and network equipment of the CSCF session management system. To guarantee the service of such loads, maximum NGN / IMS network bandwidth, minimum average delays in the IMS system, increased efficient use of SIP server resources, etc. are required.

In view of the above, an important question arises - the assessment of the quality of service of the multimedia traffic of the IMS multimedia communication, which depends both on the service access algorithm and on the efficiency of the NGN / IMS multi-service networks using the HSS home subscriber server.

In this work, a study and evaluation of the effectiveness of multiservice NGN / IMS networks in the transmission of multimedia traffic is carried out.

General problem statement and scheme of functioning the investigated model of the NGN / IMS network

To accurately describe the multimedia traffic passing between the links of the NGN / IMS network and its network elements of the CSCF session control system, an analysis of the statistical characteristics of the traffic and the choice of an adequate MM of a self-similar random process is required. Analysis of the IMS concept [1, 5, 6] showed that the further expansion of the range of multimedia services demanded by users has led to the emergence of a Triple Play Service package, which is a triple package that includes voice services, Internet access and viewing of television programs that require multi-rate service systems and a wide range of transmission rates - from 2 to 2500 Mbit / s.

As part of the study, the structural and functional diagram of the NGN / IMS network model, built on the basis of the systemic and technical analysis of multimedia services, is shown in Fig. 1 and consists of an access buffer storage (BN), a switching node using IP / MPLS, and an IMS subsystem.

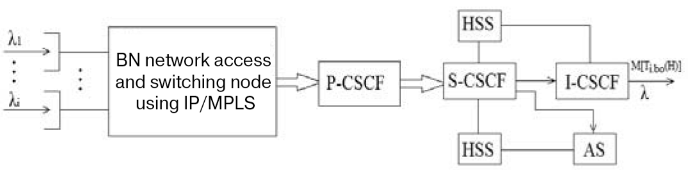


Figure 1. Structural and functional diagram of NGN / IMS network investigated model

The IMS multimedia communication subsystem as well as the Softswitch architecture, which is a set of functions connected by standard interfaces. The IMS system consists of Proxy, Interrogating and Serving CSCFs, which use SIP and Diameter IMS session establishment protocols.

In the scheme, the HSS is a user data base and provides access to personal data of subscribers associated with multimedia services. The AS communicate with the S-CSCF via SIP and maintain SIP sessions.

On the basis of studies [3, 5], it has been established that the transmitted traffic systems and protocols of NGN / IMS networks have a special structure. It was revealed in [2] that multimedia NGN / IMS traffic has self-similarity with the Hurst coefficient, H, lying in the range of 0.5≤H <1 and is equal to $H = 1 - 0{,}5\beta,\ 0 < \beta < 1$ [3].

In order to solve the problem under consideration and taking into account the importance of the interaction of the system and the NGN protocols, a mathematical model (MM) of NGN / IMS networks is proposed, taking into account the properties of self-similarity of traffic when providing multimedia services.

Description of the mathematical model and analysis of indicators of NGN / IMS networks

Recently, the analyzes carried out show [3, 4, 6] that the nature of the transmitted service and payload traffic in multiservice NGN / IMS networks differs from the Poisson one, and the service process from the exponential distribution law. Then it is necessary to turn to the results of the theory of diffusion approximation, which provides the developer with approximate, but very simple from the computational point of view, formulas [4].

Considering the last assumption and algorithms of the functional architecture of the multimedia messaging subsystem, the considered NGN / IMS networks, in the general case, is a queuing system (QS) of type $fBGI/G/N_S/N_{бн}$ with some assumptions at increased load $\rho_i \le 1,\ i = \overline{1,n}$.

Suppose that the arrival of traffic flows for service is non-Poissonian with intensity λ_i, $i = \overline{1,n}$, the distribution of the service duration is arbitrary, serving NS SIP servers have a common buffer storage (BN) with an unlimited capacity $N_{бн} \le \infty$ (where n – total number of traffic flow types).

Considered MM in the form of a general QS $fBGI/G/N_S/N_{бн}$ with queues, takes into account the performance indicators of NGN / IMS networks, the properties of self-similarity of multimedia traffic and the features of approximate methods of diffusion approximation.

Based on diffusion approximation methods and self-similarity properties of transmitted traffic in a given QS, the load of the multimedia message transmission subsystem is determined as follows:

$$\rho_i(H,\lambda_i) = [B_i^{(1)} \cdot f(H) \cdot \lambda_i / (V_i \cdot N_S)] \le 1,\ i = \overline{1,n} \qquad (1)$$

where $f(H)$ – function that takes into account the property of self-similarity of incoming traffic packets of the protocol of NGN / IMS networks; $B_i^{(1)}(t)$ – the average value of the duration of service of streams of packets of *i*-th traffic; V_i – packet stream rate of *i*-th traffic.

For the purpose of convenient calculation, expression (1) can be presented in the following edition, introducing the load intensity of the multimedia messaging subsystem for the packet stream *i*-th traffic

$$\rho(H) = \sum_{i=1}^{n} \rho_i(H, \lambda_i) \le 1, \quad i = \overline{1,n} \tag{2}$$

It should be noted that the idea and features of the diffusion approximation method described in [4, 5] is that the queue length distribution $P(n \ge 0)$ in system $fBGI/G/N_S/N_{бн}$ with queues at critical load equal to $\rho(H)$, is approximated by the following distribution:

$$P(n \ge 0) = \begin{cases} 1 - \rho(H), & при\ n = 0 \\ \rho(H)[1 - \rho(C_A, C_B)] \cdot [\rho(C_A, C_B)^{n-1}, & при\, n \ge 0 \end{cases}, \tag{3}$$

In (3) $\rho(C_A, C_B)$ is a load of the CMO type $fBGI/G/N_S/N_{бн}$, taking into account the quadratic coefficients of variation of the distribution C_A^2 intervals between incoming multimedia messages and the distribution of lengths of multimedia messages C_B^2 and is expressed as follows:

$$\rho(C_A, C_B) = \exp[-\frac{2[1 - \rho(H)]}{\rho(H) \cdot C_A^2 + C_B^2} \le 1 \tag{4}$$

Therefore, for the Poisson incoming stream of traffic packets $C_A^2 = 1$, and for the geometric distribution of message lengths is

$$C_B^2 = C_B^2(L_{сз.п}) = (\sigma_{L_{сзоп}} / L_{сз.п})^2 = p_c, \tag{5}$$

where p_c – the likelihood that the average value $L_{сз.п}$ not equal to the packet data unit for the geometric distribution of the length of multimedia messages; $\sigma^2_{L_{сз.п}}$ – variance for the geometric distribution of the length of a multimedia message.

Estimation of the transmission time of multimedia traffics

In this IMS system, we assume that all network elements are loaded identically and the load factor of one SIP server, taking into account the self-similarity property of the incoming *i*-th traffic packet $\rho_i(H)$ in a multichannel system is determined by the following expression:

$$\rho_i(H) = [\frac{L_{i.сз.п}}{V_i \cdot N_S} \cdot \mu_i \cdot \lambda_i \cdot B_i^{(1)}] \le 1, \; i = \overline{1,n}, \tag{7}$$

where $L_{i.сз.п}$ – average value of the length of the transmitted multimedia message of the *i*-th traffic.

Taking into account (2),…, (7) and the results obtained using the diffusion approximation method, it is possible to determine the probabilistic-temporal characteristics of NGN / IMS networks [4, 5].

Based on the proposed MM in the form of a general QS $fBGI/G/N_S/N_{бн}$ with queues, the average queue length is expressed as follows:

$$E[L_{i.до}, H] = 0{,}5G(\rho, N_S) \cdot \frac{C_A^2 + C_B^2}{1 - \rho_i(H)} \cdot$$
$$[B_i^{(1)} \cdot f(H) \cdot \lambda_i / (V_i \cdot N_S)], \; i = \overline{1,n} \tag{8}$$

где $G(\rho, N_S)$ – является множителем и определяет вероятность того, что потоки пакетов трафика придя в подсистему мультимедийной связи IMS, застанет все SIP-серверов занятыми и встанет в очередь и определяется как C – формула Эрланга [4].

С учетом (8) и приближенной формулы Крамера и Лангенбах-Бельца [5], среднее значение время ожидания *i*-го потока пакета трафика в очереди определяется следующим выражением:

$$E[T_{i.во}(H)] = \frac{t_S}{N_S[1 - \rho_i(H)]} \cdot G(\rho, H) \cdot \frac{C_A^2 + C_B^2}{2}, \; i = \overline{1,n} \tag{9}$$

where t_S – average service time of NGN / IMS traffic packet flows.

Figure 2 shows the dependence of the average waiting time in the queue in the NGN / IMS network on the number of SIP servers and the transmission rate of multimedia traffic.

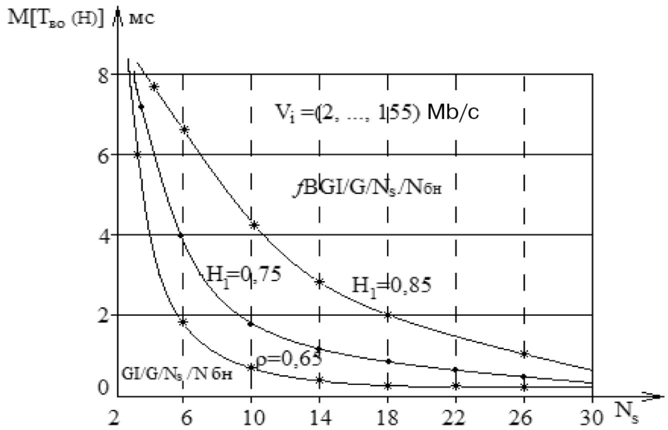


Figure 2. Dependence of the average waiting time in the queue in the NGN / IMS network on the number of SIP servers

Graphical pendency family analysis $E[T_{i.во}(H)] = F(\rho_i, N_S, V_{i.k}, H)$ shows that an increase in the limited number of SIP servers in the NGN / IMS network using HSS home subscriber servers $N_S \ge 20,...,25$, that meet the IMS fault tolerance requirements, help to minimize the average waiting time in the queue at a given speed $V_i \ge (2,...,155)$ Mbps and Hurst coefficient $H_i = 0{,}75,...,085$.

Thus, based on MM, the obtained expressions (7), ..., (9) are one of the main characteristics of the NGN / IMS network efficiency and QoS indicators, which are based on the IMS architecture.

Conclusions

The conducted research has shown that assessments and analysis of the indicators of the IMS multimedia communication platform, both the average queue length of the queue system and the average delay time of traffic transmission in the load function, are very important when studying the effectiveness of NGN / IMS multiservice networks, taking into account the properties of self-similarity of service and useful traffic.

As a result of the study, an MM was proposed in the form of a general QS $fBGI/G/N/N_{бн}$ with queues and analytical expressions were obtained that allow one to evaluate the performance indicators of NGN / IMS, taking into account the properties of self-similarity of multimedia traffic, ensuring guaranteed quality of QoS services, regulated by the ITU-T recommendations.

References

1. Fillimonov A.Yu. (2007). *Construction of multi-service Ethernet networks*. St. Petersburg: BVH – St. Petersburg, 592 p.

2. Ibrahimov B.G., Hasanov A.G., Ibrahimov R.F. (2016). *Analysis of performance characteristics of multiservice communication networksbased on the IMS subsystem*. Moscow: MTUCI, pp. 36-37. (*in Russian*)

3. Ibrahimov B.G., Hasanov A.G. (2016). *Investigation of the quality of the functioning of NGN / IMS networks during the establishment of a multimedia session*. Moscow: PFUR, 2016, pp. 21-24.

4. Basharin G.P.., Bocharov P.P., Kogan Ya.A. (1989). *Analysis of queues in computer networks. Theory and methods of calculation*. Moscow: Nauka. 336 p.

5. Gaidamaka Yu.V., Zaripova E.R., Vikhrova O.G. (2014). Approximate method of session initiation delay performance evaluation in IP multimedia subsystem. *T-Comm*, no.8, pp. 19-23.

6. Bosse J.G. Devetak F.U. (2007). *Signaling in Telecommunications Networks*. 2nd Ed. New York: Wiley. 830 p.

SEMI-INFINITE WAVEGUIDE MADE OF PIEZOELASTIC MATERIAL OF CLASS 6MM

M. V. Belubekyan, A. A. Papyan,
Institute of Mechanics National Academia Since Armenia, Yerevan, Armenia,
mbelubekyan@yahoo.com, aro088@mail.ru

DOI: 10.36724/2664-066X-2020-6-2-16-19

ABSTRACT

The problems of coupled physical fields, such as the interaction of mechanical and electromagnetic fields, piezoelectric effect, electrostriction and others, are the most urgent. The study of the issues of wave propagation in piezoelectric materials is also relevant. In this paper the propagation of a monochromatic electroelastic signal in a semi-infinite piezoelectric layer is considered. Let's consider different cases of the boundary conditions, from which localized vibrations can be obtained in the vicinity of the free edge.

KEYWORDS: *piezoelectric, wave propagation, waveguide, vibration, frequency, Bleustein-Gulyayev waves.*

INTRODUCTION

The problems of wave propagation in piezoelastic media are described in [1-9]. In [1, 8], the appearance of internal resonance in the case of propagation of a monochromatic electroelastic signal in a piezoelectric layer is considered. In [4], a system of a layer and a half-space of elastic and piezoactive materials is considered, the layer and half-space can freely slide relative to each other. The work [5] considers the identification of qualitative effects in the case of propagation of a monochromatic electroelastic signal in piezoceramic, dielectric and inhomogeneous layers over the thickness. This paper considers the propagation of a monochromatic electroelastic signal in a semi-infinite piezoelectric layer.

Variable separation method

Various variants of the combination of boundary conditions are considered, some of which allow using the method of separation of variables.

A piezoelectric layer of hexagonal symmetry of class 6 mm in a rectangular Cartesian coordinate system 0xyz is located so that the 0z axis is parallel to the symmetry axis of the piezoelectric layer, and the 0xy plane is the plane of symmetry of the piezoelectric. A piezoelectric layer in a rectangular Cartesian coordinate system occupies $0<x<\infty;\ 0\le y\le h;\ -\infty<z<\infty.$ region (Figure 1).

The equations of propagation of a plane purely shear electromechanical wave can be written in the following form [2].

$$c_t^2\Delta w=\frac{\partial^2 w}{\partial t^2},\quad \Delta\psi=0 \tag{1}$$

where

$$c_t^2=\frac{\tilde{C}_{44}}{\rho},\ \tilde{C}_{44}=C_{44}(1+\chi),$$

$$\chi=\frac{e_{15}^2}{\varepsilon C_{44}},\quad \psi=\varphi-\frac{e_{15}}{\varepsilon}w. \tag{2}$$

w – shear offset, C_{44} – shear modulus, ρ – material density, ε – dielectric constant, e_{15} – piezoelectric module, χ – electrical connection coefficient, φ – electrical potential.

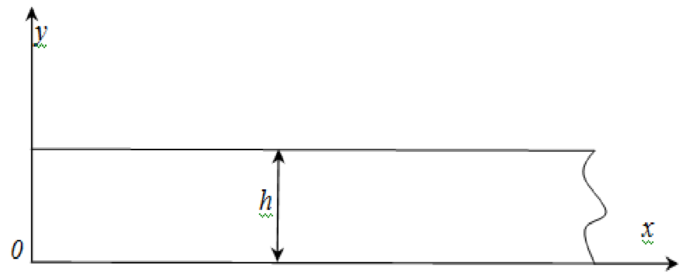


Figure 1. Location of the piezoelectric layer in a rectangular Cartesian coordinate system

For $y=0$ and $y=h$ three variants of boundary conditions are possible, which allow using the method of separation of variables:

$$w=0,\quad \varphi=0 \text{ or } w=0,\quad \psi=0 \tag{3}$$

$$\sigma_{23}=0,\quad D_2=0 \text{ or } \frac{\partial w}{\partial y}=0,\ \frac{\partial\psi}{\partial y}=0 \tag{4}$$

$$w=0,\quad D_2=0 \text{ or } w=0,\ \frac{\partial\psi}{\partial y}=0 \tag{5}$$

The second forms of the boundary conditions follow from taking into account the material equations:

$$\sigma_{31}=C_{44}\frac{\partial w}{\partial x}+e_{15}\frac{\partial\varphi}{\partial x},\ \sigma_{32}=C_{44}\frac{\partial w}{\partial y}+e_{15}\frac{\partial\varphi}{\partial y}$$

$$D_1=-\varepsilon\frac{\partial\varphi}{\partial x}+e_{15}\frac{\partial w}{\partial x},\ D_2=-\varepsilon\frac{\partial\varphi}{\partial y}+e_{15}\frac{\partial w}{\partial y} \tag{6}$$

$\sigma_{31};\sigma_{32}$ – stress tensor components, D_i – electric induction vector components.

Border conditions $\sigma_{23}=0,\quad \varphi=0$, at $y=0$ and $y=h$ do not separate variables.

Let on planes $y=0$ and $y=h$ the boundary conditions (3) hold. Then the solution to equation (1) satisfying conditions (3) will be represented in the following form:

$$w=\sum_{n=1}^{\infty}w_n(x)\sin(\lambda_n y)\exp(i\omega_n t),$$

$$\psi=\sum_{n=1}^{\infty}\psi_n(x)\sin(\lambda_n y)\exp(i\omega_n t),\quad \lambda_n=\frac{\pi n}{h} \tag{7}$$

To solve the problem of the possibility of the existence of localized oscillations (waves) in the vicinity of the free edge $x=0$ of a semi-infinite waveguide, it is necessary that the conditions:

$$\lim_{x\to\infty}w=0,\quad \lim_{x\to\infty}\psi=0 \tag{8}$$

Substitution of (2) into equations (7) leads to successive ordinary differential equations with respect to $w_n(x),\ \psi_n(x)$:

$$w_n=A_n\exp\left(-\lambda_n\sqrt{1-\eta_n}x\right),\quad \psi_n=B_n\exp(-\lambda_n x) \tag{9}$$

Here

$$\eta_n=\omega_n^2c_t^{-2}\lambda_n^{-2} \tag{10}$$

The dimensionless sought-for characterizing phase velocity must satisfy the condition:

$$0<\eta_n<1 \tag{11}$$

If at the waveguide boundary $x=0$ conditions of type (3) - (5) are given, for example:

$$w=0,\quad \frac{\partial\psi}{\partial x}=0,\quad (D_1=0) \tag{12}$$

t is easy to show that a solution satisfying condition (11) does not exist, that is, localized oscillations are not possible.

Let the conditions be given at the edge of a semi-infinite waveguide:

$$\sigma_{12} = 0, \quad \varphi = 0 \text{ при } x = 0 \tag{13}$$

Condition (13) can be reduced to the following form:

$$C_{44}(1+\chi)\frac{\partial w}{\partial x} + e_{15}\frac{\partial \psi}{\partial x} = 0, \ \psi + \frac{e_{15}}{\varepsilon} w = 0 \tag{14}$$

The requirement that solution (9) satisfies boundary conditions (14) leads to a system of algebraic equations for arbitrary constants A_n, B_n. From the condition that the determinant of this system is equal to zero, it is possible to determine the dimensionless parameter of the phase velocity.

$$\eta = 1 - \frac{\chi^2}{(1+\chi)^2} \tag{15}$$

The dimensionless quantity obtained in equation (15), which characterizes the phase velocity, coincides with the phase velocity for the Bluestein - Gulyaev waves. However, if for Bluestein - Gulyaev waves $\eta = \omega^2 c_t^{-2} k^2$ is a continuous spectrum of eigenvalues, (for an arbitrary wave number k), here this spectrum is discrete.

In the case when the boundary conditions of variant (4) are given on the semi-infinite sides of the waveguide, solution (7) will be written:

$$\begin{aligned} w &= \sum_{n=1}^{\infty} w_n(x)\cos(\lambda_n y)\exp(i\omega_n t), \\ \psi &= \sum_{n=1}^{\infty} \psi_n(x)\cos(\lambda_n y)\exp(i\omega_n t), \quad \lambda_n = \frac{\pi n}{h} \end{aligned} \tag{16}$$

It is easy to show that, similarly to the previous case, satisfying Eq. (1) and boundary conditions (14), we obtain expression (15) for the phase velocity parameter.

The option is assumed when on the planes of the plate $y = 0$ and $y = h$ different types of boundary conditions are set:

$$\begin{aligned} & w = 0, \quad \psi = 0, \text{ at } y = 0 \\ & \frac{\partial w}{\partial y}, \quad \frac{\partial \psi}{\partial x} = 0, \quad \text{at } y = h \end{aligned} \tag{17}$$

A solution to equation (1) satisfying the boundary conditions (7) can be represented as:

$$\begin{aligned} w &= \sum_{n=1}^{\infty} w_n(x)\sin(\mu_n y)\exp(i\omega_n t), \\ \psi &= \sum_{n=1}^{\infty} \psi_n(x)\sin(\mu_n y)\exp(i\omega_n t), \\ \mu_n &= \frac{(2n-1)\pi}{2h} \end{aligned} \tag{18}$$

Carrying out a procedure similar to the solution of the previous versions, we obtain for the phase velocity parameter the expressions (15) where it is necessary to replace λ_n by μ_n. Hence it follows that the phase velocity of the localized wave does not change, but the frequencies and wavelengths change.

Now let the conditions:

$$w, \quad \frac{\partial \psi}{\partial y} = 0, \quad \text{при } y = 0, \ y = h \tag{19}$$

The solution to equation (1) in this case will have the following form:

$$\begin{aligned} w &= \exp(i\omega t)\sum_{n=1}^{\infty} w_n(x)\sin(\lambda_n y), \\ \psi &= \exp(i\omega t)\sum_{n=1}^{\infty} \psi_n(x)\cos(\lambda_n y), \\ \lambda_n &= \frac{\pi n}{h} \end{aligned} \tag{20}$$

Obviously, the solution for functions w_n, ψ_n will be the same as solution (9).

However, boundary conditions (14), for $x = 0$, do not lead to separation of variables.

To satisfy the boundary conditions (14), the averaging method is used, as is done when applying the Bubnov - Galerkin method. Solution (20) is substituted into condition (14), then the first condition is multiplied by $\sin(\lambda_m y)$, second on $\cos(\lambda_m y)$, ($m = 1, 2, 3 \ldots$)and integrating over *y* in the range from 0 to *h*.

$$\begin{aligned} & \sum_{n=1}^{\infty}\Big[C_{44}(1+\chi) w_n^1(0)\int_0^h \sin(\lambda_m y)\sin(\lambda_n y)\,dy + e_{15}\psi_n^1(0)\times \\ & \times\int_0^h \sin(\lambda_m y)\cos(\lambda_n y)\,dy = 0 \Big] \\ & \sum_{n=1}^{\infty}\left[\frac{e_{15}}{\varepsilon} w_n(0)\int_0^h \cos(\lambda_m y)\sin(\lambda_n y)\,dy + \psi_n(0)\times \right. \\ & \left. \times\int_0^h \cos(\lambda_m y)\cos(\lambda_n y)\,dy = 0 \right] \end{aligned} \tag{21}$$

System (21) ($m = 1, 2, 3 \ldots$) after calculations is reduced to the form:

$$C_{44}\left(1+\chi\right)w_m^1\left(0\right)+\frac{2e_{15}}{\pi}\sum_{n=1}^{\infty}\psi_n^1\left(0\right)\frac{m\left(1-\left(-1\right)^{m+n}\right)}{m^2-n^2}=0$$

$$\frac{2e_{15}}{\pi\varepsilon}\sum_{n=1}^{\infty}\frac{n\left(1-\left(-1\right)^{m+n}\right)}{n^2-m^2}w_n\left(0\right)+\psi_m\left(0\right) \qquad (22)$$

The resulting system (22) in the approximation (*m=1*) and (*m=2*):

$$\lambda_1\sqrt{1-\eta_1}C_{44}\left(1+\chi\right)A_1-+\frac{4e_{15}}{3\pi}\lambda_2 B_2=0$$

$$\frac{8e_{15}}{3\pi\varepsilon}A_2+B_1=0$$

$$\lambda_2\sqrt{1-\eta_2}C_{44}\left(1+\chi\right)A_2-+\frac{8e_{15}}{3\pi}\lambda_1 B_1=0 \qquad (23)$$

$$\frac{4e_{15}}{3\pi\varepsilon}A_1-B_2=0$$

Hence, for the first and second frequencies, we obtain:

$$\eta_1=1-\frac{32\chi^2}{9\pi^2\left(1+\chi\right)^2},\quad \eta_2=1-\frac{16\chi^2}{9\pi^2\left(1+\chi\right)^2} \qquad (24)$$

Comparison of the values of phase velocities (24) and (15) shows that in the case of boundary condition (19) at $x=0$ the velocity of localized waves is higher, which leads to a comparatively slow decay.

Conclusion

It was found that the method of separation of variables is used for two boundary conditions. It is shown that for these conditions there are localized oscillations, the phase velocity parameter coincides with the Blyustein - Gulyaev formula, but in contrast to Bluestein - Gulyaev, here the spectrum of eigenvalues is discrete. The phase velocity of the localized wave does not change, but the frequencies and wavelengths do.

When it is impossible to apply the method of separation of variables to satisfy the conditions, the Bubnov - Galerkin method is used. Comparison of the dimensionless values of the phase velocity shows that in this case the velocities of localized waves are higher, which leads to a comparatively slow decay.

References

1. Belubekyan V.M., Belubekyan M.V. Resonant and localized shear vibrations in a layer with a rectangular cross section. *Reports of the NAS of Armenia*, 2015. Vol. 115, no. 1. P. 40-43.
2. Bardzokas D.I., Ku yavtsev B.A., Senik N.A. Propagation of waves in electromagnetoelastic media. *Editorial URSS*, 2003. 336 p.
3. Bleustein J.L. Some simple modes of wave propagation in an infinite piezoelectric plate. *Journ. of Acustical Society of America*, 1969, 45, pp. 614-620.
4. Belubekyan M.V., Belubekyan V.M. Surface waves in piezoactive elastic system of a layer on a semi-space. *Proceedings of the Yerevan State University, Physical and Mathematical Sciences. Mechanics.* 2013, no. 3, pp. 45-48.
5. Avetisyan A.S., Kamalyan A.A. Influence of transverse inhomogeneity of a piezodielectric layer and combinations of boundary conditions on the propagation of a shear electroelastic signal. *SEUA Bulletin. series "Mechanics, mechanical engineering, mechanical engineering"*. 2014. Issue 17, no. 1. P. 9-25.
6. Vilde M.V. Kaplunov Yu.D., Kossovich L.Yu. Edge and interface resonances of the phenomenon in elastic bodies. Moscow: Fizmatlit, 2010. 280 p.
7. Gulyaev Yu.V. Surface electrosonic waves in solids. *JETP Letters.* 1969. Vol.4 no. 1. P. 63-65.
8. Belubekyan M.V., Papyan A.A. On the problem of wave propagation in a layer made of piezoelastic material of 6mm class. *Actual problems of continuum mechanics: Proceedings of the IV international conference*, September 21-26, 2015, Tsaghkadzor, Armenia. P. 91-94.

ABOUT DESIGN MEDICAL DATABASES AND INFORMATION SYSTEMS FOR THE ORGANIZATION AND MANAGEMENT OF CLINICAL PROCESSES

Ahrorjon A. Abdumanonov,
Fergana branch of the Republican Scientific Center for Emergency Medical Aid, Fergana, Uzbekistan, ahror79@inbox.ru

Ruslan E. Aliyev,
Fergana branch of the Republican Scientific Center for Emergency Medical Aid, Fergana, Uzbekistan, ruslan-ltd@yandex.ru

Muhammajan K. Karabayev,
Ferghana branch of the Tashkent Medical Academy, Ferghana, Uzbekistan, externet@inbox.ru

Vokhidjon G. Hoshimov,
Fergana branch of the Republican Scientific Center for Emergency Medical Aid, Fergana, Uzbekistan, externet@inbox.ru

DOI: 10.36724/2664-066X-2020-6-2-20-26

ABSTRACT

In article questions of development of automated, integrated information systems and databases (DB), created by the authors to ensure that the information and intellectual support of organizations and management of clinical processes in a general hospital. Analyze the schema information and intellectual support for the main stages of activity of the doctor in the organization of clinical processes and proved the effectiveness of the transition to electronic document management and management of patient records. The techniques and steps in the process of designing a database of such systems. Presented data models based on relational principles and taking into account the peculiarities of information provision of medical care in a clinical hospital emergency medicine. Sanctified issues for creating a database implemented taking into account the specifics of health care regulations, standards and related problems of formalization and standardization of information, the problem of choice of database management systems (DBMS), including for medical applications, the issues of intellectualization of the database and the formation of a knowledge base. Showing crucial modeling domain – automation object articles in this case – the diagnostic and treatment process. The main requirements to the database integrated health information systems. Questions of formalization entered into the database of medical and other information, the solution of which the authors carried out through the creation of special electronic templates for all aspects of medical practice hospital. Being developed by leading medical specialists clinic, they do not only facilitate the design, but also remind us that we should find out, to ask to examine and help select the best tactics. This kind of technology is human-computer interaction – an element of the programming actions of the doctor. This option is the task of formalizing entered into the database of medical information used to provide a single standard of medical information, standardize the order of the information entered into the database. The results of the analysis of the effectiveness of the developed information system and performance of its database in the course of their use in the clinical setting. In particular, we studied the performance, response time and server load when working with different categories of users – doctors, nurses, administrators, and etc., which have shown the applicability of the developed health information system and its database to optimize the organization of clinical processes in the conditions of the experiment.

KEYWORDS: *medical information systems, electronic medical records, databases, knowledge base, intellectual support, workstations, information technology, medical-diagnostic process.*

Information about authors

Ahrorjon A. Abdumanonov, *network administrator and security of the Fergana branch of the Republican Scientific Center for Emergency Medical Aid, Fergana, Uzbekistan*

Ruslan E. Aliyev, *programmer Fergana branch of the Republican Scientific Center for Emergency Medical Aid, Fergana, Uzbekistan*

Muhammajan K. Karabayev, *Professor of "Biophysics, Information Technology and Sports" Ferghana branch of the Tashkent Medical Academy; Academician of the International Academy of Informatization at the United Nations, Fergana, Uzbekistan*

Vokhidjon G. Hoshimov, *Surgeon Fergana branch of the Republican Scientific Center for Emergency Medical Aid, Fergana, Uzbekistan*

In modern medical treatment institutions (TI), huge volumes of medical data are formed and accumulated, and the quality of the medical care they provide depends on how timely and efficiently this information is used by doctors, specialists and heads of institutions. To optimize the organization of treatment and diagnostic processes (TDP) and the management of LU activities, it is especially important not only to use the accumulated information as efficiently as possible, but also to use the implicit trends and patterns hidden in them, revealed with the help of special analysis, and in the case of emergency medicine and ensuring the required speed of information processing and transmission. The solution of these problems determines the relevance of the creation and use of automated information-analytical and communication systems in clinical practice.

Note that a medical information system (MIS) is a set of software and hardware tools, databases and knowledge that, having a number of functional capabilities, with the presence of appropriate networks, allows you to automate the organization and management of LDP, go to electronic document management and maintain an electronic medical history (EMH) [1, 13]. It also allows you to provide information and intellectual support for the adoption of medical and managerial decisions, analysis and control of the work of institutions, as well as optimize the use of its resources.

Usually such systems deal with large amounts of information with a rather complex structure. We present the recommended [14] schemes of informational and intellectual support of the physician's activity (Fig. 1), which provide for the main stages of medical activity in organizing the LDP.

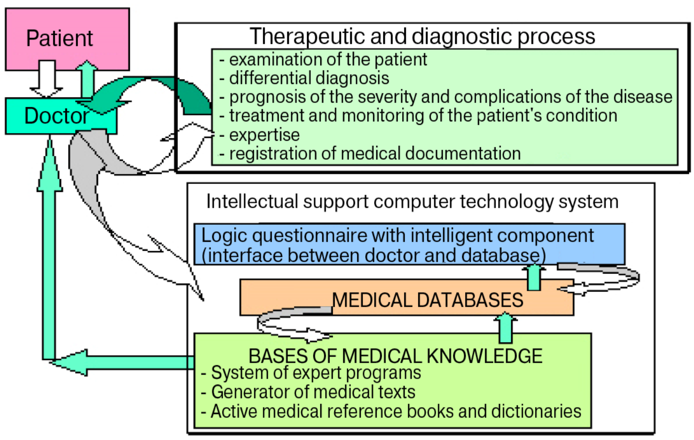


Figure 1

Almost all such systems are to one degree or another associated with the functions of long-term storage of database information (DI) - a machine tool for accumulation and storage, as well as the organization of large data arrays, on the basis of which all problems of the domain (software) are solved. Databases allow you to store patient data and collected medical data. During the observation of the patient, the records in the database are updated with data from current examinations, consultations, consultations and examinations. Therefore, the database should be able to accumulate, store and update data, as well as provide various categories of users with quick access to the required data. To do this, the data in the database must be structured and organized in accordance with some model of the subject area, which is a collection of objects, their properties and connections between them.

This paper presents the basic principles used in the design of the database of the complex information-analytical system ExterNET and its EIB, developed by us for the automated organization and management of treatment and diagnostic processes [8, 9]. This system has been introduced into clinical practice in a multi-profile hospital (350 beds) of the Fergana branch of the Republican Scientific Center for Emergency Medical Aid (Uzbekistan).

Comprehensive computerization of medical institutions, the creation of specialized integrated medical information technology systems and networks, in addition to the development of a general methodology, requires the elaboration of a large number of specific issues. These include the problems of electronic document management, implemented taking into account the specifics of medical regulations and standards, and related problems of formalization and standardization of information presentation; problems of choosing database management systems (DBMS), including those for medical applications; issues of database intellectualization; formation, on the basis of the information contained in the database, "operational" and "analytical" forms of information; reliability and safety issues; problems of transition to completely digital technologies with the possibility of automated data analysis, as well as a wide range of issues related to horizontal and vertical integration of information, etc.

Traditionally, medical records are kept in the form "Human readable" text and are designed to be read, processed, evaluated and analyzed by a person, a medical officer. This method of information processing is still the main one in the treatment and diagnostic process.

Automated processing puts forward additional requirements for the maintenance of medical information, namely, its formalization and coding. Formalization and coding of information provide tremendous advances and advantages of computer processing [3].

Formalization in this case means the structuring of information, dividing it into specific sections and subsections, up to highlighting the signs and attributes of each medical entity and a clear description of all these signs in a medical document.

Thus, the design of an integrated subject-matter, integrated into a large-sized database, which is the database of complex clinical information systems, is a difficult task [2, 4-6].

Database design is an iterative, multi-stage process of making informed decisions in the process of analyzing the software information model, data requirements from application programmers and users, synthesizing logical and physical data structures, as well as analyzing and justifying the choice of software and hardware.

The main tasks of designing a database are: ensuring the storage of all necessary information in the database; ensuring the ability to obtain data on all necessary requests; reduction of redundancy and duplication of data; ensuring data integrity (correctness of their content): elimination of contradictions in data content; elimination of data loss, etc.

In accordance with these tasks, in the design of databases and their operation, the following general requirements are imposed on them:

- the adequacy of the display of the subject area (completeness, integrity, data consistency, relevance);
- the ability to interact with users of different categories;
- ensuring high efficiency of access;
- user-friendliness of the interface;
- ensuring secrecy and confidentiality;
- ensuring the mutual independence of programs and data;
- ensuring the reliability of the database;
- data protection against accidental and deliberate destruction;
- the ability to quickly and completely restore data in the event of their destruction;
- in the case of corporate databases - ensuring maximum opportunities for each user, that is, supporting the performance of all business functions by an employee who receives the final result.

The database development process can be divided into several stages: research of the subject area; creation of an infological (information-logical) model; creation of a datalogical model; creation of a physical model.

The most important stage in the design of a database is the development of an infological model of the domain that is not focused on the DBMS. In the infological model, by means of data structures in an integrated form, the composition and structure of the data, as well as the information needs of applications (tasks and requests), are reflected. This model reflects the subject area in the form of a set of information objects and their structural links.

Information about the subject area that users work with is displayed first in the infological model, then in the datalogical and physical models. The views of individual users are formed by external infological and datalogical models. What tools can be used to compose an infological description of the subject area? There is no single answer to this question.

There are several techniques and, accordingly, different tools are used. The compiled model should be simple, clear, contain all the information for further design stages, and easily be converted into database models for common DBMS. Based on these requirements, the described design method uses a model called "entity-relationship" (or "objects-relationship").

When designing a database, methods for performing such design stages as collecting information about software, choosing a language for representing the so-called "semantic" model for fixing information about software, their subsequent analysis, and synthesizing a database model were also realized and integrated into well-structured schemes.

Analysis of the collected information about software provides for: classification, formalization and integration of structural elements of software description, formalization of both structural and procedural constraints on the integrity of elements in the future software model, determination of the dynamics of instances of software objects.

The synthesis of the conceptual model of the database includes:

- designing an integral conceptual database schema in the selected semantic modeling language;
- choice of a specific data model and DBMS for DB implementation;
- designing a logical database scheme for the selected DBMS (also called "implementation design");
- development of the physical structure of the database ("physical" or "internal" scheme, it is also - "layout"), including the placement of the database by nodes;
- development of technology and procedures for the initial creation and filling of the database;
- development of technology and procedures for database maintenance;
- development of universal database access programs and corresponding user interfaces;
- database testing, its development and improvement (tuning) of its structure.
- conceptual design is the collection, analysis and editing of data requirements. For this, the following activities are carried out:
- survey of the subject area, its study information structure;
- identification of all fragments, each of which is characterized by a user view, information objects and connections between them, processes over information objects;
- modeling and integration of all views.

At the end of this stage, we get a conceptual model that is invariant to the database structure. It is often represented as an entity-relationship model.

At the stage of DB development, the DBMS serves to describe the structure of the database, i.e.: defining tables, determining the number of fields, the type of data displayed in them, the size of the fields, determining the relationships between tables.

In addition to tables, most DBMS provide for the creation of special tools for working with data - forms, queries.

During the operation of the database, the DBMS provides editing of the database structure, filling it with data, searching, sorting, selecting data according to specified criteria, and generating reports.

Now we will try to create a complete infological model of the "Case history" problem. For this, we list the rules that the entity-relationship model must satisfy:

- the model should give a complete picture of the subject area;
- all entities necessary for the implementation of the task and their attributes should be listed, respectively;
- the names of entities must be unique;
- the names of attributes within the same entity must be unique;
- we must guarantee an unambiguous interpretation of the model;
- in each entity, an identifying set of attributes must be highlighted;
- the model must be flexible, i.e. when new tasks arise, expand without significant changes to the existing model.

The model presented in Figure 2 allows solving the main tasks of the medical history. It is one of many possible solutions.

An identifying attribute (identifying a set of attributes, ISA) is an attribute (several attributes), the value of which determines the uniqueness of an entity instance.

Almost all modern systems are based on a relational database management model [7].

In a relational database management system, all processed data are presented in the form of flat tables. Information about objects of a certain type is presented in tabular form: various attributes of objects are concentrated in the columns of the table, and rows are intended to reduce descriptions of all attributes to individual instances of objects.

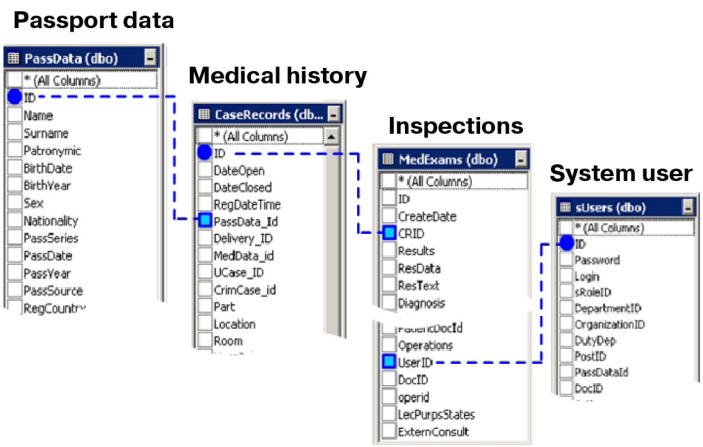


Figure 2. Key model of identifiable attributes

The model, created at the stage of infological modeling, most satisfies the principles of relativity. However, to bring this model to relational, you need to perform a procedure called normalization.

Normalization theory operates with five normal forms. These forms are designed to reduce the redundancy of information, so each subsequent normal form must meet the requirements of the previous and some additional conditions.

Rational variants of the conceptual DB schema must satisfy the third normal form, as well as the following requirements: the selected list of relations must be minimal; the relationship is used if only its necessity is caused by the tasks; the selected list of attributes should be minimal; an attribute is included in a relationship only if it will be used.

The primary key of the relationship must be minimal. That is, it is impossible to exclude any attribute from the identifying set of attributes without violating the unambiguous identification. When performing operations on data, there should be no difficulties.

Datalogical and physical models are directly implemented in the DBMS. At the same time, the physical model defines the structure of data storage on physical media.

The specifics of a particular DBMS may include restrictions on the naming of database objects, restrictions on the supported data types, etc. In addition, the specificity of a particular DBMS in physical design includes the choice of solutions related to the physical environment of data storage (choice of methods for managing disk memory, dividing the database by files and devices, methods of accessing data), creating indexes, etc.

At each design level, information is structured in such a way that at the third level information can be presented in the form of data structures implemented in the computer memory. In this way:

At the first level, which is called infological, it is determined what information about the subject area will be stored and processed in the computer, and as a result of researching the subject area, its infological model is built. Information in the infological model is presented regardless of what software and hardware will be used in the future to store and process it. At this level, the domain is described in terms of object classes and their relationships, which are understandable to end users and people working in the domain who are not familiar with the principles of database organization.

At the second level, which is called datalogical, or conceptual, information is presented in the form of data and logical connections between data, regardless of what the data is and what technical means will be used to store the data, but taking into account software tools (DBMS). There are several types of datalogical data models: network, hierarchical, relational, object, and others.

The third, physical, layer defines how and where data will be stored on the physical medium.

Our important decision was the rejection of the design of the MIS database by functional purpose, when a separate database was created for a separate task (for example, a laboratory subsystem, pharmacy, functional diagnostics, consultation, etc.). Although this approach has a number of advantages, the main one of which is to reduce the requirements for the hardware power of the server by dividing the streams of user requests to separate databases. We chose to design the MIS database in such a way that all information was collected around the patient and physically stored in one database (Fig. 3).

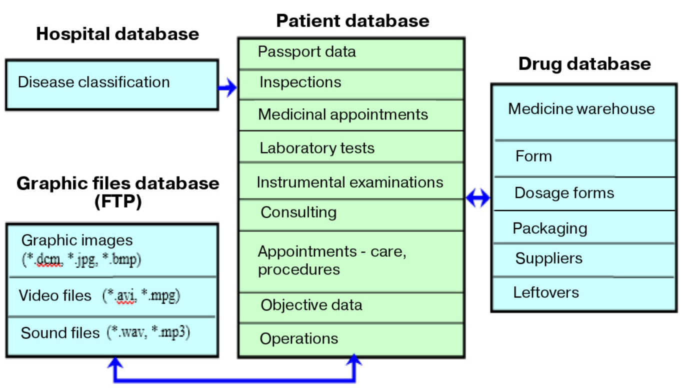


Figure 3. The enlarged scheme of the object-relational database of the medical information system

Designing the structure of the database, thus, allows you to achieve a stably small volume of the MIS database during almost the entire period of its operation, and thereby ensure the maximum possible performance of the MIS. The complexity of this technique is that the software of the information system must support any number of physical databases in the core of the system, combined into one logical structure.

Thus, it is necessary to develop algorithms for all MIS programs so that they can work correctly with a database of current documents, consisting of one or more parts. This is due to the fact that in some cases the program needs to process data not only for a separate part of the database, but also for all the information available in it.

To effectively solve this problem, it is necessary to exclude direct reference to the database in the texts of MIS programs. Instead, it is proposed to use special middleware called services middleware.

Figure 4 illustrates the interaction of the user, the DBMS and the OS when processing a request to receive data. The numbers indicate the sequence of interactions, namely:

1 - The user sends the DBMS a request to receive data from the DB.

2 - Analysis of the user's rights and the external data model corresponding to this user confirms or denies the access of this user to the requested data.

3 - If access to data is denied, the DBMS informs the user about it (arrow 12) and stops further data processing, otherwise the DBMS determines the part of the conceptual model that is affected by the user's request.

4 - The DBMS receives information about the requested part of the conceptual model.

5 - The DBMS requests information about the location of data at the physical level (files or physical addresses).

6 - The DBMS returns information about the location of the data in terms of the operating system.

7 - The DBMS asks the operating system to provide the necessary data using the operating system tools.

8 - The operating system pumps information from storage devices and sends it to

system buffer.

9 - The operating system notifies the DBMS about the end of the transfer.

10 - the DBMS selects from the delivered information,

11 - Located in the system buffer, only what the user needs, and sends this data to (12).

12 - User work area.

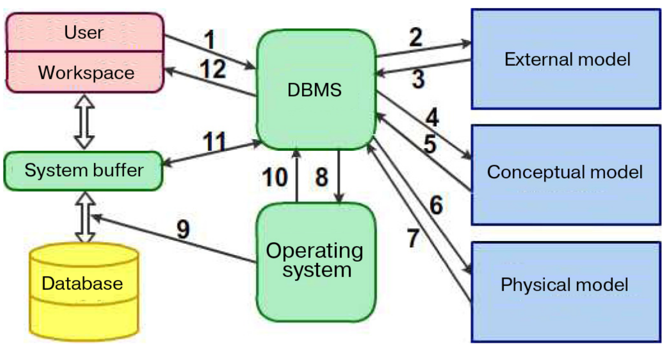


Figure 4. Scheme of passing a request to the database

Data structuring is based on concepts "Aggregation" and "generalization" (Fig. 5).

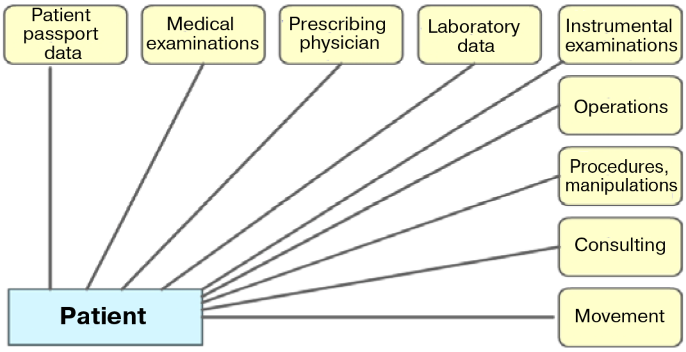


Figure 5. Data structuring

Diagram in Figure 5 defines the composition order of the data structures of this model. A data element is the smallest named unit of data (analogous to a "field" in file systems) to which the DBMS can be addressed directly and with the help of what builds all other structures. The item name is used to identify it in the high-level data structure schema. A data aggregate is a named collection of data items within a record that can be viewed as a whole. A set is a named collection of records that form a two-level hierarchical structure. To implement the possibility of presenting such information in the model, it is necessary to organize the appropriate connections between the patient and doctor entities.

Ensuring the performance of the OBD remains a challenge. At the same time, the determination of the time characteristics of working with the database and ensuring the preservation of these characteristics during the operation of the database are among the most difficult design tasks. It is well known [10] that the following main indicators are used to analyze the efficiency of an information system for any purpose:

1) performance of the information system, expressed in the number of operations performed per unit of time;

2) response time of the information system;

3) load of the information system server (percentage of server processor utilization).

Table 1

General information about the load on the MIS

Indicators	Information read (kb)	Information recorded (kb)	Updated information (kb)
General data (work of 50 users)	1239650.80 (±503499.50)	40483.20 (±29784.90)	10438.10 (±10564.50)
Per user	27493.01 (±13013.40)	809.66 (±794.264)	208.76 (±281.72)

The analysis showed that during the study period, doctors used the capabilities of MIS to 100%, nurses 80% of MIS capabilities were used, users of para-clinical department - 75%, users of administrative and managerial status - 20%, system administrators - 100%.

With a clinic occupancy of 80%, on average, doctors look through 280.03 (± 149.0) EIB per day, of which 83.20 (± 48.3) add records (examinations, appointments, manipulations, etc.). Nurses look at 417.10 (± 216.7) EIB per day, record at 178.50

(± 111.50). Taking into account the actual use of the system, these indicators for healthcare facilities with a full transition to EIB can increase to 542.23 EIB per day.

For each doctor in clinical departments, the information system processes 21176.60 (± 11526.10) kbytes of information per day, which is 68.4% of the total volume of the database. In the “pharmacy” module, 37136.40 (± 18326.2) kbytes of information were processed for each doctor.

Thus, the most resource-intensive operation is the operation of obtaining and recording information about patients by doctors and nurses. During the execution of this operation, the DBMS executes a number of complex subqueries that require significant computational resources:

1. hecking user access rights to the database;
2. ownloading programs and objects to the user's client workstation, taking into account his access level;
3. L ding of interface elements used for the operation of applications, including graphic files of instrumental and laboratory analyzes, etc .;
4. Opening the database, which is an update of all views in it, taking into account the changes that have occurred after the last execution of the command to open the database and taking into account the current user access level;
5. T nsfer of the list of documents generated when opening the database to the user's client workstation;
6. Processin of the received list, client informing and working data about patients, and displaying it on the screen.
7. Formation of a list of medicinal products from the database of pharmacy and department warehouses.
8. Formation of a template for examinations at the doctor's choice from the examination database.

As can be seen from the above list, the operation of connecting to a database widely uses the most various types of requests and operations for exchanging information between a client workstation and an information network server. Thus, the time of connecting to the database can adequately characterize the effectiveness of the applied methodology for designing the database structure.

A wide range of different types of requests are used in the work of the IIA. In order to determine their characteristics under the conditions of an operating MIS "ExterNET", a chronometric study was carried out [10], the results of which are shown in Table 2. During the study, the operation time of some basic types of requests was measured, for which, during the execution of MIS software commands, the time after submission of a command by the user and the time of outputting the result of work on the screen The difference between the indicated time intervals was recorded in a special database table using a specially created program during the study, and at the end of the study, the average value of the query execution duration and the number of queries executed per day were calculated. At the same time, the server processor load indicator was recorded during this period of time.

It was revealed that the greatest load on the information system is provided by the doctors of the departments, therefore, for the analysis, only requests from their client workstations were taken, which were executed at the time of simultaneous servicing of 50-55 users by the server. These measurements were repeated many times and their average values were taken into account. As can be seen from Table 2, the highest server load and the highest response time were recorded when executing requests for the formation of templates for inspection. However, the use of this type of request can be canceled at the request of the user, but with a template filling out the examination, the doctor needs less time to complete the examination than in the arbitrary option of filling out the examination of the patient. The template form for filling out inspections has other advantages.

Table 2

Characteristics of the main types of requests in "ExterNET"

Request Description	Average execution time, sec.	Average frequency of execution, per day	Average server processor load during query execution,%
Database connection, user identification	**5.2**	**72.1**	**19.2**
Database connection with data archives	**2.2**	**2.8**	**17.2**
Formation of the list of patients in the department	**0.5**	**152.6**	**16.3**
Formation of complete data about the selected patient	**0.5**	**588.7**	**10.8**
Recording a new inspection to the database	**0.6**	**278.7**	**11.2**
Adding or updating patient medical data to the database	**0.4**	**101.8**	**5.4**
Generate inspection templates	**10**	**278.7**	**15.8**
Database connection with laboratory data	**0.4**	**458.1**	**5.7**
Connection to the drug database	**2**	**278.7**	**8.4**
Connection to database	**0.3**	**278.7**	**7.8**
Connecting to an FTP server to receive video and graphic files in instrumental and laboratory examinations	**1**	**41.2**	**4.2**

The use of electronic media requires the formalization of the information entered. We achieved this goal by creating special electronic templates for entering information into the database at all levels of care provided, whether it is medical information, nursing, or the result of an additional examination. The term "template" does not mean either a stencil description or a stereotyped patient management.

This is a specially prepared set of medical terms and expressions for the design of a medical examination, diaries, epicrisis, description of the results of laboratory and instrumental examinations, etc. Developed by the leading medical specialists of our clinic, they not only facilitate registration, but also remind you to find out, ask, examine and help you choose the best tactics.

This is a kind of technology of human-computer interaction - an element of programming the doctor's actions. As a result, we not only achieved the maximum formalization of the input information, but also streamlined the sequence of their input into the database.

This option for solving the problem of formalization entered into the medical information database has the following advantages:

1. A unified standard of medical information used is achieved.

2. The order of information entered into the database is standardized (medical examination, prescriptions for treatment, examination, choice of treatment tactics, results of measuring the patient's objective parameters, etc.).

In connection with the above, the main condition for the application of the technology developed by us is the use of electronic templates by the doctor when registering his examination. In this case, all medical records have a strictly defined structure and contain only formalized medical information. When recording a medical record in the postoperative period, the information automatically goes through software processing through the adopted algorithm.

With the help of formalization, it is also possible to build expert systems to provide intelligent assistance to doctors in diagnosing and choosing treatment tactics. Formalized EIB makes it possible to continuously accumulate information on a specific pathology and to analyze the further course of treatment according to the recommended criteria. Over time, the MIS will accumulate a knowledge base (KB), which will be constantly replenished, and the system itself will use it in the tasks of diagnostics and the choice of treatment tactics [11].

Note that the knowledge base is, in our case, a set of units for achieving medical knowledge, logical rules and algorithms, which are formalized, using a certain method of representation, knowledge, reflecting objects of the subject area and their interrelationships [12]. They provide the ability to form and output medical opinions and recommendations based on the system of decision-making algorithms embedded in it. The medical knowledge bases contain text synthesis scripts for medical history and other medical documents, automatically generated by the generator of medical texts based on the collected patient data.

Conclusions

Thus, in the single information space of the emergency medical care center created on the basis of ICT and its database, only formalized and reliable information is formed, generated, and transported through information channels, in the required volume, in a strictly orderly manner, preventing information disorder, providing each participant in the process of providing emergency medical care information and intellectual support necessary to fulfill their functional duties, covering at the same time specialists from the lowest level to the first head of a medical institution.

References

1. Nazarenko G.I., Guliyev Y.I., Ermakov D.E. Medical Information Systems: Theory and Practice. Moscow: FIZMATLIT, 2005. 320 p.

2. Methodology of database [Electronic resource] URL: http://www.budgetrf.ru/Publications/Magazines/VestnikSF/2006/vestniksf289-1/vestniksf289-1090.htm.

3. Zingerman B.V., Emelin I.V., Lebedev G.S. Problems of definition of key terms of medical informatics. Information technologies in medicine. 2009-2010. Thematic research collection. Moscow: Radio Engineering, 2010. P. 20-33.

4. Tumanov V.E., Maklakov S.V. Design of relational data warehouses. Moscow: Dialog-MIFI. 2007. 333 p.

5. Boiko Vladimir, Savinkov V.M. Database design of information systems. Moscow: Finance and Statistics, 1989. 351 p.

6. Data, C.J. Introduction to database systems. 8th edition. Moscow, St. Petersburg, Kiev: Williams, 2005. 1327 p.

7. Meyer M. Relational database theory. Moscow: Mir, 1987. 608 p.

8. Abdumanonov A.A., Karabayev M.K., Hoshimov V.G. Information and communication technology organization of clinical processes in a hospital emergency medicine. *Int. Well. Information technology modeling and management.* 2012, no.5 (77). P. 378-385.

9. Karabayev M.K., Abdumanonov A.A. "Algorithms and technologies of information security in the health information system ExterNET". Programm products and systems, Scientific-practical. edition 1 (101), 2013. P. 150-155.

10. Grigoriev Y.A., Matyuhin V.G. Execution time assessment of a complex sql-search database in ms sql server. *The or-ganization database*. 2004. No.2 (8). P. 3-11.

11. Gavrilova T.A., Khoroshevsky V.F. Knowledge Databases Intelligent Systems. Textbook. SPb.: Peter, 2000. 384 p.

12. Abdumanonov A.A., Karabayev M.K., Mahmudov N.I. "On the intellectualization of medical information systems". *Scientific and practical journal "Modern science: actual problems of theory and practice" series of "natural and technical sciences"*. No. 9-10, 2013. Voronezh. P. 60-64.

13. Gusev A.V., Romanov F.A., Dudanov I.P., Voronin A.V. Medical information systems: Monograph. Petrozavodsk: Publishing, Petrozavodsk State University, 2005. 404 p.

14. Klyuzhev V.M., Ardashev V.N., Sablin V.M., Maltsev E.G., Sins M.M. The draft concept of automation LPU Russian Defense Ministry and telemeditsiny. 2008.

OPTIMIZATION OF MOTION OF AUTOTRANSPORT ON UNEVEN ROADS CONTAINING ASCENDS AND DESCENDS

A. I. Huseynov,
Azerbaijan Technical University, Baku, Azerbaijan, azer.h@mail.ru

DOI: 10.36724/2664-066X-2020-6-2-27-31

ABSTRACT

The impact of roads unevenness, i.e. presence of upward and downward motion zones in the road on auto transport fuel consumption is studied. The task of calculation of optimum interrelation of motion speed and acceleration at the uneven roads in sense of reaching of minimum consumed fuel and minimum emissions on environment is formulated and solved. In the study, an analysis of known works on the topic of research was carried out, which showed that the unevenness of the road, i.e. the presence of descents and ascents on the route is one of the main reasons that led to an increase in fuel consumption by vehicles on the route, while the problem of calculating the optimal functional relationship between the speed and acceleration has not yet been solved. Formulated and solved the problem of finding the optimal relationship between the speed and acceleration of vehicles in terms of achieving the minimum fuel consumption on an uneven track. Practical recommendations have been developed for the implementation of the proposed optimal driving regime on uneven terrain containing descents and ascents.

KEYWORDS: *auto transport emissions, optimization, speed of motion, acceleration, consumed fuel.*

Experimental research

Experimental studies carried out using the SMPS installation in Los Angeles to determine the concentration distribution of ultrafine aerosol in the range of particle diameters of 6-220 nm showed that both the total number of aerosol particles and their volume concentration decrease exponentially depending on the distance to motorways [4]. In this case, the normalized amount of aerosol particles decreases as follows

$$y = \exp(-0.01x) \tag{1}$$

where x is the distance to the motorway.

The volume concentration of the same aerosol particles decreases according to the law

$$y = 0.087 + 1.08\exp(-0.0093x) \tag{2}$$

Graphs of expressions (1) and (2) are shown in Figures 1 and 2.

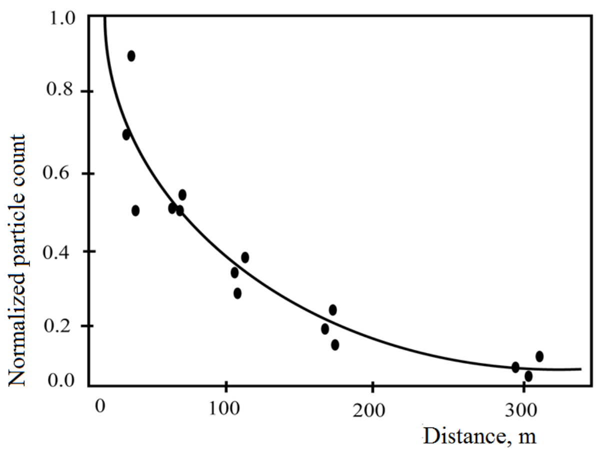


Figure 1. Distribution of the number of aerosol particles with a size of 6-220 nm, depending on the distance to the highway [4]

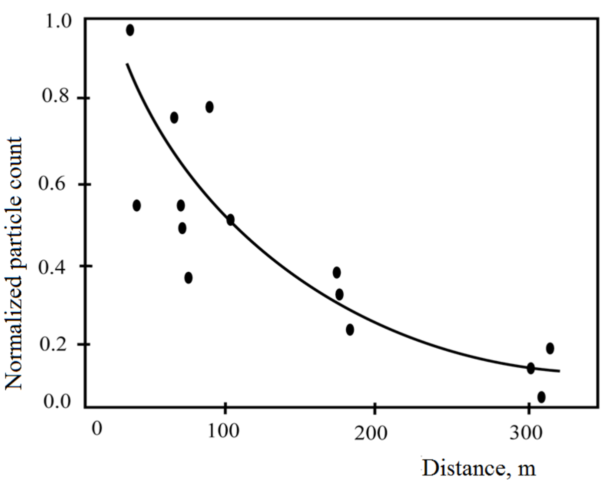


Figure 2. Distribution of the volumetric concentration of aerosol particles with a size of 6-220 nm, depending on the distance to the highway [4]

Experimental studies carried out in New York using the TSI 3039 type FMPS setup made it possible to determine the following distribution laws for the concentration of ultrafine aerosol particles [5]

$$N_1 = 13138 + 43242e^{-D/123} \tag{3}$$

during the first week, where N – is the total quantitative concentration (particle • cm^{-3}); D – Is the distance in meters, and

$$N_2 = 32665 + 22078e^{-D/69} \tag{4}$$

during the second week.

Graphs of expressions (3) and (4) are shown in Figure 3.

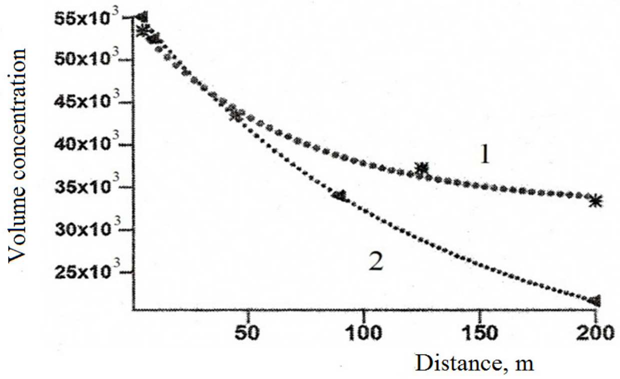


Figure 3. Graphs of the distribution of the volumetric amount of aerosol ultrafine particles depending on the distance to the highway during the first week (1) and the second week (2)

Measurements in Cincinnati, Ohio, USA with MS&T meters, Air Diagnostics Inc. and OPC, model 1.108; Grimm Technologies Inc. showed that both ultrafine (10-100 nm) and coarse (1-2.5 μm) aerosol components have an exponential distribution depending on the distance to the highway, however, the attenuation coefficient in the distribution law of coarse aerosol particles is much greater than for fine particles [6]. The corresponding particle distribution plots are shown in Fig. 4.

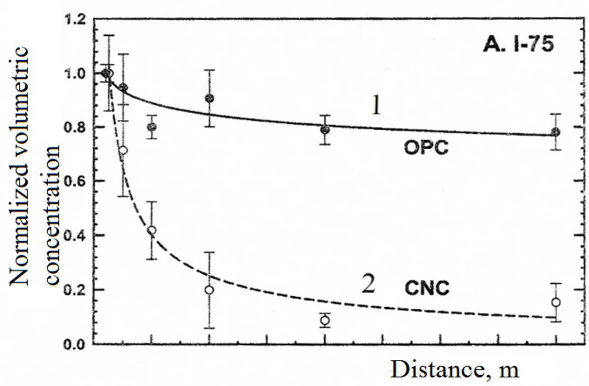


Figure 4. Curves of distribution of the volumetric concentration of aerosol particles depending on the distance to the highway for coarse particles (1) and for ultrafine particles (2)

The distribution of the volumetric concentration of aerosol near highways has a bimodal character [7]. The first maximum in the zone of particle diameters of 0.01-0.05 μm occurs in the range of 0.02-0.03 μm due to the rapid cooling of exhaust gases from automobiles [8]. The second maximum in the 0.05–1 μm zone occurs in the 0.05–0.3 μm range due to incomplete fuel combustion in diesel vehicles [9]. Moreover, the distribution laws of these two fractions of atmospheric aerosol are very different. The fraction in the range of 0.01-0.05 μm has an exponential law of rapid decrease in volumetric concentration, while the fraction in the range of 0.05-0.3 μm has a slow weakening of the volumetric concentration over distance.

At the same time, aerosol pollution of environment is not the only negative factor in impact of vehicles on environmental life support conditions in urban environment.

One of the most important factors requiring close attention of environment in terms of ensuring life safety of population is the pollution by heavy metals of adjacent to major highways. According to science [10], in accordance with modern ecotoxicological assessments, hazardous heavy metals in the soil form a series

Se > Ti > Sb > Cd > V > Hg > Ni > Cu > Cr > As > Ba

This series is very different from the series of hazard of heavy metals, adopted in GOST 17.4.1.02-83, in which the hazard of Pb, Zn and Co in soils is exaggerated, and V, Sb and Ba is underestimated [10].

As a result of measurements with a 210VGP atomic absorption spectrometer, carried out in the town of Kano, Nigeria, mainly decreasing distribution laws of such types of heavy metals as Fe, Pb, Cu, Zn and Cr were obtained [11] (Fig. 5).

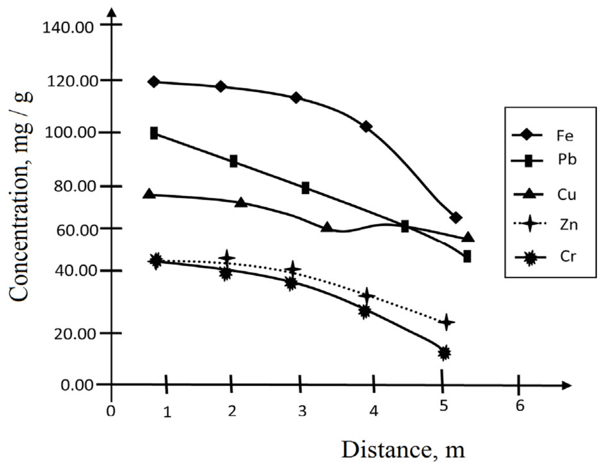


Figure 5. Distribution of hazardous metals concentration in soil depending on the distance to motorway [11]

The condition of the road surface and profile of road have a significant impact on the emission of harmful substances and fuel consumption [1]. Such a close relationship of transport and transport-road structural units leads to the presence of regression relationships between fuel consumption and various indicators of environmental pollution by vehicles. There is the following regression relationship between fuel consumption (x) and emissions CO_2 (y) [12].

$$y = 26,87x - 0,9464 \quad \text{for gasoline} \tag{5}$$

$$y = 24,173x - 2,1889 \quad \text{for diesel} \tag{6}$$

The corresponding graphs of dependencies (5), (6) are shown in Fig. 6 [12].

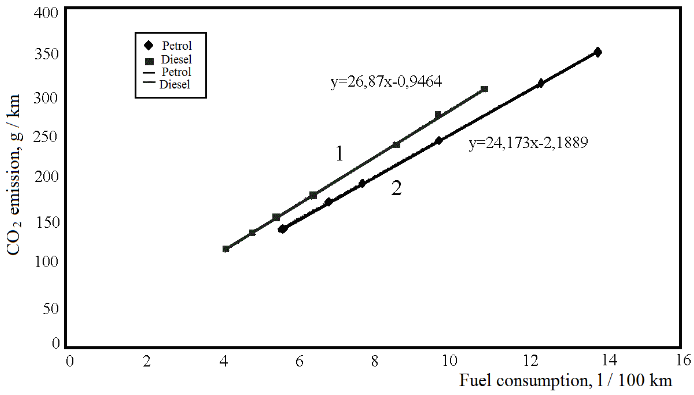


Figure 6. Regression plots of CO_2 emission (g / km) on fuel consumption (l / 100km) for a gasoline engine (5) and a diesel engine (6)

At the same time, there is the following statistical relationship between the degree of emission of the i-th pollutant (EOi) and the degree of fuel consumption (FR) [13]

$$EO_i = \lambda + \mu \cdot FR \tag{7}$$

$$EO_i = E_i \cdot FR \tag{8}$$

where E is the emission index, g / sec; FR - consumption rate fuel (g / sec); λ,μ=const.

Taking into account the above, we can conclude that degree of fuel consumption can be considered as main indicator that determines volume of emitted emissions into the atmosphere. Therefore, in order to determine integrated vehicle emissions in hilly urban areas, it is sufficient to determine the fuel consumption rate at various speeds and accelerations that occur when driving on slopes and inclines. Thus, in this article, the problem is posed of finding the optimal relationship between the driving speed and the acceleration of vehicles with some introduced restrictions in the sense of achieving minimum emissions into the atmosphere when driving on an uneven track containing descents and ascents.

Method for solving the problem

A similar research task was formulated and solved in a slightly different way in [14]. To develop a new method for solving the problem, we partially use the results of work [14], according to which the concept of a system of ecological driving of a car on a track with a descent and an ascent was developed (Fig. 7).

This system uses a vehicle dynamics model; engine fuel consumption model and road surface gradient information. This model generates a control input that maximizes fuel economy on the highway.

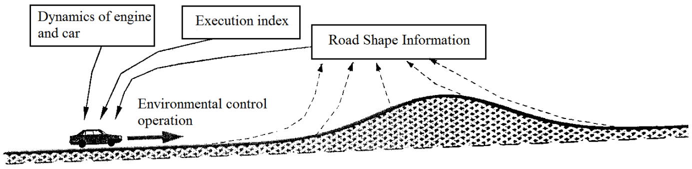


Figure 7. Eco driving system concept using road shape information and model prediction

The fuel consumed by a car at any speed and acceleration can be calculated using the formula [14]:

$$f_v = b_0 + b_1 v + b_2 v^2 + b_3 v^3 + a(c_0 + c_1 v + c_2 v^2) \qquad (9)$$

where $b_i; i = \overline{0,1,2,3}; C_i; i = \overline{0,1,2};$ a – constant values.

In future, a specific research task is formed as follows. In accordance with the proposed car model, it is assumed that in optimal mode there is a unique function

$$v = v(a) \qquad (10)$$

denoting the assumption that acceleration of the quantity a can be developed only at a speed v.

In this case, the restrictive condition is introduced

$$S_1 = \int_{-a_0}^{a_0} F(v(a))da = c; c = const \qquad (11)$$

where F – any continuous monotone function.

Taking into account expressions (9) and (10), the integral partial functional of the goal can have the form

$$f_{u.1} = \int_{-a_0}^{a_0} \left[b_0 + b_1 v(a) + b_2 v(a)^2 + b_3 \cdot v(a)^3 + a(c_0 + c_1 v(a) + c_2 v(a)^2) \right] da \qquad (12)$$

Taking into account expressions (11) and (12), compose the complete functional of unconditional variational optimization

$$f_{u.2} = \int_{-a_0}^{a_0} \left[b_0 + b_1 v(a) + b_2 v(a)^2 + b_3 \cdot v(a)^3 + a(c_0 + c_1 v(a) + c_2 v(a)^2) \right] da + \lambda \cdot \int_{-a_0}^{a_0} F(v(a))da \qquad (13)$$

where λ – Lagrange multiplier.

Thus, it is required to find such a function *v(a)* at which the value $f_{u.2}$ – the integral value of consumed fuel reaches minimum value.

We give an approximate solution to the formed optimization problem in the form of a model study using the model data given in [14].

Model research

Solving the problem of minimizing expression (13) is equivalent to minimizing the following functional (in what follows, we assume that F is the identity function):

$$f_{u.2.1} = \int_{-a_0}^{a_0} \left[b_0 + b_1 v(a) + b_2 v(a)^2 + b_3 \cdot v(a)^3 + a(c_0 + c_1 v(a) + c_2 v(a)^2) \right] da + \lambda \cdot \int_{-a_0}^{a_0} v(a)da \qquad (14)$$

According to [14], constant coefficients b and c are defined as b_0 = 0,1569; b_1 = 2,450·10^{-2}; b_2 = -7,415·10^{-4}; b_3=5,975·10^{-5}; c_0=0,07224; c_1=9,681·10^{-2}; c_2=1,075·10^{-3}.

According to Euler's method [15], the optimal function $v(a)_{opt}$ at which the functional of the goal (14) would reach an extreme value is determined by the condition:

$$\frac{d\left\{ \left[b_0 + b_1 v(a) + b_2 v(a)^2 + b_3 v(a)^3 + a(c_0 + c_1 v(a) + c_2 v(a)^2 \right] + \lambda \cdot v(a) \right\}}{dv(a)} = 0 \qquad (15)$$

Taking into account condition (15), as well as the above values of the constants a and b, obtain

$$2{,}450 \cdot 10^{-2} - 2 \cdot 7{,}415 \cdot 10^{-4} \cdot v(a) + 3 \cdot 5{,}975 \cdot 10^{-5} \cdot v(a)^2 + a \cdot 9{,}681 \cdot 10^{-2} + \\ +2a \cdot 1{,}075 \cdot 10^{-3} \cdot v(a) + \lambda = 0 \qquad (16)$$

Having performed the necessary calculations, represent expression (16) in the following form

$$v(a)^2 + v(a)(12a - 8{,}3) + 13{,}6 + 54a + \lambda = 0 \qquad (17)$$

Solving the quadratic equation (17) obtain

$$v_{1,2}(a) = \frac{8{,}3 - 12a}{2} \pm \sqrt{\frac{(8{,}3 - 12a)^2}{4} - (13{,}6 + 54a + \lambda)} \qquad (18)$$

In expression (18), one should choose addition, since otherwise, with negative accelerations, negative values of v (a) can be obtained. Therefore, we have

$$v(a) = \frac{8{,}3 - 12a}{2} + \sqrt{\frac{(8{,}3 - 12a)^2}{4} - (13{,}6 + 54a + \lambda)} \qquad (19)$$

The resulting expression (19) determines the optimal relationship between acceleration and vehicle speed at which emissions to the environment are minimal.

Discussions

As can be seen from expression (19), regardless of the sign and magnitude of the Lagrange multiplier λ, for small a we obtain large $v(a)$ and vice versa. Consequently, to achieve the extreme value of the target functional at high speed, a small positive acceleration must be realized and vice versa. With negative acceleration, everything happens the other way around, i.e. at high speed, a large negative acceleration is required, and at low speed, a small negative acceleration is required.

The extremum of functional (14) turns into a minimum under the condition

$$2v(a) > 8{,}3 - 12a \tag{20}$$

Therefore, the above conclusions are valid for

$$a > \frac{8{,}3}{12} - \frac{2v(a)}{12} \tag{21}$$

for both positive and negative a.

Otherwise, these conclusions should be reversed.

As can be seen from the above, the specific value of the Lagrange multiplier λ does not affect the obtained qualitative conclusions.

Formally, to calculate the value of λ, it suffices to write expression (18) into integral (11), carry out the integration, and for a given value C calculate the value λ_0.

Conclusions

Let's formulate the main conclusions of the study:

1. Analysis of well-known works on the topic of research showed that the unevenness of the road, i.e. the presence of descents and ascents on the highway is one of the main reasons that led to an increase in fuel consumption by vehicles on the highway.

2. Analysis of the known works on calculating the fuel consumed by vehicles on uneven terrain with descents and ascents showed that until now the problem of calculating the optimal functional relationship between the speed of movement and acceleration has not been solved.

3. Formulated and solved the problem of finding the optimal relationship between the speed and acceleration of vehicles in terms of achieving the minimum fuel consumption on an uneven track. Practical recommendations have been developed for the implementation of the proposed optimal driving regime on uneven terrain containing descents and ascents.

References

1. Erokhov V.I., Bondarenko E.V. Influence factors on traffic exhaust emissions and fuel consumption of vehicles. *Messenger OGU*. No. 4, 2005. P. 139-151.

2. Alessandrini A., Filippi F., Ortenzi F. Consumption calculation of vehicles using OBD data.

3. Baigabulova Zh., Suleimenova K., Bekmagambetova J., Suleimenov I. The ecological and economic methods of analysis of the urban transport system. *The international Archives of the Photogrammetry, Remore Sensing Information Sciences. Vol.XXXVII*. Part B8. Beijing. 2008.

4. Y. Zhu, W.C. Hinds, S. Kim, C. Sioutas. Concentration and Size Distribution of Ultrafine Particles Near a Major Highway. *Journal of the Air & Waste Management Association*. Vol. 52 September 2002. P. 1032-1042.

5. T.H. Whitlow, A. Hall, K.M. Zhang, J. Anguita. Impact of local traffic exclusion on near-road air quality: Findings from the New York City "Summer Streets" campaign. Environmental Pollution 159 (2011). 2016-2027.

6. T. Reponen, S.A. Grinshpun, S. Trakumas, D. Martuzevicius, Z.M.Wang, G.LeMasters, J.E. Lockey, P. Biswas. Concentration gradient patterns of aerosol particles near interstate highways in the Greater Cincinnati airshed. *J Environ. Monit*., 2003, No. 5. P. 557-562.

7. B. Martinenas, V. Spakauskas. Simulation of traffic pollution dispersion near roadways. Lithuanian Journal of Physics, Vol. 50, No.2, pp. 255-260. 2010. doi:10.3952/lithjphys.50212

8. N. Bukowiecki, J. Dommen, A.S.H. Prevot, R. Richer, E. Weingartner, U. Baltensperger. A mobile pollutant measurement laboratory-measuring gas phase and aerosol ambient concentrations with high spatial and temporal resolution, Atmos.Environ.36. P. 5569-5579. 2002.

9. U. Baltensperger, N. Streit, E. Weingartner, S. Nyeki, A.S.H. Prevot, R. Van Dingen, A. Virkkula, J.P. Putaud, A. Even, H.ten Brink, A. Blatter, A. Neftel, H.W. Gaggeler, Urban and rural aerosol characterization of summer smog events during the PIPAPO field campaign in Milan, Italy, J.Geophys. Res. 107(D22), 8193. 2002.

10. Vodyanitsky Yu.N. On the hazardous heavy metals/metalloids in soils. *Bulletin of Soil Institute V.V . Dokuchaev*, 2011. Vol. 68. P. 56-82.

11. J.T. Ayodele, C.D. Oluyomi. Grass contamination by trace metals from road traffic. *Journal of Environmental Chemistry and Ecotoxicolgy*. Vol.3 (3), pp. 60-67, March 2011.

12. Mickunaitis V., Pikunas A., Mackoit I. Reducing fuel consumption and CO2 emission in motor cars. Transport – 2007, Vol. XXII. No 3. P. 160-163.

13. Cappiello A., Chabini I., Nam E.K., Lue A., Zeid M.A. A Statistical Model of Emission and Fuel Consumption.

14. Kamal M.A.S., Mukai M., Murata J., Kawabe T. Ecological vehicle control on roads with up-down slopes. *Article in IEEE transactions on intelligent transportation systems*. October 2011. Impact Factor: 2.38 DOI:10.1109/TITS.2011.2112648 Source: IEEE Xplore.

15. El'sgol'ts L.P. Differential equations and calculus of variations. Moscow: Science, 1979. 340 p.

CELLULAR IoT IN THE 5G ERA

Ali Zaidi, Anders Bränneby, Ala Nazari, Marie Hogan, Christian Kuhlins,
Ericsson, Stockholm, Sweden

Ericsson, February 2020

ABSTRACT

Almost every industry can be transformed with cellular IoT. The connectivity needs of all industries can be addressed by four multi-purpose IoT segments, which efficiently co-exist in one 5G network. These segments are Massive IoT, Broadband IoT, Critical IoT and Industrial Automation IoT. This paper presents a clear evolution plan for addressing all 5G-IoT use cases, from basic to the most complex, in a cost-efficient, smooth and future-proof way.

KEYWORDS: *IoT, 5G, Ericsson, LTE.*

INTRODUCTION

The 3GPP-based global cellular networks are connecting things-to-things and things-to-persons across borders. Many industries are experiencing the benefits of cellular IoT, for example in the consumer electronics, automotive, railway, mining, utilities, healthcare, agriculture, manufacturing and transportation sectors.

There are over 1 billion cellular IoT connections today in 2020, and Ericsson forecasts around 5 billion connections by 2025[1].

With 5G in the market, almost every industry is exploring the potential of cellular connectivity for fundamentally transforming businesses. In some regions, governments are encouraging adoption of IoT via direct and indirect incentives to promote sustainability, innovation and growth.

Mobile network operators (MNOs) have long been successful in the mobile broadband (MBB) market and are also best positioned to create and capture value in the emerging IoT market with their regional and global footprint. Unlike MBB, the IoT usage scenarios have extremely diverse requirements.

For maximizing returns on investments, MNOs will have to systematically evolve cellular networks for addressing the needs of the rapidly increasing IoT use cases across multiple industries. This paper shows a clear evolution plan for addressing all 5G-IoT use cases, from basic to the most complex, in a cost-efficient, smooth and future-proof way.

Information about authors

Ali Zaidi, *Strategic Product Manager for Cellular IoT at Business Area Networks, Ericsson. He received an M.Sc. and a PhD in Telecommunications from KTH Royal Institute of Technology, Stockholm, Sweden*

Anders Brä nneby, *Strategic Product Manager for Internet of Things in Packet Core, part of Business Area Digital Services at Ericsson.*

Ala Nazari, *an expert in media delivery architecture within BDGS SA PC Architecture & Technology. He has also worked as a senior solution architect and engagement director.*

Marie Hogan, *is in charge of strategic product management of the mobile broadband and IoT areas of 4G and 5G Radio Access Networks at Ericsson. She holds an M.Sc. in Technology Management and a degree in Electronic Engineering.*

Christian Kuhlins, *Strategic Product Manager for Massive IoT and the new IoT use case areas UAV, railways and utilities at Ericsson.*

The wireless connectivity across various industries can be grouped into four distinct sets of requirements. To address these requirements, Ericsson has defined four IoT connectivity segments [2]: Massive IoT, Broadband IoT, Critical IoT and Industrial Automation IoT, as illustrated in Figure 1. Each IoT connectivity segment addresses multiple use cases in multiple industries.

Today, 4G networks are supporting Massive IoT based on Cat-M/NB-IoT and Broadband IoT based on LTE. Massive IoT continues to evolve with Cat-M/NB-IoT access in 5G-enabled networks, and Broadband IoT is being further enhanced with the introduction of 5G radio and core networks. With powerful, ultra-reliable and/or ultra-low latency capabilities, 5G networks are going to enable Critical IoT for time-critical communications. To seamlessly integrate 5G networks with Ethernet-based industrial wired communications networks, 3GPP has standardized additional capabilities that would be offered by Industrial Automation IoT connectivity.

The four IoT connectivity segments can co-exist in one 5G network, whether deployed for public or non-public access. Some devices may need multiple IoT connectivity segments for executing one or more use cases, for example, an autonomous vehicle with rich requirements [3].

Massive IoT

Massive IoT connectivity targets a large number of low-cost, narrow-bandwidth devices that infrequently send or receive small volumes of data. These devices can be situated in challenging radio conditions requiring extreme coverage and may rely solely on battery power supply. LTE-M and NB-IoT have been co-existing with LTE in 4G networks since 2017 and fulfill all 5G requirements from ITU and 3GPP for massive machine type communications [4-6].

LTE-M extends LTE to support machine-type communications, including access for the low-complexity device category series named Cat-M. NB-IoT is a standalone radio access technology based on the fundamentals of LTE. At the start of 2020, over 120 commercial networks supported NB-IoT and Cat-M access globally [7]. with millions of commercial users [1]. Forecasts indicate more than 2.5 billion connections will be in place by 2025 [1]. Commercial devices span various types of meters, sensors, trackers and wearables in many different industries, including utilities, automotive, transport, logistics, agriculture, manufacturing, healthcare, warehousing and mining [8].

There are two dominating types of Cat-M/NB-IoT modem in the market: single-mode NB-IoT modems, which are suitable for ultra-low cost devices, and dual-mode Cat-M1/NB-IoT modems, suitable for diversity of use cases with low-cost devices. The dual-mode modem combines the best attributes of the two technologies in terms of throughput, coverage, mobility, voice support and device positioning, as summarized in Figure 2.

The dual-mode devices use Cat-M1 mode in Cat-M1 signal coverage and can switch to NB-IoT access when out of Cat-M1 coverage. Cat-M1 has two coverage extension (CE) modes in the 3GPP standard: a mandatory CE mode A (for up to 10dB CE) and an optional CE mode B (for up to 20dB CE, on a par with NB-IoT).

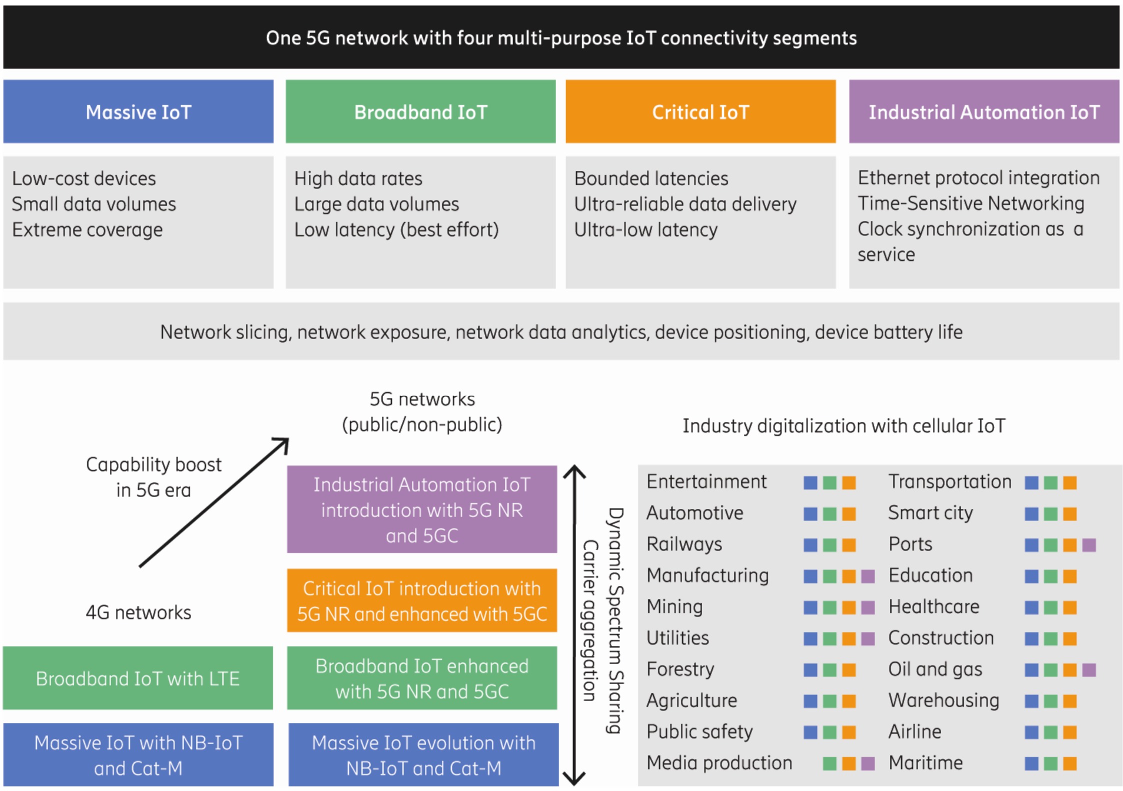


Figure 1. Industry digitalization with cellular IoT in the 5G era

Characteristics	Dual-mode Cat-M1/NB-IoT modem	Single-mode NB-IoT modem
Peak data rates	1.1Mbps (UL), 588kbps (DL)	158kbps (UL), 127kbps (DL)
Voice	Supported	Not applicable
Connected mode mobility	Supported	Not applicable
Coverage	Both modem types are on par	
Battery life	Both modem types are on par	
Guardband carrier	Guardband NB-IoT carrier can be used for both modem types	
Device positioning	Cat-M achieves better accuracy than NB-IoT due to wider bandwidth	

Figure 2. Comparison of Massive IoT modems (assuming half-duplex FDD and 3GPP Rel-16)

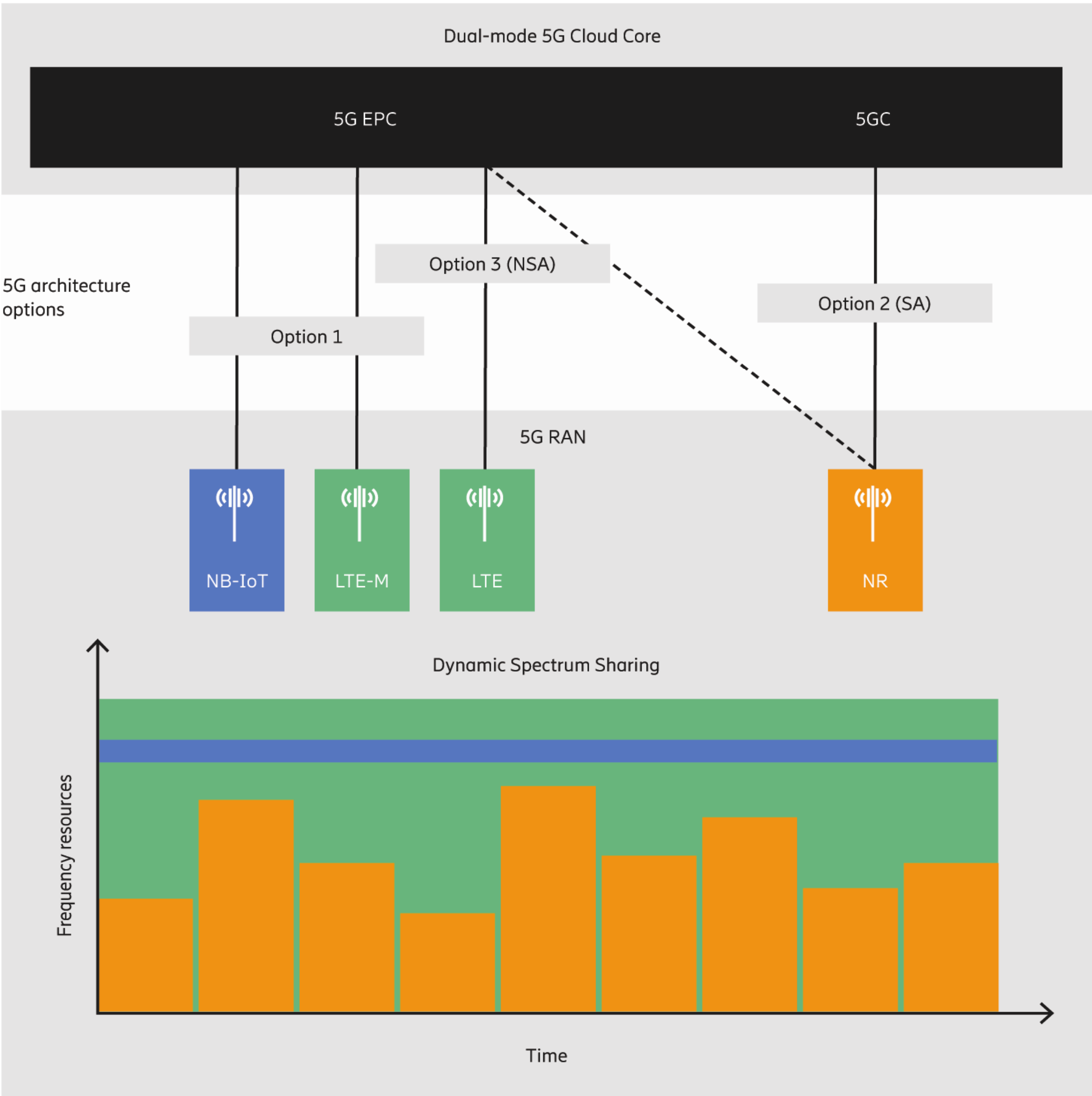


Figure 3. Cat-M and NB-IoT have a smooth and future-proof evolution in the 5G era

The performance benefits of Cat-M1 vanish when using CE mode B, due to fundamental trade-off between coverage and throughput. There is major spectrum resource consumption due to hundreds of subframe repetitions in CE mode B. Provided that the dual-mode modem can switch to NB-IoT access and leverage guardband NB-IoT carriers in extremely poor coverage scenarios, there is no commercial advantage in using CE mode B.

The peak data rates in Figure 2 are valid for commercially available Cat-M1 modems, which have 1.4MHz bandwidth. The 5MHz bandwidth Cat-M modem (known as Cat-M2) and the full-duplex operation mode have not been included, since these are currently not realized in the ecosystem due to the price-sensitive Massive IoT device market and their negative impact on device power consumption.

Cat-M1 and NB-IoT have a smooth and future-proof evolution in 5G networks when combined with Dynamic Spectrum Sharing [9], dual-mode 5G Cloud Core [10] and continued standardization in 3GPP [6]. NR is being deployed in the new 5G frequency bands, as well as in the 4G frequency bands, where Cat-M1, NB-IoT and LTE devices are operational.

With Dynamic Spectrum Sharing, all technologies co-exist efficiently, as shown in Figure 3.

Dual-mode 5G Cloud Core includes 5GC and 5G EPC. The existing and future Cat-M1/NB-IoT devices can connect to 5G EPC (known as 5G architecture Option 1) [11]. 3GPP Rel-16 is also specifying an option for connecting Rel-16 compatible Cat-M/NB-IoT devices to 5GC; however, this would be challenging for the cost-sensitive Massive IoT market, due to not only increased complexity, but also market fragmentation.

Broadband IoT

Broadband IoT connectivity adopts the capabilities of MBB for IoT to provide much higher data rates and lower latencies than Massive IoT, while enabling additional capabilities for IoT, such as extended device battery life, extended coverage, enhanced uplink data rates and enhanced device positioning precision.

Broadband IoT is relevant for all industries. There are more than 500 million Broadband IoT users in 2020, primarily with LTE access [1]. Commercial usage today is dominated by personal cars, commercial vehicles, trains, wearables, gadgets, cameras, sensors, actuators and trackers. These devices can leverage MBB connectivity; however, their requirements and traffic patterns are sometimes very different from typical MBB usage. For example, traffic patterns can be more uplink-heavy and/or periodic, while requirements on battery life, signal coverage and device positioning can be more challenging than for MBB.

LTE has a range of device categories (LTE Cat-1 and above) with wide bandwidths well suited for diverse, wide area use cases. LTE achieves Gbps data rates and RAN (best effort) latency down to around 10ms (round-trip time). With the introduction of 5G NR in old and new spectrum, Broadband IoT is set to enable tens of Gbps.

The signal coverage per base station can be enhanced if requirements are relaxed on data rate and latency; for example, an LTE device can dynamically switch between LTE and LTE- M access depending on the signal coverage. Device battery life can be significantly improved, leveraging user-specific traffic patterns. Network-based device positioning accuracy can be improved with NR since the positioning accuracy typically depends on signal bandwidth, and NR can operate in much wider bandwidths than LTE.

Uplink data rate is boosted with high-order modulation, multi-antenna transmission [12] and carrier aggregation. When using TDD, uplink and downlink capacities are fundamentally dependent on the TDD transmission pattern. Typically, MNOs have to agree on a common static TDD configuration in order to avoid interference according to regional regulations.

The TDD configurations deployed today are often downlink-heavy to optimize for MBB usage. However, with uplink-heavy IoT usage taking off, MNOs will have to reconsider mutual agreements on TDD configurations in order to achieve a good balance between uplink and downlink capacity, as well as low latency.

One of the 5G fundamentals is tight interworking between LTE and NR that allows 5G modems to simultaneously access LTE and NR carriers, known as non-standalone (NSA) 5G (Option 3 in Figure 3). A 5G-capable modem connects with NR (when in NR coverage) to experience a boost in performance while maintaining its LTE connection. This approach ensures that 5G deployments deliver value for wide area MBB and IoT users from day one [3].

Standalone (SA) 5G (Option 2 in Figure 3) is the long-term target architecture, as well as the ideal choice for usage scenarios with localized coverage needs (e.g. local industrial deployments) in the near term, from both performance and complexity standpoints.

To broaden the use cases addressable with NR, 3GPP Rel-17 will diversify NR device capabilities by introducing support for relatively less-complex modems with power-saving capabilities as part of the work on reduced capability NR devices; [13] example uses include industrial sensors and wearables.

Critical IoT

Critical IoT connectivity is for time-critical communication. It enables data delivery within desired latency bounds. It includes 5G's most powerful capabilities for ultra-high reliability and/or ultra-low latency communication at a variety of data rates. The reliability here is defined as the probability of successful data delivery within a specified time duration [14]. In contrast to Broadband IoT, which achieves low latency on best effort basis, Critical IoT can deliver data within specified latency bounds with required guarantee levels, even in heavily loaded networks.

Typical use cases with demanding combinations of reliability, latency and data rates include AR/VR, autonomous vehicles, mobile robots, real-time human machine collaboration, cloud robotics, haptic feedback, real-time fault prevention, and coordination and control of machines and processes [14-17]. Such use cases are relevant in almost every industry. Some industries are piloting these applications with 5G, for example, in the entertainment, automotive, manufacturing, mining, harbor, airport, construction and utilities sectors.

To enable demanding Critical IoT use cases, all components (networks, devices and applications) may have to step up in terms of latency and reliability. From a pure network perspective, end-to-end latency is the sum of individual latency contributions from radio, transport and core networks, and the overall reliability cannot be higher than the reliability of the weakest link.

5G NR and 5GC have been standardized for ultra-reliable and low latency communication (URLLC) from day one (Rel-15) with further evolution in Rel-16 and Rel-17 [6]. With URLLC capabilities, 5G NR can achieve latencies down to 1ms and reliability up to 99.9999 percent. Latency within the core network is typically below 1ms. The transport network can be a major contributor to the end-to-end latency. Transport network latency varies widely between regions, depending on distances and the transport solutions used. A general trend is that transport latency is being optimized by higher availability of fiber and fewer router hops. As an example, the round-trip time between 2 cities in a European country (city distance 1,300km) is today just 16ms (theoretical minimum optical fiber latency is 13ms), which is less than half of the latency of 5 years ago.

Edge computing is needed to reduce transport latency for demanding Critical IoT use cases. Distributed anchor points, local break-out or on-premise full-core deployments are key scenarios to achieve low to ultra-low latency, as illustrated in Figure 4 with typical latencies. Core user plane distribution reduces latency by keeping the user traffic as local as possible and is therefore typically distributed on more local sites than control plane network functions. For use cases that require very high reliability, the core control plane and network exposure can be further distributed to limit network disturbances, even in the rare event of a major incident or disaster scenario. Real-time mobility, group management and network monitoring could also drive decentralized control plane to optimize for reduced latency. Core control plane and network exposure are typically recommended to be located at the same site to avoid so-called signaling flow tromboning. Another driver for edge computing is data offload, which is beneficial for both MBB and IoT [18, 19].

On-premise, full-core deployment enables dedicated resources, ultra-low latency and ultra- reliability, providing autonomous operation with local subscriber data. The local area network can also be interconnected with a public network, allowing mobile IoT devices to roam between the networks.

Critical IoT can be demanding in terms of bandwidth, since any major gains in reliability and latency typically require substantial spectrum resources. NR operates in a wider range of frequencies with larger bandwidths and far greater capabilities than LTE, which makes NR the technology of choice. LTE may never be enhanced for Critical IoT due to multiple factors, such as the timing of commercial use cases, the continued capability expansion of NR URLLC in 3GPP, the momentum on NR and the available option of software upgrading existing LTE sites with NR (in the LTE spectrum bands).

With flexible spectrum assets, MNOs are best positioned to provide Critical IoT coverage not only in wide area deployments, but also for local industrial deployments. NR enables URLLC in all 5G (FDD/TDD) frequency bands [2]. Figure 5 illustrates examples of spectrum band combinations, along with key characteristics of different bands in terms of URLLC capacity and coverage for addressing wide and local area users. Considering that there is very limited bandwidth in sub-1GHz bands, these should be leveraged for high-value, wide area users.

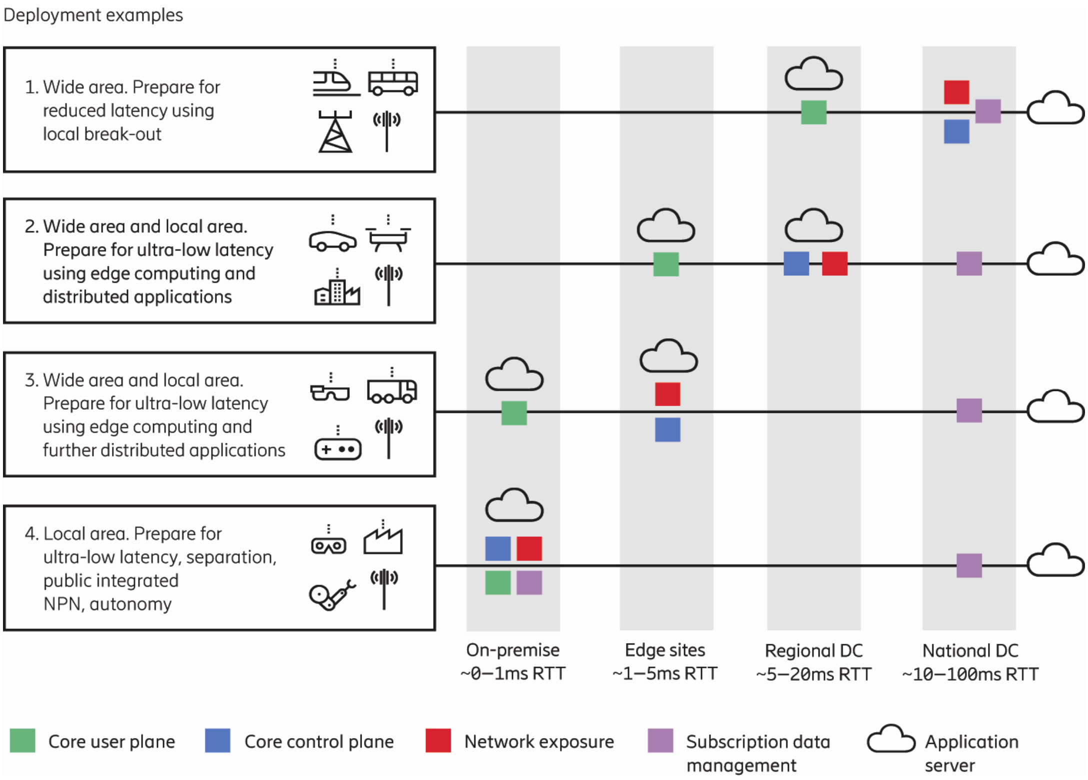


Figure 4. Core network deployment examples to support Critical IoT use cases

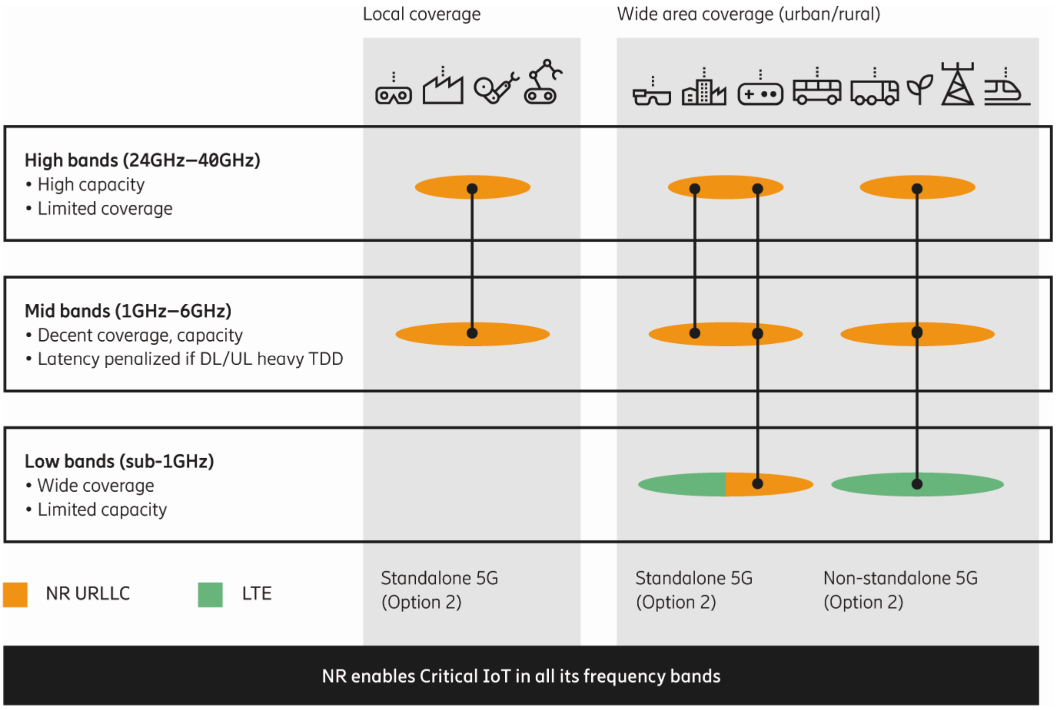


Figure 5. MNOs are best positioned to enable Critical IoT with their flexible spectrum assets

When using TDD, RAN latency is fundamentally dependent on TDD transmission patterns. A downlink- or uplink-heavy TDD configuration negatively impacts latency, especially in sub-6GHz [2, 20].

SA 5G is ideal for fulfilling the challenging Critical IoT requirements. 5GC is better than 5G EPC in terms of ultra-reliability mechanisms, advanced service differentiation, flexible edge computing, network data analytics, advanced Quality of Service (QoS), Ethernet connectivity, and end-to-end network slicing capabilities which can be important for critical use cases. Provided that LTE is also not enhanced for Critical IoT, NSA 5G does not offer full potential for URLLC from both radio access and core network perspectives. For use cases with local coverage needs, such as local industrial sites, NSA 5G deployments would not be beneficial, as discussed earlier. However, in wide area coverage, 5G deployments would be initially NSA and could leverage NR user plane capabilities for URLLC to enable less demanding Critical IoT use cases. Over time, NSA 5G deployments will transition to SA 5G, achieving the full potential of Critical IoT in wide areas.

Industrial Automation IoT

Industrial Automation IoT aims at enabling seamless integration of cellular connectivity into the wired industrial infrastructure used for real-time advanced automation. It includes capabilities for integrating 5G systems with real-time Ethernet and Time-Sensitive Networking (TSN) used in industrial automation networks.

Cellular connectivity offers great benefits in term of mobility, flexibility, cost-cutting and digitalization compared to wired communication. However, in some industrial deployments, wired networks may migrate to wireless connectivity in such a manner that different parts of an industrial system may switch to wireless connectivity gradually over time. Even if an industrial system is within 5G coverage, certain components of the system might stay connected with cables due to various factors; for example, not having a major need for a wireless solution, having long life cycles, or having extreme performance needs that are beyond 5G's current capabilities (for example, micro-second level deterministic latency). It is important that 5G supports seamless integration into the current and evolving wired infrastructure [21].

A number of industries use wired communication for advanced automation; for example, the mining, utilities, construction, ports, oil and gas sectors. There are several industrial Ethernet solutions supporting deterministic communication for real-time automation; for example, PROFINET, EtherCAT, Sercos, EtherNet/IP, Powerlink and Modbus. 3GPP has standardized support for Ethernet sessions in Rel-15. In Rel-16, Ethernet header compression has been introduced for spectral efficiency. Reliable data delivery within strict latency bounds is achieved with 5G URLLC (enabled by Critical IoT). With Rel-16, a 5G virtual network can be set up over a 5G system providing 5G LAN-type service (e.g. VLAN) by which on-demand connection of UE to UE (user equipment), multicast and broadcast private communications is supported between members of the same 5G virtual network [22]. In order to overcome challenges with the fragmented industrial Ethernet market, a common open standard is emerging: Ethernet with TSN support [23]. TSN is designed for diverse QoS requirements, including both deterministic and best-effort latencies. TSN was standardized within IEEE and its profile for industrial automation is being developed jointly by IEC and IEEE. To enable seamless integration of 5G with TSN, 3GPP has standardized a feature set in Rel-16 as part of the Industrial IoT work item.

The 5G-TSN integration is illustrated in Figure 6, where a 5G system is integrated into a TSN network as a bridge [24]. TSN Translators (TT) are introduced in the 5G system for the user plane and control plane. The user plane translators are placed at the UPF (Network TSN Translator, NW-TT) and at the UE (Device Side TSN Translator, DS-TT).

TSN is managed centrally by the TSN Central Network Controller (CNC). A TSN Application Function (AF-TT) is placed at the 5G control plane to expose 5G system capabilities (e.g. list of ports at DS-TT and NW-TT and the transfer delay between them) to the TSN CNC, for configurating and scheduling TSN flows across the 5G system bridge. Hold and forward (de-jitter) buffering at NW-TT and DS-TT is used for delivering traffic streams with deterministic latency based on the time-aware scheduling information obtained from CNC.

The TSN nodes are time-synchronized with a master clock using the generalized Precision Time Protocol (gPTP) [IEEE 802.1AS] [24]. The 5G system bridge can either support forwarding of the gPTP synchronization information or use its internal clock as a grandmaster for providing a time-reference to the TSN nodes. The 5G system can also deliver clock synchronization as a service to industrial applications that operate synchronously; for example, synchronized coordination between multiple controllers in a system.

Key enablers for cellular IoT

Network slicing

Network slicing allows creation of multiple logical networks using a common shared network infrastructure across radio, core and transport networks, which is essential for cost efficiency, scaling and flexibility [25]. An operator can either deploy end-to-end network slices per cellular IoT segment serving consumers and multiple enterprises, or create separate network slices for different enterprises where each slice may include multiple IoT connectivity segments, as shown in Figure 7.

A network slice includes required network resources configured and connected across radio, transport and core network. The resources can be physical or virtual, either dedicated to a slice or shared between slices. The slices can be dynamically created on an as-needed basis. A slice service Orchestrator automates the creation, modification and deletion of the individual slices, while also handling the assignment of the underlying resources. By continuously monitoring slice performance, the Orchestrator is able to accurately configure and adjust slice resources.

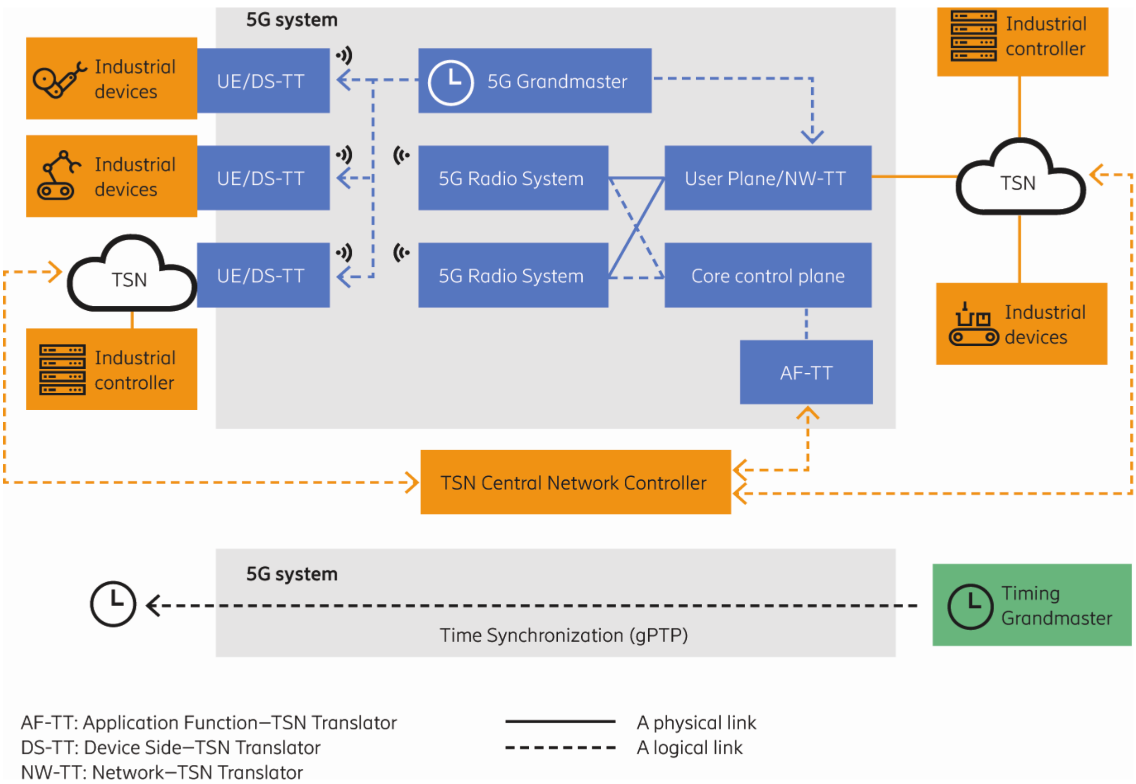


Figure 6. 5G-TSN integration and clock synchronization as a service

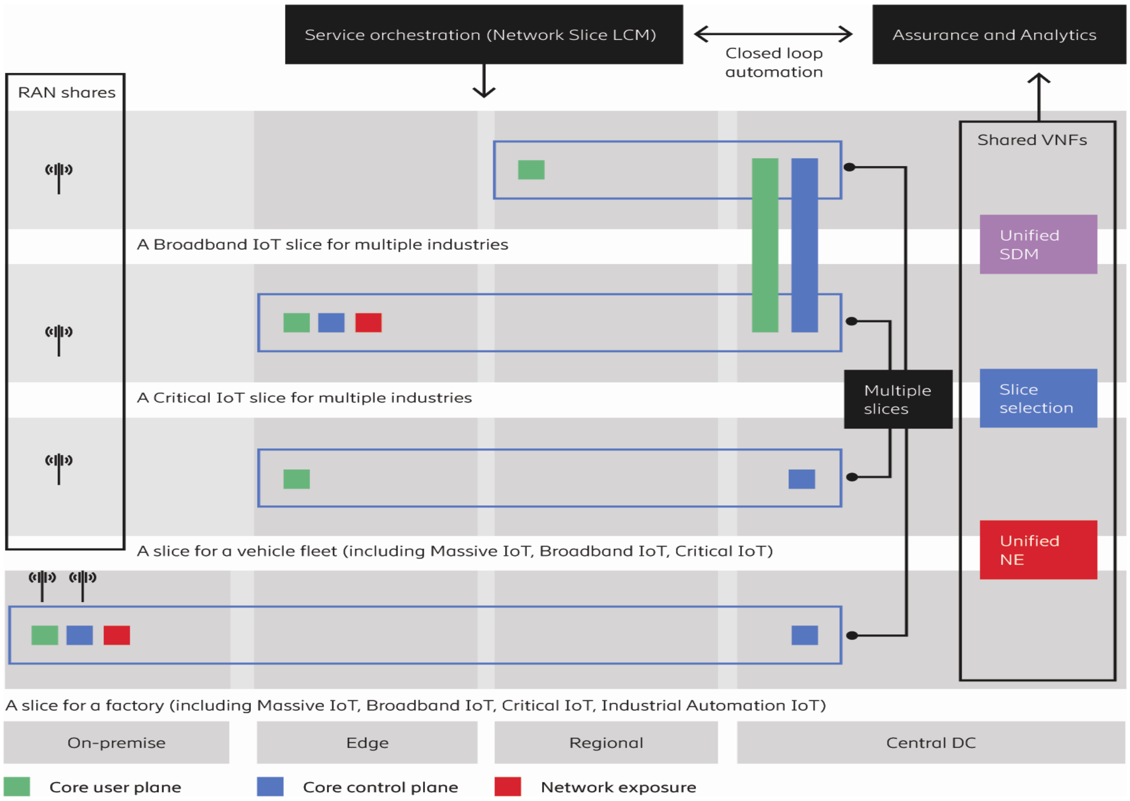


Figure 7. Dynamic network slicing with closed-loop life cycle management

Network exposure

Network exposure with RESTful APIs securely exposes network capabilities, enabling network programmability. The main types of service exposure for IoT connectivity are:

- network monitoring, including network publishing information as real-time statuses, event streams, statistics and analytics insights
- device-related APIs for provisioning, onboarding, triggering, device and connectivity management, connectivity monitoring and location information
- network control and configuration, involving control services that request configuration changes including network resource management, group management and application function influencing traffic routing (local breakout)
- payload interfaces for small data delivery, non-IP data delivery and background data transfer

Network Data Analytics

Network Data Analytics leverage the consolidated data from status monitoring of the network, applications and devices for extracting valuable network insights and optimizing the network for improved performance. The Network Data Analytics Function has been introduced in the 5G system architecture in 3GPP Rel-15 [26] with enhancements in Rel-16, which enables:

- the detection of misbehaving devices by observing abnormal traffic patterns;
- deriving a suitable policy for background data transfer by analysis of, for example, traffic volume, congestion level, load status information in the specific network area;
- dynamic traffic routing to the edge by analyzing network status (for example load information based on time and space), which service is available at the edge, and the device's location;
- assisting applications with predictable network performance by analyzing speed, direction and location of devices and network status;
- network automation by collecting and analyzing the status of network slices and assisting the network slice Orchestrator to scale up or scale down the resources for IoT network slices.

SIM flexibility

Traditionally, the cellular network subscription credentials are provided to devices using physical SIM cards. However, it can be difficult to physically access the SIM cards in various IoT devices after they are produced and sold. During production of devices, it is often not known which cellular subscription(s) would be used and the subscription(s) may also change multiple times during the device's life. To address this, GSMA has specified an Embedded SIM (eSIM) solution for remote provisioning of the subscriber credentials without physically touching the device [27]. For optimizing device cost, form factor and power consumption, an Integrated SIM (iSIM) solution embedded into a device's chipset hardware exists in the market that builds on the eSIM functionality [28].

3GPP Rel-16 has standardized non-SIM authentication (a certificate-based EAP-TLS authentication principle) for 5GC [29]. The non-SIM authentication is useful for non-public networks where devices may not need subscription. A device accessing both public and non-public networks may still use SIM functionality for the public network and the non-SIM feature for accessing the non-public network. The non-SIM authentication is also attractive for low-cost NB-IoT and Cat-M1 devices, and is therefore also relevant for 5G EPC.

Conclusion

MNOs are uniquely positioned to transform almost every industry with cellular IoT. The connectivity needs of all industries are addressed with four multi-purpose IoT segments which coexist efficiently within one 5G network, leveraging complementary capabilities of multiple radio access and core network technologies, with cost-efficiency, flexibility and scale.

Firstly, Cat-M and NB-IoT are formally 5G Massive IoT technologies with global coverage and a clear evolution plan. There is tremendous potential for realizing truly Massive IoT with continued investment in the existing Cat-M1 and NB-IoT ecosystem.

Secondly, Broadband IoT has a natural head start with 4G and initial 5G MBB roll-outs. Its long-term success depends on addressing the IoT-specific challenges, such as signal coverage, device battery life, device positioning, uplink-heavy traffic and diversifying capabilities of 5G devices.

Thirdly, almost every industry has time-critical communication needs. 5G powered with URLLC capabilities is the most suitable wireless technology for realizing Critical IoT. However, a systematic end-to-end co-development in the ecosystem is essential for realizing Critical IoT gradually over time.

Finally, Industrial Automation IoT with support for real-time Ethernet and TSN is an enabler for seamless integration of 5G into the existing and evolving industrial deterministic networks used for real-time automation. These capabilities, in conjunction with other IoT segments, enable the Industry 4.0 revolution to meet its promise for full digitalization.

Acknowledgments

The authors would like to thank Fredrik Alriksson, Joachim Sachs, Tomas Nylander, Johan Bergman, Eric Wang, Peter de Bruin, Bela Rathonyi, Christer Boberg, Robert Khello, Jan Backman, Maria Dusing, Richard Möller, Alexandra Martido, Shanqing Ullerstig, Mikael Persson, Olof Liberg, Lisa Boström, Juan-Antonio Ibanez, Thorsten Lohmar and Göran Eneroth for their valuable contributions to this paper.

References

1. Ericsson Mobility Report November 2019: https://www.ericsson.com/en/mobility-report/reports/november-2019
2. Cellular IoT Evolution for Industry Digitalization, Ericsson White Paper: https://www.ericsson.com/en/reports-and-papers/white-papers/cellular-iot-evolution-for-industry-digitalization.
3. Driving transformation in the automotive and road transport ecosystem with 5G, Ericsson Technology Review, 2019: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/ transforming-transportation-with-5g
4. IMT-2020 self-evaluation: mMTC coverage, data rate, latency & battery life. Ericsson and Sierra Wireless. 3GPP R1-1814144, November 2018: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_95/ Docs/R1-1814144.zip.
5. LTE-M and NB-IoT meet the 5G performance requirements. Ericsson blog post, December 2018: https://.www.ericsson.com/en/blog/ 2018/12/lte-m-and-nb-iot-meet-the-5g-performance-requirements.
6. Cellular Internet of Things – From Massive Deployments to Critical 5G Applications, 2nd edition, Academic Press, 2019: https:// www.elsevier.com/books/cellular-internet-of-things/liberg/978-0-08-102902-2
7. https://www.gsma.com/iot/mobile-iot-commercial-launches/
8. IoT Ecosy em: NB-IoT and LTE-M Report, GSA, 2019: https://gsacom.com/paper/iot-ecosystem-report-april19/
9. Sharing for the best performance, Ericsson White Paper, 2019: https://www.ericsson.com/en/networks/offerings/5g/sharing-spectrum-with-ericsson- spectrum-sharing/download-form
10. One core – the best of two worlds, Ericsson White Paper, 2019: https://www.ericsson.com/en/ digital-services/forms/packet-core/one-core-the-best-of-two-worlds
11. Simplifying the 5G ecosystem by reducing architecture options, Ericsson Technology Review, 2018: https://www.ericsson.com/en/ reports-and-papers/ericsson-technology-review/articles/simplifying-the-5g-ecosystem-by-reducing-architecture-options.
12. Advanced antenna systems for 5G networks, Ericsson White Paper: https://www.ericsson.com/en/reports-and-papers/white-papers/ adv antenna-systems-for-5g-networks
13. New SID on suppor f reduced capability NR devices, 3GPP, RP-193238, December 2019: :/www.3gpp.org/ftp/TSG_RAN/TSG_RAN/TSGR_86/Docs/RP-193238.zip
14. Service requirements for next generation new services and markets, 3GPP TS22.261, 2019: http://www.3gpp.org/ftp//Specs/archive/ 22_series/22.261/22261-h01.zip
15. Service requirements for cyber-physical control applications in vertical domains, 3GPP TS 22.104, 2019: http://www.3gpp.org/ftp//Specs/archive/22_series/22.104/22104-h20.zip
16. Service requirements for enhanced V2X scenarios, 3GPP TS 22.186, 2019: http://www.3gpp.org/ftp//Specs/archive/22_series/ 22.186/22186-g20.zip
17. Mobile communication system for railways, 3GPP TS 22.289, 2019: http://www.3gpp.org/ftp//Specs/archive/22_series/22.289/22289-h00.zip
18. Driving Data to the Edge: The Challenge of Traffic Distribution, AECC Technical Report, Automotive Edge Computing Consortium (AECC), Technical Solution Working Group (WG2), 2019: https://aecc.org/wp-content/uploads/2019/09/AECC_WG2_TR_v1.0.2.pdf
19. Distributed cloud – a key enabler of automotive and industry 4.0 use cases, Ericsson Technology Review, 2019: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/ distributed-cloud
20. J. Sachs et. al, 5G Radio Network Design for Ultra-Reliable Low-Latency Communication, IEEE Network, 2018: https://ieeexplore.ieee.org/document/8329620
21. 5G for Automation in Industry, 5G ACIA White Paper, 2019:https://www.5g-acia.org/fileadmin/5G-ACIA/Publikationen/5G-ACIA_White_Paper_5G_for_Automation_in_Industry/WP_5G_for_ Automation_in_Industry_final.pdf.
22. Technical Specification Group Services and System Aspects; System architecture for the 5G System (5GS), 3GPP TS 23.501
23. Boosting smart manufacturing with 5G wireless connectivity: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/boosting-smart-manufacturing-with-5g-wireless-connectivity
24. 5G-TSN integration meets networking requirements for industrial automation: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/ articles/5g-tsn-integration-for-industrial-automation.
25. Network Slicing can be a piece of cake, Ericsson: https://foryou.ericsson.com/paper-network-slicing-can-be-a-piece-of-cake-ty_1.html
26. Network Data Analytics Services, 3GPP TS29.520: https://www.3gpp.org/ftp/Specs/ archive/29_series/29.520/
27. eSIM Whitepaper, The what and how of Remote SIM Provisioning, GSMA, March 2018: http /www.gsma.com/esim/wp-content/uploads/2018/12/esim-whitepaper.pdf
28. Unlocking cellular potential for chipset makers, ARM White Paper: https://learn.arm.com/isim-for-chipset-makers.html.
29. Technical Specification Group Services and System Aspects; Security architecture and procedures for 5G system, 3GPP TS 33.501: https://www.3gpp.org/ftp/Specs/archive/33_series/33.501.

NOKIA BELL LABS' WORLD RECORDS AND INNOVATIONS IN FIBER OPTICS TO ENABLE FASTER AND HIGHER CAPACITY 5G NETWORKS OF THE FUTURE

Murray Hill, N.J., 13 March 2020 – **Nokia Bell Labs today announced that its researchers have set the world record for the highest single carrier bit rate at 1.52 Terabits per second (Tbit/s) over 80 km of standard single mode fiber -- the equivalent of simultaneously streaming 1.5 million YouTube videos – which is four times the market's current state-of-the-art of approximately 400 Gigabits per second. This world record, along with other optical networking innovations announced today, will further strengthen Nokia's ability to develop networks for the 5G era that meet the ever-growing data, capacity and latency demands of industrial Internet of Things and consumer applications.**

Several of these achievements were presented as part of Nokia Bell Labs' post deadline research papers at the Optical Fiber Communications Conference & Exhibition (OFC) being held this week in San Diego. Additionally, Nokia Bell Labs researcher Di Che was awarded the OFC Tingye Li Innovation Prize. Named after the late Tingye Li, a Bell Labs luminary in the field of optical communications, the prize is given to an early career professional who has demonstrated innovation in their research.

Marcus Weldon, Nokia CTO and President of Nokia Bell Labs, said: "It has been fifty years since the inventions of the low-loss fiber and the associated optics. From the original 45 Megabit-per-second systems to more than 1 Terabit-per-second systems of today – a more than 20,000-fold increase in 40 years – to create the fundamental underpinning of the internet and the digital societies as we know it. The role of Nokia Bell Labs has always been to push the envelope and redefine the limits of what's possible. Our latest world records in optical research are yet another proof point that we are inventing even faster and more robust networks that will underpin the next industrial revolution."

The highest single-carrier bitrate at 1.52 Terabits per second was set by a Nokia Bell Labs optical research team led by Fred Buchali. This record was established by employing a new 128 Gigasample/second converter enabling the generation of signals at 128 Gbaud symbol rate and information rates of the individual symbols beyond 6.0 bits/symbol/polarization. This accomplishment breaks the team's own record of 1.3 Tbit/s set in September 2019 while supporting Nokia's record-breaking field trial with Etisalat.

Nokia Bell Labs researcher Di Che and team also set a new data-rate world record for directly modulated lasers (DML), which are crucially important for low-cost, high-speed applications such as datacenter connections. The DML team achieved a world record data rate beyond 400 Gbit/s for links up to 15 km.

In addition to these world records, Nokia Bell Labs researchers have also recently achieved significant achievements in optical communications, including:

The first field trial using spatial-division-multiplexed (SDM) cable over a 2,000km span of 4-core coupled-core fiber was achieved by researchers Roland Ryf and the SDM team. The experiments clearly show that coupled-core fibers are technically viable, offer high transmission performance, while maintaining an industry standard 125-um cladding diameter.

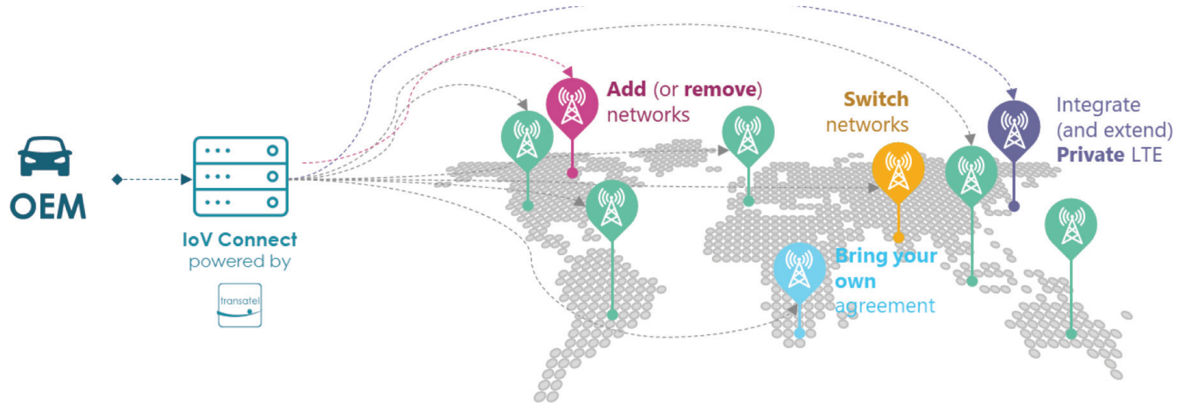
A research team led by Rene-Jean Essiambre, Roland Ryf and Murali Kodialam introduced a novel new set of modulation formats that provide improved linear and nonlinear transmission performance at submarine distances of 10,000 km. The proposed transmission formats are generated by a neuronal network and can significantly outperform traditional formats (QPSK) used in today's submarine systems.

Researcher Junho Cho and team experimentally demonstrated capacity gains of 23% for submarine cable systems that operate under electrical supply power constraints. The capacity gains were achieved by optimizing the gain shaping filters using neural networks.

The researchers that achieved the world record and research results are part of Nokia Bell Labs' Smart Optical Fabric & Devices Research Lab, which designs and builds the future of optical communications systems, pushing the state-of-the-art in physics, materials science, math, software and optics to create new networks that adapt to changing conditions and go far beyond today's limitations.

Nokia Communications

TRANSATEL (NTT GROUP) LAUNCHES IOV CONNECT, AN ULTRA AN ULTRA SPECIALIZED, GLOBAL AND SECURED CAR PLATFORM



Paris, 02 April, 2020, **aving successfully deployed connectivity services for several market leaders in Europe, such as Fiat Chrysler Automobiles and Jaguar Land Rover, ranging from vehicle manufacturers to service providers [1], Transatel announced today the launch of IoV Connect [2], a global cellular connectivity platform specifically designed for the sector of motorized vehicles.**

Available as of today in PaaS mode, the platform is dedicated to solving vehicle manufacturers' greatest challenges currently not being addressed by the market. The platform's first key differentiator is 'MNO selection at network-level'. Any addition, removal, change of host radio network anywhere in the world is programmed at network level and thus implemented instantaneously, not requiring complex Over-The-Air SIM profile update campaigns.

The second key differentiator is end user management, for which most connected car platforms come unequipped. Leveraging on Transatel's 20+ years' experience in launching mobile B2C offers around the world, IoV Connect offers the professional tools, insights and expertise to cover all aspects of pricing, marketing, billing, customer experience, customer care and regulatory compliance. Last, IoV Connect offers a fully re-brandable service for onboard Wi-Fi and infotainment, complete with marketing and end-user support services: Ubigi.

Complexity reduced: One integration for global deployment

Requiring but a single integration to cover the world (160+ countries to date), IoV Connect caters to vehicle manufacturers wishing to avoid the cost and/or complexity of multiple integrations with Mobile Network Operators. Most existing IoV platforms are feature-rich but rely on the aggregating of multiple

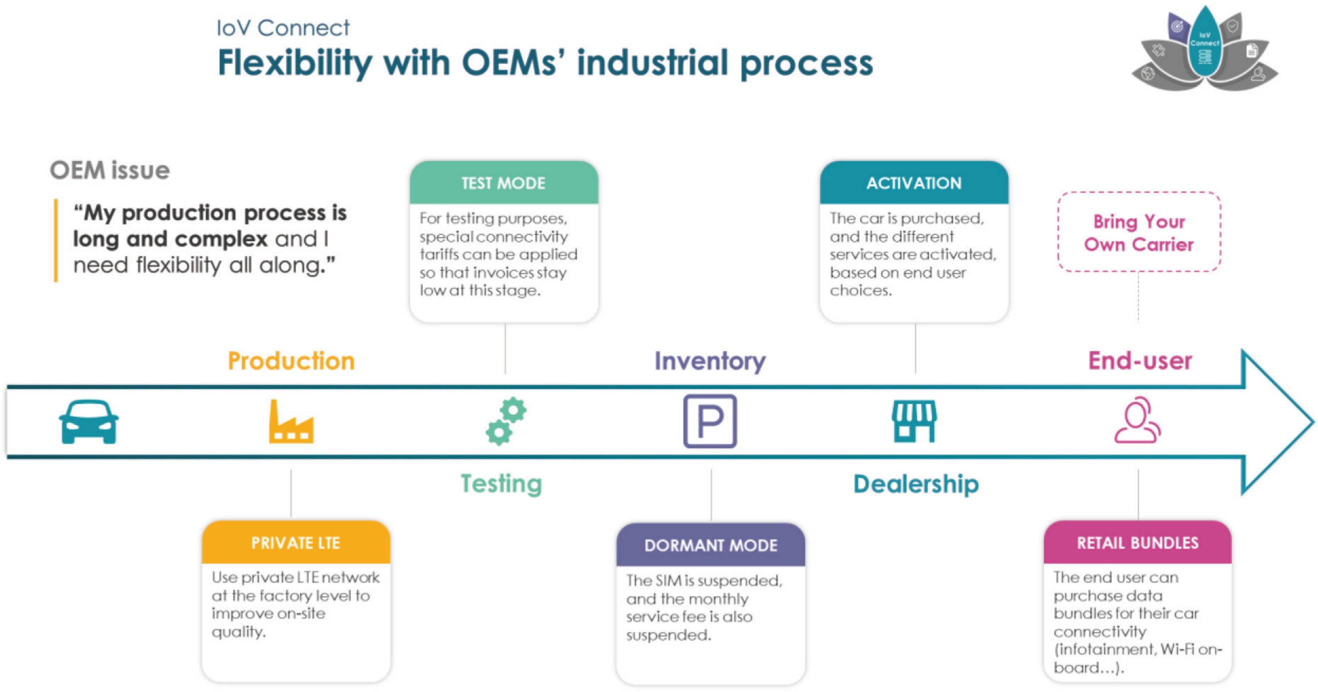
Mobile Network Operator profiles. As a vehicle manufacturer, you therefore still need a Service Agreement with the MNO in question. In the case of IoV Connect, Transatel is the MNO. Manufacturers no longer need to sign Service Agreements with Mobile Network Operators, simply Access Agreements, which greatly reduces complexity. Moreover, manufacturers benefit out-of-the-box from 160+ Access Agreements already bundled into the IoV Connect platform.

CREATIVITY ENABLED: CONFIGURATION AT NETWORK LEVEL

With most existing connected car platforms, manufacturers deciding on a change in Mobile Network Operator or a feature update must rely on costly (0,60 to 1/SIM) Over-The-Air campaigns to reconfigure the SIM profiles inside their vehicles. Any inaccessible vehicles, such as garage-parked, or cars in motion, etc., will need to proceed physically to the dealership for a SIM profile update, with all that this entails. Thanks to a configuration at network level, any switching of underlying host radio networks, or any new 'connected vehicle' feature developed, is instantly and seamlessly made available to vehicles across continents. The process, remaining perfectly transparent for the end user, leaves a lot of room for innovation at every turn.

FLEXIBILITY AT EVERY STAGE: PRODUCTIZED FEATURES, BUT TAILORED IMPLEMENTATION

IoV Connect offers a combination of productized features with full implementation, geographical scope and process flexibility.



FUTURE-PROOFED: A SECURED SOLUTION FOR VEHICLE CONNECTIVITY

Transatel has long been concerned with Security for the IoT and has developed a system for airtight connectivity. [3] IoV Connect has therefore been modeled considering the critical factor of security for vehicles, and this, at three levels.

First, as a licensed operator without spectrum, Transatel have their own non-geographic MNC (Mobile Network Code). The underlying resources used are therefore inaccessible from the exterior. For example, in the typical case of hacking with a massive brute-force attack via SMS, the only SMS accepted by the vehicle are those from the OEM's authorized servers. The unwanted SMS can't even reach the authentication process and therefore drain the vehicle's battery in the process.

Second, all the software and firmware updates are not conducted via the internet.

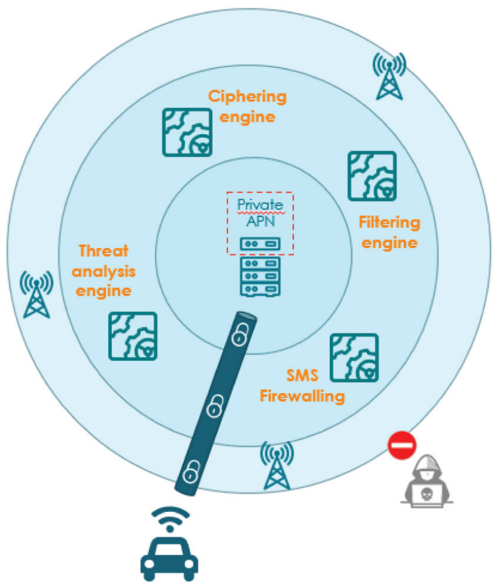
Third, IoV Connect SIM profiles are embedded with security keys for authentication purposes, to further shield vehicles from unwanted intrusions. Of course, IoV Connect is compatible with any form of additional security processes and add-ons desired by the manufacturer.

IoV Connect's security solution protects against threats and cyberattacks with a full set of services:

• A secure architecture for Data and SMS: Private APN, vehicle isolation, public non-exposure, certified architecture

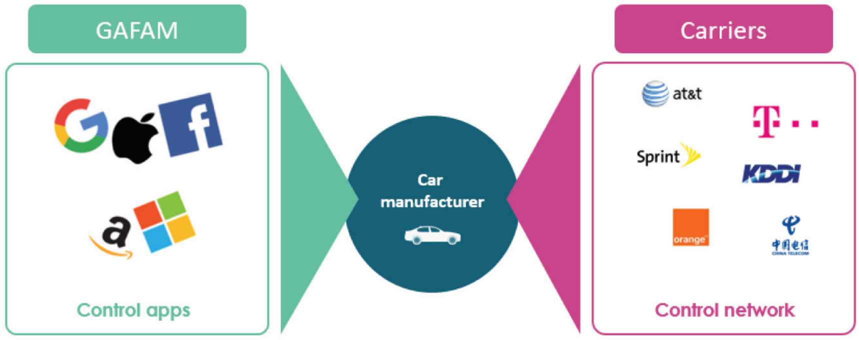
• Global security functionalities on the architecture: Threats analysis, traffic filtering, SMS firewalling

• TAC/IMEI lock.



GETTING A SHARE OF THE VALUE CREATED BY CONNECTED SERVICES

Caught between the GAFAM who control the apps, and the MNOs who control the connectivity, vehicle manufacturers are understandably trying to secure their share of the revenue created with the increasing use of apps within their vehicles.



Acting as an abstraction layer to MNO networks, IoV Connect enables OEMs to reap a share of the value generated from the connectivity services offered within their vehicles (whether from telematics, infotainment, or Wi-Fi on board). Manufacturers can also monetize the data collected and the access/visibility to their end users. Last, IoV Connect enables manufacturers to retain their independence and bargaining power vis a vis MNOs and thus control their connectivity service and financial equation.

Bertrand Salomon, Transatel Co-CEO: "I'm excited about this launch, because it's the result of years of observation and exchanges with the industry, and years of hard work from Transatel teams. IoV Connect can turn any complex, global Connected Car project, into a manageable enterprise with clear milestones and beneficial outcomes. By making things simple and global by design, we help manufacturers focus on their core activities and gain global economies of scale. We intend to prove that our platform for the Connected Car, both industry-specific and fully managed, is the safest option, and the most scalable and cost-effective business model for auto manufacturers, right up until they're ready to become fully licensed telecom operators themselves!"

About Transatel

Now a member company of NTT Group, one of the largest telecom companies in the world, Transatel offers an unparalleled, eSIM-capable, cellular solution for global and secure IoT connectivity. Since its inception in 2000, the company has launched over 170 MVNOs (Mobile Virtual Network Operators), establishing Transatel as the leading European MVNE/A (Mobile Virtual Network Enabler/Aggregator). Having acquired expertise in Machine-to-Machine connectivity, the company easily transitioned into the Internet of Things, where it now addresses the three market segments of automotive, laptops and tablets, and the industrial IoT.

[1] In Europe only for the moment, but covering EU28, Transatel signed with OEMs such as Fiat Chrysler Automobiles, Jaguar Land Rover, Scania and DAF Trucks. Transatel also signed with service providers, such as Blue Solutions (a Bollore subsidiary for car sharing solutions), Sekurity (a security after-market add-on for motorcycle users), and Xee (a fleet management platform).

[2] IoV = Internet of Vehicles

[3] Please read our whitepaper on Security for the IoT: https://www.transatel.com/security-iot/

Or cf. https://www.transatel.com/security-iot/