RZESZOW UNIVERSITY OF TECHNOLOGY

p-ISSN 3071-9003
e-ISSN 3071-9011

Advances in IT and Electrical Engineering

30

2024

THE FACULTY OF
ELECTRICAL
AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

RZESZOW UNIVERSITY OF TECHNOLOGY

Advances in IT and Electrical Engineering

30

2024

THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

p-ISSN 3071-9003
e-ISSN 3071-9011

Issued with the consent of the Rector

E d i t o r i n C h i e f
Publishing House of Rzeszow University of Technology
PhD DSc Eng. Lesław GNIEWEK, Assoc. Prof.

J o u r n a l S c i e n t i f i c B o a r d
Advances in IT and Electrical Engineering

(affiliations: Poland)

Journal editor-in-chief
PhD DSc Eng. Dominik STRZAŁKA, Assoc. Prof.

Deputy journal editor-in-chief
Prof. Lesław GOŁĘBIOWSKI

Associate editors
Prof. Robert HANUS
Prof. Jacek KLUSKA
PhD DSc Eng. Mariusz KORKOSZ, Assoc. Prof.
PhD Eng. Bartosz TRYBUS

Editorial assistant, Statistical editor
PhD Eng. Anna SZLACHTA, Assoc. Prof.

Language editor
Piotr CZERWIŃSKI

p-ISSN 3071-9003
e-ISSN 3071-9011

The electronic version of the annual Journal is the final, binding version.

Editorial Office: Rzeszow University of Technology, The Faculty of Electrical and Computer Engineering
2 Wincentego Pola Str., 35-959 Rzeszów (e-mail: aitee@prz.edu.pl)
https://journals.prz.edu.pl/aitee

Publisher: Publishing House of Rzeszow University of Technology,
12 Powstańców Warszawy Ave., 35-959 Rzeszów (e-mail: oficyna@prz.edu.pl)
https://oficyna.prz.edu.pl/

Advances in IT and Electrical Engineering

vol. 30, 2024
p-ISSN 3071-9003 e-ISSN 3071-9011

THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

SPIS TREŚCI

This is an Open Access article distributed under the terms of the CC-BY-NC-ND 3.0 PL license, which permits others to distribute the work, provided that the article is not altered or used commercially. You are not required to obtain permission to distribute this article, provided that the original work is properly cited.

Advances in IT and Electrical Engineering

vol. 30, 2024, 5-11
https://doi.org/10.7862/re.2023.x

THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

Original Research/Review

Modeling of the textronic structure using the Iterative Fundamental Solutions Method

Mariusz Mączka [2]*, Stanisław Pawłowski [1], Jolanta Plewako [3],

[1] Department of Electrodynamics and Electrical Machine Systems, Faculty of Electrical and Computer Engineering, Rzeszow University of Technology, 35-959 Rzeszow, Poland; spawlo@prz.edu.pl.
[2] Department of Electronics Fundamentals, Faculty of Electrical and Computer Engineering, Rzeszow University of Technology, 35-959 Rzeszow, Poland; mmaczka@prz.edu.pl.
[3] Department of Power Electronics and Power Engineering, Faculty of Electrical and Computer Engineering, Rzeszow University of Technology; 35-959 Rzeszow, Poland, jplewako@prz.edu.pl.

*Corresponding author. mmaczka@prz.edu.pl

Received: 30 June 2023 / Accepted: 06 February 2024 / Published online: 13 February 2024

Abstract

The paper describes an approach to modelling the conductance of a textronic structure based on the iterative fundamental solutions method. The analysis of the current density distribution in a thin conductive layer containing roughness resulting from the applied manufacturing technology is aimed at estimating its impact on the total current and the conductivity of the conductive path. The simulations showed that the current density distribution in a conductive path depends on the nature of its surface, and increasing its roughness reduces its conductance. The proposed solution makes it possible to define a structure model in three geometrical dimensions, and its numerical implementation in the form of the proposed method ensures efficiency and computational accuracy.

Keywords: Fundamental Solutions Method, Textronic Structure

1. Introduction

Textronic structures are a complex class of materials that combine the features of electronic devices and textile structures. Due to their unique properties, such as high mechanical strength, flexibility, and electrical conductivity, these materials are widely used in areas such as electronics, medicine, automation, and textiles [1].

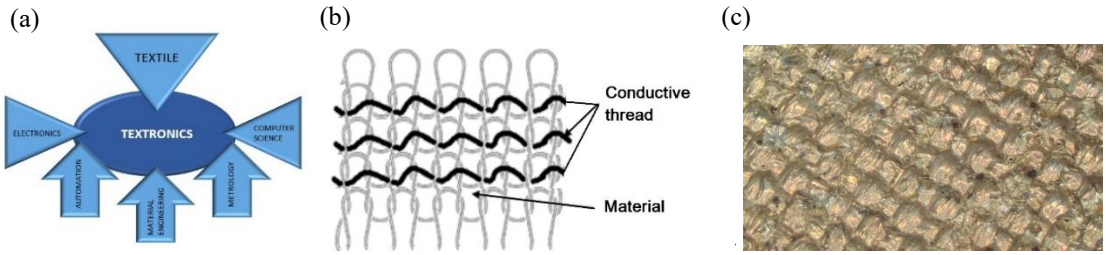

Fig. 1. The textronics area (a) [4] and typical realizations of textronic structures: conductive threads intertwined between the fibers (b) [4], conductive surfaces sputtered onto the textile material (c)

Textronic products are currently used most frequently in uniformed and rescue services and as everyday items [2]. From the textile technology side, textronic systems require the use of materials such as fibers, threads, electrically conductive fabrics, piezoelectric fabrics, magnetic fibers, optical fibers

This is an Open Access article distributed under the terms of the CC-BY-NC-ND 3.0 PL license, which permits others to distribute the work, provided that the article is not altered or used commercially. You are not required to obtain permission to distribute this article, provided that the original work is properly cited.

and textiles with shape memory as well as electroactive polymers. As electronic systems, these structures should show adequate accuracy and repeatability of response to control signals as well as resistance to external factors such as humidity or temperature. The implementation of such structures can be achieved by interlacing thin conductive threads between the fibers of the textile material [3] or using the physical vapour deposition (PVD) method [4]. Based on textronic structures, sensors and wearable electronic devices are built in the form of clothing, jewelry, watches, wristbands, glasses and others. Thanks to them, various physiological, environmental or behavioral parameters of the user, such as heart rate, blood pressure, body temperature, blood, glucose level, body position, movements, physical activity or location, can be monitored in real time [5-8]. The basis for the continuous development of the above technologies is the ability to implement new solutions using computer models that optimize project costs. In this area, numerical models dedicated to structures composed of thin conductive threads are known [9-11], but there is a still deficit of models intended for layered structures made using the PVD method. Existing models [12-13] assume infinitely thin conductive layers, limiting the model geometry to 2 dimensions. This article is the result of work aimed at expanding the possibilities of simulation in the above-mentioned area

The aim of this work is to develop a three-dimensional textronic structure model (3D TSM) realised by using the PVD method. Analysis of the distribution of current density in a thin conductive layer that contains narrowing resulting from the gravitational deposition of conductive particles on the fabric aims to estimate its impact on the total current and conductance of the conductive path.

2. Formulation of the problem

A model of a conductive layer deposited on a textile substrate is considered, which is illustrated in Fig. 2. The analysed area Ω is limited by six surfaces marked as S_1, …, S_6. Surfaces S_1, …, S_4 are flat, and the surfaces S_5 and S_6 spaced d apart have periodically distributed ridges and depressions. Its shape and arrangement can generally be different and depend on the type of textile fabric weave that is modelled. In this work, it was assumed to model them using the sine functions with a given amplitudes A_1 and A_2 for S_5 and S_6 surfaces respectively. These functions are mutually perpendicular and run in directions parallel to the surface of the fabric.

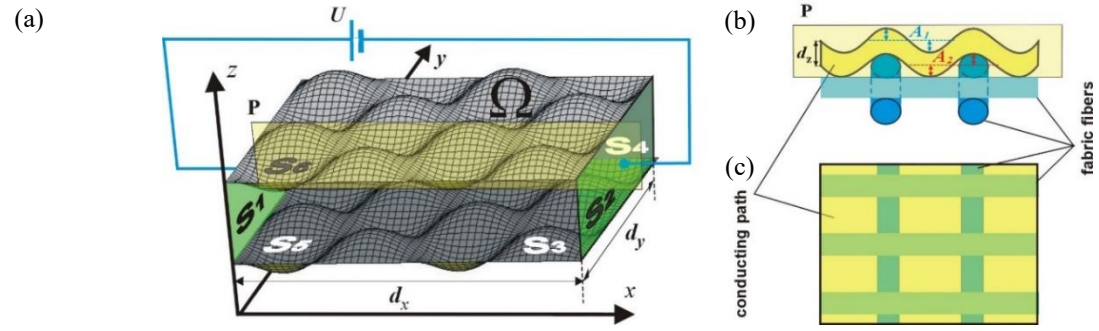

Fig. 2. Simplified concept of the 3D Textronic Structure Model (3D TSM): boundary surfaces and voltage polarization system (a); view of the model in a cross-section with plane P (b) and top view of the model (c)

The mathematical description of the area modelling the conductive layer on the textile substrate is presented by the relations:

$$\Omega:\begin{cases} 0 \le x \le d_x \\ 0 \le y \le d_y \\ A_2(1+\sin k_x x \sin k_y y) \le z \le A_1(1+\sin k_x x \sin k_y y)+d_z \end{cases} \tag{1}$$

It is assumed that the area Ω is filled with a homogeneous, isotropic and linear conductor ($\gamma = \text{const}$), in which there are no unbalanced electric charges ($\rho = 0$), its surroundings are an ideal dielectric ($\gamma_0 = 0$) and that there is a constant voltage U between the surfaces S_1 and S_2 (see Fig. 2).

For the assumptions presented above, the problem of calculating the current flow field boils down to searching for the distribution of electric potential defined as:

$$\mathbf{E} = -\operatorname{grad}\varphi \tag{2}$$

where $\vec{E}$ is the electric field strength. The potential φ meets the Laplace equation:

$$\Delta\varphi = 0 \tag{3}$$

and mixed boundary conditions:

$$\varphi = U \text{ on the surface } S_1 \tag{4}$$

$$\varphi = 0 \text{ on the surface } S_2 \tag{5}$$

$$\frac{\partial\varphi}{\partial n} = 0 \text{ on the surface } S_i,\ i = 3, 4, 5, 6 \tag{6}$$

After solving the problem formulated above, the current density distribution sought is determined on the basis of (2) and the local Ohm's law:

$$\mathbf{J} = \gamma\,\mathbf{E} \tag{7}$$

and the current from the dependence:

$$I = \iint_S \mathbf{J}\cdot \mathrm{d}\,s\,, \tag{8}$$

where s can be any section of the area Ω connecting the surfaces S_3 and S_4 (e.g. for x = const)

3. Solution of the problem

In order to solve the problem described by formulas (3) - (6), the fundamental solutions method (FSM) was applied [5]. In the next iteration steps, successive approximations of the sought potential function are calculated according to the formula:

$$\widetilde{\varphi}_k(\mathbf{r}) = \widetilde{\varphi}_{k-1}(\mathbf{r}) + \sum_{n=1}^{L_\mathrm{f}} q_{k,n} F_{k,n}(\mathbf{r}) \tag{9}$$

where: k – iteration step number,

$\mathbf{r} = (x, y, z)$ – any point within the Ω area,

$q_{k,n}$ – approximation sum coefficients calculated on the basis of boundary conditions,

$F_{k,n}(\mathbf{r}) = \dfrac{1}{|\mathbf{r} - \mathbf{r}_{k,n}|}$ – fundamental solution of the Laplace equation,

$\mathbf{r}_{k,n}$ – fixed points lying outside the Ω area (singular points of the function $F_{k,n}$),

L_f – the number of fundamental solutions considered in a single iteration step.

The function $\widetilde{\varphi}_0(\mathbf{r})$ was assumed (primary field):

$$\widetilde{\varphi}_0(\mathbf{r}) = U\left(1 - \frac{x}{d_x}\right) \tag{10}$$

This function satisfies equation (3) and boundary conditions (4) - (6) on the surfaces: S_1, S_2, S_3, S_4, which significantly accelerates the convergence of calculations. It should also be noted that regardless of the value of the coefficients $q_{k,n}$ each function described by the formula (9) satisfies equation (3) exactly, and the selection of points outside the Ω area $\mathbf{r}_{k,n}$ guarantees its finiteness in this area.

The values of the coefficients $q_{k,n}$ are determined in such a way as to obtain the best possible improvement of the fulfillment of the boundary conditions by the function $\widetilde{\varphi}_k(\mathbf{r})$ in a given iteration step in relation to the iteration calculated in the previous step. For this purpose, a measure of the accuracy of meeting the boundary conditions by the function $\widetilde{\varphi}_k(\mathbf{r})$ should be defined. On individual boundary surfaces of the area, functions of local, relative edge errors of the solution are:

$$\varepsilon_{1,k}(\mathbf{r}) = \frac{1}{U}\left(U - \widetilde{\varphi}_k(\mathbf{r})\right), \quad \mathbf{r} \in S_1 \tag{11}$$

$$\varepsilon_{2,k}(\mathbf{r}) = \frac{1}{U}\widetilde{\varphi}_k(\mathbf{r}), \quad \mathbf{r} \in \mathrm{S}_2 \tag{12}$$

$$\varepsilon_{i,k}(\mathbf{r}) = \frac{1}{E_0}\frac{\partial \widetilde{\varphi}_k}{\partial n}, \quad \mathbf{r} \in \mathrm{S}_i, \quad i = 3,...,6 \tag{13}$$

where: $E_0 = \dfrac{U}{d_x}$ is the primary electric field strength (see eq. (10), (2)), and *the boundary error functional*:

$$\delta_k = \sqrt{\sum_{i=1}^{6} \frac{1}{S_i} \iint_{\mathrm{S}_i} \varepsilon_{i,k}^2(\mathbf{r})\,\mathrm{d}s} \tag{14}$$

which is a measure, in the sense of the mean square norm, of jointly meeting the boundary conditions (4) - (6) by solution (9). Postulating the minimization of this functional with respect to the set of parameters $q_{k,n}$

$$\frac{\partial \delta_k}{\partial q_{k,m}} = 0, \quad m = 1,..., L_\mathrm{f} \tag{15}$$

a linear system of equations is obtained:

$$A_{k,m,n} q_{k,n} = B_{k,m}, \quad m, n = 1,..., L_\mathrm{f} \tag{16}$$

whose coefficients are expressed by dependencies:

$$A_{k,m,n} = \sum_{i=1}^{2} \frac{1}{S_i} \iint_{\mathrm{S}_i} F_{k,m}(\mathbf{r}) F_{k,n}(\mathbf{r})\,\mathrm{d}s + d_x^2 \sum_{i=3}^{6} \frac{1}{S_i} \iint_{\mathrm{S}_i} G_{k,m}(\mathbf{r}) G_{k,n}(\mathbf{r})\,\mathrm{d}s \tag{17}$$

$$B_{k,n} = -U\left(\sum_{i=1}^{2} \frac{1}{S_i} \iint_{\mathrm{S}_i} \varepsilon_{i,k-1}(\mathbf{r}) F_{k,n}(\mathbf{r})\,\mathrm{d}s + d_x \sum_{i=3}^{6} \frac{1}{S_i} \iint_{\mathrm{S}_i} \varepsilon_{i,k-1}(\mathbf{r}) G_{k,n}(\mathbf{r})\,\mathrm{d}s \right) \tag{18}$$

$$G_{k,n}(\vec{r}) = \frac{\partial F_{k,n}}{\partial n} = \hat{n} \cdot \operatorname{grad} F_{k,n}(\mathbf{r}) \tag{19}$$

where $\hat{n}$ is the normal to the surface S_i at $\mathbf{r}$ point.

After numerical solution of the system (16) (e.g. by Gaussian elimination method), the sought set of coefficients $q_{k,n}$ are obtained and the next iteration step follows. In each step, the local error functions (11) - (13) and the boundary error function are calculated, which allows you to control the rate of convergence of the procedure on an ongoing basis and automatically interrupt it when the required accuracy is achieved. Using the basic approximation theorem on the existence and uniqueness of the solution to the linear approximation problem [6], it can be shown that:

$$\delta_k \le \delta_{k-1} \tag{20}$$

which ensures the convergence of the described procedure.

It should be noted that the choice of functions $F_{k,n}$ in (9) boils down to determining their singular points $\mathbf{r}_{k,n}$. As follows from (20), this choice does not affect the exact fact of convergence of the procedure, but the rate of convergence depends on it. There are no general rules on how to make such a choice optimally (in principle, you can create a procedure that would find such an optimal set of points $\mathbf{r}_{k,n}$ in each iteration step, but it is unprofitable from the point of view of the speed of convergence of the method in real time). In the numerical application created to solve the problem formulated here, in each iterative step, these points are randomly selected from many pregenerated points, uniformly surrounding the Ω area.

The convergence rate of the proposed method also depends on the number of L_f of the fundamental solutions considered in each iteration step. It can be set freely - from $L_f = 1$ to a value limited only by the capacity of the computer's operating memory, which must fit the coefficients of the system of equations (16). The greater the number L_f, the smaller the number of iteration steps needed to obtain the required accuracy and the smaller the total number of fundamental solutions in the final solution (total length of the approximation sum). However, it should be noted that the value of L_f greatly affects the time of a single iteration step, which is mainly influenced by the need to calculate surface integrals occurring in (17) and (18). Numerous numerical experiments by the authors have shown that in many cases the adoption of $L_f = 1$ gives the fastest convergence of the procedure in real time. In this case, the need to numerically solve the system of equations (16) is also avoided.

A certain drawback of the described procedure (e.g. in relation to the classic, collocation version of MRF) is the need to numerically calculate many surface integrals (cf. formulas (14), (17), (18)). This difficulty can be significantly alleviated by using the proper mean square norm, the so-called *pseudonorm*, i.e. assuming the definition of the boundary error functional as:

$$\delta_k = \sqrt{\sum_{i=1}^{6} \frac{1}{L_i} \sum_{j=1}^{L_i} \varepsilon_{i,k}^2(\mathbf{r}_j)} \tag{21}$$

where $\mathbf{r}_j$ are points distributed densely and more or less evenly on the boundary surfaces of the Ω area. With this approach, formulas (17), (18) take the form:

$$A_{k,m,n} = \sum_{i=1}^{2} \frac{1}{L_i} \sum_{j=1}^{L_i} F_{k,m}(\mathbf{r}_j) F_{k,n}(\mathbf{r}_j) + d_x^2 \sum_{i=3}^{6} \frac{1}{L_i} \sum_{j=1}^{L_i} G_{k,m}(\mathbf{r}_j) G_{k,n}(\mathbf{r}_j) \frac{1}{S_i} \tag{22}$$

$$B_{k,n} = -U\left(\sum_{i=1}^{2} \frac{1}{L_i} \sum_{j=1}^{L_i} \varepsilon_{i,k-1}(\mathbf{r}_j) F_{k,n}(\mathbf{r}_j) + d_x \sum_{i=3}^{6} \frac{1}{L_i} \sum_{j=1}^{L_i} \varepsilon_{i,k-1}(\mathbf{r}_j) G_{k,n}(\mathbf{r}_j) \right). \tag{23}$$

That is, all integrals are replaced by simple sums. The use of a *pseudonorm* instead of a proper norm is formally a less strict approach, because in this case the minimisation of the boundary error applies to a discrete set of boundary points, not to the entire surface. However, it should be noted that any numerical procedure for computing integrals is burdened with a similar inaccuracy because it must also be based on a discrete, "reasonable" representation of the integrand function [7].

2. Simulation results

Using the TSM, the current density distribution in the conductive path of the tested structure was simulated and the examples results are presented in figures 3-4.

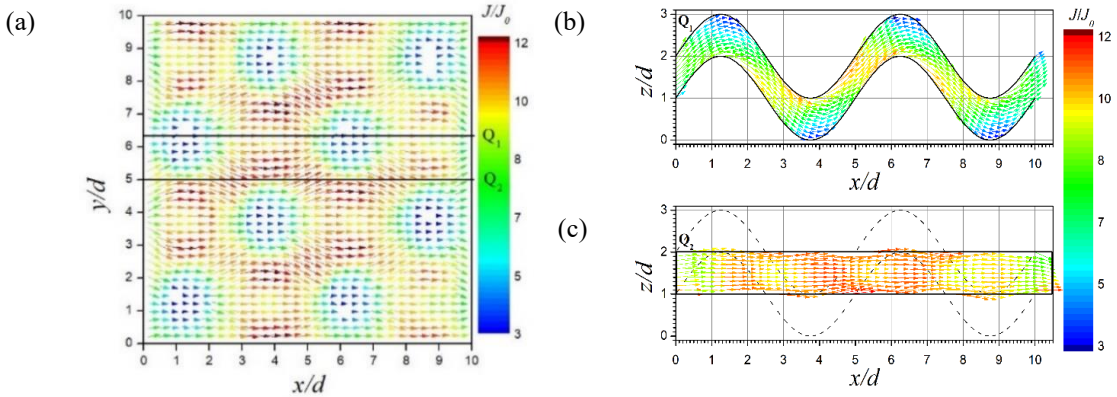

Fig. 3. The current density distribution in the form of a vector field for the 3D TSM illustrated in Fig. 2. Chart (a) illustrates the results in the top surface of the conductive path (S6), however, charts (b) and (c) in the XZ plains for $y/d = 6.25$ (Q_1) $y/d = 5$ (Q_2), respectively

Fig. 3 (a) illustrates the current density distribution in the form of a vector field on the top surface

of the conductive path (S_6). The lengths of the arrows represent the moduli of the current density vectors calculated as the ratio of the current density J at a given point to the current density J_0 at the same point in the case of a completely flat conductive path. If we compare these results with the graph shown in Fig. 3 (b) which illustrates the same vector field in the XZ plane for $y/d = 6.25$ (see line Q_1 in Fig. 3 (a)), we can see that the current flow is concentrated along the shortest path between the hills of the conductive surface. This nature of the current flow is confirmed by the results in Fig. 3 (c) showing the vector field of the current density in the XZ plane for the parameter $y/d = 5$. This plane does not intersect the hills of the conductive path and is rectangular in shape, which gives an effective cross-section for the current flow and makes the current density the highest here.

The results presented in Fig. 3 allow us to put forward the thesis that roughness on the surface of the conductive path is a significant obstacle to the current flow. To test this thesis, further calculations were carried out, the results of which are shown in Fig. 4.

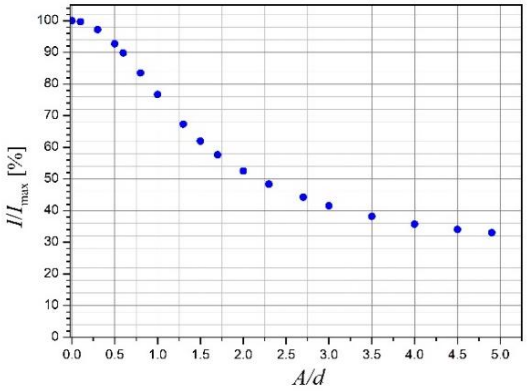

Fig. 4. Changes in the total current (I) of the TSM obtained by increasing the amplitude (A) of the conductive path roughness refer to its maximum value (I_{max}) corresponding to the flat surface of the path.

This figure shows changes in the total current (I) of the TSM obtained by increasing the amplitude ($A = A_1 = A_2$) of the conductive track roughness referring to its maximum value (I_{max}) corresponding to the flat surface of the track. The results confirm the above thesis, as can be seen for a fixed voltage U, as the amplitude of the conductive path roughness increases, the I/I_{max} ratio corresponding to the conductivity of the tested structure decreases. Qualitative confirmation of the above conclusions can be found in the results of measurements carried out in the work [17], where the authors note clear changes in resistivity caused by changes in temperature, which may be due to the smoothing of the conductive surface of the structure.

3. Conclusion

A 3D model of the textronic structure produced by the PVD method was developed. The applied Fundamental Solutions Method ensured accurate and quick calculations of the basic transport parameters of the tested structure. The simulations showed that the current density distribution in a conductive path depends on the nature of its surface, and increasing its roughness reduces its conductance. This gives inspiration for further research focused on the impact of the shape of the conductive path on the properties of ST transport parameters.

Author Contributions: Conceptualization, M.M, S.P., J.P.; methodology, S.P., M.M., J.P; software, S.P., M.M., J.P.; validation, J.P., M.M., S.P.; formal analysis, J.P., S.P., M.M.; investigation, M.M., J.P.; resources, S.P., M.M., J.P.; writing—original draft preparation, M.M. ,S.P.; writing—review and editing, J.P. All authors have read and agreed to the published version of the manuscript.

Literature

[1] Gniotek K., Stempień Z., Zięba J.: Tekstronika - nowy obszar wiedzy (in Polish), Przegląd Włókienniczy + Technik Włókienniczy, 2003, No 2, pp.17-18.

[2] Korzeniewska E., Walczak M., Rymaszewski J.: Elements of elastic electronics created on textile substrate, MIXDES-24thInternational Conference, IEEE, 2017, pp. 447-450.

[3] Jakubas A.: Badania i pomiary wybranych parametrów elektrycznych tekstylnych linii sygnałowych naniesionych metodą maszynową, Przegląd Elektrotechniczny, 2015, Vol. 91, No 12, pp.117-120.

[4] Łada-Tondyra E., Jakubas A.: Nowoczesne zastosowania systemów tekstronicznych (in Polish), Przegląd Elektrotechniczny, 2018, Vol. 94, No 12, pp. 198-201.

[5] Smith, A.A., Li, R. & Tse, Z.T.H. Reshaping healthcare with wearable biosensors. Scientific Reports 13, 4998 (2023). https://doi.org/10.1038/s41598-022-26951-z

[6] Lebioda, M., Rymaszewski, J.: Dynamic properties of cryogenic temperature sensors, Przegląd Elektrotechniczny 2015, 91, 2, 225-227, 10.15199/48.2015.02.51

[7] Tokarska M., Frydrysiak M., Zieba J., Electrical properties of flat textile material as inhomegeneous and anisotropic structure, Journal of Material Science, Materials in Electronics, (2013), 24, 5061–5068

[8] Ates, H.C., Nguyen, P.Q., Gonzalez-Macia, L., Morales-Narváez, E. Güder, F., Collins, J. J., Dincer, C., End-toend design of wearable sensors. Nature Reviews Materials 7, 887–907 (2022), doi.org/10.1038/s41578-022-00460-x

[9] Alonso-González, S. Ver-Hoeye, M. Fernández-García, C. Vázquez-Antuña and F. Las-Heras Andrés, "From Threads to Smart Textile: Parametric Characterization and Electromagnetic Analysis of Woven Structures", IEEE Access, vol. 7, pp. 1486-1501, 2019, doi: 10.1109/ACCESS.2018.2886041

[10] Surname1 N.: Exhaust System for Radial and Axial-Centrifugal Compressor with Pipe Diffuser. International Journal of Turbo and Jet Engines, 2016, Vol. 31, No 1, pp. 29-36.

[11] Dias, T., "Electronic Textiles: Smart Fabrics and Wearable Technology", 1st ed., Woodhead Publishing, Cambridge, UK, 2015)

[12] S. Pawłowski, J. Plewako, E. Korzeniewska, Przegląd Elektrotechniczny 2020, 96, 1 234-237.

[13] S. Pawłowski, J. Plewako, E. Korzeniewska, Electronics 2020, 9, 402.

[14] Kupradze V.D., Aleksidze M.A.: The method of functional equations for the approximate solution of certain boundary value problems (in Russian), USSR Computational Mathematics and Mathematical Physics, Vol. 4, 1964, pp. 82-126.

[15] Achiezer N. I.: Teoria aproksymacji (in Polish), PWN, Warszawa, 1957.

[16] Dryja M., Jankowska J., Jankowski M.: Przegląd metod i algorytmów numerycznych, (in Polish), PWN, Warszawa, 1982.

[17] Lebioda M., Korzeniewska E.: The Influence of Buffer Layer Type on the Electrical Properties of Metallic Layers Deposited on Composite Textile Substrates in the PVD Process. Materials 2023, 16, 4856. https://doi.org/10.3390/ma16134856

Advances in IT and Electrical Engineering

vol. 30, 2024, 13-24
https://doi.org/10.7862/re.2023.x

THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

Original Research/Review

Problems of estimating the uncertainty of water pH measurement

Anna Szlachta [1], Mykhailo Dorozhovets [1,2]*

[1] Department of Metrology and Diagnostic Systems, Faculty of Electrical and Computer Engineering, Rzeszow University of Technology, ul. Wincentego Pola 2A, 35-959, Rzeszów, Poland.
[2] Department of Information and Measuring Technology, National University Lviv Polytechnic, Lviv, Ukraine.

* Corresponding author. michdor@prz.edu.pl.

Received: 14 September 2023 / Accepted: 20 March 2024 / Published online: 25 March 2024

Abstract

The article analyses the main problems associated with evaluation the combined standard uncertainty of the water pH measurement by the type A and B methods. It is shown that, for a small number n of the tested water samples, the type A standard uncertainty determined by the conventional method is underestimated. Therefore, the correct expression to calculate this component of uncertainty is presented. The authors also highlighted that since in the practical measurement the influencing quantities and sensitivity coefficients are not known absolutely precisely, therefore their uncertainties often have to be taken into account when estimating the combined uncertainty. For this purpose the authors have propose their approach to correctly determine the type B components of combined standard uncertainty caused by not only the values of influencing quantities and sensitivity coefficients, but also their uncertainties. The proposed approaches are illustrated by estimating the uncertainty in the measurement of drinking water pH, presenting the corresponding components of measurement uncertainty budget.

Keywords: estimation, uncertainty, measurement, *pH*, water

1. Introduction

Drinking water, along with air, is the most important environmental factors that are a condition for life, and it has the most significant impact on human health [1], [2]. In general, life on Earth cannot exist without water. Water pollution harms human health, animals that consume water and plants that feed on water, as well as the flora and fauna of rivers, lakes, seas, and oceans, i.e., the entire biological world. Water is one of the main constituent elements in the production of agricultural products, seafood, food, medical devices, in a large number of chemical industry processes, and has an impact on the lifespan of various man-made structures, machinery, equipment, etc [1–6]. Therefore, water quality and its constant monitoring are of great social importance and are a prerequisite to ensuring an appropriate level of quality of life.

It is worth noting that the pH value of water from different sources is different. Bottled drinking water that uses reverse osmosis and ultraviolet (UV) and/or ozonation to kill organisms has a pH between 6.9 and 7.5, non-carbonated bottled mineral water has a pH between 7.1 and 7.5, while carbonated water has a pH between 5.3 and 6. Water from common household filters has a pH close to 7.5, the same as tap water, while boiled tap water has a slightly higher alkaline pH [7]

The paper shows that uncertainty in pH measurement depends on a number of factors, and discusses the key components of uncertainty that affect pH measurement.

One of the most important parameters that characterise the quality of water and other liquids and solutions is the so-called hydrogen pH [3–6]. The measurement of pH is based on the dependence of electrode potentials on the activity and concentration of hydrogen ions and is carried out by measuring

This is an Open Access article distributed under the terms of the CC-BY-NC-ND 3.0 PL license, which permits others to distribute the work, provided that the article is not altered or used commercially. You are not required to obtain permission to distribute this article, provided that the original work is properly cited.

the electromotive force of an electrometric cell, which is a set of corresponding galvanic transducers or electrodes, measuring and reference, immersed in the solution under study (here, water). Thus, to measure pH, measuring transducers are used - electrodes and a measuring device that measures the corresponding output electromotive force at the output of the electrodes [3–6].

As with other measurements, the quality of a pH measurement is determined by uncertainty [8–10]. The measurement of pH is one of the measurements that is characterised by a large number of uncertainty components of different nature. The most important uncertainty components in technical pH measurements are [8–10]:

1) uncertainty due to the heterogeneity of the measured solution, which can manifest itself in the instability of the pH measurement results of different samples of the same medium under study - this is a component of uncertainty associated with the object of measurement;
2) uncertainty due to calibration (often referred to as adjustment in the manufacturer's instructions of electrodes with buffer solutions;
3) the uncertainty component of the instrument readings and the conversion function of the measuring transducer (measuring and auxiliary electrodes) under reference measurement conditions;
4) uncertainty components caused by deviations from reference measurement conditions.

During precision pH measurements, other uncertainty components may also be taken into account, such as the component due to incomplete cleaning (rinsing) of the electrodes after the previous measurement, uncertainty due to the mismatch of the calibration conditions and the calibration procedure with reference conditions, discrete readings, a dynamic component (the result is recorded when the instrument has not yet fully settled), etc [6, 8–10] . These uncertainty components will not be discussed further, as they can be neglected in technical pH measurements.

An evaluation of the uncertainty of the pH measurement can be performed using a classical approach [11], [12], or Monte Carlo (MC) simulations [13] can be used. Commercial programmes dedicated to this purpose are also available. The paper [8] presents the advantages and disadvantages of the methods of calculating the uncertainty, i.e., the typical method, the MC method implemented in the software and the spreadsheet, and the commercial programmes; the paper considers the case of pH measurement after two-point calibration.

2. Problems in estimating pH measurement uncertainty

The main problems with determining the uncertainty of the pH measurement uncertainty are that there is no fully correct methodology for estimating the uncertainty, which would correctly calculate the uncertainty components of the Type A and Type B methods [12].

In particular, when calculating the uncertainty of the Type A method according to the accepted methodology [12] with a small number of observations, the resulting standard uncertainty is significantly underestimated, i.e., it is largely approximate. Note that although this problem is mentioned in another part of the [12], no methodology is proposed to correct the value of the standard uncertainty.

The main quantitative indicator of the heterogeneity of the pH of different samples of the same water is the unbiased estimate of the standard deviation, which is calculated using the known expression:

$$s_{pH} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(pH_i - \overline{pH}\right)^2} \tag{1}$$

where:

$$\overline{pH} = \frac{1}{n}\sum_{i=1}^{n} pH_i \tag{2}$$

is the average value of the results of n (typically n = 6... 12) of measurements (observations) of water samples: pH_1, pH_2, ..., pH_n. Based on the standard deviation estimate (1), the standard uncertainty of type A is traditionally estimated using the well-known expression [12]:

$$u_A(pH) = \frac{s_{pH}}{\sqrt{n}} = \sqrt{\frac{1}{n \cdot (n-1)} \sum_{i=1}^{n} \left(pH_i - \overline{pH}\right)^2} \quad (3)$$

However, formally, expression (3) reflects an estimate of the standard deviation of the mean, which is the measurement result. As has been shown [14], [15], according to the definition of uncertainty [12], the uncertainty is not related to the measurement result - the result (here the mean value of (2)) is known - but to the measured quantity - here pH (measurand). Therefore, the pH uncertainty should be determined according to the density of the distribution of possible pH values given the known result (mean) and an estimate of the possible spread, the standard deviation (1).

Since pH measurements are accompanied by a large number of influencing quantities, the calculation of uncertainty using the Type B method also encounters problems due to uncertainties in the values of the influencing quantities and influence factors. It is obvious that in practical measurements the value of the influencing quantity is not known exactly but with a corresponding uncertainty. The same applies to influence factors, which are also known only approximately. Therefore, these uncertainties for some variables should be taken into account when assessing measurement confidence. A detailed analysis of the problems associated with these aspects is provided in Chapter 5.

The standard DSTU 4077:2001 [9] formulates some of the main components of what it calls the measurement "error". This mainly refers to the influence on the measurement result of water sampling, electrode preparation, temperature influence, and correction of this influence.

It is obvious that the pH of drinking water depends on the source from which the water is taken, i.e., on the physicochemical and geological properties of the rocks and soils in which the source is located, as well as other factors, including the time of year, the amount and intensity of precipitation and the time that has passed since the end of precipitation, etc. This causes variability in the pH of drinking water, i.e. its different values at different times. In addition, as a result of these phenomena, various chemical and biological components and mechanical impurities may appear in the water, which change not only the pH of the water but also its composition. Therefore, when measuring the pH of water, it is also advisable to measure its composition. Such results would provide a more objective characterisation of water quality. In view of this, the method of water sampling and preparation according to DSTU 4077:2001 [9] is of great importance during measurements. This standard sets out the relevant requirements for the time intervals between sampling, sample preparation, and measurements, and recommends that water samples be treated, for example, by settling or filtering, in the event of significant contamination.

Failure to comply with these requirements can result in a significant deviation of the measurement results from the true pH value. Most importantly, subsequent estimates of the measurement uncertainty may not take into account factors related to the water itself, which is sampled from a specific source, under specific conditions, at a specific time.

The purpose of the research is to correctly assess the components of the standard uncertainty,, to find the total uncertainty and to present the uncertainty budget of water pH measurement and the measurement result in accordance with the requirements of the standard.

The physicochemical and other aspects of pH measurement are not considered here, and the studies themselves are performed on the basis of known metrological characteristics of the instruments used and known measurement conditions. Specific results relate to the determination of the pH measurement uncertainty of the manufacturer's drinking water: Lvivvodokanal in the Zolochiv direction in the village of Pluhiv. The measurements were made in accordance with: DSTU 4077:2001 [9], ISO 10523:1994 [10] and ISO 10523:2008 [11], and were carried out in the testing laboratory of VEMAKO LLC using a pH metre pH-150M [16], [17] and electrodes ESCL-08M (EKS-10610/7) [18].

3. Uncertainty from heterogeneity of water samples. Estimation of uncertainty using the Type A method

It should be noted once again that, according to the definition [12], measurement uncertainty is a parameter that characterises the dispersion of possible values of a measured quantity around the measurement result. Since all the information about the dispersion of possible values of the measured quantity is contained in the corresponding density of the distribution, the correct calculation of the standard (as well as other) uncertainty must be based on this distribution. If we assume a normal distribution model

for the observations (pH_1, pH_2, ..., pH_n) in [14], [15] was was shown that correct value of the Type A standard uncertainty of measurement is:

$$u_A(pH) = \frac{s_{pH}}{\sqrt{n}}\sqrt{\frac{n-1}{n-3}} = \sqrt{\frac{1}{n\cdot(n-3)}\sum_{i=1}^{n}\left(pH_i - \overline{pH}\right)^2} \tag{4}$$

This value directly relates to the t-Student distribution. When the number of observations is small, which is mainly the case for pH measurements, the values of (3) with [12] and (4) differ significantly. In particular, when the number of samples is $n = 6$, the value of (3) is underestimated by 29%, and when $n = 10$, it is underestimated by more than 13%. Only with a large number of observations of several dozen, which is practically impossible to obtain during pH measurements, are values (3) and (4) sufficiently close.

It should be noted that in other models of population density distributions, from which the observation is selected and which is described by two parameters, location and scale, when the standard uncertainty of the measured value is correctly calculated, the factor $\sqrt{(n-3)}$ appears in the denominator. This means that the minimum number of observations for which the standard uncertainty can be calculated correctly is $n = 4$. Therefore, methods for estimating pH measurement uncertainty should explicitly mention this minimum value, i.e., the number of samples should not be less than 4. For $n = 4$, the correct value of the standard uncertainty (4) is:

$$u_A(pH) = \sqrt{\frac{1}{4}\sum_{i=1}^{4}\left(pH_i - \overline{pH}\right)^2} \tag{5}$$

Water samples $n = 4$ were measured and the results obtained: $pH_1 = 7.06$, $pH_2 = 7.02$, $pH_3 = 7.01$, $pH_4 = 7.05$.

We take their average value as the result of the pH measurement (2):

$$pH_x = \frac{pH_1 + pH_2 + pH_3 + pH_4}{2} = 7.035 \text{ pH} \tag{6}$$

According to (4), a measure of the dispersion (heterogeneity) of pH values is their estimated standard deviation (1):

$$s_{pH} = \sqrt{\frac{1}{4-1}\sum_{i=1}^{n}\left(pH_i - \overline{pH}\right)^2} = 0.0238 \text{ pH}. \tag{7}$$

4. Estimating uncertainty using the Type B method under reference conditions

The second category of factors includes all those related to the digital meter and measuring electrodes, as well as the conditions of their use.

For the digital meter and electrodes, as well as for other measuring instruments, the manufacturer has established the maximum permissible errors of their readings $\pm\Delta_{MPE}$ under reference conditions and the coefficient of influence of temperature deviations from the reference range. Table 1 shows the maximum permissible errors of the device $\pm\Delta_{MPE,dm}$ and measuring electrodes $\pm\Delta_{MPE,el}$ under reference conditions, as well as $\pm\Delta_{MPE,cal}$ for electrode calibration.

Table 1: Maximum permissible errors of the digital evice and measuring electrodes [16], [17], [18]

MPE of digital meter:	MPE of measuring electrodes	
$\pm\Delta_{MPE,dm.}$	calibration $\Delta_{MPE,cal.}$	reference conditions, $\Delta_{MPE,el}$
±0.05 pH	± 0.01 pH	±0.02 pH

In general, the estimation of uncertainty using the Type B method for the use of measuring instruments under reference conditions does not cause problems, except for one. Since the estimation of uncertainty requires knowledge of the density of the distribution, this is the main problem for calculating

the relevant uncertainty component. The manufacturer usually does not provide information on the distribution of possible deviations of the device readings within $\pm\Delta_{MPE,dm}$. Therefore, according to the recommendation of the [12], focusing on the worst case, we assume a uniform distribution of these deviations. Then the component of the standard uncertainty of the instrument readings under reference measurement conditions for the known maximum permissible errors $\pm\Delta_{MPE,dm}$ of the instrument readings is calculated by the expression:

$$u_{dm,ref}\left(pH\right)=\frac{\Delta_{MPE,dm}}{\sqrt{3}}. \tag{8}$$

For the value of ±0.05 pH units from Table 1, the standard uncertainty associated with the measurement by the instrument under reference conditions is:

$$u_{dm,ref}\left(pH\right)=\frac{\Delta_{MPE,dm}}{\sqrt{3}}. \tag{9}$$

Similarly, we calculate the component of the standard uncertainty due to the inaccuracy of the conversion function of the measuring electrodes under reference conditions. Here, $\pm\Delta_{MPE,el}$ should be used (Table 1) and then the corresponding component of the standard uncertainty is:

$$u_{el,ref}\left(pH\right)=\frac{0.02\ pH}{\sqrt{3}}\approx 0.0115\ pH. \tag{10}$$

The uncertainty of electrode calibration with buffer solutions depends on the uncertainty $u\left(pH_{cal}\right)$ of the pH_{cal} values of the calibration solutions and on the non-compliance of the calibration conditions and the calibration procedure in the reference conditions. If the electrodes were calibrated under reference conditions, this uncertainty component is mainly determined by the uncertainty of the calibration solutions. In particular, if the maximum permissible error $\pm\Delta_{MPE,cal}$ (Table 1) of the pH_{cal} values of the calibration solutions is known, then, also assuming a uniform distribution, the corresponding component of the standard ~~measurement~~ uncertainty due to electrode calibration is:

$$u_{cal}\left(pH\right)=u\left(pH_{cal}\right)=\frac{0.01\ pH}{\sqrt{3}}\approx 0.0058\ pH. \tag{11}$$

5. Problems of estimating uncertainty components from influencing quantities

In fact, as mentioned earlier, pH measurements are measurements in which a large number of influencing factors should be taken into account. Table 2 shows the main influencing values, reference and operating conditions, and influence factors during pH measurements using a pH meter -150M with EKS-10610/7 electrodes [16], [17], [18].

Table 2: Influencing values, reference and operating conditions and influence factors (as a fraction of the basic maximum permissible error) during *pH* measurements using a *pH* meter -150M with electrodes EKC-10610/7 [16], [17], [18]

№	Influenced quantity	Operating conditions	Reference conditions	Sensitivity coefficient c_i
1	Temperature θ_m measuring medium (solution) under automatic temeperature correction	from -10°C to 100°C	from 15°C to 25°C	$c_{\theta m}$ =1.5
2	Outside temperature θ_{out} : for 10°C deviation from the reference range	from 5°C to 40°C	from 15°C to 25°C	$c_{\theta out}$ =1.5/10 °C
3	Outside relative humidity H_{rel}	from 90% to 25°C	30%÷80%	c_H =2.0
4	Device power supply voltage U_{ps}	from 198 V to 242 V	(207÷235) V	c_{Ups} =1.0

5	Resistance R_{me} of measuring electrode: the basic value of $R_{b.me}$ =500 MΩ	from 0 to 1000 MΩ	0 MΩ	c_{Rme}=1.0/500 MΩ
6	Resistance R_{re} of reference electrode: the basic value of $R_{b.ce}$ =10 kΩ	from 0 to 20 кΩ	0 кΩ	c_{Rre}=1.0/10 kΩ
7	AC voltage U_{re} (of 50 Hz frequency) in the circuit of the reference electrode	from 0 to 50 mV	0 mV	c_{Ure}=1.0/50 mV
8	DC voltage U_{s-g} in circuit of measuring solution - ground:	from -1.5 V to+1.5 V	0 V	c_{Ug-g}=1.0/1.5 V

Note 1: According to GOST 29322-92 in the current edition (2014) [19] for a 230V network, the maximum deviations (both positive and negative) in Ukraine should not exceed 10% of the nominal value, i.e. from 207 to 253 volts.

Note 2: Since the pressure value during *pH* measurements is almost always within the reference range of 84 to 106.7 kPa, the impact of this component of uncertainty is not assessed.

In Ukrainian national metrological practise [20], as a legacy from the previous system, normalisation and calculation of the uncertainty caused by an influence quantity v within the work area are most often carried out as a product of the corresponding influence factor on the standard uncertainty $u_{ref}(X)$ of the measurement under reference conditions. In this case, two options are used mainly. In the first case, the influence factor is constant; that is, the standard uncertainty component is calculated using a simple expression.

$$u_v(X) = c_v \cdot u_{ref}(X) \tag{12}$$

and does not depend on the actual value of the influencing quantity in the working area of values.

In the second case, the sensitivity coefficient has the dimension of sensitivity to the normalised deviation $\Delta v_{ref} = |v_{op} - v_{ref}|$ of the value of the quantity from the edge of the reference range v_{ref}:

$$c_{v,1} = \frac{c_v}{\Delta v_{ref}}. \tag{13}$$

Then, to calculate the corresponding uncertainty component, the deviation value Δv of the influencing quantity new from the edge of the reference range should be determined, and the standard uncertainty component is calculated by the expression:

$$u_v(X) = c_{v,1} \cdot \Delta v \cdot u_{ref}(X) = c_v \cdot \frac{\Delta v}{\Delta v_{ref}} \cdot u_{ref}(X). \tag{14}$$

In certain cases, the sensitivity factor has a sensitivity dimension to the so-called base value of the influencing quantity.

Analysing Table 2, we can note that the coefficients of influence of one part of the influencing quantities are constant, i.e., they do not take into account the actual deviation of the influencing quantity from the reference range. This applies, for example, to the temperature θ_m of the measuring medium (water) with automatic thermal compensation, the relative humidity of the environment, and the supply voltage of the device U_{ps}. For other influencing quantities, the respective influence factors take into account the corresponding normalised or reference value of the influencing quantity. In these cases, the actual value of the influencing quantity must be known to calculate the standard uncertainties.

As noted above, the manufacturers of the measuring instruments mostly provide only the values of the influence factors, as in Table 2, but do not provide their uncertainties. However, it is known that

even physical constants are characterised by uncertainty. Obviously, the values of the influence coefficients given in Table 2 are approximate, i.e., they also have uncertainty. GUM [12] draws attention to this situation. In particular, Section G.4 of [12] argues that the nonstatistical estimate of the standard uncertainty, i.e., the Type B method, is largely subjective, with values derived from scientific judgement based on the totality of available information. Often, the values that determine the uncertainty are themselves characterised by uncertainty of even a few tens of percent [12]. This approach to [12] is used to determine the so-called effective number of degrees of freedom. In this regard, based on [12], we will assume that the influence coefficients have a relative standard uncertainty of about 25-30%.

However, the value of an influential quantity is never known exactly. When measuring the influencing quantity, there is always a standard uncertainty of the measurement. Therefore, given the above, the first method of normalisation of (12) should take into account the uncertainty of the sensitivity coefficient, i.e., the standard uncertainty should be calculated using the expression:

$$u_v(X) = \sqrt{c_v^2 + u^2(c_v)} \cdot u_{ref}(X) = c_v \sqrt{1 + u_{rel}^2(c_v)} \cdot u_{ref}(X), \tag{15}$$

where $u_{rel}(c_v)$ is the relative standard uncertainty of the coefficient c_v.

In the second method of normalising (14), in addition to the uncertainty of the coefficient, the uncertainty of the influencing quantity should also be taken into account, i.e.:

$$u_v(X) = c_v \sqrt{1 + u_{rel}^2(c_v)} \cdot \frac{\Delta v_{dev} \sqrt{1 + u_{rel}^2(v)}}{\Delta v_{ref}} \cdot u_{ref}(X). \tag{16}$$

where $u_{rel}(v)$ is the relative standard uncertainty of the measured value *v*. This uncertainty can be estimated based on the characteristics of the instrument concerned, for example, the accuracy class, the measuring limit, and its indication. In some cases, it is estimated by calculation, based on an analysis of the measurement conditions or data from previous measurements.

It should be noted that, in certain cases, a seemingly paradoxical situation may arise when estimating the uncertainty components of the influential quantities. For example, let the width of the range of deviation of the influential quantity from the reference range be given as v_R. If the influential quantity is not measured, then only the expected value of the influential quantity, which is in the middle $v_o / 2$ of this range, with possible deviations $\pm v_R / 2$, is guided, then, assuming a uniform distribution and, for simplicity, neglecting the uncertainty $u_{rel}(c_v)$, in (16) we have:

$$u_v(X) \approx c_v \cdot \frac{\Delta v}{2 \cdot \Delta v_{ref}} \sqrt{1 + \frac{1}{3}} \cdot u_{ref}(X) = c_v \cdot \frac{\Delta v}{\sqrt{3} \cdot \Delta v_{ref}} \cdot u_{ref}(X). \tag{17}$$

On the other hand, if you want to improve the accuracy of measurements and for this purpose measure the influencing quantity and find that its value is close to the limit values: $|v| \approx v_{\lim}$. Then, even with a negligibly small standard measurement uncertainty of the influential quantity ($u(v) \to 0$), according to (16), the standard uncertainty of this influential quantity is:

$$u_v(X) \cong c_v \cdot \frac{\Delta v}{\Delta v_{ref}} \cdot u_{ref}(X), \tag{18}$$

which in $\sqrt{3}$ is larger than the standard uncertainty (17) without measuring this influential variable.

A priori, a one-sided triangular distribution of the deviation of the influencing quantity, for example, temperature, in a certain direction from the reference region [12] is more likely. Then in (17), instead of $\sqrt{3}$ there will be $\sqrt{6}$, i.e.:

$$u_v(X) = c_v \cdot \frac{\Delta v}{\sqrt{6} \cdot \Delta v_{ref}} \cdot u_{ref}(X). \tag{19}$$

Therefore, if the measured value of the influencing quantity is close to the limit values: $|v| \approx v_{\lim}$, then the resulting standard uncertainty will be greater $\sqrt{6}$ than the standard uncertainty without the measurement of the influencing quantity. Even if the value of the influence quantity is greater than approximately $0.4v_{\partial}$, the standard uncertainty will still be greater than the standard uncertainty without measuring the influence quantity.

The paradox is that in order to reduce the uncertainty, an additional measurement of the influencing quantity was performed, i.e., to reduce its uncertainty, but as a result, the measurement uncertainty component of this influencing quantity increased. The same effect can occur if additional research is performed to determine the actual value of the influence coefficient.

This situation can be explained by the fact that by assuming a uniform, triangular, or other distribution of the impact value (or the value of the impact coefficient) within the boundary values, we are performing an imaginary randomisation of these values, i.e. we assume that they are random. And the uncertainty estimate found with this approach is as expected when using a given type of measuring instrument under typical measurement conditions. This expected uncertainty can be called a priori. However, in a particular measurement, the value of the influencing quantity takes on a specific value (similarly, for a particular measuring instrument, the influence coefficient has a specific (though possibly unknown) value), and therefore the actual uncertainty of the measurement result when using a particular measuring instrument under specific conditions may differ from the expected one.

In general, to estimate the uncertainty caused by the influence of a quantity on the reading of a measuring instrument, one should proceed from a mathematical model:

$$\vartheta(v) = c_v \cdot v \,. \tag{20}$$

In this model, the influence coefficient c_v has a certain value c_k with a standard uncertainty $u(c_v)$, and the influencing quantity v has a value v_k with a standard uncertainty $u(v)$. Then, after performing the correction for systematic bias, the correction $p_k = -c_k v_k$, according to the requirements of the [12] and assuming the independence of the influence factor and the influencing quantity, the component of the standard uncertainty of the measurement due to this influencing quantity should be calculated by the expression:

$$u_v^2(X) \approx \sqrt{c_k^2 \cdot u^2(v) + v_k^2 u^2(c_v) + u^2(c_v) \cdot u^2(v)} \,. \tag{21}$$

Another factor that is not taken into account when estimating uncertainty components from influencing quantities is the influence of the time factor, i.e. changes in the properties (drift, "ageing") of measuring transducers, instruments, etc. after the last calibration. Unfortunately, manufacturers of pH measuring instruments do not provide any information on this issue.

6. Quantifying uncertainty from influencing quantities

The actual measurement conditions (including the values of the influencing quantities) are presented in Table 3. Standard measurement uncertainties of the influencing quantities will be evaluated using the known maximum permissible errors in the following analysis. The last two influencing quantities (AC voltage in reference electrode circuit and DC voltage in the water-ground circuit) are characterised by the so-called a priori uncertainty, i.e. calculated.

Table 3: Actual values of influencing quantities (pH measurement conditions)

№	Influence quantity	Operating conditions
1	Temperature θ_m measuring medium (solution) under automatic temeperature correction	13°C ±0.5°C
2	Outside temperature θ_{out}	29.0°C ±0.5°C
3	Outside relative humidity H_{rel}	70%±5%

4	Device power supply voltage U_{ps}	(226.0 ± 1.0) V
5	Resistance R_{me} of measuring electrode	≈750 MΩ
6	Resistance R_{re} of reference electrode	≈9 kΩ
7	AC voltage U_{re} (of 50 Hz frequency) in the circuit of the reference electrode	≈15 mV
8	DC voltage U_{s-g} in circuit of measuring solution - ground:	≈0.30 V

The temperature of test water $\theta_{\mathrm{m}} = (13 \pm 0.5)^{o}C$ (Table 3) is outside of the reference range $\theta_{\mathrm{ref}} = (20 \pm 5)^{o}C$, i.e. from 15ºC up to 25ºC. Therefore due to the 1st line of table 2, from which coefficient $c_{\theta_m} = 1.5$ (with neglected uncertainty), the standard uncertainty due to the deviation of the test water temperature from reference range is:

$$u_{\theta_m}(pH) = c_{\theta_m} \cdot u_{dm,ref}(pH) = 1.5 \cdot 0.0288 \text{ pH} = 0.0432 \text{ pH}. \tag{22}$$

The ambient (outside) temperature is measured: $\theta_{\mathrm{out}} = (29 \pm 0.5)^{o}C$ (Table 3) is outside of reference range $\theta_{\mathrm{ref}} = (20 \pm 5)^{o}C$, and temperature deviation is $\Delta\theta_{\mathrm{out}} = 29^{o}C - 25^{o}C = 4^{o}C$. Assuming triangle distribution inside MPE=±0.5°C, which causes standard uncertainty of temperature $u(\theta_{\mathrm{out}}) = 0.5^{o}C/\sqrt{6} \approx 0.20^{o}C$ or relative standard uncertainty is $u_{rel}(\theta_{\mathrm{out}}) = 0.20^{o}C/4^{o}C \approx 0.05$. This component of standard uncertainty can be neglected. According 2nd line of Table 2, from which coefficient $c_{\theta_{out}} = 1.5/10^{o}C$ (with neglected uncertainty) after substitution these values to (16) the standard uncertainty caused by deviation of the ambient temperature deviation from reference range is:

$$u_{\theta_{out}}(pH) = c_{\theta_{out}} \cdot \Delta\theta_{out} \cdot u_{dm,ref}(pH) = \frac{1.5}{10^{o}C} 4^{o}C \cdot 0.0288 \text{ pH} = 0.0173 \text{ pH}. \tag{23}$$

Because ambient humidity H_{rel}=70 % (Table 3) not deviate from the reference: 30%÷80%, therefore the component of standard uncertainty of ambient humidity deviations is not determined, i.e.:

$$u_H(\mathrm{pH}) = 0. \tag{24}$$

The supply voltage is (226.0± 1.0) V (Table 3) and reference range is (207 ÷253) V. Thus, the supply voltage is within the reference range, and for this reason, the component of the standard uncertainty due to the deviation of the device supply voltage is determined, i.e:

$$u_{Ups}(\mathrm{pH}) = 0. \tag{25}$$

The resistance of measuring electrode is $\mathrm{R_{me}} \approx 750$ MΩ (Table 3). According to 5th line of Table 2 the sensitivity coefficient c_{Rme}=1.0/500 MΩ for base value 500 MΩ. Therefore, the standard uncertainty component due to the effect of the value of the resistance of the measuring electrode is:

$$u_{Rme}(\mathrm{pH}) = c_{Rme} \cdot \frac{\mathrm{R_{me}}}{\mathrm{R_{b,me}}} \cdot u_{el,ref}(\mathrm{pH}) = 1.0 \cdot \frac{750\mathrm{M\Omega}}{500\mathrm{M\Omega}} \cdot 0.0115 = 0.0173 \text{ pH}. \tag{26}$$

The resistance of reference electrode is $\mathrm{R_{re}} \approx 9\,\mathrm{k\Omega}$ (Table 3). According to 6th line of Table 2 the sensitivity coefficient c_{Rre}=1.0/10 kΩ for the base value 10 kΩ. Therefore, the standard uncertainty component due to the effect of the auxiliary electrode resistance value is:

$$u_{Rre}(\mathrm{pH}) = c_{Rre} \cdot \frac{\mathrm{R_{re}}}{\mathrm{R_{b,re}}} \cdot u_{el,ref}(\mathrm{pH}) = 1.0 \cdot \frac{9\mathrm{k\Omega}}{10\mathrm{k\Omega}} \cdot 0.0115 = 0.0104 \text{ pH}. \tag{27}$$

The alternating voltage of frequency 50 Hz in the circuit of the reference electrode is $U_{re} \approx 15$ mV. For the limited value of 50 mV (7th line of Table 2) the sensitivity coefficient is 1.0/50 mV, so the standard uncertainty component due to the influence of the AC voltage in the reference electrode circuit is:

$$u_{Ure}(\mathrm{pH}) = c_{Ure} \cdot \frac{U_{re}}{U_{re,lim}} \cdot u_{el,ref}(\mathrm{pH}) = 1.0 \cdot \frac{15\ \mathrm{mV}}{50\ \mathrm{mV}} \cdot 0.0115 = 0.0035\ \mathrm{pH} \tag{28}$$

The DC voltage in the circuit test solution - ground is $U_{s-g} = 0.30$ V (Table 3). For the limited value of ±1.5 V (8th line of Table 2) the sensitivity coefficient is c_{Ug-g} =1.0/1.5 V 1, so the standard uncertainty component due to the influence of the DC voltage in the test solution - ground circuit is:

$$u_{Us-g}(\mathrm{pH}) = 1.0 \cdot \frac{|U_{s-g}|}{1.5\ \mathrm{V}} \cdot u_{el,ref}(\mathrm{pH}) = 1.0 \cdot \frac{0.3\ \mathrm{V}}{1.5\ \mathrm{V}} \cdot 0.0115 = 0.0023\ \mathrm{pH} \tag{29}$$

When calculating the total standard uncertainty of the technical result of the pH measurement of water, we assume that the uncertainty components are not mutually correlated, since they are caused by different factors, so the total (composite, combined) standard uncertainty of the *pH* measurement result is equal to the square root of the sum of the squares of all the components found above, i.e:

$$u_c(\mathrm{pH}) = \sqrt{\begin{aligned} &u_s^2(\mathrm{pH}) + u_{cal}^2(\mathrm{pH}) + u_{dm,ref}^2(\mathrm{pH}) + u_{nel,ref}^2(\mathrm{pH}) + u_{\theta m}^2(\mathrm{pH}) + u_{\theta out}^2(\mathrm{pH}) + \\ &+ u_H^2(\mathrm{pH}) + u_{Ups}^2(\mathrm{pH}) + u_{Rme}^2(\mathrm{pH}) + u_{Rre}^2(\mathrm{pH}) + u_{Ure}^2(\mathrm{pH}) + u_{Us-g}^2(\mathrm{pH}). \end{aligned}} \tag{30}$$

Since all the components of the uncertainty from the influencing quantities are normalised relative to the uncertainty of the digital meter and measuring electrode under reference conditions, the combined standard uncertainty (28) can be written in a different form:

$$u_c(\mathrm{pH}) = \sqrt{\begin{aligned} &u_s^2(\mathrm{pH}) + u_{cal}^2(\mathrm{pH}) + u_{dm,ref}^2(\mathrm{pH}) \cdot \left[1 + c_{\theta m}^2 + c_{\theta out}^2 \frac{\Delta\theta_{out}^2}{\Delta\theta_{ref}^2} + c_H^2 + c_{Ups}^2\right] + \\ &+ u_{el,ref}^2(\mathrm{pH}) \cdot \left[1 + c_{Rme}^2 \frac{R_{me}^2}{R_{b,me}^2} + c_{Rre}^2 \frac{R_{re}^2}{R_{b,re}^2} + c_{Ure}^2 \frac{U_{re}^2}{U_{re,lim}^2} + c_{Up-э}^2 \frac{U_{s-g}^2}{U_{s-g,lim}^2}\right] \approx 0.0644\ \mathrm{pH} \end{aligned}} \tag{31}$$

The expanded uncertainty $U_p(pH)$ of the technical pH measurement is calculated as the product of the combined standard uncertainty $u_c(\mathrm{pH})$ calculated above by the coverage factor k_p for a given confidence level *p*:

$$U_p(pH) = k_p \cdot u_c(pH). \tag{32}$$

Since the number of non-zero uncertainty components is 10 (two components are equals to zero: $u_H(\mathrm{pH}) = 0$ and $u_{Ups}(\mathrm{pH}) = 0$) and components of standard uncertainty have approximately similar values and are independent, the value of coverage factor can be taken as the corresponding quantile of the normal distribution [12] , i.e. for $p = 0.95$ $k_{0.90} \approx 1.96$ [12].

The calculated components are recorded in the *pH* measurement uncertainty budget table (Table 4).

Table 4. Uncertainty budget for water *pH* measurement.

N	Quantity, parameter, *q*	Quantity value	Type. (A or B)	PDF *p(q)*	Sensitivity coef. c_i	Stand. uncert. *u(q),* pH
1	Solution heterogeneity v_{pH}	± 0.06 pH	B	Triangle	1.0	0.0238
2	Electrode calibration, $pH_{cal.}$, MPE_{cal}, $\Delta_{cal,lim}$	± 0.01 pH	B	Uniform	1.0	0.00577

3	Electrode (reference conditions), MPE_{el}, $\Delta_{el,ref,lim}$	±0.02 pH	B	Uniform	1.0	0.0115
4	Digital meter (reference conditions), MPE_{dm}, $\Delta_{dm,ref,lim}$	±0.05 pH	B	Uniform	1.0	0.0288
5	Temperature of measuring solution	$\theta_m = 27^oC$ $\theta_{ref,lim}=25^oC$	B	Uniform	1.5	0.0432
6	Outside temperature	$\theta_{out} = 29^oC$, $\theta_{ref,lim}=25^oC$ $\Delta\theta_{out} = 4^oC$	B	Uniform	1.5/10°C	0.0173
7	Relative humidity	H=70 %, $H_{ref,lim}$=80 %	B	Uniform	1.0	0
8	Power supply.	(226.0±1.0) V $U_{ps.lim}$=207 V÷253 V	B	Triangle	1.0	0
9	Measuring electrode resistance o	R_{re}=750°MΩ, $R_{b.re}$=500°MΩ	B	Uniform	1.0	0.0173
10	Reference electrode resistance	R_{me}=9°kΩ, $R_{b.me}$=10°kΩ	B	Uniform	1.0	0.0104
11	AC voltage in circuit of reference electrode	U_{re}=15 mV $U_{re,lim}$=50 mV	B	Triangle	1.0	0.0035
12	DC voltage in circuit solution - ground	U_{s-g}=0.3 V $U_{s.g,lim}$=1.5 V	B	Triangle	1.0	0.00192
13	Combined standard uncertainty					0.0644
14	Expanded uncertainty(p=0,95, $k_{0,95}$=1,96)					0.126 pH
15	Relative combined standard uncertainty					0.915%
16	Effective number degrees of freedom, ν_{eff}					33
17	Expanded coefficient (coverage factor), k_p					2.035
18	Refined value of expanded uncertainty (p=0.95, ν_{eff}=33, k_p=2.035)					0.131 pH
19	Relative expanded uncertainty					1.86%
20	Result of measurement (short presentation,,): pH=(7.04±0.13) pH, p = 95, k_p=2.035					1.86%

7. Conclusion

Based on the analysis of the measurement conditions performed and the evaluation of uncertainty, we can confirm that pH measurements of drinking water refer to measurements in which multiple components of uncertainty must be taken into account.

These are measurements in which one of the important components is the uncertainty associated with the test object itself, namely, the uncertainty caused by the heterogeneity of the pH of different samples of drinking water under test from the same source.

The second important component of pH measurement uncertainty is due to the inaccuracy of the measurement tools (digital meter and electrodes) directly involved in the measurements.

In addition, there are a large number of uncertainty components in these measurements from the interactions of influencing quantities. Among them are influences from: temperature of tested water and outside temperature, humidity of air, voltage supply (in these measurements, the values of relative humidity of the air and value of voltage supply were within reference limits), values or resistances of measuring and reference electrodes, AC voltage in circuit of reference electrode and DC voltage in circuit test water – ground.

As a result of the evaluation of the uncertainty in the measurement of drinking water pH, the relative expanded uncertainty (confidence level 0.95) was found to be about 1.9%. This value can be accepted. In measurements with significant impact of influencing quantities and therefore required correction of their impact, because the correction is never perfect, for correct estimation of the uncertainty caused by these influences, in addition to the values of sensitivity coefficients, it is necessary to have the values of their uncertainty. Of course, the measurement uncertainty of the influencing quantities must also be estimated.

Literature

[1] Spellman FR. The Drinking Water Handbook. 3rd ed. Boca Raton: CRC Press; 2017
[2] Gray NF. Drinking Water Quality: Problems and Solutions. 2nd ed. Cambridge: Cambridge University Press; 2008
[3] Emerson process measurement. Manual: Theory and Practice of pH Measurement. Emerson process measurement. PN 44-6033/rev. D December 2010
[4] Nayla Hassan Omer. Quality Parameters - Science, Assessments and Policy. DOI: http://dx.doi.org/10.5772/intechopen.89657
[5] Alley ER. Water Quality Control. Handbook. Vol. 2. New York: McGraw-Hill; 2007
[6] Shah C. Which Physical, Chemical and Biological Parameters of Water Determine Its Quality? 2017
[7] Kulthanan K., Nuchkull P., Varothai S. The pH of water from various sources: An overview for recommendation for patients with atopic dermatitis. Asia Pacific. Allergy. 2013;3:155–160. doi: 10.5415/apallergy.2013.3.3.155.
[8] Wiora J, Wiora A. Measurement Uncertainty Calculations for pH Value Obtained by an Ion-Selective Electrode. Sensors (Basel). 2018 Jun 12;18(6):1915. doi: 10.3390/s18061915.
[9] ДСТУ4077:2001 (ISO 10523:1994, MOD) «Якість води. Визначення pH».
[10] Basic laboratory skills training guide: Measurement of pH. VAM Guide, LGC, VAM, 2001.
[11] EURACHEM/CITAC Guide Quantifying Uncertainty in Analytical Measurement. Third Edition. Editors S L R Ellison (LGC, UK) A Williams (UK), QUAM:2012. 133p.
[12] Evaluation of measurement data—Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100. 2008.
[13] JCGM 101:2008. Evaluation of measurement data—Supplement 1 to the 'Guide to the Expression of Uncertainty in Measurement'— propagation of distributions using a Monte Carlo method'. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML
[14] Dorozhovets, Forward and inverse problems of type A uncertainty evaluation, Measurement, Volume 165, 2020 (1-16), 108072.
[15] Dorozhovets M.., A. Szlachta. Uncertainties of the estimators and parameters of distribution in measurements with multiply observations. Measuring Equipment and Metrology, Volume 81, no.4, 2020. Pp. 1-8. DOI: https://doi.org/10.23939/istcmtm 2020.04.003/
[16] РУП «Гомельский завод измерительных приборов», Республика Беларусь, 246001, г. Гомель, Интернациональная, 49, тел. (10-375-232) 53-09-75, факс (10-375-232) 53-47-03, Email: zip@mail.gomel.bv, www.zipgomel.com.
[17] pH-МЕТР pH-150М. Руководство по эксплуатации: 1Е2.840.858РЭ. Библиотека Ладовед, OCR Войкин Ю. В., 2008 г.
[18] Электроды стеклянные комбинированные лабораторные ЭСКЛ-08М. Паспорт Ш2.840.696 ПС.
[19] GOST 29322-92 Standard voltages, 2014.
[20] Дорожовець М., Б.Стадник В.Мотало, В.Василюк, А.Ковальчик, Р.Борек: Електричні вимірювання. Підручник для студентів. Основи метрології і вимірювальна техніка. Том 2. Вдавництво НУ "Львівська політехніка", Львів, 2005.-654 с.

Advances in IT and Electrical Engineering

vol. 30, 2024, 25-32
https://doi.org/10.7862/re.2023.x

THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

Original Research/Review

Analysis of the harmonic structure of the vowel /a/ taking into account the age and gender of the speaker

Szymon Drywa [1*]

[1] 1st High School with Bilingual Departments named after Hieronim Derdowski in Kartuzy

* Corresponding author. szmekdrywa@wp.pl

Received: 15 February 2024 / Accepted: 22 August 20224 / Published online: 19 September 2024

Abstract

Sound waves are disturbances propagating through an elastic medium that, upon reaching the ear, elicit auditory sensations. Sounds generated by the surroundings can be captured by a transducer (microphone), which transforms them into an electrical signal. The signal from the microphone is then transmitted to a computer, where software allows for the extraction and analysis of individual tones. This process enables the description of psychoacoustic characteristics of sound as perceived by the human ear, including pitch, loudness, and timbre. Specific sound signals can be transformed into a sound spectrum, facilitating the examination of voice harmonics, which are integer multiples of fundamental tone (lowest voice frequency). Harmonics possess various attributes, including intensity and frequency (loudness and pitch). This article provides an overview of the human voice and related topics, along with insights from studying differences in psychoacoustic features of sound based on gender and age. The objective of the paper is to show harmonic structure of vowels and emphasize the importance of studying the human voice.

Keywords: sound wave, intensity, harmonic frequencies

1. Introduction

The formation of the human voice is an exceedingly complex process that relies on various organs and mechanisms. We can distinguish two stages in the creation of the voice. The first stage is phonation, which occurs in the larynx in conjunction with the respiratory system [3]. This process is responsible for producing the fundamental frequency, which is the primary acoustic signal generated by the vocal cords. The next phenomenon is articulation, where the voice gains resonance and amplification. Articulation takes place through sound-shaping resonators located, among other places, in the nasal and oral cavities. The specific sites responsible for voice production and the corresponding processes are illustrated in Figure 1. The sound produced within the human body is interpreted as a sound wave.

2. Sound wave

A sound wave is a disturbance involving the transfer of mechanical energy due to the vibrations of particles in a medium, such as air. Particles undergo compression and rarefaction. The number of these compressions and rarefactions occurring in one second is referred to as the frequency of the wave. The human ear can detect a wide range of sound frequencies, from 16 Hz to even 20 kHz [5]. Values lower and higher than this range are respectively known as infrasound and ultrasound, with both being inaudible to humans. Another characteristic of a sound wave is its velocity. The physicist who first experimentally determined the speed of sound was M. Mersenne [1]. His measurement in 1636 indicated a speed of approximately 450 m/s, differing by around 120 m/s from modern measurements. Subsequent

This is an Open Access article distributed under the terms of the CC-BY-NC-ND 3.0 PL license, which permits others to distribute the work, provided that the article is not altered or used commercially. You are not required to obtain permission to distribute this article, provided that the original work is properly cited.

derivation of the formula for the speed of sound by Newton, and its refinement by Laplace, facilitated further developments in acoustics. Newton's-Laplace formula (1):

$$u = \sqrt{\frac{kp_o}{\varrho}}(1 + at) \tag{1}$$

In this formula:
u – the velocity of sound,
k – a constant,
p_o – the initial pressure of the medium,
ρ – the density of the medium,
a – another constant,
t – the time.

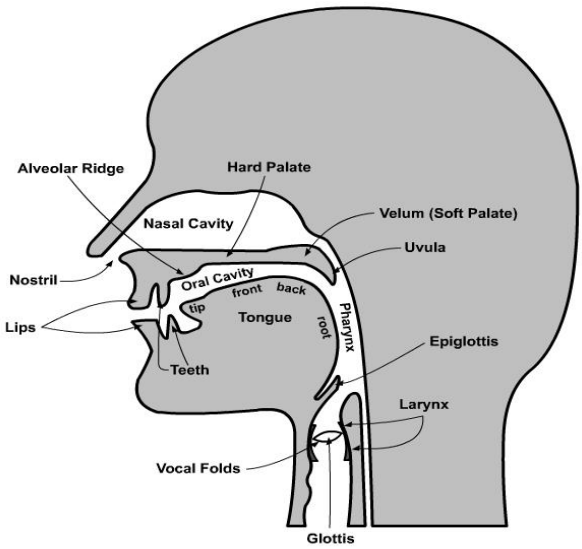

Fig. 1. Speech organs https://alic.sites.unlv.edu/chapter-11-2-speech-organs/ [9]

Measurement methods, including the use of an oscilloscope, have allowed for even more precise measurements of the speed of sound. By knowing the wavelength and frequency of the wave, it was possible to calculate its velocity according to formula (2):

$$u = \upsilon\lambda \tag{2}$$

In this formula:
u – the velocity of sound,
v – the frequency of the sound wave,
λ – the wavelength.

3. Psychoacoustic Characteristics of Sound

3.1. Intensity

Sound is measured as a sound pressure level (SPL), in units of decibels, with a sound level meter (SLM). A decibel is a ratio between a measured level and a reference level [8].

Sound intensity is the power carried by a wave per unit area, as defined by formula (3):

$$I = \frac{P}{S} \tag{3}$$

In this formula:
I – sound intensity,
P – the power of the sound wave,
s – the surface area.

Sound intensity can be defined in various ways, all of which are related to acoustic pressure and the velocity of the sound wave. The human ear can detect a wide range of sound intensities. To improve understanding and the clarity of results, the concept of sound intensity level, determined by equation (4), was introduced [4]:

$$L = 10lg\frac{I}{I_0} \tag{4}$$

In this equation:
L – the sound intensity level,
I – the sound intensity being measured,
I_0 – the reference intensity (the threshold of hearing, typically 10^{-12} watts per square meter).

3.2. Pitch

Pitch is dependent on the frequency of vibrations. In music, sequences of pitch define melody, and simultaneous combinations of pitch define harmony [7]. Lower frequencies result in lower pitch, while higher frequencies lead to higher pitch. Frequency is expressed in hertz, which corresponds to one vibration per second. The name of this unit originates from the discoverer of electromagnetic waves, the German physicist Heinrich Hertz (1857-1894).

The human ear is sensitive to a wide range of sound frequencies [2]. Different individuals can hear different frequency ranges, and this can depend on age and lifestyle. To assess one's hearing ability, audiometric tests can be conducted to determine the values of sound intensity and pitch at which one stops hearing.

3.3. Timbre

This is an individual characteristic of each person, setting us apart from others in society. It allows us to distinguish between voices with the same intensity, frequency, and duration. The timbre of a human sound is variable and can be modulated, among other methods, through adjustments in the larynx and the activation of specific resonant spaces. Physically, timbre refers to the number of harmonic tones and their amplitudes. In other words, timbre is dependent on the structure of the acoustic spectrum (the chart of component tones of the sound based on their pitch).

4. Psychoacoustic Characteristics of Sound

The experiment involved 100 teenagers in the age range of 14 to 17 years. Each participant was tasked with reading aloud the appearing vowels displayed on a computer screen, continuously, for about 3 seconds. After collecting the data, individual vowels were isolated from all the audio files. Subsequently, using the Sound Laboratory program at the Gdańsk University of Technology, sound spectra were created for these vowels, from which harmonic frequencies were extracted for further analysis. For certain groups, the sound intensity level (volume) was also recorded. All the collected data was used to generate tables, charts, and formulate conclusions.

5. Men - Conclusions

The collected data was grouped and presented in tables. Subsequently, for each vowel, the average harmonic frequency at the point of spectrum energy increase was determined, labelled as H[N]. In the next step, the ratio of the average frequency and a given N (representing the sequential number of the harmonic in the spectrum) was calculated. This provided the value H[N]/N, which is crucial for data interpretation and further analysis. Some of the results are presented in Tables 1-3.

Table 1. Average of the harmonic frequencies for vowel /a/ for the 2007-2008 cohort

Gender	Year of birth	Vowel	N	H[N]	H[N]/N
M	2007/2008	A	1	111.583	111.583
M	2007/2008	A	2	224.750	112.375
M	2007/2008	/a/	3	337.000	112.333
M	2007/2008	/a/	4	448.916	112.229
M	2007/2008	/a/	5	562.166	112.433

Table 2. Average harmonic values for vowel /a/ for the 2006-2007 cohort

Gender	Year of birth	Vowel	N	H[N]	H[N]/N
M	2006/2007	/a/	1	101.666	101.666
M	2006/2007	/a/	2	202.777	101.388
M	2006/2007	/a/	3	304.000	101.333
M	2006/2007	/a/	4	405.666	101.416
M	2006/2007	/a/	5	509.222	101.844

Table 3. Average harmonic values for vowel /a/ for the 2005-2006 cohort

Gender	Year of birth	Vowel	N	H[N]	H[N]/N
M	2005/2006	/a/	1	109.111	109.111
M	2005/2006	/a/	2	215.666	107.833
M	2005/2006	/a/	3	332.777	110.925
M	2005/2006	/a/	4	441.222	110.305
M	2005/2006	/a/	5	548.111	109.622

The obtained values of H[N]/N indicate that the study group consisted of males, as there are regular increments in energy around every 101-112 Hz. Values of fundamental frequency (H[1], also in the literature written as F0) are very similar as those in the article [11]. D. R. Feinberg, B. C. Jones, A. C. Little, D. M. Burt & D. I. Perrett received values of H[1] for men aged 20-22 equal 124.6 ± 20.5 Hz. In the research conducted within the age range of 2005-2008, there don't appear to be significant differences in the average harmonic values. For instance, the data in Table 2 is approximately 11 Hz lower than the data in Chart 1.

To provide a clearer illustration of the results, a graph (Fig. 2) was created to show the relationship between the average sound intensity level and the average harmonic frequencies for male participants for vowel /a/ (2007-2008 cohort). Figure 2 also maintains an increasing trend, although it doesn't increase uniformly.

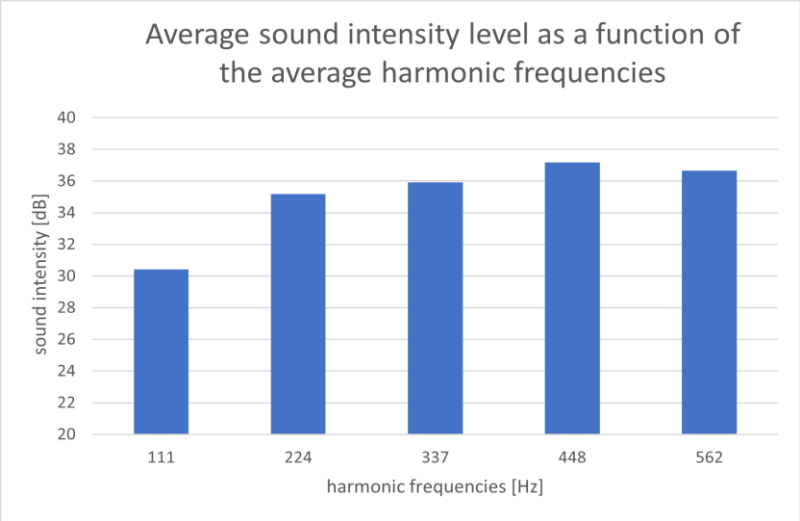

Fig. 2. Average sound intensity level as a function of the average harmonic frequencies for male participants for vowel /a/, 2007-2008 cohort.

6. Women - Conclusions

The next group of participants examined were women. The study was conducted in the same manner as for men. The data was collected and organized into tables. Sample results for vowel /a/ are presented in Tables 4-6.

Table 4. Average harmonic values for vowel /a/ for the 2007-2008 cohort

Gender	Year of birth	Vowel	N	H[N]	H[N]/N
W	2007/2008	/a/	1	219.857	219.857
W	2007/2008	/a/	2	439.071	219.535
W	2007/2008	/a/	3	662.607	220.086
W	2007/2008	/a/	4	884.035	221.008
W	2007/2008	/a/	5	1102.786	220.557

Table 5. Average harmonic values for vowel /a/ for the 2006-2007 cohort

Gender	Year of birth	Vowel	N	H[N]	H[N]/N
W	2006/2007	/a/	1	208.500	208.500
W	2006/2007	/a/	2	416.500	208.250
W	2006/2007	/a/	3	625.500	208.500
W	2006/2007	/a/	4	829.500	207.375
W	2006/2007	/a/	5	1038.250	207.650

Table 6. Average harmonic values for vowel /a/ for the 2005-2006 cohort

Gender	Year of birth	Vowel	N	H[N]	H[N]/N
W	2005/2006	/a/	1	199.000	199.000
W	2005/2006	/a/	2	400.500	200.250
W	2005/2006	/a/	3	602.312	200.770
W	2005/2006	/a/	4	801.625	200.406
W	2005/2006	/a/	5	1003.313	200.662

For women, the ratio H[N]/N ranges from 199 Hz to even 221 Hz, which is nearly twice as high as the values for male harmonics. The presented results demonstrate that human voices differ based on gender. The energy clusters in the female voice spectrum are shifted towards higher frequencies significantly more than in male voices. Another noticeable difference in male and female voices is the varying relationship between the sound intensity level and frequency. The graph of female data shown in Fig. 3 significantly differs from the male graph (Fig. 2). A female voice is characterized by a higher sound intensity level than a male voice, and the sound intensity level graph does not exhibit an increasing trend.

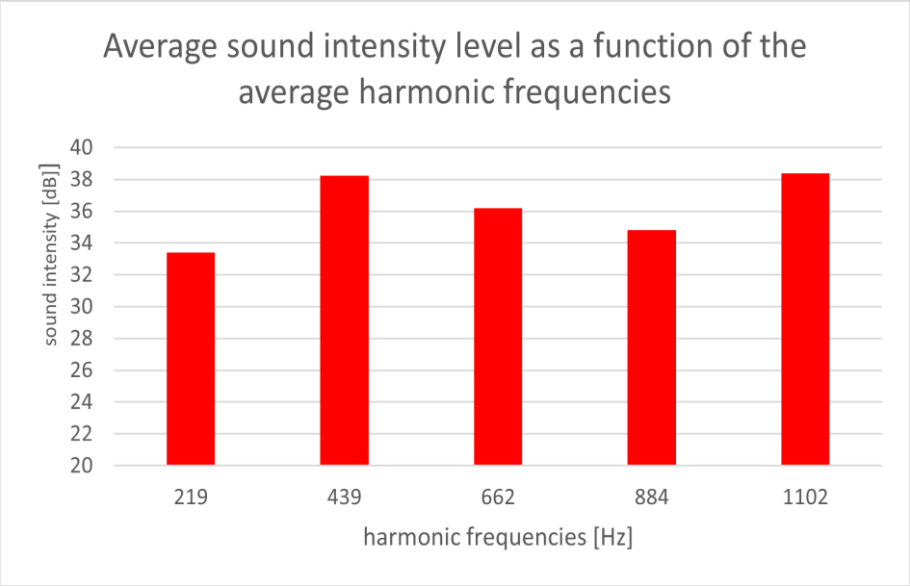

Fig. 3. Average sound intensity level as a function of the average harmonic frequencies for female participants for vowel /a/ in the 2007-2008 cohort.

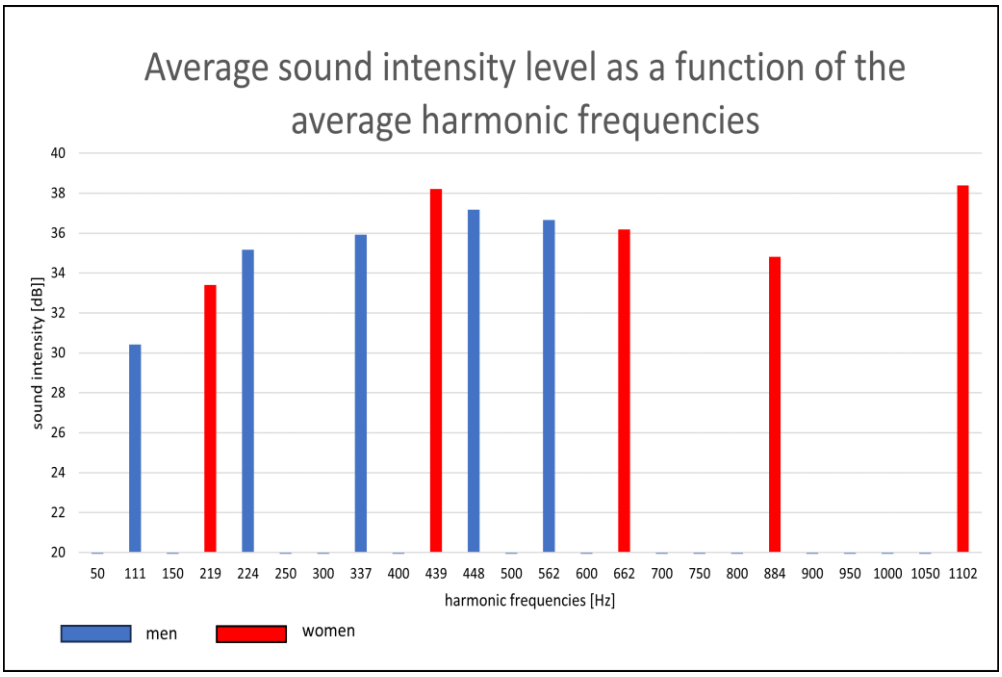

Fig. 4. Average sound intensity level as a function of the average harmonic frequencies for both male and female participants together for vowel /a/ in the 2007-2008 cohort.

Fig. 4 shows the average harmonic frequencies for both male and female participants together in one graph. Male harmonic frequencies are marked in blue whereas female are marked in red. As mentioned in the article, the harmonics of women are shifted towards higher frequencies compared to men. To illustrate these differences more effectively, a graph (Fig. 4) was created, where it's clear that there is a larger concentration of blue bars (male harmonics) and a smaller concentration with a shift of up to 1100 Hz in the red bars (female harmonics).

7. Comparison

To verify the conclusion that human voices differ based on gender, the presented study was compared to the work of a student from the Lodz University of Technology titled [6]. In this study, the author

also examined two groups, namely men and women. The results of this experiment were presented in graphs, namely Fig. 5 and Fig. 6. Graphs come from work [6] and were obtained using computer program.

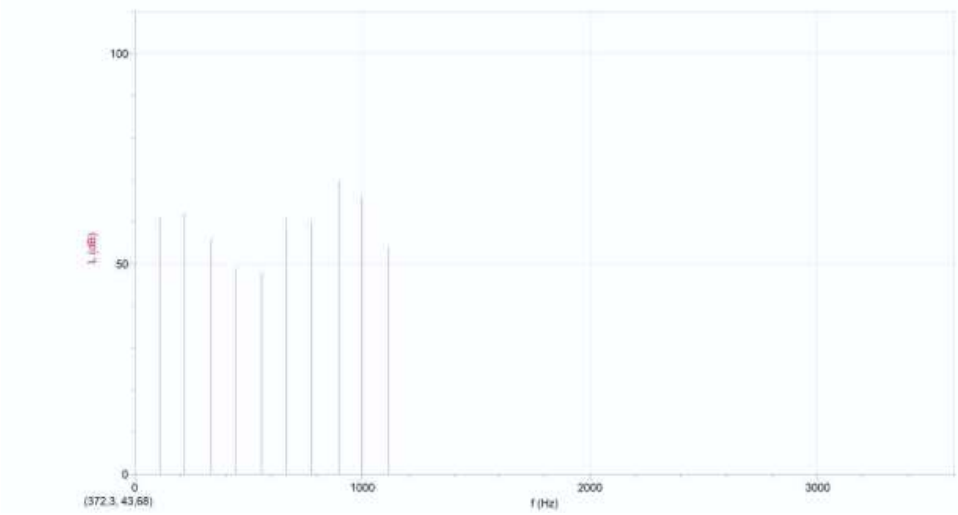

Fig. 5. Average sound intensity level as a function of the average harmonic frequencies for male participants.

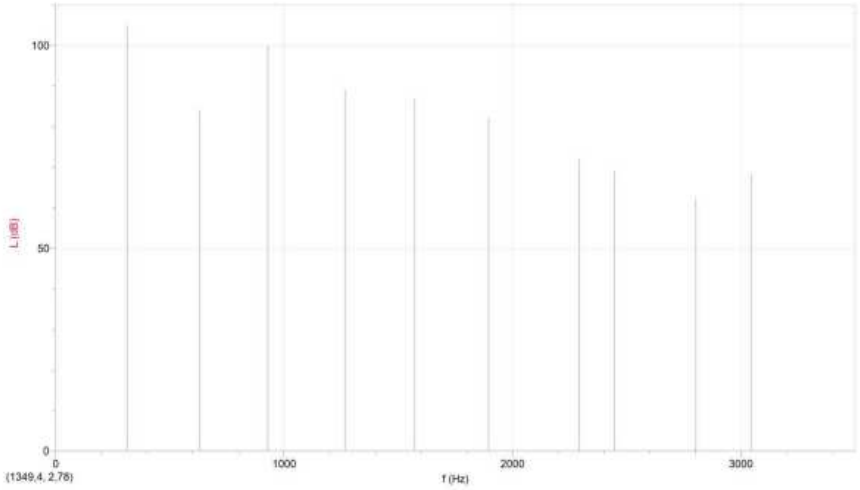

Fig. 6. Average sound intensity level depending on the average harmonic frequencies for women.

In Fig. 5, it can be observed that male harmonics are closely clustered and occur regularly at approximately every 100 Hz, while in Fig. 6 depicting female harmonics, the frequency values are almost three times higher than those of males. The sound intensity level is also different, with higher values for females.

The results presented in the paper [6] confirm that the voice differs depending on gender. Women's harmonics are shifted towards higher frequencies, occur less frequently, and achieve a higher sound intensity level compared to male harmonics.

Human voice is very complicated but at the same time inspiring. Conducting research about psychoacoustic features could contribute to the development of science and technology. Research may also be conducted for other age groups in future. It will show voice differences over the years just like in the [10] paper where authors studied differences of fundamental tone depending on the age of responders.

Literature

[1] Urbański B. Elektroakustyka w pytaniach i odpowiedziach, Wyd. 2 uzup. i uaktual. Wydawnictwa Naukowo-Techniczne, 1984.

[2] Szczeniowski S. Fizyka doświadczalna część I Mechanika i Akustyka, Państwowe Wydawnictwo Naukowe, Warszawa, 1972.

[3] Łopatka J., Kotus J. Teoria wytwarzania dźwięków mowy, formanty, modelowanie wytwarzania dźwięków mowy. Katedra Systemów Multimedialnych, Politechnika Gdańska.

[4] Terlecki J. et al. Ćwiczenia laboratoryjne z biofizyki medycznej dla studentów medycyny i farmacji, Gdańsk : AM, 1996

[5] Mycek B. et al. Ćwiczenia laboratoryjne z biofizyki. Skrypt dla studentów Wydziału Farmaceutycznego Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków 2018.

[6] Anonimowy student Politechniki Łódzkiej. M6 analiza harmoniczna dźwięku, Sprawozdanie, Fizyka Politechnika Łódzka, 2020/2021. https://www.studocu.com/pl/document/politechnika-lodzka/fizyka/m6-analiza-harmoniczna-dzwieku/11917094 published : 09.09.2023.

[7] Oxenham A. J. Pitch perception. Journal of Neuroscience, 2012; 32(39):13335–8. https://doi.org/10.1523/JNEUROSCI.3815-12.2012

[8] Lawson N, Thompson K, Saunders G, Saiz J, Richardson J, Brown D, Ince N, Caldwell M, Pope D. Sound intensity and noise evaluation in a critical care unit. Am J Crit Care. 2010 Nov;19(6), 88-98. doi: 10.4037/ajcc2010180.

[9] ALIC - Analyzing Language in Context, Students understanding the complexity of language. Fluida & WordPress https://alic.sites.unlv.edu/chapter-11-2-speech-organs/ (Access 17.11.2023)

[10] Borkowska B., Pawlowski B. Female voice frequency in the context of dominance and attractiveness perception, Animal Behaviour, Volume 82, Issue 1, 2011, Pages 55-59, https://doi.org/10.1016/j.anbehav.2011.03.024.

[11] Feinberg D. R., Jones B. C., Little A. C.,. Burt D. M & Perrett D. Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices, Animal Behaviour, Volume 69, Issue 3, 2005, Pages 561-568, https://doi.org/10.1016/j.anbehav.2004.06.012.

Advances in IT and Electrical Engineering

vol. 30, 2024, 33-48
https://doi.org/10.7862/re.2023.x

THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

Original Research

Custom Perimeter Alarm System: Enhancing Surveillance Across Multiple Checkpoints

Patryk Jaskuła [1], Mariusz Węglarski [2,*]

[1] Department of Electronic and Telecommunications Systems, Rzeszów University of Technology, ul. Powstańców Warszawy 12, 35-959 Rzeszów, Poland, p.jaskulski41@gmail.com
[2] Department of Electronic and Telecommunications Systems, Rzeszów University of Technology, ul. Powstańców Warszawy 12, 35-959 Rzeszów, Poland, wmar@prz.edu.pl

* Corresponding author. wmar@prz.edu.pl

Received: 06 September 2024 / Accepted: 24 November 2024 / Published online: 23 December 2024

Abstract

This work develops an innovative perimeter loop configuration aimed at surpassing the limitations of conventional alarm systems. The proposed design enables the integration of a large number of sensors into a single monitoring loop, using only two connecting wires, and is compatible with the commonly used normally closed (NC) connection. The custom perimeter system features simplified installation with the use of standardized parametrizing resistors, and includes a calibration method to enhance detection accuracy. A prototype was tested with 20 magnetic contacts, demonstrating that wire resistance and the tolerance of parametrizing resistors are critical factors influencing system performance. In practical trials, a maximum of 15 sensors could be reliably detected. However, the introduction of calibration procedure allowed for an increased detection of up to 20 sensors. The system, which can handle numerous closely spaced control points, also features battery backup capability to ensure operation during power failures. Overall, the system effectively manages complex monitoring requirements and provides reliable performance under various conditions.

Keywords: alarm system, perimeter loop, parametrized line, end-of-line resistor, magnetic contacts, reed switches

1. Introduction

Alarm systems play a crucial role in ensuring security in privet residences, workplaces, stores, various industrial facilities, etc. Their primary function is twofold: they act as a deterrent to potential burglars and provide real-time alerts to property owners in the event of a breach of the protected area [1]. Modern systems can also be used for monitoring the current status of various points within the property (e.g., whether doors or windows are open or closed, or the current value of a measured physical parameter, such as temperature in a room). This allows property owners to have greater control and awareness of their environment, even when they are away from the premises. In contemporary technology-driven world, manufacturers are continually developing more advanced solutions that go beyond mere anti-burglary functions [2]. Many alarm systems are now integrated with smart home automation features, offering enhanced convenience and functionality [3, 4]. For instance, such systems can be connected to lighting, heating, ventilation or air conditioning systems, allowing users to remotely control and monitor these functionalities through mobile apps or other interfaces.

In the case of perimeter systems, a breach of the designated monitoring zone is detected by utilizing an appropriate number of so-called intrusion sensors [5, 6]. There are various types of these access devices, each designed to meet specific security needs depending on the environment and the level of protection required. These include, among others, contact sensors (magnetic contacts, reed switches)

This is an Open Access article distributed under the terms of the CC-BY-NC-ND 3.0 PL license, which permits others to distribute the work, provided that the article is not altered or used commercially. You are not required to obtain permission to distribute this article, provided that the original work is properly cited.

that detect the opening of windows or doors [7], motion detectors, sound and vibration sensors [8] or infrared barriers [9]. Each of these sensors is carefully selected and positioned to cover potential points of entry or vulnerable areas within the perimeter. The choice and number of sensors depend on several factors, including the size of the area to be monitored, the nature of the threats, and the specific security requirements of the property.

Equally important is the method of connecting intrusion sensors to the central control unit, which oversees the operation of the alarm system [10]. The way in which these connections are established can significantly impact the reliability, responsiveness, and maintainability of the entire security system. In traditional alarm systems, all access devices are typically hardwired directly to dedicated inputs on the control panel or in series with an existing monitoring loop. It should be noted that wired connections are generally more reliable, they are less prone to interference from external sources, such as wireless signals or physical obstacles that can disrupt communication. They are less vulnerable to hacking or signal jamming, which can be a concern with wireless systems. Tampering with a wired connection would require physical access to the cables, making it more difficult for an intruder to disable the system without triggering an alarm. Moreover, in a wired system, sensors typically draw power directly from the central control unit, eliminating the need for individual batteries [11]. This reduces the maintenance required to keep the system operational, as there are no batteries to replace. However, the traditional wired approach also presents certain challenges. Installing a wired alarm system can be labour-intensive and requires careful planning, particularly in large or complex buildings. Wired systems are less flexible when it comes to relocating sensors or making changes to the layout of the system. From an economic standpoint concerning installation costs, both wired and wireless systems have their advantages and disadvantages [12]. In this aspect, only considerations specific to a particular scenario provide a measurable comparison.

In scientific literature, the topic of wired perimeter systems in the scope of their construction and operation is rarely addressed in detail. However, certain recent research activities addressed the reliability of reed switches. In these studies, the magnetic contacts were exposed to varying temperatures [13] or multiple switching repetitions [14]. Conversely, there is a substantial body of research focused on wireless solutions, which is quite understandable given the advancements in radio communications technology [15]. Wireless systems offer numerous advantages over traditional wired counterparts, such as greater flexibility in installation and easier adaptation to changing environments. However, they come with their own set of challenges, including the need for regular maintenance to replace power sources (as they operate remotely and are battery-powered) [11], which can be a significant inconvenience, particularly in large protected areas with numerous access points. Additionally, the costs for radio frequency components in sensors and central receiving units must be taken into account in installation projects. Electromagnetic interference generated in the environment [16] or cyber security of communication networks [17] also have a significant impact on the reliable operation of the wireless alarm systems. Nevertheless, there are numerous examples in the literature of research efforts aimed at effectively replacing the components of typical wired perimeter systems by creating novel and unconventional components. For instance, these include the contact glass-break detector [18], which provides a specialized mechanism for detecting vibrations, or a highly sophisticated solution involving autonomous drones working in tandem with a set of video cameras that remain inactive until an event is detected [19]. Another notable example is the use of radar sensors to enable early warning capabilities and track the path of an intruder [20]. These innovative approaches highlight the ongoing quest to enhance security systems by leveraging advanced technologies and optimizing operational efficiency.

Based on the identified limitations and drawbacks of existing alarm systems, an innovative perimeter monitoring installation has been developed by the authors. In the presented work, special attention was given to the number of possible control points that a user can connect to a single input on the control panel or expansion module, as well as the complexity of the wiring required with an increasing number of nodes in a single loop. The analysis led to the development of the custom perimeter system solution that addresses these shortcomings and allows for the connection of the maximum number of intrusion sensors using the fewest possible wires. The primary objective of the presented study is to ensure the system can accurately detect which of the numerous monitored control points has been triggered within an extensive monitoring loop and to communicate this information to the user. In the view of the authors, their proposed solution simplifies the installation process while maintaining the system's reliability.

2. Research Assumptions

2.1. Effectiveness and Limitations of Perimeter Line Configurations

A typical manufacturer of alarm devices offers a range of more or less advanced control panels, designed for applications ranging from the simplest home installations to extensive industrial surveillance areas. These devices primarily differ in the number of inputs, outputs, and built-in functions. This allows the user to select a device that suits their specific needs while taking economic considerations into account.

As an example, the most cutting-edge products from manufacturers active in the local alarm control panel market are present in Table 1 [21-27]. The most advanced control panel, supported by expanders, can handle up to 256 sensors. It is important to note that the manufacturer provides the number of inputs available on the main board alone and additionally the maximum number of inputs that can be supported through additional expansion modules. It should also be noted that the main control panel board can manage only up to 16 wired sensors, while each expander can handle only 8. To utilize the 256 inputs claimed by the manufacturer, it is necessary to use as many as 30 expansion modules. The cost of one expander is about 1/5 of the price of the control panel itself. Therefore, a significant portion of the costs incurred in building a multi-point surveillance system is allocated to input expanders. Consequently, a desirable solution is a system that allows the connection of a large number of alarm sensors without the use of additional expansion modules.

From the perspective of creating a perimeter installation, manufacturers of control panels adhere to established standards that define the types of inputs designed for connecting wired sensors (Fig. 1). These inputs include those operating as NC (Normally Closed), NO (Normally Open), as well as parameterized configurations such as EOL/NC (End of Line/Normally Closed), 2EOL/NC, 3EOL/NC, and their corresponding EOL/NO (End of Line/Normally Open), 2EOL/NO or 3EOL/NO versions [7, 28, 29]. Standards involving more complex resistor configurations, such as 4EOL, 5EOL, etc., are not commonly used in commercial products; whereas 3EOL type is rarely available and only found in the most advanced devices (e.g. in Integra 256 Plus).

Table 1. Comparison of control panels with potential for use in wired perimeter alarm systems

Parameters	Control panel			
	ROPAM [21, 22] NeoGSM-IP-64	**DSC [23, 24] HS2128**	**Paradox [25, 26] Spectra SP7000**	**Satel [27] Integra 256 Plus**
Number of physical inputs for wired sensors on main board of control panel	16	8	16	16
Maximum number of wired inputs that can be managed through expanders	64	128	32	256
Number of partitions / zones	4	8	2	8 / 32
Number of physical outputs on the control panel's main board	8	4	4	16
Number of programmable outputs that can be managed through expanders	40	148	16	256
Permissible configuration of input lines	NC, NO, EOL/NC, EOL/NO, 2EOL/NC,2EOL/NO and 3EOL/NC*, 3EOL/NO* (*only Integra 256 Plus)			

Parameterization involves incorporating additional resistors into the circuit with the sensor's detector / switch [28]. From an electrical standpoint, switching to the additional resistance changes the equivalent resistance of the alarm circuit. In this way, the additional state of sensor activity can be distinguished. Manufacturers of alarm system use this technique to integrate security functions (e.g., tamper detection, beam obstruction, or removal of the device from its mounting surface) directly into the alarm sensor. This allows for monitoring the integrity of the perimeter lines, thereby increasing the system's reliability by preventing tampering or damage to the connection wires.

In alarm systems, NC connections (Fig. 1-a) are more commonly used due to the additional security they provide against potential damage to the alarm loop. Interrupting the continuity of the wire immediately triggers a response from the alarm control panel. In the case of NO inputs (Fig. 1-b), when the circuit's continuity is interrupted, it is not possible to determine if the sensor has been triggered to the close state. Unfortunately, the NC solution also has an additional drawback: if the circuit is normally closed by the sensor's detector / switch (SWx in Fig. 1), current always flows through the alarm loop

during standby. The only way to limit this current is to use the resistors. By adding these resistors, the perimeter loop is converted into an EOL parameterized line (Fig. 1-c).

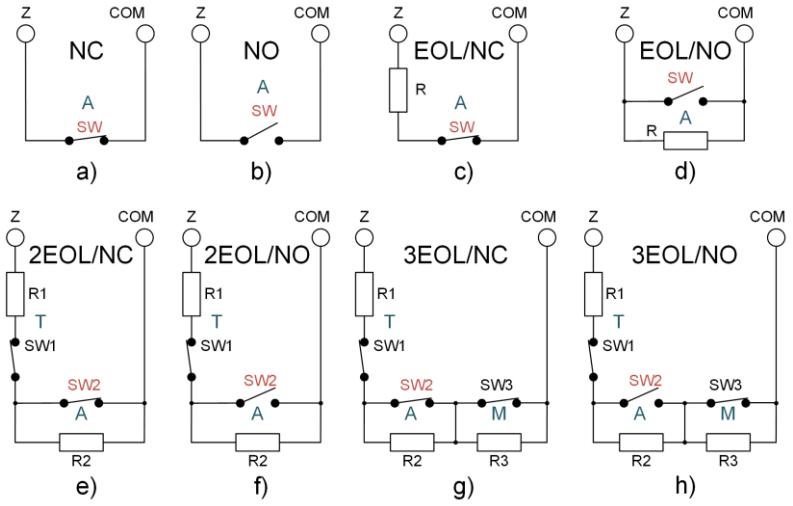

Fig. 1. Types of connections in parametrized lines: a) NC; b) NO; c) EOL/NC; d) EOL/NO; e) 2EOL/NC; f) 2EOL/NO; g) 3EOL/NC; h) 3EOL/NO; A – alarm; T – tamper; M – masking

The parametrizing resistor R is connected either in series (Fig. 1-c) or parallel (Fig. 1-d), depending on the required default standby state (NO or NC sensor). For the NC line (Fig. 1-c), the primary standby state is the closed circuit through the resistor R and switch SW. The second activation state (open state) occurs when the switch SW is open and no current flows. The third state can be triggered if a short circuit, caused by damage to the connection wires, happens between the sensor and the input of the control panel. In the event of wire breakage, it is impossible to distinguish this state from the triggered state, as in both cases, no current flows through the circuit. This approach is advantageous because, as mentioned earlier, it allows for the detection of a short circuit in the installation, caused by accidental reasons or tampering attempts. In the case of the EOL/NO line (Fig. 1-d), the default state is reversed. Additionally, for an NO sensor, it is possible to distinguish the sensor's activation state from the state of a broken connection. Thus, in the EOL/NC configuration, it is possible to detect a short circuit in the perimeter loop, whereas in the EOL/NO configuration, it is possible to detect broken wires.

Another variation is the 2EOL configuration (Fig. 1-e, f), which uses two resistors R1 and R2 and two switches SW1 and SW2. Typically, the additional switch is used by the alarm sensor as an extra protection to detect tampering attempts, such as opening the enclosure or detaching it from the wall. Each subsequent state is distinguished by a different equivalent resistance R_Z value observed between the ‘Z’ and ‘COM’ terminals and for 2EOL/NC configuration (Fig. 1-e) it is equal:

- $R_Z = R_1$ – SW1 closed, SW2 closed => standby state,
- $R_Z = \infty\Omega$ – SW1 opened, SW2 closed or opened => temper state or broken line,
- $R_Z = 0\Omega$ – short circuit in the installation, state of SW1 and SW2 any => fault state,
- $R_Z = R_1 + R_2$ – SW1 closed, SW2 opened => alarm state.

Similar considerations can also be applied to the 2EOL/NO line (Fig. 1-f):

- $R_Z = R_1 + R_2$ – SW1 closed, SW2 opened => standby state,
- $R_Z = \infty\Omega$ – SW1 opened, SW2 closed or opened => temper state or broken line,
- $R_Z = 0\Omega$ – short circuit in the installation, state of SW1 and SW2 any => fault state,
- $R_Z = R_1$ – SW1 closed, SW2 closed => alarm state.

As can be observed, the standby and alarm states for the 2EOL/NC circuit are reversed compared to the 2EOL/NO circuit. In both circuits, it is not possible to detect a separate case of a break in the continuity of the connection wiring. In either case, this state is always associated with tampering. It should also be noted that the values of both resistors can be the same, as this does not affect the ability to distinguish between the individual states. In one case, the equivalent resistance of the circuit is equal to R_1, while in the other case it is the sum of the series connection of R_1 and R_2. Regardless of the resistor values used, it is always possible to differentiate between these two states.

The 3EOL configuration (Fig. 1-g, h) is supplemented with an additional switch and resistor, which allows for the differentiation of an extra state. However, this modification introduces some complications. When resistors with the same values are used, it becomes impossible to distinguish between the activation states of switches SW2 and SW3. In both cases, the equivalent resistance is equal to the sum of the resistances. Therefore, manufacturers recommend that resistor R3 should have a value equal to the sum of R_1 and R_2 in series. Nevertheless, to correctly distinguish between states, it is sufficient to apply different values to the mentioned resistors and for 2EOL/NC configuration (Fig. 1-g) the resistance equivalent R_Z is equal:

- $R_Z = R_1$ – SW1 closed, SW2 closed, SW3 closed => standby state,
- $R_Z = \infty\Omega$ – SW1 opened, SW2 and SW3 closed or opened => temper state or broken line,
- $R_Z = 0\Omega$ – short circuit in the installation, state of SW1 and SW2 any => fault state,
- $R_Z = R_1 + R_2$ – SW1 closed, SW2 opened, SW3 closed => alarm state,
- $R_Z = R_1 + R_3$ – SW1 closed, SW2 closed, SW3 opened => antimasking state,
- $R_Z = R_1 + R_2 + R_3$ – SW1 and SW2 and SW3 opened => alarm with antimasking state.

Similar considerations can also be applied to the 3EOL/NO line (Fig. 1-h):

- $R_Z = R_1 + R_2$ – SW1 closed, SW2 opened, SW3 closed => standby state,
- $R_Z = \infty\Omega$ – SW1 opened, SW2 and SW3 closed or opened => temper state or broken line,
- $R_Z = 0\Omega$ – short circuit in the installation, state of SW1 and SW2 any => fault state,
- $R_Z = R_1$ – SW1 and SW2 and SW3 closed => alarm state,
- $R_Z = R_1 + R_2 + R_3$ – SW1 closed, SW2 opened, SW3 opened => antimasking state,
- $R_Z = R_1 + R_3$ – SW1 closed, SW2 closed, SW3 opened => alarm with antimasking state.

As with the 2EOL line, the alarm and standby states for NO and NC are reversed. This also applies to the additional states present only in the 3EOL line. This type of parameterized line is most commonly used by manufacturers for infrared barriers and motion detectors, incorporating anti-masking to detect attempts to obstruct the emitted infrared beam.

2.2. Perimeter Loop with Numerous Intrusion Sensors

In the alarm system market, it is difficult to find products that efficiently handle a large number of intrusion sensors installed across multitude windows and doors within a single monitoring zone. If the requirements specify that each window in a hall must be protected by contact sensors or more advanced monitoring units (such as a vibration, glass break, or ultrasonic sensors), and if it is necessary to distinguish which access point has been breached, there is no other option but to use separate inputs on the alarm control panel for each sensor. When designing a perimeter alarm circuit under such requirements, several challenges have to be addressed. The first one is the complexity and significant expansion of the wiring circuit, as separate cables must be run to each access point (Fig. 2). For example, in parallel perimeter loop, a bundle of wires coming directly from the alarm control panel in a configuration with multiple windows have a large diameter due to the high number of individual cores within the cable. The second, and more serious, problem is the limited availability of alarm inputs, especially in budget versions of commercial control panels.

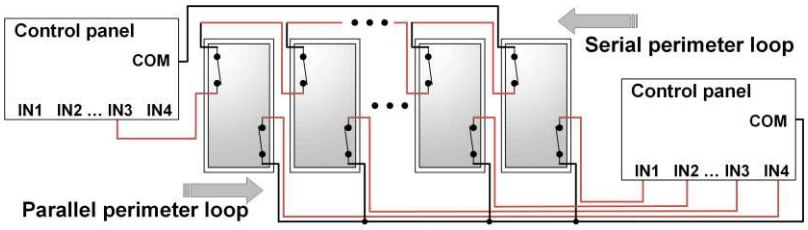

Fig. 2. Example of perimeter alarm line diagram for multiple windows protected by reed switches: parallel perimeter loop – each sensor connected separately to one input of control panel; serial perimeter loop – connection of sensors in series to one input of control panel

A potential solution to the above-mentioned problems is to connect the intrusion sensors in series (Fig. 2). This approach reduces the number of inputs used on the alarm control panel. It also simplifies

the wiring installation, as only two wires extend from the control panel. The drawback of this configuration is the inability to determine which sensor in the series is triggered at alarm moment. Typically, this limitation is of minor importance, as the most crucial information for the user is that the perimeter loop has been interrupted, possibly indicating the monitored zone or room, rather than the precise location of the triggered sensor. This holds true for a simple installation with only a few sensors. However, in the case of a complex installation with a large number of control points, this system limitation can become problematic. For example, servicing the entire installation becomes challenging. In the event of a sensor failure, each sensor has to be checked individually, which is time-consuming. The ideal configuration is a circuit in which, despite connecting multiple sensors using only two wires, it is still possible to identify which monitoring point is triggered during an alarm. This capability is partially provided by the 3EOL parameterized line (Fig. 1-g, h), which could be further modified with additional sections to an nEOL type. The expansion method, allowing to manage a larger number of sensors, is illustrated in Figure 3. The modification involves implementing additional sections of switches with resistors connected in parallel to the existing setup. Each switch corresponds to each alarm detector.

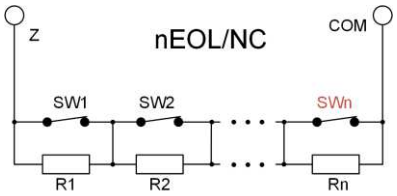

Fig. 3. The modified nEOL/NC configuration, which includes additional sections

The equivalent resistance R_Z of the nEOL circuit depends on which of the switches is open. When one of the switches is open, the resistor connected in parallel with it is no longer short-circuited by the circuit branch. This modification requires that each resistor has a different resistance value, because only then it is possible to determine which switch has been opened. The need to use different resistor values is not the only challenge here. When two or more switches are open, the value of R_Z equals the sum of the series-connected resistors that are not short-circuited. It is possible that the obtained sum could match the resistance value of another open section. This would result in an incorrect identification of the breach point within the supervised zone. To ensure accurate detection when multiple monitoring points are triggered simultaneously, each subsequent section should be equipped with resistors of progressively increasing resistance. Mathematically, this relationship can be expressed as follows:

$$R_Z > \sum_{i=1}^{n-1} R_i \,, \tag{1}$$

where n denotes the index of successive resistance R_i in the system.

In order to determine which sensor has been activated, each possible state must be defined and distinguished. Therefore, the nEOL configuration is challenging to implement due to the large number of possible combinations 2^n (e.g., 1024 for 10 points) that need to be distinguished. Additionally, there is a problem during the installation of the alarm system. Each loop section requires different resistor values, which complicates the assembly process and increases the likelihood of errors by installers.

2.3. Concepion of Custom Perimeter Alarm System

The custom perimeter alarm system (Fig. 4-a), as proposed by the authors, is structurally similar to the 3EOL configuration. The states of the perimeter loop are distinguished by changes in the equivalent resistance R_Z observed at nodes A and B (Fig. 4-b).

In the performed analysis, zero resistance R_x is assumed for the wires and switches in the branches. When all switches are closed, the equivalent resistance R_Z of the circuit is infinitely small because all resistors are bypassed. If any switch is open, the electrical potential at node A differs from that at node B. All resistors from the terminal Z to the open switch are not bypassed and actively contribute to the equivalent resistance. Meanwhile, all resistors to the right of the open switch remain unchanged – they are still bypassed and do not affect the value of R_Z. In the example (five monitored access node, the third section triggers alarm) shown in Figure 4-b, the equivalent resistance of the circuit is equal to the parallel combination of R_1, R_2, and R_3.

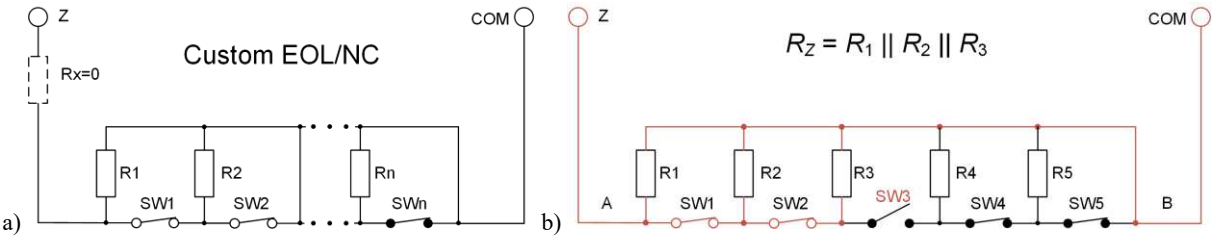

Fig. 4. Diagram of custom perimeter loop: a) Idea; b) Example of five nodes; SW3 open; current flow is marked in red

In the Custom EOL configuration, the equivalent resistance R_Z for any open switch, regardless of its number, is described by the following relationship:

$$R_Z = \frac{1}{\sum_{i=1}^{n} \frac{1}{R_i}} \tag{2}$$

where n denotes the index of the switch that is open, and R_i represents the resistances of the various resistors in the circuit. If resistors with equal values are used instead of different ones, the relationship (2) simplifies to the form:

$$R_Z = \frac{1}{\frac{n}{R}} = \frac{R}{n}, \tag{3}$$

where R represents the common resistance value for all resistors. Using identical pull-up resistors to the COM terminal means that for the n-th active switch, the equivalent resistance of the circuit is equal to the resistance of a single pull-up resistor divided by the index number of the open switch, counted from the terminal Z.

Each section in the perimeter loop corresponds to a single access point. Each subsequent switch has a lower priority compared to the previous one, meaning that when multiple points are activated simultaneously, the circuit's state is defined by the section with the smallest index. For example, if switches SW1, SW3, and SW5 are open, the equivalent resistance of the circuit corresponds only to the open state of SW1. This is a limitation of the Custom EOL configuration, as it is not possible to differentiate multiple monitored nodes being activated at the same time. Nevertheless, it is possible to detect each triggered alarm individually, one by one, according to their ascending numbers. At the moment the alarm loop is first interrupted, it is possible to determine where the breach occurred, which is crucial. This information allows for addressing the breach and moving on to manage next access points if they have also been activated.

The developed configuration of perimeter loop is simple to service. If an access point is damaged or the installation wires are interrupted, locating the fault is straightforward. It is only necessary to check the indicated access point and the wires between adjacent points. An additional advantage is the simplicity of implementing this solution compared to the 3EOL system and its extended modification (Fig. 3). The Custom EOL configuration does not require resistors with various values, also simplifying the assembly process. The limitation of these extended configurations is that they only support sensors with NC (normally closed) outputs.

3. System Evaluation

3.1. Design of Custom Perimeter Alarm System

The method of installation in real conditions (supervised zone secured with a perimeter loop of NC-type reed switches) is shown in Figure 5. The developed Custom EOL configuration requires additional pull-up resistors between the switches and the neutral point of the COM circuit. However, it is important to emphasize that, unlike in nEOL, the resistance value is the same for all resistors.

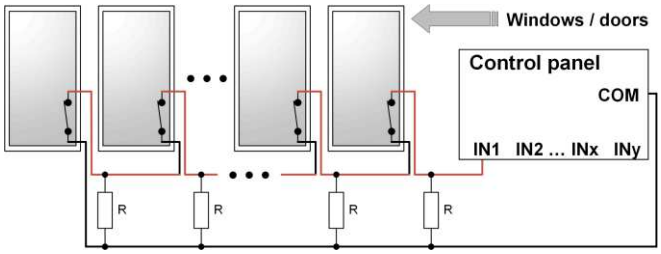

Fig. 5. Diagram of Custom EOL configuration of perimeter line secured with type NC reed switches

Because the alarm state of any access point is determined by changes in the equivalent resistance R_Z of the perimeter loop, it is necessary to propose an appropriate measurement method. Values of R_Z can be obtained using a voltage divider circuit combined with an analog-to-digital converter (ADC, Fig. 6-a). By carefully selecting the resistance ratio, it is possible to maximize the voltage change for the alarm state of each subsequent intrusion sensor.

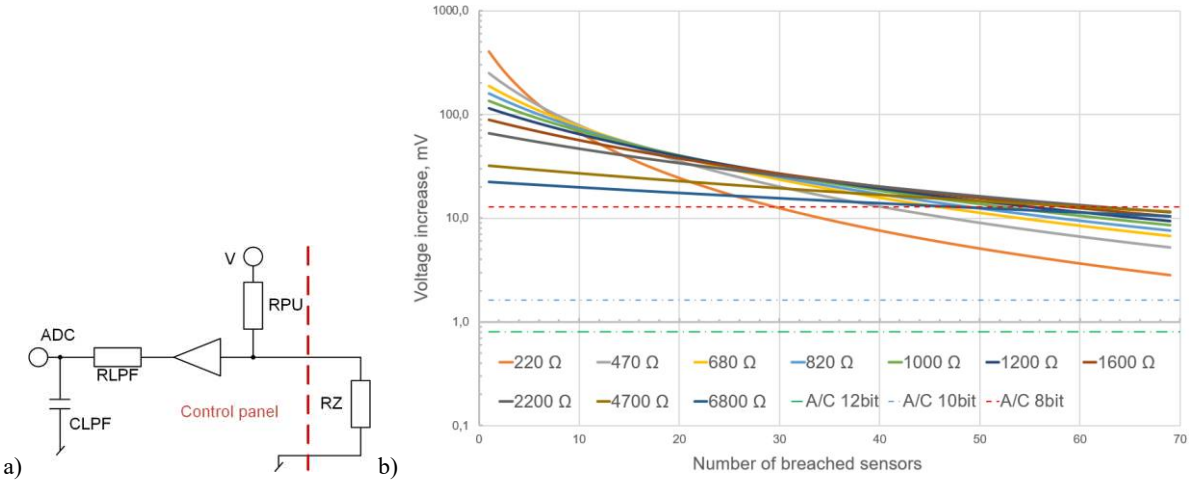

Fig. 6. Measurement of equivalent resistance R_Z: a) Signal conditioning circuit diagram; RPU – pull-up resistor; RLFP and CLPF– low-pass filter resistor and capacitor; V – supply voltage; ADC – analog-to-digital converter; b) Voltage change increment at divider's output relative to the next activated sensor for various resistor R values in perimeter loop; $R_{PU} = 47\ \Omega$

The change in voltage increment relative to the next activated sensor for several different resistance values R added to each section of the perimeter loop is presented in Figure 6-b. Thus, all resistors that make up the equivalent resistance R_Z have the same resistance value, and the upper resistor RPU in the voltage divider circuit has a constant value for all cases. As can be observed, the voltage increment is negative and changes in an exponential manner. The voltage decrement for sensor in the alarm state that is closer to the beginning of the loop is significantly larger than for that one that is closer to the end.

Therefore, the shape of the increment curve can be modelled by adjusting the resistance values in the divider. The optimal value is the one that provides the greatest voltage change for a given number of sections. In most cases, the method for selecting resistance involves finding (Fig. 6-b) the point on the curve corresponding to the greatest increment. If two or more curves intersect near the found point, the one that provides the greatest voltage change for the initial sensors is chosen. For example, for a loop with 30 sections, the optimal resistance is $R = 1600\ \Omega$. Nevertheless, the smallest possible values are preferred due to their lower susceptibility to interference, which can be induced in the connecting wires, for example, through inductive or capacitive coupling. Therefore, the upper resistor in the divider should be small, on the order of several tens of ohms. For this reason, $R_{PU} = 47\ \Omega$ is used in the calculations. The R_Z resistance can also be included in the voltage divider (Fig. 6-a) as the upper resistor. By maintaining the resistance ratios, only the sign of the voltage increment will change.

Additional lines, corresponding to the minimum voltage values that can be distinguished by ADC converters with resolutions of 8, 10, and 12 bits, are included in Figure 6-b. As can be observed, the 8-bit converter does not provide sufficient resolution for correct operation across the entire intended range. Its line intersects the voltage increment curves in all cases. If the loop configuration is limited to 20 sections, an 8-bit converter could theoretically suffice. However, the difference between the voltage change upon sensor activation and its voltage resolution is too small for proper alarm state detection,

especially with higher resistance, such as R = 6800 Ω. The 10-bit and 12-bit ADC converters have sufficient resolution and can be used in the considered operating scenarios. The best results are obtained using the 12-bit converter, as it allows for a more precise determination of the voltage difference.

3.2. System Efficiency in Real Conditions

The measured voltage value for a given state of perimeter loop is slightly different from the ideal value obtained from calculations. This is due to the influence of real-world parameters such as resistor tolerance, non-zero wire resistance, uncertainties in the ADC, and interferences. For this reason, it is necessary to determine the voltage ranges within which the value must fall in order to correspond to the perimeter loop state associated with the activated intrusion sensor. The boundaries of these ranges are calculated by dividing the increment between two adjacent states in half and adding the resulting value to the previous state (Fig. 7-a). The width of these ranges defines the correct operating limits of the system.

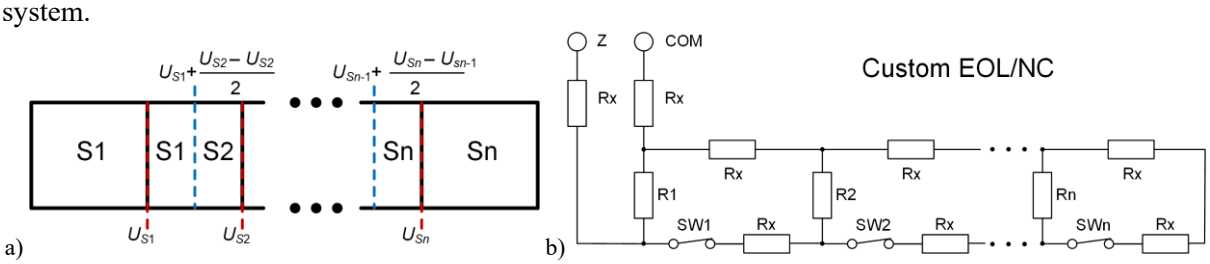

Fig. 7. System efficiency in real conditions: a) Method for calculating the voltage ranges for each state; b) Electrical equivalent of perimeter loop, taking into account resistance of wires R_x

To examine the impact of real-world parameters, the effect of wire non-zero resistance and resistor tolerances is considered for the worst-case scenarios (Fig. 7-b). The worst-case scenario is defined as the combination of extreme tolerance values for both the upper resistor in the divider and the resistors on the alarm loop side. The extreme resistance values are determined by the common resistors' tolerance.

The resistance of the connecting wires R_x depends on the type of wire used in the alarm installation. Typically, multi-core cables of type YTDY 6x0.5 mm (with a single wire diameter of 0.5 mm) are used. Knowing the diameter and material of the wire, it is possible to calculate the resistance for a given length using the following mathematical formula:

$$R = \rho \frac{l}{A}, \tag{4}$$

where ρ means resistivity of copper, l – length of conductor, A – cross-sectional area. Thus, for YTDY 6x0.5 mm, the resistance per meter is equal 89.35 mΩ.

Since resistance depends on the length of the wire between successive sensors, numerical calculations were conducted for several lengths: 1 m, 3 m, and 5 m. Tolerances of 1% and 5% were also considered. Simulations were conducted for a perimeter loop with 30 sensors, and the results were compared with those of the ideal case without accounting for wire resistance. The obtained results are presented in four graphs in Figure 8. The curves on the graphs labelled 'Upper boundary value' and 'Lower boundary value' define the range within which the voltage can vary from the ideal value for each activation state. If the voltage for a given state falls outside this range, the alarm control panel will incorrectly determine the number of the breached sensor. The intersection points define the maximum number of sensors that can theoretically be handled without inaccuracies.

The obtained results are summarized in Table 2. Since two extreme values were considered for each case, the maximum number of sensors that can be handled is taken as the smaller value obtained for a given tolerance. This approach ensures that no value will exceed the boundary for correct detection. Such an occurrence might happen because, in the minimal extreme case, the wire resistance for the middle sensors in the line partially offsets the tolerance effect, artificially increasing maximum number of sections.

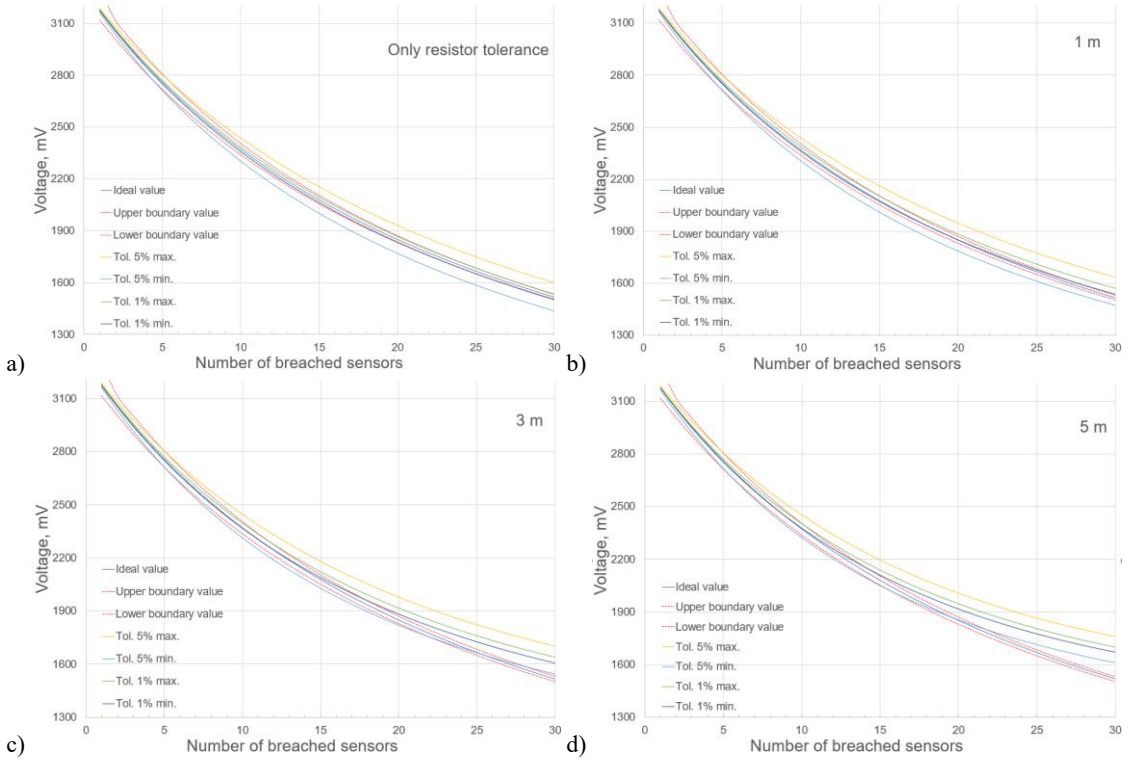

Fig. 8. Voltage variation graph depending on the active sensor; red dashed line – curves indicating boundaries within which the voltage can vary: a) case without accounting for resistance of wires; b) case for wires with length of 1 m between sensors; c) case for wires with length of 3 m between sensors; d) case for wires with length of 5 m between sensors

Table 2. Points of intersection with the boundary curves

Wire length / Tolerance	Without wire resistance	1 m	3 m	5 m
	Theoretical number of sections			
max. @ tol. 5%	6	5	5	5
min. @ tol. 5%	5	5	5	6
max. @ tol. 1%	26	15	12	10
min. @ tol. 1%	25	28	17	15

The wire resistance and resistor tolerance significantly limit the proper operating of the perimeter alarm system. Using resistors with a lower tolerance improves the system's performance, but it does not eliminate the negative impact of the wire resistance. The length of the wires depends on the size of the alarm installation: the size of the building, the layout of the rooms, the placement of windows and doors, and the wires routing method. The complexity of the installation is a disadvantage because, for sensors closer to the end, the wire resistance has a dominant effect on the difference between the ideal and actual values. This is clearly seen when comparing the graphs without accounting for wire resistance to those for a 5 m length. The characteristic bending of the ideal calculated curves for the end sensors is evident.

In the simulations conducted, an equal wire length between successive intrusion sensors was assumed. In a real perimeter alarm system, the distribution of sensors in the protected zone is uneven, resulting in varying wire lengths between them. A possible way to improve the system's performance is through a calibration method. This involves measuring the actual voltage values for all activation states and using them to calculate the operating ranges. This approach allows for the elimination of the impact of wire resistance and the tolerance of resistors used in the alarm loop. Based on the obtained values, the operating ranges can be calculated as shown in Figure 7-a.

To implement the calibration method practically, for a finished installation with any number of sensors, it is necessary to measure the voltage for each activation state. Most of this procedure can be performed by the alarm control panel. The role of the installer is merely to activate the specified access point and communicate this to the control system of the panel. The measurement procedure must be carried out for all sensors in the perimeter loop. In the subsequent steps of the calibration procedure, the following actions should be performed:

- approaching the access point indicated by the alarm control panel and activating it;
- return to the alarm control panel and confirm the activation of the specified sensor;
- the alarm control panel indicates the next sensor to activate;
- deactivate the sensor that was previously activated;
- repeat the entire procedure for all sensors in the alarm circuit.

However, remote access to the control panel significantly reduces the time required for this process. Moreover, the calibration procedure is performed only once at the beginning, during the installation. The measured voltage values can be recorded and then used for monitoring the status of the alarm circuit. This approach allows installers to eliminate the effects of wire resistance and resistor tolerance, enabling the handling of a greater number of sensors.

3.3. Measurement Stand

To practically test the developed system, a special alarm control panel and manipulator with an alphanumeric display and keyboard were designed. The separation of functions into two devices allows the installer to place manipulator in any visible location and easily accessible location for the user, such as near the building's entrance doors. Meanwhile, the control panel should be located in the secured part of the monitored zone to make it more difficult for potential intruders to locate and disable the entire alarm system.

Both devices operate under the control of an ARM Cortex-M4 STM32F303CBT6 microcontrollers, which can operate at frequencies up to 72 MHz and features four 12-bit ADCs with support for up to 39 channels. Communication between the modules is carried out via a CAN 2.0a bus. The use of the CAN serial link allows for the connection of a larger number of devices and provides high resistance to interferences. This enables the system to be expanded in the future with additional modules, e.g., GSM wireless communication. The designed system also includes emergency operation capability on battery power, since it has to be resistant to power supply interruptions. The block diagrams of the main PCB boards are shown in Figure 9 and their visualisation in figure 10.

The control software was written in C language using high-level HAL (Hardware Abstraction Layer) libraries. Through the user interface, it is possible to change the settings of the alarm control panel and communicate the system's status to the user. When the alarm system is activated, the access panel retrieves the saved configuration settings from FLASH memory. After synchronization is complete, the main menu appears on LCD with information on the status of:

- alarm zones – the system has two independent zones;
- inputs – which input data is activated and calibrated;
- power supply – operating on battery or mains power, battery charge level.

A settings menu was also developed, allowing the following options:

- display configuration – adjustment of the brightness and contrast level;
- input configuration – selection of the input type NC, NO, EOL, 2EOL, 3EOL, custom connection system; selection of the number of sensors for the custom system;
- output configuration – selection of the default output state;
- zone configuration – assignment which inputs trigger a specific zone and which outputs are controlled by that zone.

If the settings are not configured correctly, the system will not allow access to the arming menu. If the custom connection system is selected, the calibration process is begun upon entering the arming menu (Fig. 11). This process involves measuring the voltages for all sensor activation states and using them to calculate the correct detection ranges.

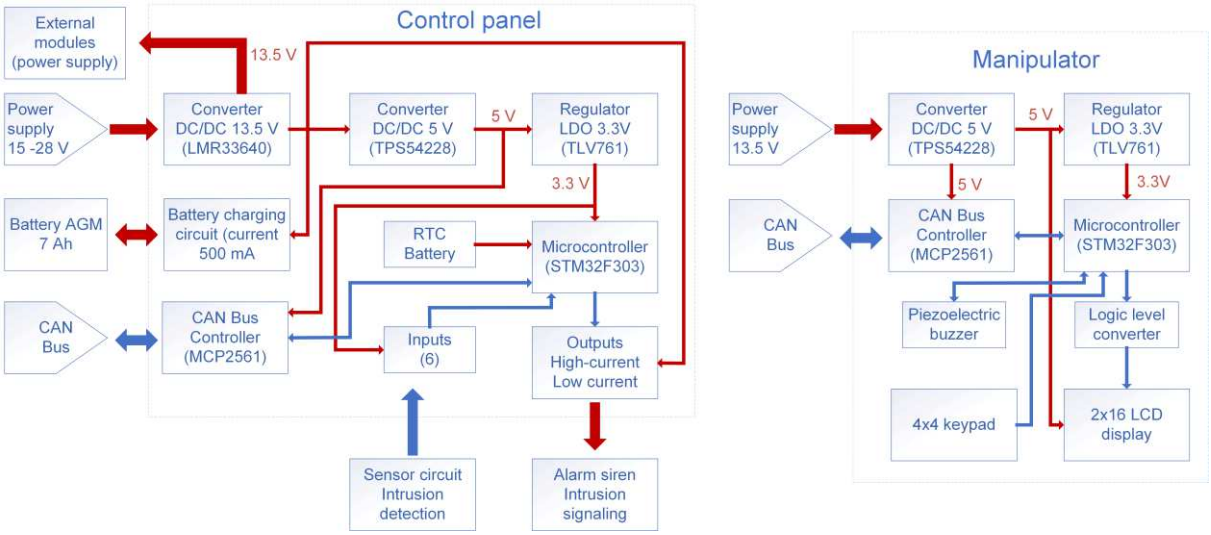

Fig. 9. Block diagram of alarm control panel and manipulator: power supply connections are indicated in red; signal and control connections are in blue

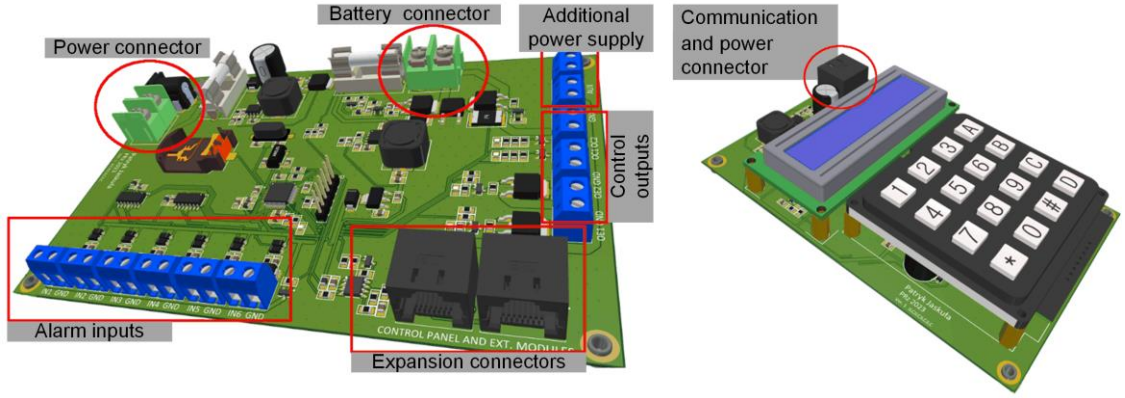

Fig. 10. 3D model of printed circuit boards

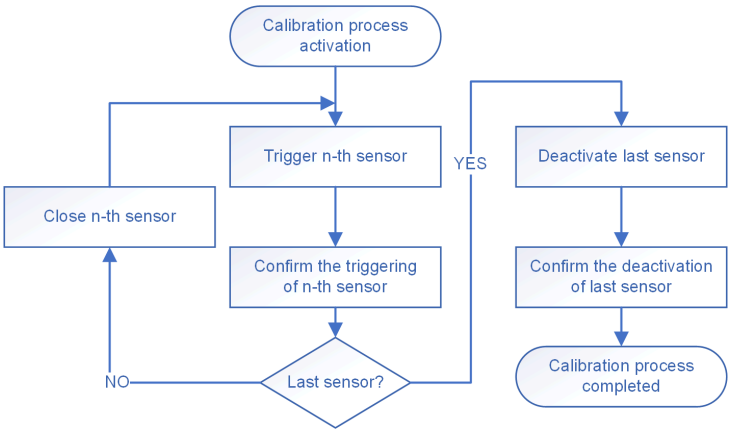

Fig. 11. Block diagram of calibration process

After the calibration process is completed, the measured voltage values are saved in FLASH memory to avoid the need for re-calibration in the event of a power loss. Then, the user can select the zone to be armed or disarmed. The system has two independent zones, which are armed with an 8-digit code consisting only of numbers. If an incorrect code is entered during disarming, an alarm is triggered. If any sensor on a line is triggered for a given input, a message appears showing the input number and sensor number.

3.4. System Tests

To evaluate the performance of the developed perimeter configuration, a test circuit was created. It consists of 20 sensors, each spaced 1 meter apart. The circuit loop in assembled using 2x0.5 mm YDY cable as well as parametrizing resistors with a value of 470 Ω and a tolerance of 1%. On the control panel side, a 47 Ω resistor is used as a pull-up resistor to the 3.3 V power supply in the voltage divider. The test circuit, along with the alarm system, is shown in Figure 12.

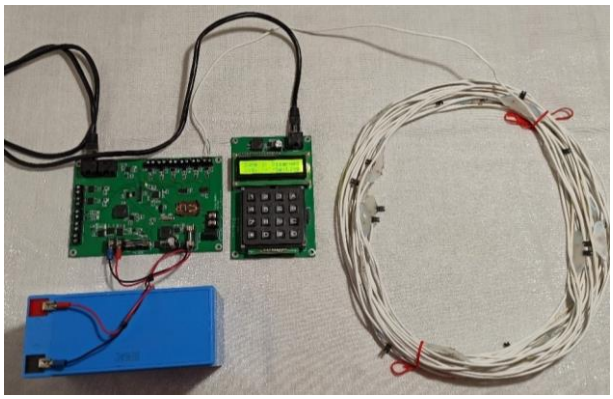

Fig. 12. Test perimeter loop, along with the alarm system

The obtained results of voltage measurement are presented in Figure 13, where they are plotted alongside the ideal values calculated for a resistor tolerance of 1%, taking into account the resistance of the cables.

The point where the curve based on the measurement intersects with the dashed lines determines the detection limits in relation to the ideal case (Fig. 13). Therefore, the maximum number of sensors that can be handled without errors for the designed circuit is 15. The intersection points for the curve calculated at the tolerance limit values are listed in Table 3. As can be observed, the intersection points for the measured values occurred for sensors located farther than the calculations for the worst-case scenario indicated. The measured voltage curve is situated between the curves calculated for extreme tolerances, confirming that the theoretical considerations closely approximate the system's behaviour. This allows the detection limits to be determined for installations with any number of sensors and cable lengths if a calibration method is not applied.

Table 3. Intersection points for curve calculated at tolerance limit values

	Calculations for 1 m and 1% tolerance		**Measurement**
	Max.	**Min.**	
Intersection points	13	19	15

To examine the noise level in the test perimeter loop, voltage fluctuations were measured by the A/D converter for the test circuit with 20 sensors. A total of 12,500 samples were collected at a sampling frequency of 12,500 Hz for three sensor activation cases in the circuit. The collected data is presented in four histograms in Figure 14.

For all cases, the read value varies by four quantization levels. The ADC voltage resolution is approximately 0.805 mV, which corresponds to a peak-to-peak value of the measured signal of 3.22 mV. This value represents the smallest signal change that the ADC can accurately distinguish. Anything below this is considered as the ADC's own noise and disturbances induced on the cable.

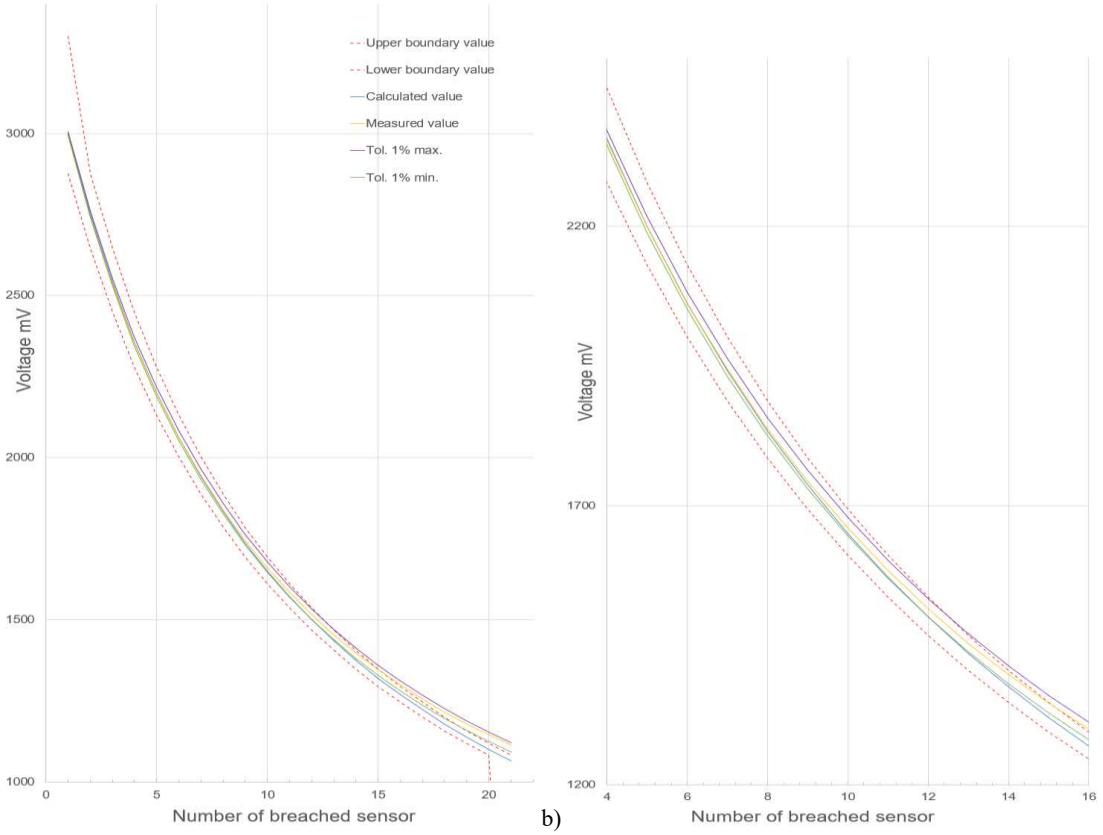

Fig. 13. Voltage at output of perimeter loop as a function of activated sensor; dashed curves indicate voltage variation limits for each activation state: a) Main graph; b) Enlarged section of graph

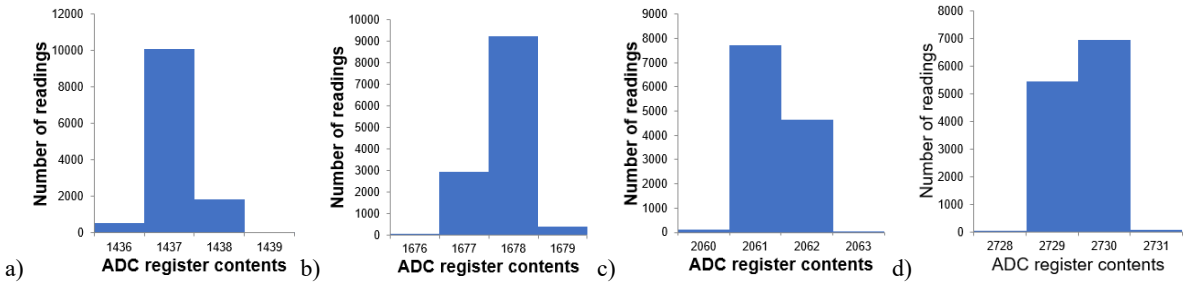

Fig. 14. Variation in values read from ADC: a) For 20th breached sensor; b) For 15th breached sensor; c) For 10th breached sensor; d) For 5th breached sensor

The level of noise depends on the electromagnetic environment in which the alarm system operates and cannot be estimated without specialized measurement equipment. However, during the calibration process, it is possible to determine the minimum voltage value below which the voltage on the perimeter line should not be reduced. This approach provides a margin for potential disturbances. Measurements from the test circuit allowed the determination of the minimum voltage value at which the system operates correctly. The smallest voltage change occurs with the activation of the last sensor and is 41 mV relative to the penultimate sensor. The maximum voltage change from the midpoint of the range for the last sensor's activation state is calculated as shown in Figure 7-a. This value is 20.5 mV and represents the threshold for the designed system using the calibration method. This confirms the proper operating of the system for a circuit with 20 sensors under real conditions. The chosen value is greater than the measured voltage changes to ensure a margin for additional disturbances.

4. Summary

In this work, an attempt was made to develop a proprietary configuration for a perimeter alarm system, which addresses the limitations of commercially systems available on the market. A parametrized configuration line was proposed that allows for the handling of a large number of sensors within a single monitoring loop. The developed perimeter loop ensures the ability to distinguish the activation of each detector using only two connecting wires. The developed system is compatible with any NC (normally closed) type sensors, such as reed switches. Additionally, the proposed connection method is straightforward to install, thanks to the standardization of the parametrizing resistors used in the circuit.

Furthermore, a calibration method was proposed to enhance the capabilities of the developed alarm system. The limit on the number of sensors depends on the accurate detection of individual states. Any disturbances in the circuit can lead to incorrect identification of which sensor has been triggered. Sensors near the end are most susceptible to this issue due to the small voltage change between consecutive activation states. Nevertheless, the calibration procedure significantly increases the number of access points in the perimeter loop.

To test the performance of the developed configuration for perimeter line, a demonstrator was built, consisting of an alarm control panel and an access manipulator. The effectiveness of the proposed connection configuration was tested in an alarm circuit with 20 sensors. Measurements showed that the most significant factor affecting the system's performance is the resistance of the connecting wires.

Thus, the Custom EOL configuration, utilizing the calibration method, can handle 20 access sensors in a single monitoring loop and allows for detecting where a breach has occurred. The system can successfully be used in facilities where multiple control points are required to be supervised and they are located close to each other. Moreover, the alarm system was designed with the possibility of expanding it with additional modules, such as a GSM wireless communication panel. An additional feature was the ability to operate even in case of a power failure, through the use of an emergency battery.

Literature

[1] G. Vardakis, G. Hatzivasilis, E. Koutsaki, N. Papadakis “Review of Smart-Home Security Using the Internet of Things,” *Electronics* 2024, vol. 13, no. 3343. doi: 10.3390/electronics13163343.

[2] T. Shi, P. Guo, R. Wang, Z. Ma, W. Zhang, W. Li, H. Fu, H. Hu, “A Survey on Multi-Sensor Fusion Perimeter Intrusion Detection in High-Speed Railways,” *Sensors* 2024, vol. 24, no. 5463, doi: 10.3390/s24175463.

[3] Andreas, C. R. Aldawira, H. W. Putra, N. Hanafiah, S. Surjarwo, A. Wibisurya, “Door Security System for Home Monitoring Based on ESP32,” *Procedia Computer Science* 2019, vol. 157, pp. 673-682, doi: 10.1016/j.procs.2019.08.218.

[4] M. H. Assaf, R. Mootoo, S. R. Das, E. M. Petriu, V. Groza, S. Biswas, “Sensor based home automation and security system,” *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, Graz, Austria, 2012, pp. 722-727, doi: 10.1109/I2MTC.2012.6229153.

[5] M. Kijima, Y. Miyagaw, H. Oshita, N. Segawa, M. Yazawa, and M. Yamamoto, “Poster Abstract: Multiple Door Opening/Closing Detection System Using Infrasound Sensor,” *2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, IEEE, Apr. 2018, pp. 126-127, doi: 10.1109/IPSN.2018.00026.

[6] S. Smith, J. Ellis, R. Abrams, “Chapter 8 - Central Alarm Stations and Dispatch Operations”, Editor(s): IFPO, “The Professional Protection Officer”, Butterworth-Heinemann, 2010, Pages 89-103, ISBN 9781856177467, doi: 10.1016/B978-1-85617-746-7.00008-0.

[7] George Risk Industries, Inc., “Magnetic Contacts with Built-In E.O.L Resistors and Resistor Packs,” *www.grisk.com*, access on 2024-11-12.

[8] H. Ali, H. Medjadba, L. M. Simohamed and R. Chemali, “Intrusion detection and classification using optical fiber vibration sensor,” *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, Tlemcen, Algeria, 2015, pp. 1-6, doi: 10.1109/CEIT.2015.7233060.

[9] M. Verma, R. S. Kaler, and M. Singh, “Sensitivity enhancement of Passive Infrared (PIR) sensor for motion detection,” *Optik (Stuttg),* Oct. 2021, vol. 244, p. 167503, doi: 10.1016/j.ijleo.2021.167503.

[10] M. E. Kalinkina, A. G. Korobeynikov, O. I. Pirozhnikova, N. A. Shmakov, and V. L. Tkalich, "Designing of reed switches for sensors and security alarm devices," *IOP Conf Ser Mater Sci Eng*, Feb. 2021, vol. 1100, no. 1, p. 012009, doi: 10.1088/1757-899X/1100/1/012009.
[11] T. Li, D. Han, J. Li, A. Li, Y. Zhang, R. Zhang, Y. Zhang, "Your Home is Insecure: Practical Attacks on Wireless Home Alarm Systems," IEEE INFOCOM 2021 – IEEE Conference on Computer Communications, Vancouver, BC, Canada, 2021, pp. 1-10, doi: 10.1109/INFOCOM42981.2021.9488873.
[12] S. Ramadhani and D. P. Putri, "Design of a Home Door Security System Based on NodeMCU ESP32 Using a Magnetic Reed Switch Sensor and Telegram Bot Application," S*inkron*, Oct. 2023, vol. 8, no. 4, pp. 2059-2068, doi: 10.33395/sinkron.v8i4.12688.
[13] M. Boroš, A. Vel'as, V. Šoltés, J. Dworzecki, "Influence of the Environment on the Reliability of Security Magnetic Contacts," *Micromachines* 2021, vol. 12, no. 401, doi: 10.3390/mi12040401.
[14] M. Boroš, A. Vel'as, Z. Zvaková, V. Šoltés, "New Possibilities for Testing the Service Life of Magnetic Contacts," *Micromachines* 2021, vol. 12, no. 479, doi: 10.3390/mi12050479.
[15] A. S. Devi, A. K, B. C, A. Shali, D. Kavitha, and S. Hemavathi, "Smart Security System," in *2022 1st International Conference on Computational Science and Technology (ICCST)*, IEEE, Nov. 2022, pp. 1–3, doi: 10.1109/ICCST55948.2022.10040301.
[16] K. Jakubowski, J. Paś, A. Rosiński, "The Issue of Operating Security Systems in Terms of the Impact of Electromagnetic Interference Generated," *Energies* 2021, vol. 14, no. 8591, doi: 10.3390/en14248591.
[17] G. Uçtu, M. Alkan, İ. A. Doğru, M. Dörterler, "Perimeter Network Security Solutions: A Survey," *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT),* Ankara, Turkey, 2019, pp. 1-6, doi: 10.1109/ISMSIT.2019.8932821.
[18] V Mach, A. Mizera, P. Stoklasek, M. Karhankova, M. Adamek, M. Bednarik, "Development of a Contact Glass-Break Detector for the Highest Security Level," *Sensors* 2024, vol. 24, no. 97, doi: 10.3390/s24010097.
[19] P. Teixidó, J. A. Gómez-Galán, R. Caballero, F. J. Pérez-Grau, J. M. Hinojo-Montero, F. Muñoz-Chavero, J. Aponte, "Secured Perimeter with Electromagnetic Detection and Tracking with Drone Embedded and Static Cameras," *Sensors* 2021, vol. 21, no. 7379, doi: 10.3390/s21217379.
[20] H. Xu, Y. Li, C. Ma, L. Liu, B. Wang, J. Li, "A Combined Sensing System for Intrusion Detection Using Anti-Jamming Random Code Signals," *Sensors* 2022, vol. 22, no. 4307, doi: 10.3390/s22114307.
[21] Ropam Elektronik, "Porównanie central i terminali Ropam Elektronik", (Product Catalogue), *ropam.com.pl*, access on 2024-11-12.
[22] Ropam Elektronik, "NeoLTE-IP-64, Neo-IP-64, NeoGSM-IP-64 Centrale alarmowe z komunikacją LTE/IP," (Installation Instructions), Document version: 2.0, 2023-01-23*, ropam.com.pl*, access on 2024-11-12.
[23] A Tyco International Company DSC, "PowerSeries Neo HS2016/HS2016-4/HS3032/HS2064/HS2064 E/HS2128/HS2129 E Alarm Controller Reference Manual," (Product Catalogue), *docs.johnsoncontrols.com/dsc*, access on 2024-11-12.
[24] A Tyco International Company DSC, "CENTRALE ALARMOWE HS2016/HS2032/HS2064/ HS2128," (Installation and Programming Instructions), Document version 1.1, *www.montersi.pl*, access on 2024-11-12.
[25] Paradox, "SP Spectra. Control Panel Comparision Chart," *paradox.ee*, access on 2024-11-12.
[26] Paradox, "Magellan/Spectra SP. Reference & Instalation" *paradox.ee*, access on 2024-11-12.
[27] Satel, "Made to Protect," (Product Catalogue), 2024, *www.satel.pl,* access on 2024-11-12.
[28] Editorial Staff, "Why we use End of Line (EOL) Resistor in Fire and Gas System?", Control and Instrumentation, *www.controlandinstrumentation.com*, access on 2024-11-12.
[29] System Automatycznej Kontroli Obiektu, "Parametryzacja NO/NC, EOL i 2EOL w systemach kontroli dostępu i alarmowych: Zalety parametryzacji wejść i stany konfiguracji" (NO/NC, EOL and 2EOL parameterization in access control and alarm systems: Advantages of input parameterization and configuration states), GMP Power, 21 November 2023, *sakokd.pl/blog*, access on 2024-11-12.

Advances in IT and Electrical Engineering

vol. 30, 2024, 49-59
https://doi.org/10.7862/re.2023.x

THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING
RZESZOW UNIVERSITY OF TECHNOLOGY

Original Research/Review

Impact of Artificial Intelligence on Computer Networks

Kacper Zdrojewski [1*]

[1] Department of Information Technology Networks, Regional Information Technology Center Warsaw, 00-909 Warsaw, Poland.

* Corresponding author. kacper.zdrojewski.1997@gmail.com

Received: 24 October 2024 / Accepted: 23 December 2024 / Published online: 31 December 2024

Abstract

The integration of artificial intelligence (AI) into computer networks has rapidly evolved, influencing network architecture, security measures, and traffic management. This paper explores AI's transformative impact on these areas, focusing on advancements in machine learning (ML), deep learning (DL), and reinforcement learning. These innovations are reshaping network security by improving threat detection and anomaly identification, as well as enhancing traffic management through predictive and adaptive routing. AI-driven systems are also making strides in automating network management tasks, allowing for more efficient resource allocation and self-healing networks. Despite these advancements, challenges remain, particularly concerning the integration of AI with legacy infrastructures and the ethical implications of AI decision-making processes.

Keywords: artificial intelligence, machine learning, network security, deep learning

1. Introduction

In recent years, the integration of artificial intelligence into various domains has revolutionized industries, with computer networks being a notable beneficiary of these advancements. As global internet traffic grows exponentially, driven by the proliferation of Internet of Things (IoT) devices, cloud computing, 5G technology, and the ever-increasing demand for bandwidth, the complexity of managing and securing networks has become a critical challenge. Traditional network management techniques, often based on manual configuration and rule-based systems, struggle to cope with this increasing complexity and the dynamic nature of modern networks.

Simultaneously, the number and sophistication of cyber threats continue to grow. Traditional intrusion detection systems (IDS) rely heavily on predefined signatures or rules to detect known threats. While effective against previously encountered attacks, these systems often fail to identify novel or evolving threats, such as zero-day vulnerabilities or advanced persistent threats (APTs) [2]. This has led to a paradigm shift towards more adaptive, AI-driven approaches.

AI, particularly machine learning and deep learning, offers promising solutions to these challenges. By leveraging vast amounts of historical and real-time data, AI models can learn traffic patterns, detect anomalies, and make decisions autonomously. For instance, AI-driven systems can automatically adjust traffic routing based on real-time congestion data, ensuring optimal performance and minimizing packet loss. In terms of security, AI systems can detect and mitigate threats more efficiently by identifying anomalous behavior that might indicate the presence of malicious activity [2, 3].

Moreover, reinforcement learning (RL) has enabled networks to adapt in real-time by optimizing routing paths and network configurations dynamically. These RL-based systems learn by continuously interacting with the network environment, making them ideal for highly dynamic network scenarios, such as mobile ad hoc networks (MANETs) or multi-cloud architectures.

This is an Open Access article distributed under the terms of the CC-BY-NC-ND 3.0 PL license, which permits others to distribute the work, provided that the article is not altered or used commercially. You are not required to obtain permission to distribute this article, provided that the original work is properly cited.

However, the deployment of AI in network environments also presents unique challenges. The integration of AI systems with legacy network infrastructure is often difficult due to hardware limitations or the lack of necessary computational resources. Additionally, the growing reliance on AI for critical network functions raises concerns regarding accountability, transparency, and the potential for bias in AI decision-making processes [1]. Despite these challenges, the potential benefits of AI integration into computer networks - such as enhanced security, more efficient traffic management, and autonomous network operations - are vast and continue to drive research and development in this area.

This paper delves into the multifaceted impact of AI on computer networks, examining how AI can address current networking challenges and predict future trends. Artificial intelligence refers to the simulation of human intelligence by machines, encompassing a broad range of techniques and approaches. These include machine learning, where systems learn from data to make predictions or decisions, and deep learning, a subset of ML that uses neural networks to model complex patterns in data. Computer networks refer to interconnected systems of devices and communication technologies that enable data exchange. These networks can be viewed from different perspectives, including physical infrastructure (e.g., wired or wireless networks), logical architecture (e.g., client-server or peer-to-peer models), and functional layers such as transport, application, or network layers in the OSI model. This article focuses on how AI can enhance the management and optimization of such networks across various layers, particularly in areas like traffic management, security, and resource allocation.

2. AI in Network Security

AI has become indispensable in securing modern computer networks, where the volume and complexity of cyberattacks are constantly growing. Traditional network security methods, such as firewalls and signature-based intrusion detection systems, are increasingly insufficient in the face of sophisticated threats. AI enhances network security by enabling dynamic threat detection and rapid response to anomalies. ML-based IDS can analyze historical network traffic data to identify malicious patterns, while DL systems improve this capability by learning from unstructured data and detecting novel attacks.

2.1. Machine Learning in Intrusion Detection

Intrusion detection focuses on identifying unauthorized access and abnormal activities within network environments. The primary challenge lies in distinguishing between legitimate and malicious traffic in real-time, a task complicated by the diversity and scale of modern networks. The effectiveness of an IDS is typically measured using metrics such as detection rate, false positive rate, precision, recall, and F1 score [12, 13]. Traditional methods, relying on predefined signatures (such as Snort[1] or AIDE[2]), struggle to detect novel threats such as zero-day attacks, highlighting the need for adaptive, intelligent systems. Zero-day attacks are cyber exploits targeting unknown vulnerabilities in software or hardware, leaving no time for defensive measures.

Machine learning has revolutionized intrusion detection by providing more accurate and scalable solutions compared to rule-based systems. Supervised learning algorithms, such as Support Vector Machines (SVM) and Random Forests, are trained on labeled datasets to distinguish between legitimate and malicious traffic. However, one limitation is their reliance on large datasets for training, which can result in performance issues when faced with zero-day attacks. On the other hand, unsupervised learning, including anomaly detection, enables the identification of previously unknown threats by analyzing deviations from normal traffic behavior.

Detecting malicious traffic can be achieved by analyzing its behavior, such as deviations in packet flow, unusual connection frequencies, or irregular user activity patterns, a principle employed by behavioral Intrusion Detection Systems [14]. The authors of [14] designed a multi-level intrusion detection method for identifying abnormal network behaviors using machine learning techniques. Their approach integrates multiple classifiers to detect anomalies at various levels of analysis, starting with broad detection of unusual traffic patterns and proceeding to detailed evaluation of specific threats, such as Distributed Denial of Service (DDoS) attacks or port scans. The method demonstrated improved accuracy

[1] https://www.snort.org

[2] https://aide.github.io

in identifying zero-day attacks while significantly reducing false positives, showcasing the effectiveness of hierarchical analysis in intrusion detection systems.

Building on this solution, the authors of [12] developed a specialized IoT crawler integrated into the fog layer (network layer), designed to prioritize critical nodes for inspection based on their significance within the network. The IoT crawler utilizes a behavioral analyzer with a machine learning core to differentiate between malicious and legitimate nodes based on data streams collected from IoT devices. The proponents of this idea evaluated this system using machine learning algorithms like Random Forest, AdaBoost, and Extra Tree, achieving a remarkable 98.3% accuracy with the Extra Tree classifier. This result highlights the system's ability to process IoT-specific data streams efficiently, adapting dynamically to threats without overwhelming resource-constrained IoT nodes. This work demonstrates the potential of integrating multi-level machine learning frameworks within IoT ecosystems to enhance real-time intrusion detection. By leveraging fog computing for reduced latency and prioritizing critical nodes, the IoT crawler exemplifies the practical application of advanced machine learning in securing modern, heterogeneous networks.

Mukkamala [6] described approaches to intrusion detection and audit data reduction using support vector machines and Neural Networks, highlighting their effectiveness in analyzing high-dimensional datasets. The primary goal was to create classifiers capable of distinguishing normal network traffic from various types of attacks. The researchers compared the performance of SVM with a radial kernel to Neural Networks with two intermediate layers. The results demonstrated that SVM with a radial kernel outperformed Neural Networks in terms of hit ratio and processing time for both model training and prediction tasks. This research underscores the potential of SVM in efficiently processing and classifying network traffic, making it a robust solution for real-time intrusion detection in environments characterized by large-scale, complex datasets.

Similarly, Kim [7] introduced a hybrid intrusion detection model that integrates decision trees (DT) with one-class SVM to combine the strengths of anomaly detection and misuse detection. This approach addresses the limitations of standalone methods by leveraging the high accuracy of misuse detection for known attack patterns and the adaptability of anomaly detection to identify previously unknown threats. The researchers proposed a two-stage framework. In the first stage, DT, as a supervised learning algorithm, are employed to quickly and effectively classify known attack patterns. This ensures reliable identification of previously encountered threats. In the second stage, a one-class SVM is used to analyze the residual data (traffic not classified in the first stage) for potential anomalies. This step enhances the system's ability to detect zero-day attacks and other novel intrusions. The study evaluated the hybrid method using benchmark datasets, demonstrating that the integration of these techniques improves both detection accuracy and processing efficiency. Specifically, the model achieved a significant reduction in false positives while maintaining high detection rates, particularly for mixed traffic scenarios.

The article [16] categorizes various IDS based on deep learning techniques. The authors explore how models like autoencoders and Long Short-Term Memory (LSTM) networks can detect anomalies in network traffic more effectively than traditional signature-based systems. Autoencoders, as unsupervised learning models, excel at detecting anomalies in network traffic by reconstructing input data and identifying deviations that signify potential threats. LSTM networks, on the other hand, are particularly effective in modeling sequential data, such as network logs or traffic flows, enabling the detection of complex temporal patterns associated with sophisticated attacks. The study showcases that deep learning enables real-time analysis and reduces false positives significantly, addressing a key limitation of conventional IDS.

2.2. Deep Learning for Enhanced Security

Deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), offer superior performance in threat detection by analyzing more complex features. For example, RNNs, with their ability to capture temporal dependencies, are effective in detecting attacks like Distributed Denial of Service, where the timing of events is crucial. Furthermore, DL-based systems can operate autonomously, learning from new attack patterns without requiring frequent human intervention.

Kou in [4] proposed a network security situational element recognition method combining a deep-stacked encoder with the backpropagation (BP) neural network algorithm. The method leverages unsupervised learning algorithms to train each layer of the network individually, allowing for a hierarchical representation of the data. By stacking the encoders, the method creates a deep-stacked network capable

of extracting high-dimensional features from raw network data. In the initial stage, unsupervised training is conducted using the stacked encoders to learn meaningful features from unlabeled data. These encoders utilize reconstruction-based loss functions to ensure that the most relevant information is retained during feature extraction. Once the unsupervised training phase is complete, the BP algorithm is applied to fine-tune the network using labeled data, optimizing it for classification tasks. The authors conducted simulation studies to evaluate the effectiveness of this method in enhancing situational awareness in network security. The results indicated that the deep-stacked encoder significantly outperformed traditional models in terms of precision and recall when recognizing situational elements such as potential threats, vulnerabilities, and network states. Moreover, the method demonstrated resilience against noisy and incomplete data, which are common in real-world network environments.

In addition, Fu in [5] proposed to use an adaptive genetic algorithm to effectively optimize the traditional APT attack prediction model, thereby improving prediction accuracy. This model's ability to accurately predict risk nodes that may be present in the network system as well as to track the progress of APT attacks in real time and determine the attack path through sequence attacks greatly enhances the network system's security [1].

In 2018, Radford [17] presented an anomaly detection model using a LSTM network to analyze network traffic logs for cybersecurity applications. The model was designed to leverage the temporal sequence learning capabilities of LSTM networks, enabling it to identify anomalies in network behavior that might indicate potential security threats.

In [18] authors introduced RawPower, a DL architecture designed to analyze raw bytestream data for network anomaly detection. This approach eliminates the need for extensive feature engineering by directly processing raw traffic measurements. The experimental results demonstrate that RawPower achieves exceptional performance, surpassing traditional anomaly detection systems in terms of detection accuracy, robustness, and scalability. Specifically, it excels in scenarios involving high-speed networks, where traditional methods struggle to keep up with the sheer volume and diversity of data.

The studies [19, 20] investigate the application of deep learning techniques for identifying various types of malicious network activities, such as malware communication. By utilizing CNNs to analyze packet-level features, the method achieves superior performance in detecting previously unseen threats. The authors compare the results with conventional techniques and highlight significant improvements in precision and recall metrics.

The authors of [15] provides a thorough review of how deep learning techniques are applied to various domains of network security, highlighting their significant advantages over traditional, rule-based systems. The authors analyze several deep learning architectures, including CNNs and RNNs, demonstrating their ability to automatically extract and learn meaningful patterns from raw network data without the need for extensive manual feature engineering. The survey emphasizes the limitations of rule-based systems, which rely on predefined rules or signatures, making them ineffective against zero-day attacks and adaptive threats. In contrast, deep learning models are dynamic and capable of generalizing to unseen attack scenarios. By analyzing both known and novel attack patterns, these methods significantly enhance detection rates and reduce false positives, which are common drawbacks of conventional systems.

The article [21] highlights the transformative role of deep learning in securing 5G networks. The authors discuss how the high complexity and dynamic nature of 5G architectures, including virtualization, software-defined networking (SDN), and network slicing, demand advanced security mechanisms capable of real-time threat detection and mitigation. The study concludes that integrating deep learning into 5G security frameworks significantly enhances the adaptability, precision, and scalability of network defenses, making them more resilient to evolving threats. However, it also underscores the need for addressing challenges such as computational overhead and the interpretability of deep learning models to ensure their effective deployment in real-world 5G applications.

3. Traffic Optimization with AI

Efficient traffic management is critical in maintaining network performance, particularly as the number of connected devices and the amount of data traffic continue to grow. AI's ability to predict traffic patterns and optimize routing decisions in real-time has significantly improved the efficiency of networks. AI-based systems analyze historical and real-time data to dynamically adjust traffic routes and manage bandwidth, thereby reducing congestion and packet loss.

3.1. Predictive Traffic Routing

AI can predict network traffic fluctuations by analyzing historical data and recognizing patterns that suggest future behavior. By leveraging ML models, networks can proactively re-route data to avoid congestion before it occurs, improving Quality of Service (QoS) for end-users. QoS refers to the ability of a network to provide predictable performance levels, typically measured by parameters such as bandwidth, latency and packet loss. It ensures that network resources are allocated efficiently to maintain the reliability and quality of specific applications or services, such as video streaming or VoIP. Predictive routing is particularly beneficial in large-scale networks, such as data centers or cloud infrastructures, where traffic load balancing is essential for maintaining optimal performance.

The study presented in [30] focus on predicting the network traffic by using the different prediction regression models such as K-Nearest Neighbors, Random Forest, Gradient Boosting and DT with different sub-parameters. Using real-world network traffic data, the authors train and test the models to predict key traffic parameters, such as bandwidth demand and packet flow rates. The results demonstrate that Gradient Boosting outperforms the other algorithms in terms of accuracy and error metrics, such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

One notable study [8] presents a framework for traffic flow classification based on deep learning models. The authors train deep neural networks (DNNs) on real-world network traffic data to predict characteristics such as flow throughput and duration. Their approach moves beyond binary classifications like "mice" (small) and "elephant" (large) flows [23], opting instead for a multi-class quantization strategy. This methodology classifies flows into a range of categories based on their characteristics (flow throughput, duration, or packet interarrival times), rather than relying on rigid binary distinctions. The proposed system is intended to enhance network traffic management by predicting flow behaviors, ultimately improving routing decisions in real time.

The authors of [9, 31] explored the application of AI in predicting network traffic patterns to enhance routing efficiency in smart networks. It reviews various AI methodologies, including machine learning techniques, and discusses their roles in optimizing resource allocation and reducing latency. Techniques such as RNNs and LSTM networks are discussed for their ability to analyze temporal traffic data and accurately forecast future traffic demands. These predictions enable dynamic adjustments to routing protocols, reducing congestion and enhancing QoS across diverse network environments, including 5G, IoT, and edge computing. Despite the advancements, the article identifies several challenges associated with AI integration in network traffic management, such as scalability in large-scale networks, maintaining prediction accuracy in highly dynamic environments, and computational overhead. The authors suggest that future research should focus on lightweight AI models, federated learning to address data privacy concerns, and explainable AI (XAI) to improve the interpretability and trustworthiness of predictive systems.

3.2. Reinforcement Learning for Dynamic Traffic Management

Reinforcement learning, a type of machine learning where agents learn by interacting with their environment, has been applied to dynamic traffic management. RL agents learn optimal routing strategies through trial and error, adjusting decisions based on rewards, such as reduced latency or higher throughput. This approach allows networks to adapt in real-time, adjusting to changing conditions without the need for human intervention.

The article [10] explores the application of reinforcement learning for adaptive routing in networks subject to dynamic changes. The authors present an RL framework that dynamically learns optimal routing policies by interacting with the network environment, thereby facilitating efficient traffic management and minimizing delays, even under unpredictable traffic fluctuations and varying network topologies. The study highlights the potential of RL to enhance routing performance in scenarios where traditional algorithms may struggle due to variability in network conditions.

Abrol in [22] presents a framework leveraging deep reinforcement learning (DRL) to optimize network traffic management dynamically. The authors propose a model that integrates a deep graph convolutional neural network with a reinforcement learning agent to predict and adapt to real-time traffic demands. This approach is particularly suited for next-generation networks, which face challenges such as high data volumes, dynamic topologies, and diverse service requirements. By modeling the network as a graph, the DRL agent learns optimal routing policies by interacting with the network environment and receiving feedback through reward signals. These signals are designed to reflect key performance metrics, such as throughput, latency, and packet loss. Over time, the model identifies patterns in traffic

behavior and dynamically adjusts routing decisions to prevent congestion and maximize resource utilization. The approach minimizes packet delays and reduces congestion, leading to improved QoS for users and applications.

Q-Learning [24] (QL) uses unsupervised RL to determine optimal behaviour to maximise performance when interacting with its environment. The method has also found its way into network traffic management and optimization in SDN. The authors of [25] addressed network congestion in SDN by reselecting flow paths and changing flow table using predefined threshold. The researchers in [26] introduced fairness function in SDN for load-balancing in peak traffic conditions. Harewood-Gill [27] proposed three Q-routing algorithms [28] with distinct performance metrics to enhance traffic management in SDN environments and conducted a comparative analysis of their effectiveness against the K-Shortest Path algorithm. A more detailed description of these articles, along with additional examples of QL applications in SDN, is provided by the authors in [29].

4. AI in Network Management

AI has also revolutionized network management by automating routine tasks such as configuration management, fault detection, and network monitoring. AI-driven network management systems can identify and resolve issues autonomously, reducing human workload and minimizing downtime.

In traditional networks, configuration management is a labor-intensive process prone to human error. AI tools automate the configuration process, ensuring consistency and reducing the risk of misconfigurations. By analyzing network requirements, AI systems can automatically apply optimal settings and adjust them as network demands evolve.

Moreover, AI systems excel at detecting anomalies in network performance, which can be indicators of hardware failure, security breaches, or performance degradation. By using ML models, these systems can predict failures before they occur and take preventive action, such as rerouting traffic or initiating backup systems. Some AI-driven networks even exhibit self-healing capabilities, where the system automatically corrects issues without human intervention. Examples of the use of AI solutions in network management are described in [8, 9, 11, 32, 39].

The article [33] discusses the integration of machine learning techniques into cognitive network management systems to enhance decision-making processes and automate network operations. The authors emphasize the segmentation of network management into distinct areas - Fault, Configuration, Accounting, Performance, and Security (FCAPS) - and the assignment of specific ML algorithms to address challenges in each domain. Furthermore, they underscore that developing an integrated network management system is a highly complex yet indispensable task, particularly in light of the rapid expansion of computer networks in recent years.

Li in [34] explores the transformative role of AI and ML in enhancing the management of data center networks. It provides a detailed analysis of how ML techniques are being employed to address the growing complexity of modern data centers by enabling adaptive, automated, and efficient network management solutions. A notable contribution of the survey is the introduction of a quality assessment criterion called REBEL-3S, designed to impartially evaluate the strengths and weaknesses of the proposed research approaches.

Kadiyala in [38] examines the groundbreaking potential of AI in network automation, emphasizing its ability to predict and prevent network issues, optimize resources, and enable self-healing capabilities. It presents real-world case studies demonstrating AI's effectiveness in enhancing network reliability and reducing downtime.

An existing technological solution utilizing AI for network infrastructure management is Cisco AI Network Analytics[3]. This application is designed to enhance network management by leveraging AI and ML to provide proactive insights and automated solutions [35]. The platform collects and analyzes vast amounts of telemetry data from network devices, enabling it to identify anomalies, predict potential performance issues, and optimize network configurations in real time. A key advantage of Cisco AI Network Analytics is its ability to automate routine tasks, such as identifying misconfigurations or reallocating bandwidth, reducing the need for manual intervention and minimizing operational costs.

[3] https://www.cisco.com/c/en/us/td/docs/cloud-systems-management/network-automation-and-management/dna-center-assurance/2-3-7/b_cisco_dna_assurance_2_3_7_ug/b_cisco_dna_assurance_2_3_6_ug_chapter_010.pdf

Similarly, Juniper Networks' AI-driven network management solution, Marvis, employs ML to proactively predict network issues and deliver actionable insights [36]. By analyzing data from various sources - including network devices, applications, and user behavior - Marvis identifies patterns and anomalies with high accuracy. In a customer deployment, Marvis successfully predicted 90% of network issues, reducing the mean time to resolution (MTTR) by 70% and increasing network uptime by 25% [37], showcasing its impact on operational efficiency and reliability.

5. Challenges and Future Directions

Despite the numerous advantages of integrating AI into computer networks, there are several challenges that must be addressed. One key issue is the difficulty of integrating AI into existing legacy network infrastructure, which often lacks the computational power or flexibility required for AI systems. Additionally, the deployment of AI in critical networks raises ethical concerns, including questions of accountability, transparency, and bias in decision-making processes.

5.1. Ethical and Accountability Concerns

As AI systems become more autonomous, determining accountability in the event of network failures or security breaches becomes increasingly complex. The opaque nature of some AI algorithms, particularly deep learning models or decision trees, makes it difficult to understand how decisions are made. This is why many AI systems operate as "black boxes", making it difficult for users and stakeholders to understand how decisions are made. This lack of transparency raises accountability issues, particularly in scenarios where AI decisions impact user safety or privacy. Establishing clear accountability frameworks is essential; organizations must define who is responsible for AI-driven decisions, whether it be the developers, the companies that deploy these systems, or designated administrators. This clarity can help foster trust and facilitate ethical governance of AI technologies.

The findings reported in [40] highlight practical issues of developing and operating ML-based solutions in real networks. The authors discuss concerns related to data quality and availability, emphasizing that the effectiveness of AI systems heavily relies on large, high-quality datasets, which are often difficult to obtain or maintain due to privacy regulations and dynamic network environments. It is worth mentioning that gathering sufficient data from diverse network environments is often hindered by logistical limitations, such as incompatible data formats, the high cost of data acquisition, and varying network configurations. One more significant challenge lies in data labelling. Supervised learning algorithms, which are commonly used in network management tasks like traffic classification and anomaly detection, require labeled datasets for training. Labeling network data is both time-consuming and resource-intensive, often requiring expert knowledge to correctly identify patterns or categorize flows. This bottleneck can slow down the development cycle of AI-driven solutions and limit the scope of their applicability.

Another major issue is algorithmic bias, which can emerge from the data sets used to train AI systems. If these data sets reflect historical biases or imbalances, the resulting AI algorithms may reinforce or exacerbate these biases in decision-making processes. For example, an AI system employed for network traffic management might unintentionally prioritize data flows from certain applications or user groups, leading to unequal access to bandwidth and resources. Addressing this concern requires ongoing scrutiny of training data and the implementation of measures that ensure fairness and inclusivity in AI applications, promoting equitable service distribution across all users.

Furthermore, the ethical implications of data privacy in computer networks cannot be overlooked. AI systems often require vast amounts of network traffic data to function effectively, raising concerns about how this data is collected, stored, and used. Users may be unaware of the extent to which their online activities are being monitored and analyzed, leading to potential violations of privacy rights. Organizations must prioritize ethical data management practices, ensuring that users are informed about data usage and that consent is obtained for data collection. Establishing regulatory frameworks that protect user privacy while allowing for innovation in AI technologies will be a crucial step forward in fostering trust and accountability within network systems.

5.2. The Future of AI-Driven Networks

The future of AI-driven networks is set to revolutionize the way we manage and interact with digital infrastructures. One of the most promising advancements lies in the development of self-optimizing networks. These networks will leverage AI algorithms to analyze real-time data, enabling them to dynamically adjust their performance based on current conditions. This capability could significantly enhance efficiency, as networks become adept at reallocating resources and bandwidth in response to changing demands. As a result, users can expect faster, more reliable connections, improving overall user experiences across various applications.

Moreover, AI-driven networks are likely to play a pivotal role in enhancing security measures. The integration of AI in cybersecurity can enable proactive threat detection and response, as systems learn to identify and mitigate potential vulnerabilities before they can be exploited. By analyzing patterns in network traffic, AI can differentiate between legitimate user behavior and suspicious activities, significantly reducing the risk of cyberattacks. However, this increased reliance on AI also necessitates the development of robust safeguards to ensure that these systems themselves do not become targets for manipulation or exploitation.

When discussing the future of computer networks, it is impossible to overlook the transformative role of generative AI. This technology is set to revolutionize network optimization, security, and automation through its innovative applications. One such use case is synthetic data generation, where models like Generative Adversarial Networks (GANs) [41] and diffusion models create realistic network traffic patterns. These synthetic datasets address the challenges of limited or biased real-world data, enhancing the robustness and adaptability of AI systems in managing network operations.

In the area of network security, generative AI has proven its utility in simulating complex cyberattacks, such as DDoS or phishing scenarios [42]. By proactively testing security systems against these simulated threats, networks can better anticipate and counter emerging vulnerabilities. However, this dual-use potential also introduces risks, as attackers might exploit generative AI to craft more sophisticated and hard-to-detect malicious traffic, necessitating effective protections and ethical guidelines.

6. Conclusion

The impact of artificial intelligence on computer networks is profound and multifaceted, offering both significant benefits and challenges. AI technologies have the potential to revolutionize network management, enhancing efficiency, reliability, and security. Through the use of advanced algorithms, networks can achieve greater self-optimization and self-healing capabilities, leading to improved performance and reduced downtime. This transformation is particularly crucial in an era where demand for bandwidth and responsiveness continues to grow, necessitating innovative solutions to manage complex network environments. Table 1 provides a summary of artificial intelligence applications and their transformative impact on key areas of computer networks.

Table 1. AI Applications and Their Impact on Key Network Areas

Computer Network Area	Traditional Challenges	AI Approaches	Impact of AI
Traffic Management	Network congestion, inefficient routing	Predictive analytics, reinforcement learning	Improved routing, reduced latency, dynamic traffic optimization
Network Security	Threat detection delays, zero-day attack detection, advanced persistent threats	Anomaly detection, generative AI, ML classifiers, neural networks	Faster threat detection, adaptive defense mechanisms, reduced false positives, effective prediction of APT and zero-day attacks
Resource Allocation	Static resource distribution, underutilization	AI-based optimization models, deep learning	Efficient bandwidth management, adaptive resource distribution
Fault Detection	Manual monitoring, delayed detection of hardware or software failures, delayed troubleshooting	Predictive maintenance, neural networks	Early failure detection, minimized downtime, automated troubleshooting
Quality of Service	Packet loss, inconsistent service quality	AI-driven traffic prioritization, reinforcement learning	Enhanced user experience, optimized service delivery

Network Design and Planning	Complex manual configurations, Complexity in multi-cloud or MANET scenarios	Generative models (e.g., GANs), optimization techniques	Automated network topology design, scalability, reduced human error

However, the integration of AI into computer networks is not without its drawbacks. Ethical considerations, particularly around data privacy and algorithmic bias, pose serious challenges that must be addressed to foster trust among users and stakeholders. As AI systems become increasingly autonomous in their decision-making processes, ensuring accountability becomes critical. Organizations must implement transparent practices and ethical guidelines that govern AI usage, ensuring that user data is handled responsibly and equitably. A groundbreaking step toward the responsible and ethical development of this technology is the EU Artificial Intelligence Act[4], which represents the world's first comprehensive legal regulation for artificial intelligence systems and models.

Looking ahead, the future of AI-driven networks necessitates a collaborative approach that integrates the insights of technologists, ethicists, and policymakers. This interdisciplinary cooperation is crucial for establishing standards that promote technological advancement while safeguarding users' rights and classified information. By balancing innovation with ethical considerations, the full potential of artificial intelligence can be harnessed for applications in computer networks, ensuring that these technologies enhance their integrity, accessibility, and security.

Literature

[1] Wang M., Song G., Zhang Y.: The Current Research Status of AI-Based Network Security Situational Awareness. Electronics (2023), 12, 2309. https://doi.org/10.3390/electronics12102309.

[2] Amrollahi M., Hadayeghparast S., Karimipour H., Derakhshan F., Srivastava G.: Enhancing Network Security Via Machine Learning: Opportunities and Challenges, (2020), In: Choo, KK., Dehghantanha, A. (eds) Handbook of Big Data Privacy. Springer, Cham. https://doi.org/10.1007/978-3-030-38557-6_8.

[3] De Lucia, M.J., Srinivasan, A.: Artificial Intelligence and Machine Learning for Network Security: Quo Vadis?, (2024), In: Chen, Y., Wu, J., Yu, P., Wang, X. (eds) Network Security Empowered by Artificial Intelligence. Advances in Information Security, vol 107. Springer, Cham. https://doi.org/10.1007/978-3-031-53510-9_3.

[4] Kou G., Wang S., Zhang D.: Recognition of network security situation elements based on depth stack encoder and back propagation algorithm. J. Electron. Inf. Technol. 2019, 41, 2187–2193.

[5] Fu T., Lu Y., Zhen W.: APT attack situation assessment model based on optimized BP neural network. In Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (IT-NEC), IEEE, Chengdu, China, 15–17 March 2017; pp. 2108–2111.

[6] Mukkamala S., Janoski G., Sung A.: Intrusion detection using neural networks and support vector machines. Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290), Honolulu, HI, USA, 2002, pp. 1702-1707.

[7] Kim G., Lee S., Kim S.: A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection With Misuse Detection. Expert Systems with Applications, vol. 41, no. 4, pp. 1690–1700, 2014.

[8] Hardegen C., Pfülb B., Rieger S., Gepperth A.: Predicting Network Flow Characteristics Using Deep Learning and Real-World Network Traffic. IEEE Transactions on Network and Service Management, vol. 17, no. 4, pp. 2662-2676, (2020).

[9] Chen A., Law J., Aibin M.: A Survey on Traffic Prediction Techniques Using Artificial Intelligence for Communication Networks. Telecom 2021.

[10] Khodayari S., Yazdanpanah M.: Network routing based on reinforcement learning in dynamically changing networks. 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05), 2005.

[11] Sivalingam K.: Applications of Artificial Intelligence, Machine Learning and related techniques for Computer Networking Systems. (2021).

[12] Albdour L., Manaseer S., Sharieh A.: IoT Crawler with Behavior Analyzer at Fog layer for Detecting Malicious Nodes. International Journal of Communication Networks and Information Security, vol. 12, pp. 83-94. https://doi.org/10.17762/ijcnis.v12i1.4459, (2020).

[13] Deng X., Liu Q., Deng Y., Mahadevan S.: An improved method to construct basic probability assignment based on the confusion matrix for classification problem. Information Sciences, vol. 340–341, pp. 250–261, May 2016.

[4] https://artificialintelligenceact.eu/

[14] Ji S.-Y., Jeong B.-K., Choi S., Jeong D.: A multilevel intrusion detection method for abnormal network behaviors. J. Netw. Comput. Appl., vol. 62, pp. 9–17, February 2016. https://doi.org/10.1016/j.jnca.2015.12.004.
[15] Berman D. S., Buczak A. L., Chavis J. S., Corbett C. L.: A Survey of Deep Learning Methods for Cyber Security. Information (2019), vol. 10(4), 122. https://doi.org/10.3390/info10040122.
[16] Lansky J. et al.: Deep Learning-Based Intrusion Detection Systems: A Systematic Review. IEEE Access, vol. 9, pp. 101574-101599, 2021, doi: 10.1109/ACCESS.2021.3097247.
[17] Radford B., Apolonio L., Trias A., Simpson J.: Network Traffic Anomaly Detection Using Recurrent Neural Networks. (2018), 10.48550/arXiv.1803.10769.
[18] Marín G., Casas P., Capdehourat G.: RawPower: Deep Learning based Anomaly Detection from Raw Network Traffic Measurements. In Proceedings of the ACM SIGCOMM 2018 Conference on Posters and Demos (SIGCOMM '18). Association for Computing Machinery, New York, NY, USA, pp. 75–77. https://doi.org/10.1145/3234200.3234238.
[19] Marín G., Casas P., Capdehourat G.: Deep in the Dark - Deep Learning-Based Malware Traffic Detection Without Expert Knowledge. IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 2019, pp. 36-42, doi: 10.1109/SPW.2019.00019.
[20] Marín G., Casas P., Capdehourat G.: DeepMAL - Deep Learning Models for Malware Traffic Detection and Classification. (2021), doi: 10.1007/978-3-658-32182-6_16.
[21] Haider N., Baig M. Z., Imran M.: Artificial Intelligence and Machine Learning in 5G Network Security: Opportunities, advantages, and future research trends. (2020), doi: 10.48550/arXiv.2007.04490.
[22] Abrol A., Mohan P. M., Truong-Huu T.: A Deep Reinforcement Learning Approach for Adaptive Traffic Routing in Next-gen Networks. (2024), https://doi.org/10.48550/arXiv.2402.04515.
[23] Mori T., Uchida M., Kawahara R., Pan J., Goto S.: Identifying elephant flows through periodically sampled packets. Proc. 4th ACM SIGCOMM Conf. Internet Meas., 2004, pp. 115–120.
[24] Watkins C.: Learning From Delayed Rewards. Ph.D. dissertation, King's College, Cambridge, UK, (1989).
[25] Kim S., Son J., Talukder A., Hong C. S.: Congestion prevention mechanism based on Q-leaning for efficient routing in SDN. 2016 International Conference on Information Networking (ICOIN), Kota Kinabalu, Malaysia, pp. 124-128, doi: 10.1109/ICOIN.2016.7427100.
[26] Tennakoon D., Karunarathna S., Udugama B.: Q-learning Approach for Load-balancing in Software Defined Networks. 2018 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, pp. 1-6, doi: 10.1109/MERCon.2018.8421895.
[27] Harewood-Gill D., Martin T., Nejabati R.: The Performance of Q-Learning within SDN Controlled Static and Dynamic Mesh Networks. (2020), pp. 185-189, doi: 10.1109/NetSoft48620.2020.9165530.
[28] Boyan J., Littman M.: Packet Routing in Dynamically Changing Networks: A Reinforcement Learning Approach. Advances in Neural Information Processing Systems, vol. 6, (1999).
[29] Dake D., Gadze D., Klogo G., Nunoo-Mensah H.: Traffic Engineering in Software-defined Networks using Reinforcement Learning: A Review. International Journal of Advanced Computer Science and Applications, vol. 12, (2021), doi: 10.14569/IJACSA.2021.0120541.
[30] P S., Kamboj A., Shete V., R H.: Machine Learning Based Network Traffic Predictive Analysis. Review of Computer Engineering Research, vol. 9(2), pp. 96–108, (2022). https://doi.org/10.18488/76.v9i2.3065.
[31] Vashishth T., Sharma V., Kumar B., Chaudhary S., Panwar R., Sharma S.: ARTIFICIAL INTELLIGENCE-ENABLED NETWORK TRAFFIC OPTIMIZATION: A COMPREHENSIVE SURVEY. Journal of Industrial Engineering, vol. 52, pp. 26-34, (2023).
[32] Sivalingam K.: Applications of Artificial Intelligence, Machine Learning and related techniques for Computer Networking Systems. (2021), https://arxiv.org/abs/2105.15103.
[33] Ayoubi S., Limam N., Salahuddin M., Shahriar N., Boutaba R., Estrada-Solano F., Caicedo M.: Machine Learning for Cognitive Network Management. IEEE Communications Magazine, vol. 56, (2018), doi: 10.1109/MCOM.2018.1700560.
[34] Li B., Wang T., Yang P., Chen M., Yu S., Hamdi M.: Machine Learning Empowered Intelligent Data Center Networking: A Survey. (2022), https://doi.org/10.48550/arXiv.2202.13549.
[35] Cisco: Cisco DNA Center: Intent-Based Networking for the Enterprise. Solution Overview, 2020.
[36] Juniper Networks: Marvis: AI-Driven Virtual Network Assistant. Datasheet, 2021.
[37] Juniper Networks: Global Retailer Achieves Network Efficiency and Uptime with Juniper Marvis. Case Study, 2021.
[38] Kadiyala C., Chilukoori S., Gangarapu S.: AI-Powered Network Automation: The Next Frontier in Network Management. Journal of Advanced Research Engineering and Technology, vol. 3, pp. 223-233, (2024).

[39] Ge J., Li T., Wu Y.: AI and Machine Learning for Network and Security Management. Wiley-IEEE Press, (2022).
[40] Liu Q., Zhang T., Hemmatpour M., Qiu H., Zhang D., Chen C. S., Mellia M., Aghasaryan A.: Operationalizing AI in Future Networks: A Bird's Eye View from the System Perspective. (2024), https://doi.org/10.48550/arXiv.2303.04073.
[41] Nukavarapu S. K., Ayyat M., Nadeem T.: MirageNet - Towards a GAN-based Framework for Synthetic Network Traffic Generation. GLOBECOM 2022 - 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 2022, pp. 3089-3095, doi: 10.1109/GLOBECOM48099.2022.10001494.
[42] Sai S., Yashvardhan U., Chamola V., Sikdar B.: Generative AI for Cyber Security: Analyzing the Potential of ChatGPT, DALL-E and Other Models for Enhancing the Security Space. IEEE Access, (2024), vol. 12, pp. 53497-53516, doi: 10.1109/ACCESS.2024.3385107.