Харківський національний університет радіоелектроніки
Державне підприємство
"Південний державний
проектно-конструкторський та науково-дослідний
інститут авіаційної промисловості"

Kharkiv National University of Radio Electronics
State Enterprise
"Southern National Design & Research Institute of Aerospace Industries"

Сучасний стан наукових досліджень та технологій в промисловості

Innovative technologies and scientific solutions for industries

Щоквартальний науковий журнал

№ 3 (33), 2025

Затверджений до друку
Науково-технічною Радою
Харківського національного університету радіоелектроніки
(Протокол № 8 від 26 вересня 2025 р.)

ЗАСНОВНИКИ
Харківський національний університет радіоелектроніки,
Державне підприємство "Південний державний проектно-конструкторський та науково-дослідний інститут авіаційної промисловості"

АДРЕСА РЕДАКЦІЇ:
Україна, 61166, м. Харків, проспект Науки, 14
Інформаційний сайт: http://itssi-journal.com
https://journals.uran.ua/itssi
E-mail редколегії: journal.itssi@gmail.com

Quarterly scientific journal

No. 3 (33), 2025

Approved for publication
by the Scientific and Technical Council
of the Kharkiv National University of Radio Electronics
(Protocol No. 8 d.d. 26 September 2025)

ESTABLISHERS
Kharkiv National University of Radio Electronics,
State Enterprise
"National Design & Research Institute
of Aerospace Industries"

EDITORIAL OFFICE ADDRESS:
14 Nauky Ave., Kharkiv, Ukraine, 61166
Information site: http://itssi-journal.com
https://journals.uran.ua/itssi
E-mail of the editorial board: journal.itssi@gmail.com

Ідентифікатор медіа R30-03878
Витяг з реєстру суб'єктів у сфері медіа – реєстрантів від 25.04.2024 № 1410

Журнал включено до Переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата наук наказом Міністерства освіти і науки України від 16.07.2018 №775 (додаток 7)

Харків – 2025

СУЧАСНИЙ СТАН НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ В ПРОМИСЛОВОСТІ

№ 3 (33), 2025

ЗМІСТ

Інформаційні технології

Інженерія та промислові технології

За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор

INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES

No. 3 (33), 2025

CONTENTS

Information Technology

Engineering & Industry Technologies

The author is responsible for the accuracy of the facts, quotations and other information

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

UDC 004.932: 004.5
DOI: https://doi.org/10.30837/2522-9818.2025.3.005

O. BARKOVSKA

TWO-FACTOR AUTHENTICATION BASED ON KEYWORD SPOTTING AND SPEAKER VERIFICATION

The **subject** matter of the article is the development and evaluation of a two-factor speaker authentication method based on voiceprint identification and keyword spotting (KWS), designed for secure voice-based access in human-machine interfaces, especially for users with limited mobility. The **goal** of the work is to create a method for managing speaker authentication using convolutional neural networks (CNNs), comparing the efficiency of two widely used spectral feature extraction techniques – Mel-Frequency Cepstral Coefficients (MFCC) and Short-Time Fourier Transform (STFT) spectrograms. The following tasks were solved in the article: a model of a two-factor authentication method is proposed, which includes speaker identification and voice password recognition; the quality of MFCC and STFT spectrograms features is compared; the influence of the number of epochs, CNN architecture and training parameters on the system accuracy is evaluated; the effect of the sampling rate on the performance of the models was investigated. The following **methods** are used: deep learning methods with CNN architecture, fine-tuning, MFCC, and STFT feature extraction, mathematical and statistical analysis of training efficiency, and system performance metrics. The following **results** were obtained: the method achieved 97.95% accuracy in speaker identification using MFCCs after 60 training epochs, and 99.82% accuracy in voice password verification using the same CNN structure after 20 epochs. The average accuracy of the entire authentication process was 98.75%. Moreover, using MFCC features reduced training time by a factor of 23 and memory consumption by a factor of 7 compared to STFT spectrograms. **Conclusions**: the effectiveness of a two-factor voice authentication method that combines speaker identification by acoustic voice characteristics and voice password verification was implemented and studied. Further research directions include studying the impact of alternative spectral features (in particular, CQCC, GFCC, prosodic parameters) on improving accuracy and resistance to spoofing. Special attention will be paid to optimizing the model for energy-efficient use on portable devices.

Keywords: voice authentication, identification, voice password, MFCC, spectrogram, CNN, MHI, biometrics.

Introduction

Interaction between humans and computers has become a common form of communication in the modern world. Human-machine interaction (HMI) can be facilitated through hardware devices such as keyboards, touchscreens, or mice, as well as via a more natural communication modality for humans – voice [1–2]. The proliferation of voice-based interfaces has led to significant scientific progress in areas such as speech modeling, linguistic pattern analysis, and acoustic signal processing. Voice technologies are emerging as a key component of the Fourth Industrial Revolution and are expected to have a growing impact on how people interact with machines.

Technological advancements in natural language processing (NLP), text-to-speech (TTS) systems, real-time speech pattern recognition, noise filtering, and multi-speaker separation (e.g., for conference calls), as well as system personalization based on speaker-specific attributes such as accent, speech rate, or physiological speech impairments, represent major scientific challenges with considerable practical value.

Voice and speech recognition methods, along with voice-driven human-machine interfaces, have particular practical significance in the following domains:

– providing accessibility for individuals with disabilities (e.g., voice control for those unable to use a keyboard) [3–4];

– supporting globalization and cultural preservation through universal translators and tools for low-resource languages [5].

A voice sample contains rich information – from the speaker's gender and age to, in some cases, emotional state. Speaker recognition and voice authentication tasks typically aim to identify or verify a speaker based on one or more biometric parameters. In contrast to passwords or PIN codes, which can be guessed, stolen, or observed, voice biometric data is inherently unique, making it much harder to spoof. As such, voice-based identification and authentication offer an effective and secure method for protecting sensitive data and personal information.

Speaker identification can also serve as a component of multifactor authentication (Figure 1), alongside other modalities such as fingerprint recognition, facial identification, or PIN codes [6–8]. This layered approach

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

enhances security, requiring an attacker to overcome multiple verification barriers to gain unauthorized access.

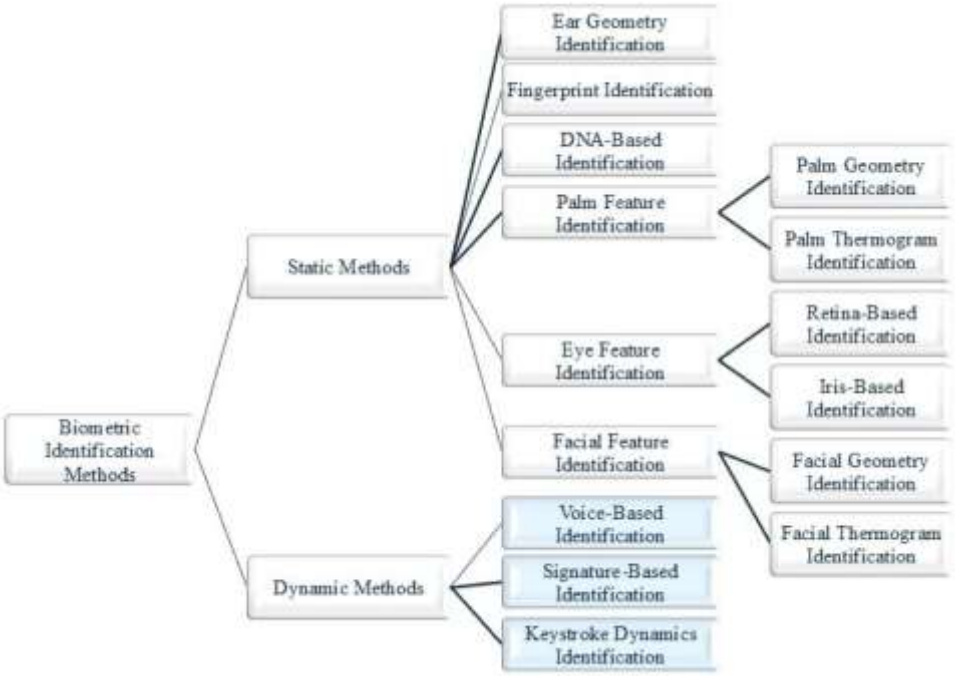


Fig. 1. The role of voice identification among other biometric authentication methods

The analysis of a human voice sample enables the extraction of a wide range of psychophysiological, social, and biometric information, extending far beyond the basic task of speaker identification (Figure 2).

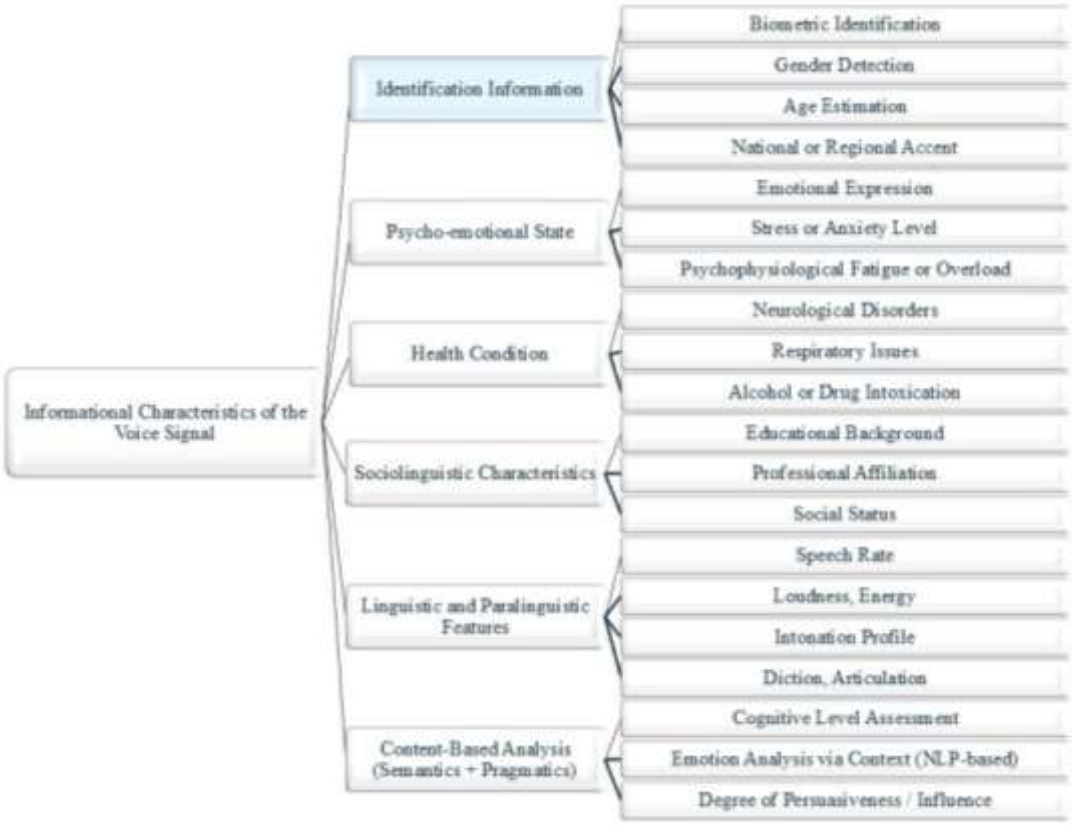


Fig. 2. Categorization of tasks based on voice sample analysis

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Analyzing a person's voice sample enables the extraction of detailed information not only about the individual but also about their psychophysiological state at the moment of speaking. Primarily, the voice is used for biometric identification, as each person possesses unique acoustic characteristics, including timbre, pitch, and articulatory movement patterns. These features allow for not only speaker identification, but also the estimation of gender, approximate age, and even linguistic background based on accent or dialect.

Advanced voice analysis can reveal the speaker's psycho-emotional state: parameters such as intonation, speech rate, loudness, pauses, and rhythm provide insight into emotions (e.g., joy, anxiety, anger), stress level, or even fatigue. Additionally, the voice contains significant health markers – neurological, respiratory, and even viral disorders may manifest through vocal tremor, hoarseness, irregular breathing, or dysarthria. These indicators are increasingly studied in the context of conditions such as Parkinson's disease and COVID-19.

The sociolinguistic dimension of speech further enables assessment of the speaker's educational level, professional affiliation, or social status through analysis of vocabulary, stylistic choices, and linguistic behavior. Paralinguistic features such as speech tempo, loudness, articulation clarity, and emotional expression offer additional cues regarding the speaker's intentions and confidence level.

Finally, content-level speech analysis can provide insight into the cognitive complexity of utterances, strength of argumentation, emotional polarity, and semantic richness – particularly when supported by natural language processing (NLP) tools. All these data can be captured and interpreted using modern acoustic signal processing techniques (e.g., formant analysis, spectrograms, MFCC), as well as deep learning methods, including neural network architectures such as CNNs, RNNs, LSTMs, and transformers, enabling the voice to function as a multidimensional channel of personal information.

Analysis of last achievements and publications

It should be noted that speaker identification and verification are different processes that can complement each other in voice authentication systems. According to the ISO/IEC 19794-13:2018 standard, identification is the process of determining a person from a set of registered users based on the acoustic characteristics of speech. Instead, verification (or authentication) checks whether the user's voice matches the reference sample, confirming the right to access. At the same time, authentication systems can use a fixed phrase (voice password), combining keyword recognition (KWS) and voice verification. In this way, voice identification determines who is speaking, and verification determines whether this person has access rights.

A review of the literature confirms that spectrograms (based on STFT) and MFCCs are the de facto standards in the architectures of modern deep learning models for speaker identification and verification tasks (e.g., x-vector, ECAPA-TDNN, CNN-KWS).

In the study by [9], the practical efficiency of MFCCs for speech feature extraction is demonstrated. The authors used 13 cepstral coefficients obtained from speech signals segmented with 50 ms frames and 50% overlap at a 16 kHz sampling rate. Despite limitations of the Madaline Type I neural network used for classification, the recognition accuracy within the database reached 61%, and rejection accuracy for unknown utterances reached 84%. This highlights the robustness of MFCCs in encoding speaker-specific spectral patterns independently of lexical content, which is critical for real-time speaker identification systems.

In the work by [10], the ECAPA-TDNN architecture is introduced as an enhanced version of the x-vector model. The authors implement channel attention and aggregation mechanisms that allow the model to better capture speaker-relevant features. The paper also notes that the use of MFCC as input features is a common practice in such architectures.

The study by [11] explores automatic speaker identification based on features extracted from spectrograms using a convolutional neural network (CNN). The authors demonstrate that spectrograms are effective input features for speaker identification tasks, as they preserve local spectral properties of the speech signal and result in high recognition accuracy.

A systematic review by [12] analyzes key feature extraction methods for speaker identification. The authors conclude that MFCCs and spectrograms are among the most effective and widely used approaches due to their capability to represent acoustically meaningful features relevant for distinguishing between speakers.

Other features – such as LPCC (Linear Predictive Cepstral Coefficients), PLP (Perceptual Linear Prediction), GFCC (Gammatone Frequency Cepstral Coefficients), CQCC (Constant-Q Cepstral Coefficients), as well as prosodic features (e.g., speech rate, intonation, pauses) and wavelet-based representations (DWT, CWT) –

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

have significant scientific value, particularly in multimodal systems, noisy conditions, or cross-domain applications [13–14].

In future work, the analysis will be extended to include CQCC and prosodic features, with particular attention to cross-lingual and intermodal verification scenarios.

This paper focuses on the most widely adopted feature types, as evidenced by their presence in state-of-the-art toolkits and frameworks such as Kaldi, SpeechBrain, and the PyTorch Speaker Verification Toolkit.

The extracted features serve as the input for the next stage of the standard audio sequence analysis pipeline – the analyzer or classifier [15–17].

The purpose of this study is to evaluate the authentication accuracy of the proposed two-factor authentication (2FA) method based on keyword spotting (KWS) and speaker verification by analyzing the impact of STFT-based spectrograms and MFCC features on the accuracy of speaker identification and verification. The system is implemented using a convolutional neural network (CNN) architecture with the application of fine-tuning techniques.

To achieve this goal, the following objectives were addressed:

– a model of a two-factor authentication method is proposed, which includes speaker identification and voice password recognition;

– the quality of MFCC and STFT spectrograms features is compared;

– the influence of the number of epochs, CNN architecture and training parameters on the system accuracy is evaluated;

– the effect of the sampling rate on the performance of the models was investigated.

Materials and methods

There are two main groups of methods on which voice identification and authentication systems are based (Figure 3) [18]:

– the reference method is based on comparing some voice features (these can be either physiological or articulatory features) with some reference. A group of individual words is used as a reference;

– the phoneme-oriented method is based on the extraction of individual phonemes from the speech stream.

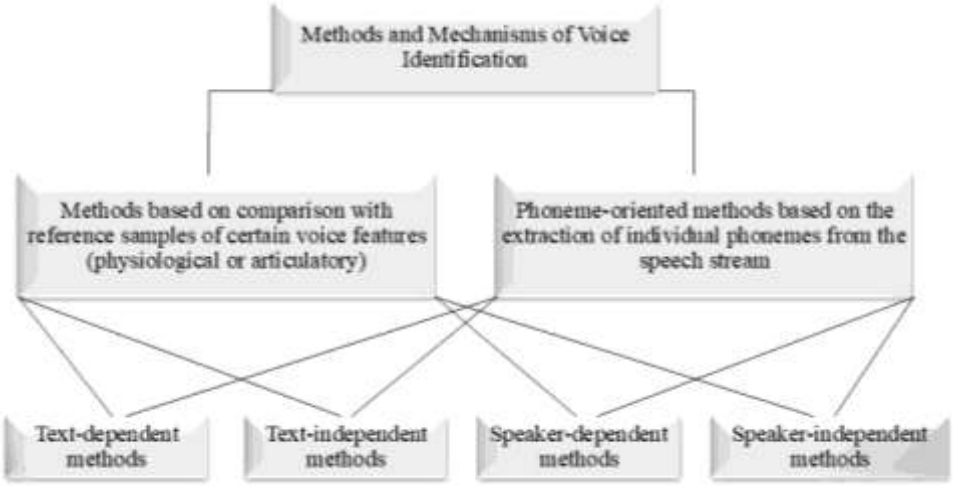


Fig.3. Methods and mechanisms of speaker identification

A speaker-dependent model requires preliminary training on a specific speaker. It generates a speaker embedding, which is subsequently used to compare new voice samples. In contrast, a speaker-independent model is trained on a large, diverse corpus of speakers and can estimate the degree of similarity between samples with high accuracy, even when the speaker is not included in the training data.

Text-dependent identification is effective when the spoken content is known in advance (e.g., a predefined passphrase), as this reduces variability in the linguistic content. Text-independent models, on the other hand, are more flexible and capable of handling a wider range of input, but they require larger datasets and more complex architectures that can generalize across varying speech content.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

These categories are closely related to challenges posed by acoustic variability (e.g., background noise, microphone quality), speech variability (e.g., emotional tone, intonation), and contextual variability (e.g., speaker stress, fatigue).

A generalized standard pipeline for speaker identification – across both speaker-dependent/independent and text-dependent/independent approaches – includes the following stages: preprocessing, feature extraction, feature selection, and decision-making or classification (Figure 4) [19].

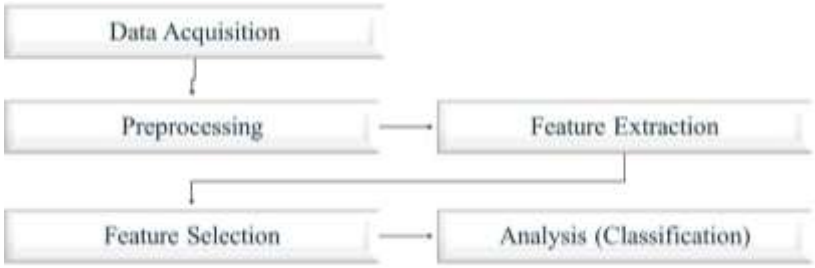


Fig.4. A standard pipeline for audio analysis (keyword spotting and voice verification)

Pre-processing is required to remove background noise and convert the input signal into a form suitable for feature extraction [20, 21]. The following feature extraction methods are commonly considered: MFCC (Mel-Frequency Cepstral Coefficients) – regarded as the standard for speaker identification tasks, though it requires signal normalization and pre-processing, LPC/LPCC (Linear Predictive Coding / Linear Predictive Cepstral Coefficients) – often used in combination with MFCC to enhance classification accuracy, PLP/RASTA-PLP (Perceptual Linear Prediction) – well-suited for noisy environments due to perceptual spectrum filtering, GFCC (Gammatone Frequency Cepstral Coefficients) – an alternative to MFCC, designed for more challenging acoustic conditions, DNN/CNN-based features – enable automatic, data-driven feature learning but demand significant computational resources, Prosodic features – serve as complementary inputs to spectral features, capturing suprasegmental information (intonation, rhythm, stress), STFT-based features (Short-Time Fourier Transform) – serve as a foundational representation for time-frequency analysis and underlie most spectral methods (e.g., MFCC, GFCC). Typically used as a pre-processing step rather than as standalone features, CQCC (Constant-Q Cepstral Coefficients) – apply a logarithmic frequency scale that aligns with musical tones and low-frequency speech components; frequently used in modern biometric security systems (Table 1).

The strengths and limitations of these techniques allow for selecting an appropriate trade-off tailored to specific application requirements and computational constraints [20].

The generalization of the table indicators leads to the following conclusions regarding the recommended applications and research directions for feature extraction methods in audio analysis:

– speech recognition: MFCC, LPCC, GFCC;

– music processing: CQCC, STFT;

– biometric authentication: CQCC for speaker recognition;

– medical diagnostics: DWT, CWT for ECG/EEG signal analysis;

– audio compression: DWT, e.g., in MPEG-4 encoding.

This work focuses on the analysis of deep neural architectures such as x-vector, ECAPA-TDNN, ResNet, CRNN, among others, as these models represent the state of the art (SOTA) in speaker verification and identification in both current research and practical systems.

Traditional classifiers such as CNN, RNN, and MLP are considered as building blocks within larger composite models and, therefore, are not separately featured in the comparative analysis (Table 2).

Classical machine learning classifiers such as SVM, K-NN, and HMM are excluded from the comparison due to their limited scalability and effectiveness in large-scale speaker recognition tasks.

The observed balance between accuracy and computational efficiency of STFT-based spectrograms and MFCCs ensures high performance with moderate resource consumption, which is essential for real-time applications and deployment on resource-constrained devices.

Moreover, the compatibility of these feature types with convolutional neural network (CNN) architectures motivates continued research on CNN-based models incorporating fine-tuning techniques on domain-specific or task-specific datasets.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Table 1. *Comparison of local feature extraction methods for speaker identification*

Type / Principle	Typical Parameters / Input Data	Primary Application Area	Advantages	Limitations
MFCC (Mel-Frequency Cepstral Coefficients)				
Spectral analysis using Mel-scale filterbanks, followed by cepstral transformation; short-time spectral analysis	Frame length: 20–40 ms; number of coefficients: typically 13–39	Application areas: speaker identification, speech recognition, voice password systems (e.g., ASV) (наприклад, ASV).	simplicity, computational efficiency, established standard in ASR systems, good performance on clean speech, incorporates perceptual characteristics of human hearing	sensitive to noise, limited temporal dynamics
LPCC, Linear Prediction Cepstral Coefficients				
Modeling of the vocal tract as a linear system for signal prediction	10–16 coefficients per frame	speaker identification	simplicity, effective in clean environments	low robustness to noise, limited relevance in modern ASR
PLP, Perceptual Linear Prediction				
Incorporates psychoacoustic properties (e.g., Bark scale critical bands) and spectral filtering	13–20 coefficients per frame	speaker identification, speech recognition	improved noise robustness compared to MFCC; more closely aligned with human auditory perception	still limited in temporal resolution; higher computational complexity than MFCC
GFCC, Gamma-tone Frequency Cepstral Coefficients				
Uses gammatone filterbanks to simulate the human auditory system	13–20 coefficients per frame	speaker identification in noisy environments, voice password systems	high noise robustness; well-suited for challenging acoustic conditions due to high frequency resolution	higher computational complexity compared to MFCC
CQCC, Constant Q Cepstral Coefficients				
Constant-Q transform with a logarithmic frequency scale	29–40 coefficients; long time windows	voice password systems, anti-spoofing, synthetic speech detection, speaker identification under adversarial conditions	high resolution at low frequencies; effective against spoofing and synthetic attacks	high computational cost; less commonly used in traditional ASR systems
Prosodic Features (e.g., pitch, duration, intensity, rhythm)				
Suprasegmental temporal features	computed over longer utterances (500 ms – 2 s)	speaker profiling, disfluency detection in ASR	convey speaking style and emotional state; complement spectral features; useful for speaker discrimination	not suitable for short utterances or passphrases due to context dependency
Spectrogram (based on STFT – Short-Time Fourier Transform)				
Time-frequency representation based on STFT; amplitude spectrogram	frame length 20–40 ms; FFT windowing	speaker identification, voice password systems	visual representation of energy distribution; rich information content	sensitive to noise; requires CNNs for feature learning
Wavelet Transform (DWT – Discrete Wavelet Transform, CWT – Continuous Wavelet Transform)				
Multilevel time-frequency analysis	choice of mother wavelet; multi-scale decomposition	speaker identification, anti-spoofing	captures short-term phenomena; adaptive localization in time and frequency	sensitive to parameter selection; lack of standardization

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Table 2. *Analytical comparison of ai-based voice feature extractors*

Model	Principle	Input Features	Sensitivity To	Accuracy	Remarks and Resource Usage
Speaker Identification					
x-vector (TDNN)	Embeddings from Time Delay Neural Network	MFCC/STFT spectrogram	Noise, speech distortion	~90–95% (VoxCeleb1)	Stable model, well integrated in Kaldi. Does not capture long-term context. Moderate resource usage.
d-vector (LSTM)	Averaging LSTM-based embeddings	MFCC	Noise, signal length	~85–92% (VoxCeleb)	Captures temporal structure. Slow, less scalable. Moderate resource usage.
Siamese CNN	Speech spectrum comparison (embedding-distance)	STFT spectrogram	Quality of positive/negative pairs	~85–90%	Few-shot capability. Depends on pair generation. Moderate resource usage.
Voice Password (KWS – Keyword Spotting)					
CNN-KWS	CNN for keyword classification	MFCC/STFT spectrogram	Noise, pronunciation, speech rate	95–97% (Google Speech Commands)	Easy implementation, good performance. Moderate resource usage.
DS-CNN	Mobile-optimized deep CNN	MFCC	Background noise, limited vocabulary	92–95%	Optimized for mobile devices. Lower accuracy. Low resource usage.
CRNN	CNN + recurrent layers for temporal modeling	STFT spectrogram	Utterance length, tempo variation	96–98%	Context-aware. More complex training. Moderate resource usage.

Research results

The proposed voice access system enables both speaker identification (i.e., determining who is speaking) and speaker verification (i.e., confirming whether the correct passphrase has been spoken). The authentication process is organized in two sequential stages: first, the system verifies the speaker's identity, and then it checks the correctness of the spoken password.

The study focuses on evaluating the impact of two commonly used acoustic representations – the STFT-based spectrogram and the Mel-Frequency Cepstral Coefficients (MFCC) – on the accuracy of speaker identification and verification. The extracted features are analyzed using a convolutional neural network (CNN) architecture with the application of fine-tuning techniques to improve model generalization and performance (Figure 5).

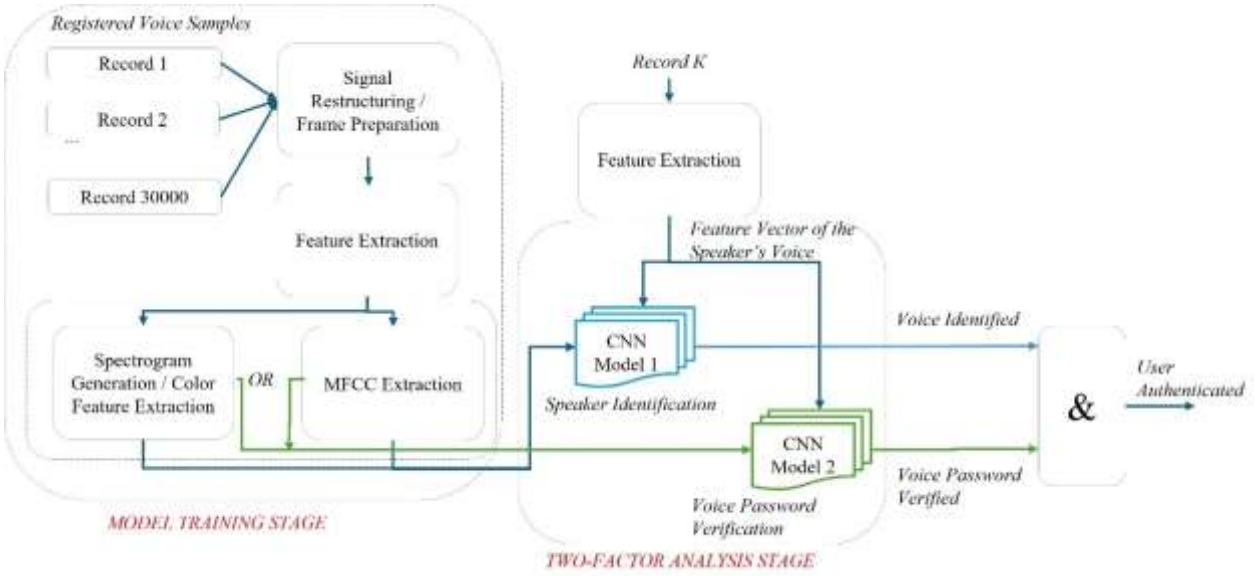


Fig. 5. Functional model of the proposed two-factor speaker authentication method

A series of experiments was conducted based on the proposed method to validate and analyze its performance. In particular, the effectiveness of two audio signal processing methods – spectrogram-based representation

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

and Mel-Frequency Cepstral Coefficients (MFCC) – was evaluated and experimentally confirmed for the task of extracting salient acoustic features.

The study further examined the design of convolutional neural network (CNN)-based analyzers, the impact of dataset partitioning into training, validation, and test subsets, as well as the influence of training duration and hardware requirements on the system's stability and real-time applicability.

Neural network models with convolutional architectures were trained using a dataset organized as described in Table 3:

– Speaker_ID – speaker identifier; the dataset includes 60 speakers (labeled Speaker_1 through Speaker_60);

– Digit – the digit spoken by the speaker (from 0 to 9);

– File_ID – unique identifier for each audio file;

– Volume – recording volume level (low, medium, high);

– Pronunciation – articulation type (clear, muffled, fast, slow);

– Duration – length of the recording in milliseconds (ranging from 500 ms to 2499 ms).

Experiment 1 focuses on comparing the accuracy of speaker identification based on Mel-Frequency Cepstral Coefficients (MFCC) and STFT-based spectrograms. The processing sequence for implementing the system using MFCC features (Figure 6) is as follows:

– the initial dataset, consisting of WAV-format audio recordings of digits (0–9) pronounced in English 50 times by each of 60 speakers, is loaded into memory for further processing;

– each audio sample undergoes a resampling procedure to standardize the sampling rate across the dataset;

– MFCC features are extracted from each audio sequence, and the first 13 coefficients are stored in a data array;

– a feature set is formed by pairing the extracted MFCC vectors with their corresponding speaker identifiers. The resulting dataset is then split into three subsets: training, validation, and test;

– a convolutional neural network (CNN) is trained using the training and validation data subsets;

– the trained model is evaluated on the test set to determine the speaker identification accuracy.

Table 3. *Dataset characteristics for training neural network analyzers*

Speaker_ID	Digit	File_ID	Volume	Pronunciation	Duration
Speaker_1	0	File_1	low	clear	500ms
Speaker_1	1	File_2	medium	muffled	501ms
Speaker_1	2	File_3	high	fast	502ms
Speaker_1	3	File_4	low	slow	503ms
Speaker_1	4	File_5	medium	clear	504ms
...	...	...	...	...	...
Speaker_60	9	File_30000	high	slow	2499ms

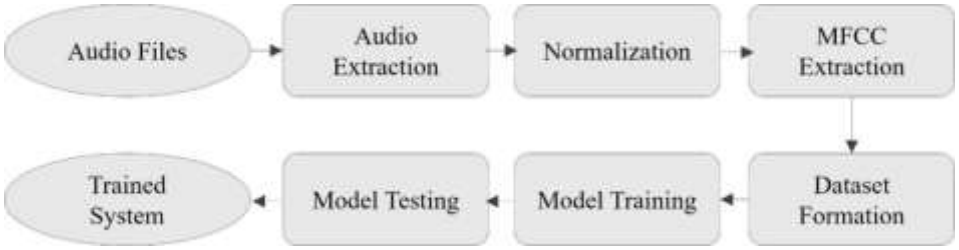


Fig. 6. Sequential steps for implementing the system using mel-frequency cepstral coefficients (MFCC)

The workflow for the system based on spectrograms (Fig. 7) differs in the initial preprocessing steps. Instead of extracting MFCC features, the following operations are performed:

– each audio signal is converted into a spectrogram using the Short-Time Fourier Transform (STFT), and the resulting image is resized to 305 × 184 pixels and stored for further use;

– the spectrogram image is then loaded and its color features are extracted across three channels (RGB) (Fig. 7);

– the extracted color channel data from the spectrograms are used to construct the training dataset.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

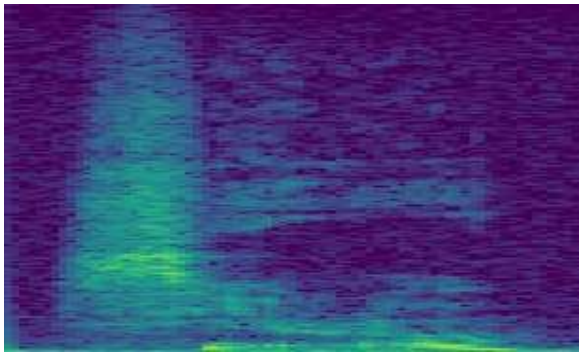


Fig. 7. Example of a spectrogram

To compare the accuracy of systems based on Mel-Frequency Cepstral Coefficients (MFCC) and spectrograms, two neural networks with identical configurations, architectures, and training parameters were developed and trained. The only difference between the two systems lies in the type of input features used by the neural network: MFCCs in the first case and spectrograms in the second.

Both models were trained using the same initial dataset, consisting of 30,000 WAV-format audio files. Each file contains the recording of one of ten digits, spoken by one of 60 different speakers, with each digit repeated 50 times per speaker under varying conditions of volume, pronunciation, and duration. The training was performed over the same number of epochs for both systems (Fig. 8). To evaluate the identification accuracy, both models were tested using a hold-out test dataset that was not included in the training process.

The results of evaluating training time and speaker identification accuracy depending on the type of feature representation (spectrogram or Mel-Frequency Cepstral Coefficients) using a convolutional neural network are presented in Table 4.

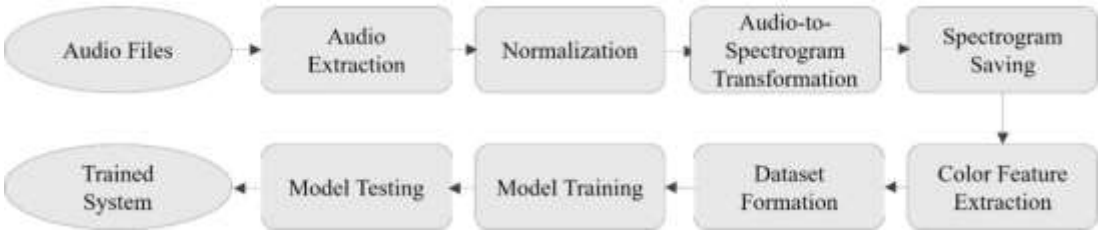


Fig. 8. Sequence of steps for implementing the system using spectrograms

Table 4. *Training results of systems using mel-frequency cepstral coefficients and spectrograms*

	Spectrogram			**Mel-Frequency Cepstral Coefficients (MFCC)**		
Number of Epochs	Training Time	Memory Usage	Accuracy	Training Time	Memory Usage	Accuracy
20	7 год. 48 хв.	15 Гб	73.13%	6 хв. 43 сек.	2 Гб	96.84%
60	23 год. 57 хв.	15 Гб	80.9%	14 хв. 58 сек.	2 Гб	97.64%

The model trained using Mel-Frequency Cepstral Coefficients (MFCCs) demonstrated 17% higher accuracy compared to the model trained on spectrograms. In addition, the training time for the MFCC-based model was 23 hours and 42 minutes shorter.

Accordingly, further experiments focused on analyzing the impact of different training/validation/test splits within the working dataset were conducted exclusively using MFCC features.

The proportion of validation data was gradually increased by reducing the training portion, in order to assess the relationship between dataset composition, training time, and model performance.

The results of these experiments (see Figure 9) show that training time decreases as the size of the training set decreases. This outcome is expected, as fewer training samples require less processing time per epoch.

However, the model accuracy also declines with a reduction in training data. This indicates that a larger training set improves the model's ability to make accurate predictions.

The highest accuracy, 97.4%, was achieved with a 70% training / 15% validation / 15% test split.

It is important to consider that training time and computational resources may be limited. Therefore, the choice of data split should be based on a trade-off between the desired accuracy and acceptable training time. The conducted experiment enables the observation of accuracy trends depending on the ratio between training, validation, and test subsets.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Fig. 9. Impact of data split proportions on model accuracy and training time

In the course of this work, an additional experiment was performed to determine the optimal sampling rate that ensures the highest accuracy of speaker identification by voice. The effect of various sampling rate values on model accuracy was investigated. The initial dataset consisted of audio files recorded at a sampling rate of 48 kHz. For comparative evaluation, the most common alternative sampling rates were also tested: 8 kHz, 16 kHz, 20.05 kHz, and 32 kHz. The results of this experiment are presented in Table 5.

Table 5. *Model accuracy as a function of audio sampling rate*

Sampling Rate (kHz)	Training Time	Accuracy (%)
8	14 хв. 57 сек.	97.62
16	15 хв. 13 сек.	98.02
20.05	16 хв. 59 сек.	97.84
32	17 хв. 05 сек.	98.06
48	17 хв. 52 сек.	98.28
88.2	19 хв. 35 сек.	97.77

Based on the results of the conducted experiments, it was determined that the most optimal sampling rate is 48 kHz, which provides the highest accuracy, as well as efficient training time and faster audio loading. Higher sampling rates capture more detailed information from the signal. In audio processing tasks, a higher sampling rate may help the model detect subtle nuances and fine-grained acoustic patterns, potentially improving accuracy. Although such high sampling rates are not traditionally used in speaker recognition tasks, each dataset and task may have unique characteristics that influence optimal system settings.

The experiments were also conducted to determine the fine-tuning configuration of the convolutional neural network (CNN) architecture – specifically, the number of convolutional layers, training epochs, and hyperparameter settings – to achieve the highest possible speaker identification accuracy while avoiding overfitting.

During the course of this work, a CNN was tested with varying numbers of convolutional layers, different training durations, and regularization techniques to mitigate overfitting.

The process began with a baseline architecture of three convolutional layers, each followed by MaxPooling and BatchNormalization layers to enhance learning stability and improve feature extraction accuracy. Each subsequent convolutional layer doubled the number of filters compared to the previous one, increasing the model's representational capacity and enabling it to capture more complex patterns.

Throughout the testing phase, the number of convolutional layers was gradually increased in order to determine the optimal architecture and number of training epochs. The initial model was trained on a dataset consisting of 60 speakers, using the optimal data split: 70% for training, 15% for validation, and 15% for testing.

The results of these experiments are summarized in Table 6.

Based on the results of preliminary testing, the optimal neural network architecture was determined to consist of four convolutional layers and 60 training epochs, providing the best trade-off between model accuracy and training time.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Table 6. *Evaluation of CNN architectures and training epoch counts*

Number of Layers	Number of Epochs	Training Time	Accuracy (%)
3	20	4 хв. 45 сек.	94.3
3	60	12 хв.	96.1
3	150	30 хв.	96.4
4	20	5 хв. 32 сек.	95.66
4	60	14 хв. 58 сек.	97.95
4	150	39 хв.	97.64
5	20	6 хв. 24 сек.	94.6
5	60	19 хв. 25 сек.	96.18
5	150	45 хв.	97.1
6	20	7 хв. 43 сек.	93.4
6	60	22 хв. 57 сек.	96.2
6	150	53 хв.	97.05

The achieved speaker identification accuracy was 97.95%, with a total training time of 14 minutes and 58 seconds.

A detailed description of the neural network architecture is provided in Table 7.

Table 7. *CNN architecture and configuration details*

Layer Type	Output Dimension	Number of Parameters
Conv2D	(None, 87, 13, 32)	320
MaxPooling2D	(None, 44, 7, 32)	0
BatchNormalization	(None, 44, 7, 32)	128
Conv2D	(None, 44, 7, 64)	18496
MaxPooling2D	(None, 22, 4, 64)	0
BatchNormalization	(None, 22, 4, 64)	256
Conv2D	(None, 22, 4, 128)	73856
MaxPooling2D	(None, 11, 2, 128)	0
BatchNormalization	(None, 11, 2, 128)	512
Conv2D	(None, 11, 2, 256)	295168
MaxPooling2D	(None, 6, 1, 256)	0
BatchNormalization	(None, 6, 1, 256)	1024
Flatten	(None, 1536)	0
Dense	(None, 128)	196736
Dropout	(None, 128)	0
Dense	(None, 61)	7869

As a result of fine-tuning the model on a dataset of 3,000 audio recordings from six new speakers, the recognition accuracy significantly improved due to the increased diversity of training data.

An accuracy level of 99.5% was achieved, while the training time remained nearly unchanged, enabling efficient integration of new users within a relatively short period of time.

The second component of the proposed system (Figure 5) is responsible for voice password verification. The authentication principle is as follows: the user is considered authenticated only if both the voice is correctly identified and the spoken password matches the one stored in the database; otherwise, access is denied.

At this stage, the analyzer is implemented as a convolutional neural network (CNN) with four convolutional layers, trained over 20 epochs.

The main difference from the architecture used for speaker identification lies in the output layer – in the verification model, it contains ten output neurons, corresponding to the ten digits that need to be recognized as part of the personal voice password (a digit sequence). A shared dataset consisting of 30,300 audio files from 66 different speakers was used for training. In this case, each audio file corresponds to one of ten digits. The training results of the digit recognition model are summarized in Table 8.

Table 8. *Training results of the neural network for spoken digit recognition*

Number of Layers	Number of Epochs	Training Time	Accuracy (%)
4	20	6 хв. 22 сек.	99.82
4	60	18 хв. 4 сек.	99.4

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

The model trained with fewer epochs demonstrated higher accuracy, as the network trained over 60 epochs encountered overfitting on the initial dataset. The result of the complete multifactor authentication system for one of the users – who was successfully identified by voice and verified through correct password recognition – is illustrated in Figure 10.

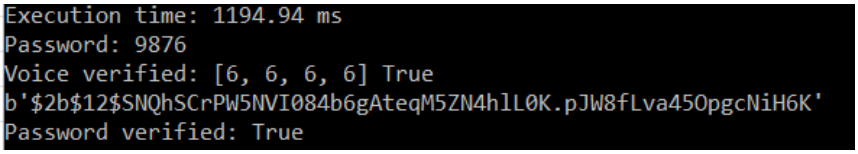


Fig. 10. Output of the multifactor authentication method

Based on the console output, a general conclusion can be drawn confirming the successful execution of all experimental stages:

– the total processing time was 1194.94 ms, indicating high execution efficiency;

– the password "9876" was successfully verified;

– voice verification was successfully completed, yielding output values of [6, 6, 6, 6], which confirm the authenticity of the speaker;

– the password was securely hashed using the bcrypt algorithm, ensuring a high level of cryptographic security;

– the password was successfully matched and validated, confirming its correctness and consistency with the expected value.

These results demonstrate that the proposed system effectively performs both password processing and voice verification, delivering a high degree of security and reliability. Among the key advantages of the proposed solution is the maximum voice password verification accuracy, which reached 99.82% using a convolutional neural network with four layers, trained for 20 epochs over a duration of 6 minutes and 22 seconds, with MFCC features used for acoustic feature extraction. The maximum speaker identification accuracy achieved was 97.95%, using the same CNN architecture with four convolutional layers, trained for 60 epochs over 14 minutes and 58 seconds. The average speaker authentication accuracy of the proposed two-factor biometric authentication system is 98.75%.

Conclusions

The relevance of the research is due to the growing number of people with visual impairments and the need to create As a result of this work, a two-factor voice authentication method was developed and evaluated. It combines speaker identification based on acoustic voice characteristics with keyword verification (voice password recognition). It is built using convolutional neural networks (CNNs) and analyzes two types of spectral features: STFT-based spectrograms and Mel-Frequency Cepstral Coefficients (MFCCs).

A series of experiments was conducted to compare the accuracy, training time, and computational resource requirements associated with each feature type.

The highest speaker identification accuracy – 97.95% – was achieved using MFCC features in a four-layer CNN architecture trained for 60 epochs. The voice password verification model demonstrated even greater effectiveness, reaching 99.82% accuracy with the same architecture trained over 20 epochs. The overall average accuracy of the proposed two-factor method was 98.75%, which is a competitive result for modern biometric security applications.

The use of MFCCs was shown to reduce training time by a factor of 23 and memory consumption by a factor of 7, compared to models using STFT spectrograms. This makes MFCCs the preferred choice for real-time applications, especially on resource-constrained devices.

The method was validated on a synthetic dataset of over 30,000 audio recordings from 60 speakers, and was further fine-tuned with new user data, achieving an identification accuracy of up to 99.5%, demonstrating its adaptability.

Analysis of sampling rate impact revealed that 48 kHz offers the best trade-off between accuracy (98.28%) and computational efficiency.

The method is of particular practical value in the context of human-computer interaction for individuals with limited mobility or bedridden patients, as it enables contactless and keyboard-free access to digital devices via voice alone.

The findings confirm that the proposed method can be successfully integrated into assistive technologies or home automation solutions to enable reliable, accurate, and adaptive biometric voice authentication.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

The scientific novelty of the study lies in the comprehensive analysis of the effectiveness of spectral features – MFCC and STFT spectrograms – for the simultaneous tasks of speaker identification and keyword-based voice verification (KWS) within a unified convolutional neural network architecture, with a focus on fine-tuning techniques.

The practical significance of the research lies in the fact that its results can be directly applied to the development of real-world contactless authentication systems, particularly:

– in access systems for individuals with physical disabilities, where traditional input methods are not feasible;

– in medical or household devices that require hands-free interfaces;

– in mobile or resource-constrained environments, where computational efficiency is critical – the use of MFCC reduces training time by a factor of 23 and memory usage by a factor of 7 compared to STFT;

– in biometric security systems where high accuracy is essential (97.95% for identification and 99.82% for verification), along with adaptability to new users (achieving up to 99.5% accuracy after fine-tuning).

Future research directions aim to expand the functionality and robustness of the proposed method. In particular, the study will focus on implementing few-shot learning techniques for dynamic adaptation to new users, and enhancing method noise robustness using spectral denoising and data augmentation strategies. Additionally, it is proposed to integrate alternative spectral and suprasegmental features (e.g., CQCC, GFCC, and prosodic parameters) to improve accuracy and resistance to spoofing attacks.

Further efforts will be directed toward optimizing the model for energy-efficient deployment on portable and embedded devices, and extending the method's capabilities toward multimodal biometric authentication, combining voice with gaze tracking or emotion analysis. The proposed method is also planned to be tested in real-world human-machine interaction scenarios, particularly for users with limited mobility or speech impairments.

References

1. Mourtzis, D., Angelopoulos, J., Panopoulos, N. (2023), "The Future of the Human–Machine Interface (HMI) in Society 5.0". *Future Internet*, № 15, 162 p. DOI: https://doi.org/10.3390/fi15050162
2. Grobelna, I., Mailland, D., Horwat, M. (2025), "Design of Automotive HMI: New Challenges in Enhancing User Experience, Safety, and Security". *Appl. Sci*. № 15, 5572 p. DOI: https://doi.org/10.3390/app15105572
3. Esquivel, P. et al. (2024), "Voice Assistant Utilization among the Disability Community for Independent Living: A Rapid Review of Recent Evidence", *Human Behavior and Emerging Technologies*, Vol. 2024, №. 1, 6494944 p. DOI: https://doi.org/10.1155/2024/6494944
4. Semary, H. E., Al-Karawi, K. A. (2024), "Abdelwahab M. M. Using voice technologies to support disabled people", *Journal of Disability Research*, 2024. Vol. 3. №. 1. DOI: https://doi.org/10.57197/jdr-2023-0063
5. Lawrence, I. D., Pavitra, A. R. R. (2024), "Voice-controlled drones for smart city applications", *Sustainable Innovation for Industry 6.0*. P. 162–177. DOI: DOI: 10.1109/ICUFN.2017.7993759
6. Ryu, R., Yeom, S., Kim, S. H., Herbert, D. (2021), "Continuous multimodal biometric authentication schemes: a systematic review", *IEEE Access*. Vol. 9. P. 34541–34557. DOI: 10.1109/ACCESS.2021.3061589
7. Barkovska, O., Liapin, Y., Muzyka, T., Ryndyk, I., Botnar, P. (2024), "Gaze direction monitoring model in computer system for academic performance assessment. Civil law aspect", *Information Technologies and Learning Tools*, Vol 99, №1, P. 63–75. DOI: 10.33407/itlt.v99i1.5503
8. Shaheed, K., Mao, A., Qureshi, I. et al. (2021), "A Systematic Review on Physiological-Based Biometric Recognition Systems: Current and Future Trends". *Arch Computat Methods Eng 28*, P. 4917–4960. DOI: https://doi.org/10.1007/s11831-021-09560-3
9. Sasongko, S. M. A., Tsaury, S., Ariessaputra, S., Ch, S. (2023), "Mel Frequency Cepstral Coefficients (MFCC) Method and Multiple Adaline Neural Network Model for Speaker Identification". *International Journal on Informatics Visualization*, № 7(4), P. 2306–2312. DOI: https://doi.org/10.30630/joiv.7.4.1376
10. Desplanques, B., Thienpondt, J., & Demuynck, K. (2020), "ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification". *In Interspeech 2020*, P. 3830–3834. DOI: https://doi.org/10.21437/Interspeech.2020-2650
11. Jahangir, R., Alreshoodi, M., Alarfaj, F. K. (2025), "Spectrogram Features-Based Automatic Speaker Identification for Smart Services". *Applied Artificial Intelligence*, № 39(1). DOI: https://doi.org/10.1080/08839514.2025.2459476
12. Tirumala, S. S., Shahamiri, S. R., Garhwal, A. S., Wang, R. (2017), "Speaker Identification Features Extraction Methods: A Systematic Review". *Expert Systems with Applications*, № 90, P. 250–271. DOI: https://doi.org/10.1016/j.eswa.2017.08.015
13. Iliev, Y.; Ilieva, G. (2023), "A Framework for Smart Home System with Voice Control Using NLP Methods". *Electronics 2023*, № 12, 116 p. DOI: https://doi.org/10.3390/electronics1201011614
14. Kim, Y., Hyon, Y., Lee, S., Woo, S. D., Ha, T., Chung, C. (2022), "The coming era of a new auscultation system for analyzing respiratory sounds", *BMC Pulmonary Medicine*, Vol. 22, №. 1. 119 p. DOI: 10.1186/s12890-022-01896-1

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

15. Barkovska, O, Havrashenko, A. (2024), "Research of the impact of noise reduction methods on the quality of audio signal recovery", *Information and control systems at railway transport*, 2024, Vol. 29, №. 3. P. 57–65. DOI: https://doi.org/10.18664/ikszt.v29i3.313606
16. Zaman, K., Sah, M., Direkoglu, C., Unoki, M. (2023), "A Survey of Audio Classification Using Deep Learning", *IEEE Access*, Vol. 11, P. 106620–106649. DOI: 10.1109/ACCESS.2023.3318015
17. Xie, X., Cai, H., Li, C., Wu, Y., Ding, F. (2023), "A Voice Disease Detection Method Based on MFCCs and Shallow CNN", *Journal of Voice*, Oct. 2023, DOI: https://doi.org/10.1016/j.jvoice.2023.09.024
18. Tu, Y., Lin, W., Mak, M. W. (2022), "A survey on text-dependent and text-independent speaker verification", *IEEE Access*. Vol. 10. P. 99038–99049. DOI: DOI: 10.1109/ACCESS.2022.3206541
19. Luitel, Sophina, Mohd, Anwar. (2022), "Audio Sentiment Analysis Using Spectrogram and Bag-of-Visual- Words", *IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI), IEEE,* P. 200–205. DOI: https://doi.org/10.1109/IRI54793.2022.00052
20. Singh, V. K., Sharma, K., Sur, S. N. (2023), "A survey on preprocessing and classification techniques for acoustic scene", *Expert Systems with Applications*, Vol. 229, 120520 p. DOI: https://doi.org/10.1016/j.eswa.2023.120520
21. Labied, M., Belangour, A., Banane, M., Erraissi, A. (2022), "An overview of Automatic Speech Recognition Preprocessing Techniques", *2022 International Conference on Decision Aid Sciences and Applications (DASA),* Chiangrai, Thailand, P. 804–809, DOI: 10.1109/DASA54658.2022.9765043

Надійшла (Received) 25.05.2025

Відомості про авторів / About the Authors

Barkovska Olesia – Ph.D (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor of the Department of Electronic Computers, Kharkiv, Ukraine; e-mail: olesia.barkovska@nure.ua; ORCID ID: https://orcid.org/0000-0001-7496-4353

Барковська Олеся Юріївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри Електронних обчислювальних машин, Харків, Україна.

ДВОФАКТОРНА АВТЕНТИФІКАЦІЯ НА ОСНОВІ МЕТОДУ KWS ТА ГОЛОСОВОЇ ВЕРИФІКАЦІЇ

Предметом статті є розробка та оцінка двофакторного методу автентифікації мовця на основі ідентифікації голосового відбитка та верифікації ключових слів (KWS), призначеного для безпечного голосового доступу в інтерфейсах «людина-машина», особливо для користувачів з обмеженою мобільністю. **Метою** роботи є створення методу управління автентифікацією мовця з використанням конволюційних нейронних мереж (CNN), порівняння ефективності двох широко використовуваних методів вилучення спектральних ознак – спектрограм Мел-частотних кепстральних коефіцієнтів (MFCC) та короткочасного перетворення Фур'є (STFT). У статті вирішено такі **завдання**: запропоновано модель двофакторного методу автентифікації, що включає ідентифікацію мовця та розпізнавання голосового пароля; порівняно якість ознак спектрограм MFCC та STFT; оцінено вплив кількості епох, архітектури CNN та параметрів навчання на точність системи; досліджено вплив частоти дискретизації на продуктивність моделей. Використовуються такі **методи**: методи глибокого навчання з архітектурою CNN, точне налаштування, вилучення ознак MFCC та STFT, математичний та статистичний аналіз ефективності навчання та показники продуктивності системи. Отримано такі **результати**: метод досяг 97,95% точності в ідентифікації мовця за допомогою MFCC після 60 епох навчання та 99,82% точності в перевірці голосового пароля за допомогою тієї ж структури CNN після 20 епох. Середня точність всього процесу автентифікації становила 98,75%. Більше того, використання MFCC-ознак дозволило скоротити час навчання в 23 рази, а споживання пам'яті – в 7 разів порівняно зі спектрограмами STFT. **Висновки**: було реалізовано та досліджено ефективність двофакторного методу голосової автентифікації, що поєднує ідентифікацію мовця за акустичними характеристиками голосу та перевірку голосового пароля. Подальші напрямки досліджень включають вивчення впливу альтернативних спектральних характеристик (зокрема, CQCC, GFCC, просодичних параметрів) на підвищення точності та стійкості до підробки. Особлива увага буде приділена оптимізації моделі для енергоефективного використання на портативних пристроях.

Ключові слова: олосова автентифікація, ідентифікація, голосовий пароль, MFCC, спектрограма, CNN, МНІ, біометрія.

Бібліографічні описи / Bibliographic descriptions

Барковська О. Ю. Двофакторна автентифікація на основі методу KWS та голосової верифікації. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 5–18. DOI: https://doi.org/10.30837/2522-9818.2025.3.005

Barkovska, O. (2025), "Two-factor authentication based on keyword spotting and speaker verification", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 5–18. DOI: https://doi.org/10.30837/2522-9818.2025.3.005

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 658.012.23 DOI: https://doi.org/10.30837/2522-9818.2025.3.019

М. Кікоть, Ю. Малєєва

МОДЕЛЬ ЛОГІСТИЧНОЇ ІНФРАСТРУКТУРИ ДВОХЕТАПНОГО ПРОЦЕСУ РЕСАЙКЛІНГУ СКЛАДНОЇ ТЕХНІКИ

Предмет дослідження – моделі формування логістичної інфраструктури ресайклінгу складної техніки з огляду на двохетапний процес перероблення та багатокомпонентного складу вторинної сировини. **Метою роботи** є створення моделі формування логістичної інфраструктури підприємств ресайклінгу складної техніки, упровадження якої дасть змогу зменшити витрати на транспортування вторинної сировини та розбудову інфраструктури. **Завдання**: дослідити сучасний стан і проблематику сфери перероблення компонентів складної техніки в Україні та за кордоном; визначити особливості логістичних процесів ресайклінгу складної техніки; розробити модель формування логістичної інфраструктури ресайклінгу складної техніки. **Методи**: математичне моделювання, методи розміщення-розподілу, системний підхід. **Результати**: проаналізовано сучасний стан і проблематику перероблення компонентів складної техніки, зокрема недосконалість інфраструктури, низький рівень сортування та нерівномірність розміщення підприємств; проаналізовано підходи до виконання завдання розміщення об'єктів інфраструктури ресайклінгу; побудовано дворівневу структуру системи ресайклінгу складної техніки, що містить мережу локальних пунктів збирання та сортування та переробних підприємств; запропоновано модель формування логістичної інфраструктури, у якій узято до уваги виробничі потужності, транспортні витрати та багатокомпонентність ресурсів. **Висновки**. Запропоновано модель формування логістичної інфраструктури ресайклінгу складної техніки, у якій узято до уваги двохетапний характер перероблення складної техніки – від збирання та сортування до кінцевого перероблення компонентів, а також транспортування декількох видів ресурсу. Модель поєднує економічні, просторові та технологічні аспекти, що дає змогу мінімізувати загальні витрати на транспортування вторинної сировини та розбудову інфраструктури ресайклінгу. Використання моделі забезпечує ефективне планування мережі підприємств ресайклінгу з огляду на виробничі потужності, логістичні маршрути та обсяги різних типів вторинної сировини.

Ключові слова: ресайклінг; перероблення вторинної сировини; складна техніка; оптимізаційна модель; інфраструктура; логістика; двохетапна система.

Вступ

Через війну Україна зіштовхнулася з безпрецедентними руйнуваннями житлових, громадських приміщень, різних інфраструктурних споруд, транспортних об'єктів, а отже, і з величезними обсягами великогабаритних відходів. Влада громад, населених пунктів, особливо тих, що постраждали від окупації, не має стратегії щодо зберігання та утилізації відходів. Через відсутність достатньої кількості складів, сортувальних ліній та ефективних механізмів поводження з відходами в країні зростає кількість нелегальних сміттєзвалищ, що загрожує довкіллю та здоров'ю населення. Наявні потужності недовантажені й не охоплюють навіть базовий рівень побутових відходів, що унеможливлює ефективне оброблення більш складних фракцій [1].

Варто зазначити, що проблема поводження з відходами в країні існувала і до повномасштабного вторгнення: переповнені звалища та полігони, слабка культура сортування й перероблення відходів.

Дослідження питання перероблення будівельних відходів і відходів руйнації розглянуто в праці [2]. У цій статті приділяється увага насамперед проблемам перероблення та повторного використання компонентів складної техніки (СТ).

До компонентів СТ належать: метал, пластик, скло, мікросхеми, кабелі та провідники, рідини й мастила, поліетилен, акумулятори тощо [3].

Необхідно виокремити кілька напрямів ресайклінгу складної техніки.

1. Знищення та поховання техніки. Напрям визначається значним негативним впливом на довкілля. Особливістю також є відсутність прибутку від реалізації проєкту.

2. Комерціалізація ресайклінгу. Передбачає розбирання СТ на складники, оцінювання технічного стану, ремонт, модернізацію, зберігання та продаж. Цей напрям найбільш раціональний у процесі ресайклінгу СТ.

3. Перероблення СТ у вторинну сировину, яку можна використати повторно для створення нових зразків СТ.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Важливим питанням, що впливає на економічний складник процесу ресайклінгу, є розміщення об'єктів інфраструктури ресайклінгу, сучасні підходи до розв'язання якої розглянемо в наступному розділі.

Аналіз останніх досліджень і публікацій

Для оптимального розміщення об'єктів ресайклінгу і підвищення логістичної ефективності дослідники зосереджують увагу на розробленні математичних моделей та алгоритмів. Розв'язання задач розташування підприємств, складів та інших інфраструктурних підприємств є складним багатоаспектним процесом – необхідно брати до уваги географічні умови, особливості земельного фонду, логістичну доступність, економічну ефективність, екологічні обмеження та соціальні критерії. У широкому контексті такі задачі розміщення належать до оптимізаційних, що описують логістичні мережі та транспортні системи. Для формалізації використовують як дискретні моделі, так і неперервні.

Задачі розміщення об'єктів широко застосовуються в різних сферах логістики: вибір місця для центрів обслуговування або дистрибуційних центрів; розташування сервісних пунктів у межах регіону. У багатьох дослідженнях розглядалися "центричні" моделі розміщення, які гарантують покриття попиту, на зразок задач розміщення лікарень чи пожежних станцій [4]. Подібні підходи застосовують також у плануванні розроблення територій, зрошувальних систем чи розташування базових станцій мобільного зв'язку.

Через значну кількість факторів для вибору локації часто застосовують методи багатокритеріальної оптимізації або експертне оцінювання. Це дає змогу додавати до моделі якісні чинники поряд із кількісними [5]. Проте залученість експертів привносить суб'єктивність і заздалегідь обмежує простір пошуку можливих місць. Відомо, що такий підхід спрощує задачу, але може знижувати об'єктивність результатів.

У нелінійних моделях використовують метод множників Лагранжа, щоб брати до уваги обмеження завантаженості виробничих ліній та мінімально необхідний обсяг переробки для досягнення економічної ефективності. Застосування p-медіани дає змогу мінімізувати середню відстань між джерелами відходів та пунктами їх збирання. У разі нерівномірного просторового розподілу споживачів та складної геометрії території застосовують адаптовані алгоритми для неперервної p-медіани, які демонструють значне зниження транспортних витрат за великих обсягів інформації та заданих просторових обмежень. У статті [6] описано підхід до розв'язання задачі розміщення об'єктів на неперервному просторі способом знаходження p-оптимальних центрів (медіан), які мінімізують сумарну відстань до точок попиту. Запропонований авторами еволюційний алгоритм демонструє високу ефективність за умов значного обсягу інформації та складної геометрії простору. Метод неперервної p-медіани є особливо корисним для задач, де попит або джерела відходів розподілені нерівномірно або без чіткої дискретизації, що робить його придатним для моделювання систем логістики у сфері ресайклінгу. Результати дослідження продемонстрували, що метод дає змогу досягти суттєвого зменшення транспортних витрат і є гнучким до адаптації під особливі просторові обмеження.

Нині існує низка алгоритмів і методів, призначених для розв'язання багатовимірних і складних задач розміщення. Зокрема методи на основі генетичних алгоритмів і евристичних підходів активно використовуються завдяки їх здатності швидко знаходити ефективні рішення без потреби в складних теоретичних обґрунтуваннях. К. Хак, Дж. Джойнс та М. Кей [7] порівнюють генетичні алгоритми з іншими евристичними методами в процесі розв'язання масштабних задач розміщення та розподілу ресурсів. У дослідженні [8] запропоновано вдосконалений цей самий алгоритм із застосуванням теорії сірих систем для розв'язання задачі оптимального розміщення об'єктів охорони здоров'я в умовах невизначеності.

Серед українських дослідників можна згадати праці Л. Коряшкіної та Д. Лубенця [9] – алгоритми територіальної сегментації, що формулюють неперервні задачі оптимального мультиплексного розбиття множин з обмеженнями, формалізацією яких є математична модель задачі розміщення сервісних центрів і територіального зонування ресурсів. О. Кісєльова, С. Ус і О. Притоманова [10] запропонували двохетапну математичну модель оптимального розбиття та розподілу ресурсів у системах, що мають поєднану неперервну й дискретну структуру, яка може бути адаптована для задач розміщення в логістиці, разом з екологічними системами.

У контексті прикладних моделей для ресайклінгу важливими є дослідження О. Станіної, яка запропонувала

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

методи двохетапного розміщення виробництва з неперервно розподіленим ресурсом [11]. Історично важливими для теорії залишаються роботи, що сформували основу класичних моделей розміщення. Також варто віддати належне українським ученим, які досліджують розміщення об'єктів у логістичних і регіональних системах.

З огляду на багатофакторний характер задачі та наявність обмежень у науковій літературі запропоновано низку підходів до її розв'язання. Серед актуальних методів було обрано п'ять, що демонструють ефективність у задачах оптимального розміщення складів у межах логістичних мереж ресайклінгу складної техніки.

У праці [12] запропоновано гібридний симевристичний підхід для розв'язання задачі розміщення підприємств в умовах невизначеності. Поєднання *Tabu Search* (методу для поступового покращення рішень із запам'ятовуванням попередніх кроків, щоб не повертатися до вже перевірених варіантів) і *Path Relinking* (стратегії, що дає змогу знаходити ще кращі рішення способом поєднання двох успішних) дає змогу ефективно знаходити рішення, що мінімізують загальні витрати з огляду на можливі коливання попиту та витрат на обслуговування.

Дослідження [13] демонструє використання імітаційного моделювання для розв'язання задачі розміщення складів у ланцюгах постачання. Застосовано програмні засоби FLEXSIM та SIMMAG3D для моделювання різних сценаріїв розміщення, що допомагає оцінити вплив розташування на ефективність логістичної системи.

Крім того, щоб зважати на такі чинники, як рельєф, стан земель, наявність забудови, доступність транспорту тощо, нерідко звертаються до GIS-аналізу. У роботі [14] розроблено математичні моделі та методи зонування й розміщення об'єктів у системах екстреної логістики з використанням геоінформаційних систем (ГІС). Зокрема застосовано діаграми Вороного для визначення зон обслуговування та оптимального розміщення центрів реагування в умовах надзвичайних ситуацій.

Нижче в таблиці наведено порівняльну характеристику методів, описано їх інструменти, ключові властивості, переваги й обмеження (табл. 1).

Таблиця 1. Підходи до розв'язання задачі розміщення складів

№	Назва підходу	Основні інструменти / методи	Ключові властивості	Переваги	Обмеження
1	Гібридний симевристичний підхід	Tabu Search, Path Relinking	Оптимізація мережі в умовах невизначеності та змінного попиту	Дає квазіоптимальні рішення за великої розмірності; адаптивний до змін	Потребує налаштування параметрів; високі обчислювальні витрати
2	Імітаційне моделювання логістики	FLEXSIM, SIMMAG3D	Моделювання функціювання мережі складів за різних сценаріїв	Дає змогу оцінити наслідки змін у структурі логістичної мережі	Потребує наявності достовірної інформації; складність валідації
3	ГІС з діаграмами Вороного	ArcGIS, QGIS, Python	Просторове зонування та аналіз доступності для оптимізації покриття територій	Візуалізація та інтеграція реальних картографічних показників	Не бере до уваги потужність об'єктів; обмежена точність у щільній міській забудові
4	Математичне моделювання (нелінійна оптимізація, теорія двоїстості)	Метод множників Лагранжа	Розв'язання задачі з обмеженнями за потужністю складів і транспортом	Зважає на реальні виробничі обмеження; точність	Складність формулювання та інтерпретації для значної кількості об'єктів
5	Геометрична оптимізація, інтегральні методи	Метод неперервної p-медіани	Оптимізація розміщення об'єктів у просторі з неперервним розподілом відходів	Зменшує середні витрати на транспортування; бере до уваги розсіяну географію	Вимагає інтегральних розрахунків; не підходить для малих територій

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Для вибору оптимальних місць розташування складів і виробничих майданчиків ресайклінгу зазвичай застосовують комплексні оптимізаційні моделі, що беруть до уваги транспортні витрати, виробничі потужності, географічні особливості та соціально-економічні обмеження. Проте через складність реальних логістичних систем одних лише математичних методів недостатньо. Тому все більшої ваги набуває імітаційне моделювання як ефективне доповнення до оптимізаційних підходів, що дає змогу гнучко аналізувати різні сценарії, зважати на динаміку змін середовища й адаптувати рішення у взаємодії з децидентом [15].

Постановка завдання дослідження

Аналіз сучасних досліджень свідчить, що більшість робіт зосереджено на базових моделях розміщення, які розширюються внаслідок долучення додаткових умов, таких як обмеження на обсяги виробництва, кількість видів продукції, багатокритеріальність тощо.

Проте значна частина досліджень обмежується одноетапними задачами або розглядає транспортування одного виду ресурсу, що не відповідає складності реальних систем ресайклінгу, де необхідно брати до уваги багатоступеневі процеси та різноманітність видів відходів. Отже, моделі, що зважають на багатопродуктивність і багатоетапність, залишаються недостатньо розробленими.

Зазначимо, що існує обмежена кількість досліджень, присвячених оптимальному розміщенню сміттєсортувальних підприємств у контексті інтегрованих систем управління відходами. Більшість таких робіт присвячено перерозподілу твердих побутових відходів, незважаючи на багатокомпонентність процесу ресайклінгу СТ та особливості інфраструктури.

Мета дослідження – створення моделі формування логістичної інфраструктури підприємств ресайклінгу СТ, упровадження якої дасть змогу зменшити витрати на транспортування вторинної сировини та розбудову інфраструктури ресайклінгу СТ.

Для здійснення окресленої мети необхідно виконати такі завдання:

1) проаналізувати сучасний стан сфери перероблення компонентів складної техніки;

2) визначити особливості логістичних процесів ресайклінгу складної техніки;

3) розробити модель формування логістичної інфраструктури ресайклінгу СТ.

Зі свого боку створення оптимізаційної моделі потребує розв’язання таких питань:

– визначення типу оптимізаційної задачі;

– формування цільової функції, що бере до уваги основні види витрат на формування інфраструктури;

– формулювання обмежень оптимізаційної задачі.

Сучасний стан сфери перероблення компонентів складної техніки

Проаналізуємо сучасний стан сфери перероблення вторинної сировини в Україні та за кордоном.

До повномасштабної війни Україна мала низький рівень перероблення відходів на тлі домінування захоронення: перероблялося лише близько 5–7% побутових відходів, а решта потрапляла на численні легальні й нелегальні сміттєзвалища. Це призводило до дефіциту якісної вторинної сировини, оскільки система роздільного збирання відходів фактично була відсутня. Унаслідок цього вітчизняні підприємства вимушено імпортували макулатуру, полімерні матеріали та склобій для забезпечення своїх потреб у сировині [16].

Станом на 2020 р. в Україні працювало 17 підприємств із перероблення макулатури, 39 – полімерів, 19 – ПЕТФ-сировини, 16 – склобою (рис. 1). Щорічно на переробні підприємства України потрапляє 160 тис. т полімерів, 50 тис. т ПЕТ-пляшок. Проте потужність підприємств із перероблення полімерів та ПЕТ-пляшок, наприклад, завантажена тільки на 65%. У 28 населених пунктах працюють 34 сортувальні лінії, у 17 населених пунктах упроваджуються сортувальні комплекси. У 1462 населених пунктах впроваджується роздільне збирання побутових відходів [17]. Серед підприємств є один сміттєспалювальний – київський завод "Енергія", що спалює близько 25 % загальної кількості сміття столиці, житомирський сміттєпереробний завод, відкритий 2023 р. з потужністю перероблення 85000 т на рік, та ще один у Львові, який планують запустити 2025 р. з потужністю 250000 т на рік. Також міністерство захисту довкілля та природних ресурсів сформувало план розміщення нових заводів, який передбачає 207 проєктів, що мають з’явитися по всій країні до 2034 р.

У Європі працює понад 500 сміттєспалювальних заводів, але країни ЄС поступово відмовляються

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

від спалювання через екологічні ризики, високу вартість та інші фактори. Спалювання – один із найдорожчих способів виробництва енергії – не знищує відходи повністю, а лише зменшує їх об'єм і підвищує токсичність, створюючи небезпечні побічні продукти, що потребують захоронення.

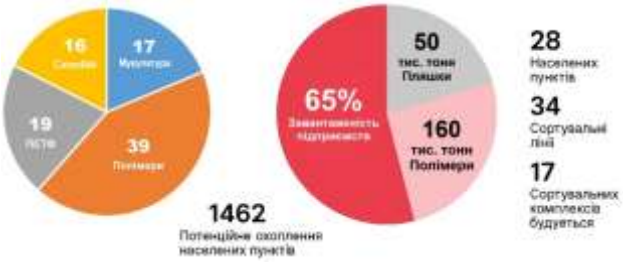


Рис. 1. Інфраструктура перероблення вторинної сировини в Україні станом на 2020 р.

Серед європейських країн, де розвинена інфраструктура ресайклінгу, однією з найбільш чистих є Швеція, яка за офіційною інформацією переробляє майже 100% твердих побутових відходів (ТПВ). Проте ці показники суперечать дійсності, оскільки майже 50% шведських відходів спалюється, а це абсолютно інший процес від перероблення, і в Україні діє 32 сміттєспалювальні заводи, де з 2000 р. до 2015 р. спалювали в середньому майже 50% усіх своїх ТПВ. Щодо перероблення, 2015 р. Швеція переробила 32% від загального обсягу ТПВ (48% разом з органічними відходами), хоча це і дуже далеко від загальної мети Європейської комісії щодо перероблення ТПВ у 65% до 2030 р.

2020 р. Швеція запровадила податок 75 крон за тонну спалених відходів, зокрема імпортований RDF із Великобританії. Метою є скорочення спалювання для досягнення кліматичних цілей і ефективнішого управління відходами [18].

Станом на 2017 р. в Італії функціонувало 285 підприємств із компостування, які обробляли близько 3,3 млн т твердих побутових відходів. У Литві з 2011 р. зафіксовано суттєве зростання обсягів компостування – з приблизно 2,5% до понад 20%. Водночас Ірландія демонструє стратегічний підхід до управління окремими потоками відходів, встановивши мету: до 2030 р. забезпечити перероблення або компостування 70% усіх пакувальних матеріалів [19].

В Іспанії діє 137 сміттєпереробних підприємств, які щороку обробляють близько 13 млн т відходів. Вони поділяються на шість категорій: чотири працюють із змішаними побутовими відходами (сортування, компостування, біосушіння, біометанізація), а дві – з роздільно зібраними біовідходами. Із змішаних відходів вдається вилучити близько 5% вторинної сировини (переважно пластик, метали, папір), а решта йде на виробництво біогазу або у відходи [20].

Відповідно до статистичних показників [21], серед заводів, що фізично переробляють відходи (зокрема пластикові, електронні, металеві тощо) на нові матеріали, лідирує Німеччина з кількістю 1840 од., а 10-те місце посідає Бельгія з 250 од. (рис. 2).

#	Країн	Одиниці (Підприємства)	Минулого року		
1	Німеччина	1840	2023 рік	+4,54%	+1,53%
2	Велика Британія	1290	2023 рік	+1,74%	+1,75%
3	Словеччина	1100	2023 рік	+10,18%	+13,08%
4	Італія	1090	2023 рік	+1,87%	+4,33%
5	Франція	773	2023 рік	+0,26%	+3,95%
6	Іспанія	564	2023 рік	+1,62%	+3,32%
7	Польща	449	2023 рік	+1,13%	+0,77%
8	Австрія	312	2023 рік	0%	+0,19%
9	Чехія	305	2023 рік	-2,24%	-2,26%
10	Бельгія	250	2023 рік	+0,81%	+4,56%

Рис. 2. Кількість підприємств із перероблення в країнах ЄС

Після проведення аналізу було взято до уваги приклади інфраструктури розвинених у сфері ресайклінгу країн. Порівнявши їх зі статистичними показниками України, можна прийти до висновку, що в країні необхідно розробити оптимізаційні рішення для створення системи ефективного перероблення СТ.

Особливості логістичних процесів ресайклінгу складної техніки

Загалом процес ресайклінгу СТ є багатоетапним.

На першому етапі організовують збирання та сортування відпрацьованої СТ, визначають основні джерела утворення відходів, серед яких промислові підприємства, комунальні служби, торговельні мережі, будівельні компанії та приватні домогосподарства. Аналіз обсягів, складу й періодичності утворення відходів дає змогу сформувати стратегію оптимального розташування пунктів збирання. Вони розміщуються з огляду на територіальну щільність відходів, транспортну доступність, можливість розширення під'їзних шляхів до сортувальних центрів.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Після збирання відходи піддаються первинному сортуванню, під час якого відокремлюють різні матеріали, такі як пластик, метал, скло, органічні відходи та електроніка. Додатково матеріали пресують та ущільнюють для оптимізації транспортування, подрібнення великих об'єктів, а також вилучення небезпечних компонентів, що не підлягають подальшому обробленню.

Другий етап ресайклінгу СТ охоплює транспортування відсортованих відходів із пунктів збирання до переробних підприємств, безпосередньо технологічні процеси перероблення, а також подальший розподіл отриманої вторинної сировини між кінцевими споживачами.

На початку цього етапу необхідно забезпечити організацію транспортних процесів між пунктами збирання та сортування до переробних центрів. Оптимізація маршрутів транспортування ґрунтується на тому, щоб брати до уваги територіальне розташування об'єктів, обсяги відходів, які підлягають перевезенню, та пропускну спроможність кожного підприємства.

Після цього переробляють відходи, що надходять на підприємства. Важливим аспектом є контроль за відповідністю обсягів відходів потужностям підприємств, що реалізується за допомогою математичних моделей балансування потоків сировини.

Після перероблення отримана вторинна сировина або енергетичні ресурси розподіляються між кінцевими споживачами. Цей процес передбачає аналіз попиту на перероблені матеріали, оптимізацію мережі постачань і вибір найефективніших транспортних засобів для їх доправлення. Оскільки місця споживання можуть бути територіально розподіленими, важливим є вибір місць розташування стратегічних центрів зберігання та перевантаження.

Отже, узагальнений процес ресайклінгу СТ містить два етапи, і, відповідно, структура системи передбачає два рівні (рис. 3):

1 рівень – локальні підприємства зі збирання й сортування (ЛПЗС), що займаються первинним обробленням відходів: збиранням, сортуванням і підготовкою до подальшого перероблення;

2 рівень – підприємства з перероблення (ПП) відходів. Вони можуть бути вузькоспеціалізованими або широкого профілю, що переробляють різні види відходів.

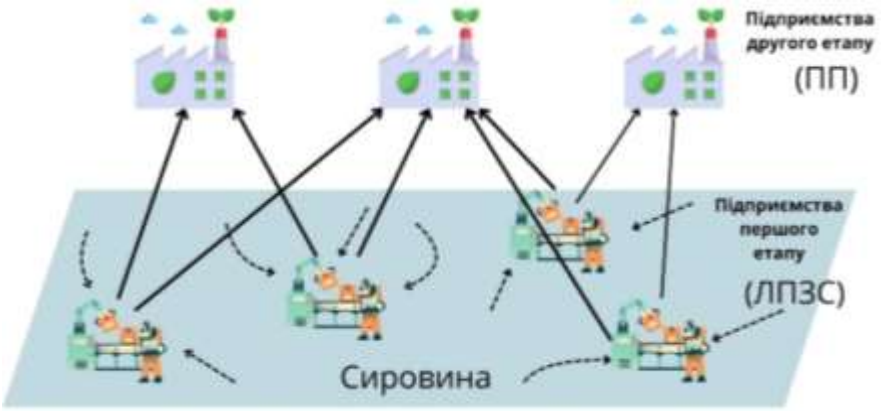


Рис. 3. Схема двохетапного процесу ресайклінгу СТ

Отже, у системах, що поєднують логістичні й виробничі процеси, ключовими компонентами є два типи підприємств: перші (підприємства першого рівня) призначені для збирання просторово розосереджених матеріальних ресурсів, другі (підприємства другого рівня) – для подальшого використання або перероблення ресурсів. Зазвичай підприємства, що здійснюють первинний збір, обслуговують закріплену за ними територію. Рух матеріальних потоків у межах такої системи відбувається поетапно: спочатку – від джерел виникнення ресурсу до пунктів збирання, а далі – від попередньо оброблених чи відсортованих матеріалів до спеціалізованих заводів, що забезпечують кінцеве оброблення або споживання. У контексті мережі підприємств, що управляють відходами, ресурсом є відходи у вигляді СТ, непридатної до використання та з вичерпаним ресурсом [22].

Зазначимо, що витрати в системі ресайклінгу та перероблення вторинної сировини тісно пов'язані з виконанням завдання розташування та визначення оптимальної кількості ЛПЗС та ПП. Зі збільшенням

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

кількості перших зростають витрати на транспортування відходів до других, однак водночас скорочуються витрати на збирання відходів у населених пунктах, що може загалом зменшити витрати на логістику. Крім того, підвищується зручність для населення та інших джерел відпрацьованої СТ, оскільки пункти збирання стають більш доступними. Водночас зростає загальний обсяг проміжних запасів відходів у системі, хоча обсяг, що накопичується на кожному окремому ЛПЗС, зменшується. Експлуатаційні витрати на одне локальне підприємство можуть бути нижчими, однак загальні витрати на утримання мережі таких об'єктів і їх управління зростають, хоча темпи цього зростання поступово сповільнюються. Водночас скорочення кількості ЛПЗС збільшує відстань між джерелами утворення відходів і місцями збирання, що може ускладнити охоплення територій, доступність для населення та знизити ефективність системи перероблення. Тому для мінімізації витрат важливо не лише оптимізувати кількість підприємств обох етапів, а й правильно обрати їх розташування, зважаючи на відстань до населених пунктів, обсяг вантажообігу та наявну транспортну інфраструктуру.

Отже, для розв'язання завдання розміщення підприємств першого й другого етапів ресайклінгу з метою мінімізації витрат, що передбачають витрати на розбудову нових заводів і витрати на транспортування відходів між ПП та ЛПЗС, необхідно сформувати математичну постановку двохетапної задачі оптимального розподілення ресурсу, що має брати до уваги декілька видів ресурсу (вантажу).

Модель формування логістичної інфраструктури ресайклінгу складної техніки

У процесі формування логістичної інфраструктури ресайклінгу компонентів СТ можливі декілька стратегій:

1) використання мережі наявних підприємств, якщо їх потужності покривають потреби перероблення;

2) розбудова мережі наявних заводів із долученням нових, щоб задовольнити потреби перероблення.

У цьому разі постає питання досягнення компромісу між транспортними витратами та витратами на будівництво й експлуатацію цих підприємств.

Одним із ключових підходів до моделювання та оптимізації розміщення логістичних об'єктів є задачі оптимального розбиття множин (ОРМ). Ці задачі передбачають поділ простору на підмножини таким чином, щоб мінімізувати визначений цільовий функціонал, наприклад, витрати на транспортування, оброблення або утилізацію ресурсів.

Загалом **задача розміщення-розподілу** полягає у визначенні кількості нових об'єктів, їх просторового розташування та оптимального розподілу потоків між новими й наявними пунктами з огляду на мінімізацію загальних витрат на транспортування до кінцевих споживачів.

Зважаючи на вказану особливість ресайклінгу СТ, це дослідження присвячено формулюванню двохетапної виробничо-транспортної задачі, яка є окремим випадком задачі розміщення-розподілу. Така постановка відтворює послідовний процес збирання, первинного сортування та подальшого перероблення відходів, що дає змогу оптимізувати логістичні витрати та підвищити ефективність управління ресурсами.

Постановка узагальненої **двохетапної задачі розміщення-розподілу** може бути задана в такий спосіб. Необхідно обрати розташування заводів першого й другого етапів та визначити маршрути, за якими перевозиться визначена кількість продукту, що постачається від кожного підприємства першого етапу до підприємств другого етапу і від підприємств другого етапу споживачам таким чином, щоб сумарні витрати на доправлення продукції були мінімальні.

Для ефективної реалізації двохетапної виробничо-транспортної моделі необхідно застосувати оптимізаційні методи, що дають змогу мінімізувати логістичні витрати, оптимізувати розміщення пунктів збирання та перероблення, а також забезпечити баланс між обсягами зібраних і перероблених відходів.

Зважаючи на мережну структуру системи підприємств ресайклінгу, запропонуємо познаки, що дадуть змогу розглянути її складники [3]:

l – порядковий номер виду відходу, $l = 1 \ldots L$;

L – кількість видів відходів, що підлягають збиранню, сортуванню та переробленню;

i – порядковий номер пункту відправлення (ПВ) відходів, $i = 1 \ldots I$;

I – кількість пунктів відправлення відходів;

j – порядковий номер локального пункту зі збирання й сортування (ЛПЗС), $j = 1 \ldots J$;

J – кількість локальних пунктів ЛПЗС;

k – порядковий номер ПП, $k = 1 \ldots K$;

K – кількість підприємств з перероблення.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Подамо інфраструктуру ресайклінгу СТ, запропоновану на рис. 1, у формалізованому вигляді (рис. 4). Познаки на рис. 4 відповідають тим, що наведені вище.

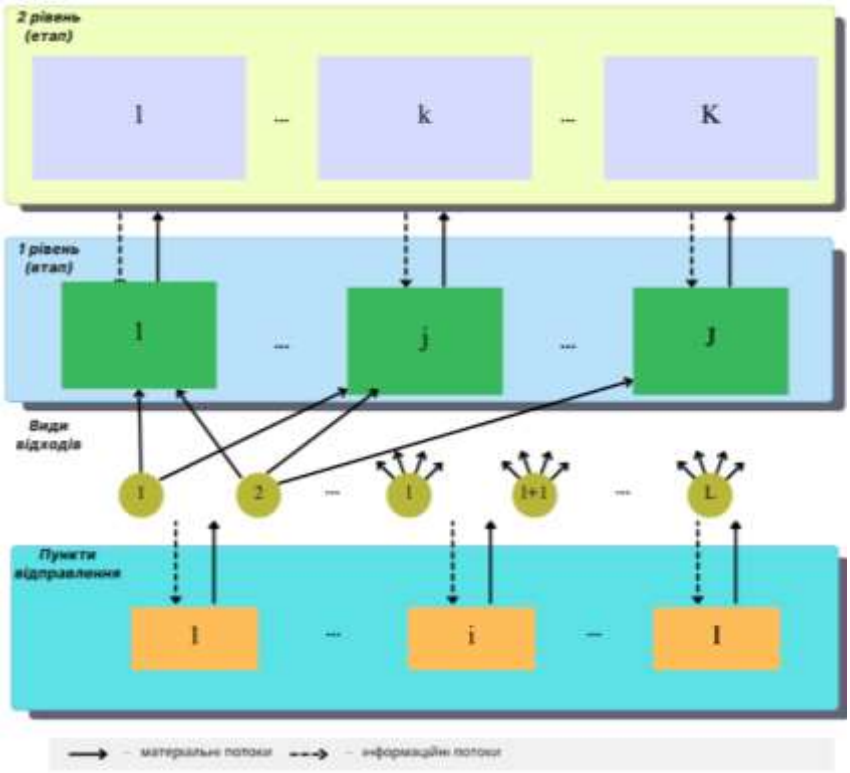


Рис. 4. Інфраструктура ресайклінгу СТ

Зауважимо, що транспортування відходів із пунктів відправлення до ЛПЗС може здійснюватись за допомогою джерел, що утворюють відходи, тому необхідність брати до уваги витрати на це транспортування має вирішуватись у кожному разі окремо. Якщо основними постачальниками відпрацьованої техніки є великі компанії та процес відбувається на регулярній основі, тоді зважати на витрати на транспортування необхідно обов'язково. У роботі розглянуто загальний випадок з огляду на витрати на транспортування СТ з пунктів відправлення до ЛПЗС.

Крім того, здебільшого за умови існування достатньої кількості підприємств другого етапу (ПП), задача полягатиме в пошуку місць розташування підприємств ЛПЗС. До ПП належать спеціалізовані заводи з перероблення металу, пластику, скла тощо, у яких відбуваються складні технологічні процеси, а отже, є потреба в дорогому обладнанні. Тому важливо розглядати сучасні переробні підприємства та розв'язувати завдання розміщення ЛПЗС з огляду на наявне розміщення ПП. Зазначимо, що додатковим обмеженням є те, що не береться до уваги попит кінцевого споживача продукції переробних підприємств.

Відповідно, основні фактори в моделі формування логістичної інфраструктури ресайклінгу СТ подамо таким чином:

v_{lj} – об'єм відходів l-го виду, що збирається та сортується в j-му ЛПЗС;

v_{lij} – об'єм відходів l-го виду, що транспортується з i-го пункту відправлення до j-го ЛПЗС;

v_{ljk} – об'єм відходів l-го виду, що транспортується з j-го ЛПЗС до k-го ПП;

v_{lk} – об'єм відходів l-го виду, що переробляється на k-му ПП;

d_{ij} – відстань від i-го пункту відправлення до j-го ЛПЗС;

d_{jk} – відстань від j-го ЛПЗС до k-го ПП, км;

c_{lj} – витрати на збирання та сортування одиниці об'єму відходів l-го виду в j-му ЛПЗС;

c_{lij} – питомі витрати (на 1 км) транспортування одиниці об'єму l-го виду відходів з i-го пункту відправлення до j-го ЛПЗС;

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

c_{ljk} – питомі витрати на 1 км транспортування одиниці об'єму l-го виду відходів з j-го ЛПЗС до k-го ПП;

c_{lk} – витрати на перероблення одиниці об'єму відходів l-го виду на k-му ПП.

1. Розглянемо спочатку транспортні витрати на перевезення відходів.

Сумарні витрати на транспортування з пунктів відправлення до ЛПЗС:

$$C_{tr}^{ПВ\text{-}ЛПЗС} = \sum_{l=1}^{L}\sum_{i=1}^{I}\sum_{j=1}^{J} c_{lij}\, d_{ij} v_{lij} x_{ij},$$

де x_{ij} – цілочисельна (булева) змінна:

$$x_{ij} = \begin{cases} 1, \text{ якщо наявний маршрут перевезень} \\ \quad \text{від } i\text{-го пункту відправлення до } j\text{-го ЛПЗС}, \\ 0, \text{ якщо маршрут відсутній.} \end{cases}$$

Сумарні витрати на транспортування з ЛПЗС до ПП:

$$C_{tr}^{ЛПЗС\text{-}ПП} = \sum_{l=1}^{L}\sum_{j=1}^{J}\sum_{k=1}^{K} c_{ljk}\, d_{jk} v_{ljk} x_{jk},$$

де x_{jk} – цілочисельна (булева) змінна:

$$x_{jk} = \begin{cases} 1, \text{ якщо наявний маршрут перевезень} \\ \quad \text{від } j\text{-го ЛПЗС до } k\text{-го ПП}, \\ 0, \text{ якщо маршрут відсутній.} \end{cases}$$

Тоді сумарні витрати на перевезення відходів та напівфабрикатів дорівнюють

$$C_{tr} = C_{tr}^{ПВ\text{-}ЛПЗС} + C_{tr}^{ЛПЗС\text{-}ПП} =$$
$$= \sum_{l=1}^{L}\sum_{i=1}^{I}\sum_{j=1}^{J} c_{lij}\, d_{ij} v_{lij} x_{ij} + \sum_{l=1}^{L}\sum_{j=1}^{J}\sum_{k=1}^{K} c_{ljk}\, d_{jk} v_{ljk} x_{jk}.$$

2. Наступним кроком розглянемо витрати, пов'язані з розміщенням підприємств першого та другого рівнів, тобто ЛПЗС та ПП відповідно.

Нехай $M_r = (M_1, M_2)$ – множина можливих пунктів розміщення r-го рівня, $r = 1, 2$;

M_1 – множина можливих місць розміщення підприємств першого рівня (ЛПЗС);

M_2 – множина можливих місць розміщення підприємств другого рівня (ПП);

g_j – витрати на розміщення ЛПЗС у пункті j;

g_k – витрати на розміщення ПП у пункті k.

Задача розв'язується за умови, що кожне підприємство першого рівня має свою зону обслуговування, ці зони не перетинаються та покривають всю територію.

Потрібно обрати підмножини пунктів розміщення кожного рівня (етапу) $S^r \subset M_r$, $r = 1, 2$ та здійснити призначення обраних підприємств на пункти відправлення вторинної сировини так, щоб мінімізувати сумарні витрати на розміщення всіх обраних підприємств і на транспортування продукту.

Додамо цілочисельні (булеві) змінні y_j, $(j \in M_1)$ та y_k $(k \in M_2)$, що визначають необхідність розміщення підприємств першого та другого рівнів відповідно:

$$y_j = \begin{cases} 1, \text{ якщо ЛПЗС розміщується в пункті } j, \\ 0, \text{ в іншому разі.} \end{cases}$$

$$y_k = \begin{cases} 1, \text{ якщо ПП розміщується в пункті } k, \\ 0, \text{ в іншому разі.} \end{cases}$$

Витрати на розміщення підприємств першого (ЛПЗС) та другого (ПП) етапів розраховуються таким чином:

$$G = \sum_{j \in M_1} g_j y_j + \sum_{k \in M_2} g_k y_k.$$

Цільовою функцією є мінімізація всіх сумарних витрат, пов'язаних із розміщенням підприємств, виробництвом і транспортуванням сировини (відходів, відпрацьованої СТ) і напівфабрикатів (після сортування та попереднього оброблення) з ЛПЗС до ПП.

Отже, математична модель двохетапної задачі розміщення-розподілу формулюється таким чином: знайти такі значення змінних x_{ij}, x_{jk}, y_j, y_k, що надають мінімального значення цільовій функції

$$\sum_{j \in M_1} g_j y_j + \sum_{k \in M_2} g_k y_k +$$
$$+ \sum_{l=1}^{L}\sum_{i=1}^{I}\sum_{j=1}^{J} c_{lij}\, d_{ij} v_{lij} x_{ij} + \sum_{l=1}^{L}\sum_{j=1}^{J}\sum_{k=1}^{K} c_{ljk}\, d_{jk} v_{ljk} x_{jk} \to \min,$$

з огляду на вказані нижче обмеження.

Зазначимо, що в статті проаналізовано задачу розміщення за наявності обмежень на потужність підприємств. Насамперед розглянемо обмеження виробничої потужності підприємств із перероблення.

1. Кожний j-й локальний пункт ЛПЗС має обмеження за об'ємом відходів, що можуть бути перероблені за певний період:

$$\sum_{i=1}^{I}\sum_{l=1}^{L} v_{lij} x_{ij} \le V_{j\max},$$

де $V_{j\max}$ – максимальний об'єм відходів, який може бути перероблений на j-му ЛПЗС за певний період (потужність j-го ЛПЗС).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Іншими словами, сумарні запаси ресурсу в зоні обслуговування j-го підприємства першого етапу не мають перевищувати потужності цього підприємства.

2. Наступна умова передбачає, що кількість продукту, який вивозиться з j-го ЛПЗС на k-ті ПП, дорівнює потужності цього ЛПЗС:

$$\sum_{l=1}^{L}\sum_{k=1}^{K} v_{ljk} x_{jk} = V_{j\max} .$$

3. Переробні заводи мають обмеження за загальним об'ємом перероблення:

$$\sum_{j=1}^{J}\sum_{l=1}^{L} v_{ljk} x_{jk} \le V_{k\max} ,$$

де $V_{k\max}$ – максимальний об'єм відходів, що може бути перероблений на k-му ПП за певний період (потужність k-го ПП).

Ця умова відповідає тому, що кількість продукту, що доправляється до k-го підприємства другого етапу, не має перевищувати виробничу потужність цього підприємства.

Додаємо також обмеження, властиві для задачі розміщення.

4. Кожен пункт відправлення має бути під'єднаний до одного ЛПЗС:

$$\sum_{j\in J} x_{ij} = 1, \forall i \in I .$$

5. Пункт відправлення може бути під'єднаний лише до відкритого ЛПЗС:

$$x_{ij} \le y_j, \forall i \in I, \forall j \in J .$$

6. ЛПЗС може бути під'єднаний лише до відкритого переробного підприємства:

$$x_{jk} \le y_k, \forall j \in J, \forall k \in K.$$

Умова бінарності змінних:

$$x_{ij}, x_{jk}, y_j, y_k \in \{0,1\} .$$

Необхідно обрати розташування підприємств першого та другого етапів і визначити маршрути, якими транспортується певна кількість продукту з пунктів відправлення до кожного підприємства першого етапу й від кожного підприємства першого етапу до підприємств другого етапу таким чином, щоб сумарні витрати були мінімальні.

Наразі існують алгоритми й методи для розв'язання такого виду задач з огляду на більшу розмірність і складність, які докладно описано в роботі [11, с. 44–48]. Переважно це методи розв'язання задач дискретної оптимізації, зважаючи на лінійні та нелінійні обмеження. Види обмежень для розв'язання задачі в конкретних умовах впливатимуть на можливість і вибір методу розв'язання.

Результати досліджень та їх обговорення

Задачі розміщення-розподілу вирізняються високою складністю як на етапі формалізації, так і під час практичного впровадження запропонованих рішень. Моделювання таких задач потребує участі досвідчених аналітиків, оскільки зміна місця розташування об'єкта після реалізації проєкту є надзвичайно затратною та практично незворотною процедурою. Вартість помилки на етапі проєктування системи може суттєво вплинути на ефективність її подальшого функціювання, особливо в умовах нестабільного середовища або обмежених ресурсів.

Математичні моделі, у яких взято до уваги етапи перероблення та просторовий розподіл сировини, дають змогу суттєво знизити витрати на транспортування в межах комплексних логістичних систем. Це особливо важливо для організації виробничо-переробних ланцюгів, де джерела ресурсів розміщені нерівномірно, як, наприклад, у процесі збирання великогабаритних відходів після руйнації інфраструктури. Оптимізація переміщення матеріальних потоків на кожному етапі системи забезпечує зниження як витрат, так і часових затримок у логістиці.

Для побудови ефективного математичного забезпечення задач цього типу використовуються положення теорії неперервних задач оптимального розбиття множин, методи розміщення центрів підмножин, концепції двоїстості та підходи, основані на класичних алгоритмах лінійного програмування транспортного типу.

Варто зазначити, що в разі, коли ресурс має непередбачуваний характер розподілу, а також коли кількість джерел постачання надмірно велика й можуть виникнути проблеми зростання обчислювальної складності, необхідно застосовувати методи імітаційного моделювання.

Висновки

У сучасних умовах глобалізації зростає потреба в оптимізації логістичних процесів у сфері поводження з відходами, зокрема в системах ресайклінгу. Раціональне розміщення інфраструктурних об'єктів, пов'язаних зі збиранням, обробленням і переробленням відходів, є складною багатокритеріальною задачею,

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

що вимагає брати до уваги економічні, просторово-географічні, соціальні та інші фактори. Ефективне територіальне планування такої мережі сприяє зниженню логістичних витрат, підвищенню ефективності обслуговування територій та зменшенню негативного впливу на довкілля.

У статті проаналізовано сучасні підходи до задачі розміщення об'єктів інфраструктури ресайклінгу, зокрема математичних моделей p-медіани, центричних моделей, методів розбиття множин, а також імітаційного моделювання та GIS-аналізу. Виокремлено п'ять ефективних методів і наведено їх порівняльну характеристику. На основі аналізу методів установлено, що більшість наявних підходів зосереджені на одноетапні задачі ресайклінгу й не зважають на багатокомпонентність відходів.

Проаналізовано проблеми організації логістичної інфраструктури ресайклінгу СТ. Порівняно приклади інфраструктури розвинених країн і України у сфері ресайклінгу та визначено, що в нашій країні необхідно розробити оптимізаційні рішення для покращення системи ефективного перероблення СТ.

Описано структуру двохетапної системи перероблення відходів СТ, що містить локальні пункти збирання й сортування на першому рівні та підприємства з перероблення на другому. Було зазначено, що важливим є знаходження балансу між кількістю перших і других. Надмірна концентрація ЛПЗС покращує контроль і зручність для постачальників відпрацьованої СТ, але ускладнює логістику. Водночас їх нестача призводить до зниження доступності та охоплення джерел відходів.

Сформовано математичну модель двохетапної задачі розміщення-розподілу, що бере до уваги виробничі обмеження, транспортування декількох видів ресурсу та логістичні витрати.

Отже, **наукову новизну** дослідження становить модель формування логістичної інфраструктури ресайклінгу СТ, яка зважає на двохетапний характер перероблення компонентів СТ та транспортування декількох видів ресурсу, впровадження якої дасть змогу зменшити витрати на транспортування вторинної сировини та розбудову інфраструктури ресайклінгу СТ унаслідок оптимізації мережі підприємств.

Актуальними завданнями, що можуть бути виконані в подальших дослідженнях, є: розроблення методів моделювання, які зважають на невизначеність та ризики (нестабільність об'єму та складу вторинної сировини, законодавчі зміни, технологічні інновації); взяття до уваги динамічної зміни потоків відходів у часі; інтеграція ГІС-технологій для точного визначення просторових особливостей територій; створення інтегрованих моделей, що дають змогу оптимізувати одночасне розміщення кількох взаємопов'язаних типів інфраструктури.

Список літератури

1. Lytvynenko O., Pavlov K., Mykhailova L. Waste management in Ukraine: organizational aspects. *E3S Web of Conferences*, Vol. 280, No. 11004. 2021. DOI: 10.1051/e3sconf/202128011004
2. Шишкін Е., Гайко Ю., Черноносова Т. Шляхи рециклінгу будівельного сміття під час післявоєнної відбудови зруйнованих міст. *Містобудування та територіальне планування*. 2024. Т. 85. С. 679–697. DOI: https://doi.org/10.32347/2076-815x.2024.85.679-697
3. Кікоть М., Малєєва Ю. Моделі формування логістичної інфраструктури ресайклінгу складної техніки. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. Т. 3, № 29. С. 15–28. DOI: https://doi.org/10.30837/2522-9818.2024.3.015
4. Wang H., Luo P., Wu Y. Research on the location decision-making method of emergency medical facilities based on WSR. *Sci Rep*. 2023. Vol. 13, No. 18011. DOI: 10.1038/s41598-023-44209-0
5. Dissanayake I., Gunathilake M. A comprehensive survey of recent advances in facility location problems: Models, solution methods, and applications. *Journal of Sustainable Development of Transport and Logistics*. 2024. Vol. 9, No. 2. P. 78–90. DOI: 10.14254/jsdtl.2024.9-2.6
6. Kazakovtsev L., Rozhnov I., Shkaberina G. Self-Configuring (1 + 1)-Evolutionary Algorithm for the Continuous p-Median Problem with Agglomerative Mutation. *Algorithms.* 2021. Vol. 14, No. 130. DOI: 10.3390/a14050130
7. Christopher R. Houck, Jeffrey A. Joines, Michael G. Kay. Comparison of Genetic Algorithms, Random Restart, and Two-Opt Switching for Solving Large Location-Allocation Problems. *Computers & Operations Research*. 1996. Vol. 22, No 6. P. 587–596

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

8. Zhang Y., Li J., Zhu Y., Xu H. An Improved Genetic Algorithm for Location Allocation Problem with Grey Theory in Public Health Emergencies. *International Journal of Environmental Research and Public Health*. 2022. Vol. 19, No. 15(9752). DOI: 10.3390/ijerph19159752
9. Коряшкіна Л., Лубенець Д. Математичні моделі та методи мультиплексного розбиття і багатократного покриття множин для задач розміщення-розподілу. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2023. Т. 4. С. 20–31. DOI: https://doi.org/10.32782/IT/2023-4-3
10. Kiseleva A. I., Prytomanova O. M., Us S. A. Solving a Two-Stage Continuous-Discrete Problem of Optimal Partition–Allocation with a Given Position of the Centers of Subsets. *Cybern Syst Anal.* 2020. Vol. 56. P. 1–12. DOI: https://doi.org/10.1007/s10559-020-00215-y
11. Станіна О. Д. Моделі та методи розміщення двоетапного виробництва з неперервно розподіленим ресурсом: дис. канд. техн. наук.: 01.05.02. Харків, 2019. 188 с.
12. Peidro, D., Martin, X.A., Panadero, J. Solving the uncapacitated facility location problem under uncertainty: a hybrid tabu search with path-relinking simheuristic approach. *Appl Intell*. 2024. Vol. 54. P. 5617-5638. DOI: 10.1007/s10489-024-05441-x
13. Szczepanski E., Jachimowski R., Izdebski M., Jacyna-Gołda I. Warehouse location problem in supply chain designing: a simulation analysis. *Archives of Transport*. 2019. Vol. 50, no. 2. P. 101-110. DOI: 10.5604/01.3001.0013.5752
14. Коряшкіна Л. С., Дзюба С. В. Математичні моделі та методи зонування і розміщення об'єктів в системах екстреної логістики. *Системні технології*. 2023. Т. 6, № 149. с. 107-122. DOI: https://doi.org/10.34185/1562-9945-6-149-2023-09
15. Катренко А. В., Антоняк Т. І. Розв'язання задач оптимального розміщення об'єктів методом імітаційного моделювання. *Вісник. Інформаційні системи та мережі*. 2011. № 715. С. 150–162
16. Kostiuchenko L.V., Marchuk V.Ye., Harmash O.M. Development of recycling infrastructure in Ukraine. *Electronic Scientific Journal "Intellectualization of Logistics and Supply Chain Management"*, 2021. Vol. 9. P. 44–52. DOI: 10.46783/smart-scm/2021-9-4
17. Із третього світу в перший. Реформа управління відходами в Україні. *PWC*. URL: https://www.pwc.com/ua/en/survey/2020/waste-management.pdf (дата звернення: 15.04.2025).
18. Мартиненко А. Міф про безпечне сміттєспалювання. *ГС Український Альянс Нуль Відходів*. URL: https://zwsociety.org/wp-content/uploads/mif-pro-bezpechne-smittyespalyuvannya.pdf (дата звернення: 22.04.2025).
19. Xiaoxia Cao, Paul N. Williams, Yuanhang Zhan, Scott A. Coughlin, John W. McGrath, Jason P. Chin, Yingjian Xu. Municipal solid waste compost: Global trends and biogeochemical cycling. *Soil & Environmental Health*. Vol. 1 No. 4. 2023. DOI: 10.1016/j.seh.2023.100038
20. Edo-Alcon N, Gallardo A, Colomer-Mendoza FJ, Lobo A. Efficiency of biological and mechanical-biological treatment plants for MSW: The case of Spain. *Heliyon*. 2024. Vol. 10 no. 4. DOI: 10.1016/j.heliyon.2024.e26353
21. European Waste Treatment and Disposal Number of Enterprises by Country. *ReportLinker*. URL: https://www.reportlinker.com/dataset/18a291d77e387c50e181b0c0e68e784476cc0ca5 (date of access: 07.05.2025).
22. Us S. A., Koriashkina L. S., Stanina O. D. An optimal two-stage allocation of material flows in a transport-logistic system with continuously distributed resource. *Radio Electronics, Computer Science, Control*. (2019). Vol. 1. DOI: 10.15588/1607-3274-2019-1-24
23. Kiseleva, E., Prytomanova, O., Hart, L. Solving a Two-stage Continuous-discrete Problem of Optimal Partitioning-Allocation with Subsets Centers Placement. *Open Computer Science*. 2020. Vol. 10, No. 1. P. 124-136. DOI: https://doi.org/10.1515/comp-2020-0142

References

1. Lytvynenko, O., Pavlov, K., Mykhailova, L. (2021), "Waste management in Ukraine: organizational aspects". *E3S Web of Conferences*, 280, 11004 p. DOI: https://doi.org/10.1051/e3sconf/202128011004
2. Shyshkin, E., Haiko, Y, Chernonosova, T. (2024), "Approaches to Recycling Construction Waste during the Post-War Reconstruction of Devastated Cities". *Urban Planning and Territorial Development*, 85, P. 679–697. DOI: https://doi.org/10.32347/2076-815x.2024.85.679-697
3. Kikot, M., Maleieva, Y. (2024), "Models of forming logistics infrastructure for complex equipment recycling". *Current State of Scientific Research and Technologies in Industry*, 3(29), P. 15–28. DOI: https://doi.org/10.30837/2522-9818.2024.3.015
4. Wang, H., Luo, P., Wu, Y. (2023), "DOI:Research on the location decision-making method of emergency medical facilities based on WSR". *Sci Rep*., 13(18011). DOI: https://doi.org/10.1038/s41598-023-44209-0

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

5. Dissanayake, I., Gunathilake, M. (2024), "A comprehensive survey of recent advances in facility location problems: Models, solution methods, and applications". *Journal of Sustainable Development of Transport and Logistics*, 9(2), P. 78–90. DOI: https://doi.org/10.14254/jsdtl.2024.9-2.6
6. Kazakovtsev, L., Rozhnov, I., Shkaberina, G. (2021), "Self-Configuring (1 + 1)-Evolutionary Algorithm for the Continuous p-Median Problem with Agglomerative Mutation". *Algorithms*, 14(130). DOI: https://doi.org/10.3390/a14050130
7. Christopher R. Houck, Jeffrey A. Joines, Michael G. Kay. (1996), "Comparison of Genetic Algorithms, Random Restart, and Two-Opt Switching for Solving Large Location-Allocation Problems". *Computers & Operations Research*, 22(6), P. 587–596.
8. Zhang, Y., Li, J., Zhu, Y., Xu, H. (2022), "An Improved Genetic Algorithm for Location Allocation Problem with Grey Theory in Public Health Emergencies". *International Journal of Environmental Research and Public Health*, 19(15(9752)). DOI: https://doi.org/10.3390/ijerph19159752
9. Koriashkina, L, Lubenets, D. (2023), "Mathematical Models and Methods of Multiplex Partitioning and Multiple Coverage of Sets for Location-Allocation Problems". *Information Technology: Computer Science, Software Engineering and Cyber Security*, 4, P. 20–31. DOI: https://doi.org/10.32782/IT/2023-4-3
10. Kiseleva, A. I., Prytomanova, O. M., Us, S. A. (2020), "Solving a Two-Stage Continuous-Discrete Problem of Optimal Partition–Allocation with a Given Position of the Centers of Subsets". *Cybern Syst Anal*., 56, P. 1–12. DOI: https://doi.org/10.1007/s10559-020-00215-y
11. Stanina, O. D. (2019), "Models and Methods for Locating Two-Stage Production with Continuously Distributed Resource" [Doctoral dissertation, National Technical University "Dnipro Polytechnic"]. Karazin University. 2019. 188 p.
12. Peidro, D., Martin, X.A., Panadero, J. (2024), "Solving the uncapacitated facility location problem under uncertainty: a hybrid tabu search with path-relinking simheuristic approach". *Appl Intell.*, 54, P. 5617–5638. DOI: https://doi.org/10.1007/s10489-024-05441-x
13. Szczepanski, E., Jachimowski, R., Izdebski, M., Jacyna-Gołda, I. (2019), "Warehouse location problem in supply chain designing: a simulation analysis". *Archives of Transport*, 50(2), P. 101–110. DOI: https://doi.org/10.5604/01.3001.0013.5752
14. Koriashkina, L. S., Dziuba, S. V. (2024), "Mathematical Models and Methods for Facility Location and Territory Zoning in Emergency Logistics Systems". *System Technologies*, 6(149), P. 107–122. DOI: https://doi.org/10.34185/1562-9945-6-149-2023-09
15. Katrenko, A. V., Antonyak, T. I. (2011), "Solving Optimal Facility Location Problems by Simulation Modeling". *Bulletin. Information Systems and Networks*, (715), P. 150–162.
16. Kostiuchenko, L.V., Marchuk, V. Ye., Harmash, O. M. (2021), "Development of recycling infrastructure in Ukraine". *Electronic Scientific Journal "Intellectualization of Logistics and Supply Chain Management"*, 9, P. 44–52. DOI: https://doi.org/10.46783/smart-scm/2021-9-4
17. PWC. (2020), "From Third World to First. Waste Management Reform in Ukraine". *PWC*. available at: https://www.pwc.com/ua/en/survey/2020/waste-management.pdf
18. Martinenko, A. (2020), "The Myth of Safe Waste Incineration". *Ukrainian Zero Waste Alliance NGO*. available at: https://zwsociety.org/wp-content/uploads/mif-pro-bezpechne-smittyespalyuvannya.pdf
19. Xiaoxia, Cao, Paul, N. Williams, Yuanhang, Zhan, Scott, A. Coughlin, John, W. McGrath, Jason, P. Chin, Yingjian Xu. (2023), "Municipal solid waste compost: Global trends and biogeochemical cycling". *Soil & Environmental Health*, 1(4). DOI: https://doi.org/10.1016/j.seh.2023.100038
20. Edo-Alcon, N, Gallardo, A, Colomer-Mendoza, FJ, Lobo, A. (2024), "Efficiency of biological and mechanical-biological treatment plants for MSW: The case of Spain". *Heliyon*, 10(4). DOI: https://doi.org/10.1016/j.heliyon.2024.e26353
21. ReportLinker. (2023), "European Waste Treatment and Disposal Number of Enterprises by Country". *ReportLinker*. available at: https://www.reportlinker.com/dataset/18a291d77e387c50e181b0c0e68e784476cc0ca5
22. Us, S. A., Koriashkina, L. S., Stanina, O. D. (2019), "An optimal two-stage allocation of material flows in a transport-logistic system with continuously distributed resource". *Radio Electronics, Computer Science, Control*, 1. DOI: https://doi.org/10.15588/1607-3274-2019-1-24
23. Kiseleva, E., Prytomanova, O., Hart, L. (2020), "Solving a Two-stage Continuous-discrete Problem of Optimal Partitioning-Allocation with Subsets Centers Placement". *Open Computer Science*, 10(1), P. 124–136. DOI: https://doi.org/10.1515/comp-2020-0142

Надійшла (Received) 05.04.2025

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

Відомості про авторів / About the Authors

Кікоть Максим Сергійович – аспірант, Національний аерокосмічний університет "ХАІ", кафедра комп'ютерних наук, Харків, Україна; e-mail: maxym1111111@gmail.com; ORCID ID: https://orcid.org/0009-0005-7554-9246

Малєєва Юлія Анатоліївна – кандидат технічних наук, Національний аерокосмічний університет "ХАІ", доцент кафедри комп'ютерних наук та інформаційних технологій, Харків, Україна; e-mail: juliabelokon84@gmail.com; ORCID ID: https://orcid.org/0000-0003-3553-9156

Kikot Maksym – PhD Student, National Aerospace University "Kharkiv Aviation Institute", Department of Computer Science, Kharkiv, Ukraine.

Malieieva Julia – PhD (Engineering Sciences), National Aerospace University "Kharkiv Aviation Institute", Associate Professor at the Department of Computer Science and Information Technologies, Kharkiv, Ukraine.

MODEL OF LOGISTICS INFRASTRUCTURE FOR THE TWO-STAGE RECYCLING PROCESS OF COMPLEX EQUIPMENT

Subject matter: models for the formation of the logistics infrastructure for complex equipment recycling, taking into account the two-stage processing system and the multicomponent composition of secondary raw materials. **This study aims** to develop a model for the formation of a logistics infrastructure for complex equipment recycling enterprises, the implementation of which will reduce transportation costs of secondary raw materials and infrastructure development expenses. **Objectives**: to investigate the current state and challenges of complex equipment recycling in Ukraine and abroad; to identify the characteristics of logistics processes in complex equipment recycling; to develop a model for the formation of recycling logistics infrastructure for complex equipment. The following **methods** were used during the research: mathematical modeling, location-allocation methods, and systems approach. **Research results**: the current state and challenges of complex equipment recycling have been analyzed, including infrastructure deficiencies, a low level of sorting, and uneven distribution of processing facilities; approaches to solving the facility location problem for recycling infrastructure have been reviewed; a two-level structure of the complex equipment recycling system has been developed, consisting of a network of local collection and sorting centers and processing enterprises; a mathematical model for forming the logistics infrastructure has been developed, considering production capacities, transportation costs and the multicomponent nature of resources. The **conclusions** state that the proposed optimization model for the logistics infrastructure of complex equipment recycling considers a two-stage processing sequence, from collection and sorting to final recycling. The model integrates economic, spatial, and technological factors, enabling the minimization of total transportation and infrastructure development costs. The application of the proposed model ensures effective planning of the recycling network, taking into account processing capacities, logistical routes, and the volume of different types of secondary raw materials.

Keywords: recycling; secondary raw material processing; complex equipment; optimization model; infrastructure; logistics; two-stage system.

Бібліографічні описи / Bibliographic descriptions

Кікоть М. С., Малєєва Ю. А. Модель логістичної інфраструктури двохетапного процесу ресайклінгу складної техніки. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 19–32. DOI: https://doi.org/10.30837/2522-9818.2025.3.019

Kikot, M., Malieieva, Y. "Models of forming logistics infrastructure for complex equipment recycling", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 19–32. DOI: https://doi.org/10.30837/2522-9818.2025.3.019

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.94:005.8:355.45(477) DOI: https://doi.org/10.30837/2522-9818.2025.3.033

Д. Кобилкін, О. Зачко, Р. Мицько, С. Захарчишин, Chetin Elmas

УПРАВЛІННЯ КРИТИЧНИМИ ПАРАМЕТРАМИ ЛОГІСТИЧНИХ ТА ІНФРАСТРУКТУРНИХ ПРОЄКТІВ ЗАСОБАМИ КОМП'ЮТЕРНОГО ЕКСПЕРИМЕНТУ (НА ПРИКЛАДІ СЕКТОРУ БЕЗПЕКИ ТА ОБОРОНИ В УМОВАХ ВІЙНИ)

Предметом дослідження в статті є управління критичними параметрами інфраструктурних та логістичних проєктів на прикладі сектору безпеки та оборони в умовах війни засобами комп'ютерного експерименту з використанням сучасного інструментарію штучного інтелекту, зокрема мультиагентних систем. **Метою дослідження** є проведення комп'ютерного експерименту з дослідження критичних параметрів функціонування інфраструктурних та логістичних проєктів на базі інтелектуальних моделей, що застосовують теорію мультиагентних систем. У статті вирішуються **завдання:** формалізації технічних параметрів транспортної інфраструктури; формалізації предметної області управління логістичними та інфраструктурними проєктами, моделювання критичних параметрів функціонування інфраструктурних та логістичних проєктів в умовах воєнного стану, формування ризиків логістичних та інфраструктурних проєктів на основі аналізу критичних місць. Використовуються такі **методи:** методи дослідження операцій; методи комп'ютерного моделювання, технології штучного інтелекту, зокрема мультиагентні системи; теорії ризиків. **Результати дослідження:** розглянуто методику проведення комп'ютерного експерименту моделювання критичних параметрів функціонування інфраструктурних та логістичних проєктів в умовах воєнного стану. Проведено аналіз реалізації логістичних проєктів в залежності від таких критеріїв як відстань, час проєкту, вантажопідйомність, критичні зони логістичних проєктів та якісні переваги від реалізації. Розроблено матрицю маршрутів логістичних проєктів в умовах воєнного стану. Проведено якісний та кількісний аналіз ризиків логістичних та інфраструктурних проєктів в умовах воєнного стану. **Висновки:** отримані наукові результати доповнюють методологію управління інфраструктурними проєктами та дають змогу реалізувати логістичні проєкти в умовах воєнного стану та погіршення ситуації з продовольчою безпекою, зокрема як альтернативи Чорноморської зернової ініціативи, а також інфраструктурні та логістичні проєкти у секторі безпеки та оборони в умовах війни. Також отримані результати можуть бути застосовані при плануванні логістичних проєктів постачання озброєнь в умовах воєнного стану.

Ключові слова: інфраструктурні проєкти; логістичні проєкти; управління безпекою; сектор безпеки та оборони.

Вступ

Управління проєктами, як наука є в галузі технічних наук та дає змогу розроблення інструментарію, механізмів, моделей, методів т а інформаційних технологій управління складними організаційними соціотехнічними системами [1–3].

Умови воєнного стану кардинально змінили специфіку реалізації логістичних та інфраструктурних проєктів, що в свою чергу вимагає розроблення нових підходів до управління. Зокрема, нові тренди реалізації інфраструктурних проєктів полягають в майбутніх програмах поствоєнного відновлення інфраструктури України, в т.ч. об'єктів критичної інфраструктури, які вимагають нових механізмів бюджетування, проєктного фінансування [4]. Логістичні проєкти є підґрунтям успішної реалізації інфраструктурних проєктів, адже від раціонально спланованої логістики залежить час реалізації проєктів, їх бюджет [5–7]. Як окрема ланка логістичні проєкти зараз мають нові виклики в ситуації з продовольчою безпекою, секторі безпеки та оборони.

Аналіз останніх досліджень і публікацій

Аналізуючи існуючі стандарти PMBOK [1] та P2M [3] узагальнено, що в них систематизовані процеси управління проєктами, зокрема є визначеними рамки та параметри для планування, моніторингу і контролю великих інфраструктурних програм, що в свою чергу формує базову методологічну основу управління складними проєктами. Однак дані стандарти недостатньо адаптовані до умов воєнного стану і не враховують вплив динамічних воєнних ризиків на життєвий цикл проєктів.

У своїх дослідженнях А. Івко та С. Бушуєв [4, 8, 9] розробили синкретичну методологію управління для цифрових і відновлювальних інфраструктурних

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

проєктів, що поєднує як гнучкі так і класичні інструменти управління, що в свою чергу дозволяє підвищити ефективність управління у швидкозмінному середовищі.

P. Jiang, H. Zhang, M. Bertolini, R. Guido, S. Liao, M. Tseng у наукових дослідженнях [5–7] розглянуто процес застосування IT систем локалізації в режимі реального часу, автоматизації і обробки великого обсягу даних, що в умовах мирного часу підвищує рівень ефективності логістичних операцій. Однак дані результати дослідження не адаптовані для використання у безпековому секторі, і не аналізують відповідну специфіку поведінки і функціонування логістичних проєктів в умова війни.

У наукових працях B. Flyvbjerg, N. Bruzelius, W. Rothengatter [10, 11] обґрунтувано підходи до оцінки ризиків у мегапроєктах, зокрема в контексті системних відхиленнях у термінах і вартості, що стало основою для розвитку карт ризиків та їх інтеграції в процес управління. Однак недоліком є те, що недостатня увага приділена просторово-часовій динаміці ризиків, а ризик розглядається переважно статично, тоді як у воєнних умовах критичні параметри інфраструктурних проєктів можуть швидко змінюватися.

Науковці K. Hayat, M. Hafeez, K. Bilal, M. Shabbir у праці [12] особливо виокремлюють ролі організаційних структур та роботи команди проєкту у забезпеченні успіху проєктів, що в свою чергу є важливим для управління складними логістичними процесами та проектами.

У роботах О. Зачка [13–15] досліджувалися процеси інтелектуального моделювання інфраструктурних проєктів, процесам управління безпекою та портфельного управління. Це сформувало базу для розробки інструментарію аналізу параметрів інфраструктурних проєктів у системі цивільному захисті, однак без інтеграції зі сценаріями воєнних дій.

Як підсумок, на сьогодні науковцями проведено ґрунтовні дослідження та напрацьований значний доробок у галузі управління інфраструктурними та логістичними проєктами. Однак дані наукові і практичні результати не адаптовані до умов сектора безпеки та оборони в умовах війни, що і становить актуальність проведення дослідження процесу управління критичними параметрами логістичних та інфраструктурних проєктів засобами комп'ютерного експерименту за даних умов та параметрів.

Аналіз проблеми й наявних методів

Наукова проблема, яка є невирішеною досі – це проведення натурного експерименту, оскільки відгук складної соціотехнічної системи на зміну параметрів управління може сягати багатьох років, що робить таке дослідження економічно не рентабельним та не реалістичним. Особливо це стосується логістичних та інфраструктурних проєктів, які характеризуються великим та складним життєвим циклом, масштабним бюджетом, великими ризиками та складністю управління [8, 9].

Під час проведення досліджень використовувалися методи дослідження операцій – для формалізації технічних параметрів транспортної інфраструктури; методи комп'ютерного моделювання – для формалізації предметної області управління логістичними та інфраструктурними проєктами, а також технології штучного інтелекту, зокрема мультиагентних систем – для моделювання критичних параметрів функціонування інфраструктурних та логістичних проєктів в умовах воєнного стану; теорії ризиків – формування ризиків та проведення їх кількісного та якісного аналізу в логістичних на інфраструктурних проєктів [10–12]. Суть комп'ютерного експерименту полягає у описі властивостей продуктів інфраструктурних та логістичних проєктів імітаційною моделлю [13–15].

Метою статті є розроблення підходів, методик, алгоритмів та інформаційних технологій з проведення комп'ютерного експерименту моделювання критичних параметрів функціонування логістичних та інфраструктурних проєктів в умовах воєнного стану.

Постановка наукового завдання

Для досягнення цієї мети ставляться такі завдання:

- вивчення сучасних технологій та методів, що використовуються для проведення комп'ютерного експерименту в складних соціотехнічних організаційних системах, в т.ч. безпекових, де відгук системи на зміну параметрів управління сягає років та не має можливості провести натурний експеримент;
- сформулювати гіпотезу та план комп'ютерного експерименту моделювання критичних параметрів функціонування логістичних та інфраструктурних проєктів в умовах воєнного стану;
- на основі результатів комп'ютерного експерименту розробити практичні рекомендації

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

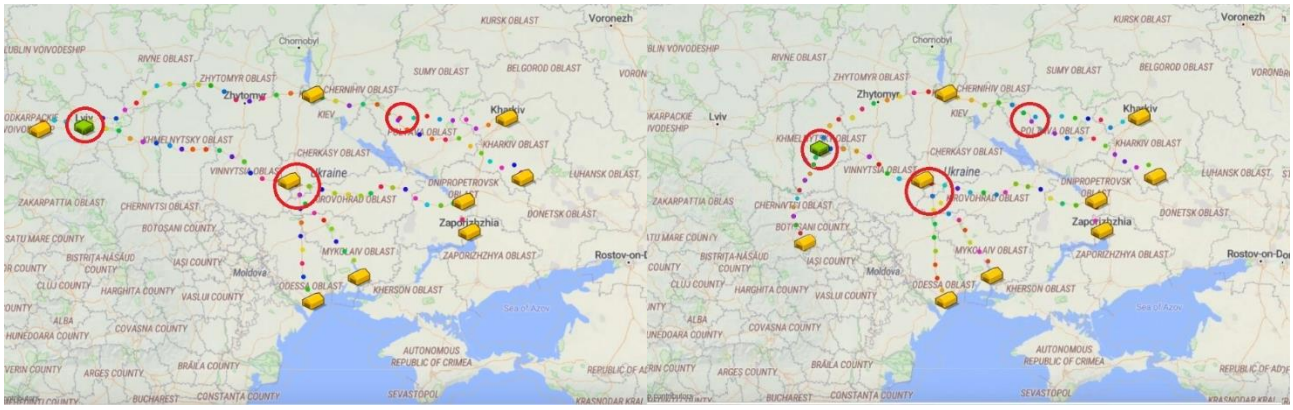
управління критичними параметрами інфраструктурних та логістичних проєктів на прикладі сектору безпеки та оборони в умовах війни;

- визначення перспектив розвитку та майбутніх тенденцій у галузі управління логістичними та інфраструктурними проєктами в умовах воєнного стану згідно отриманих наукових результатів.

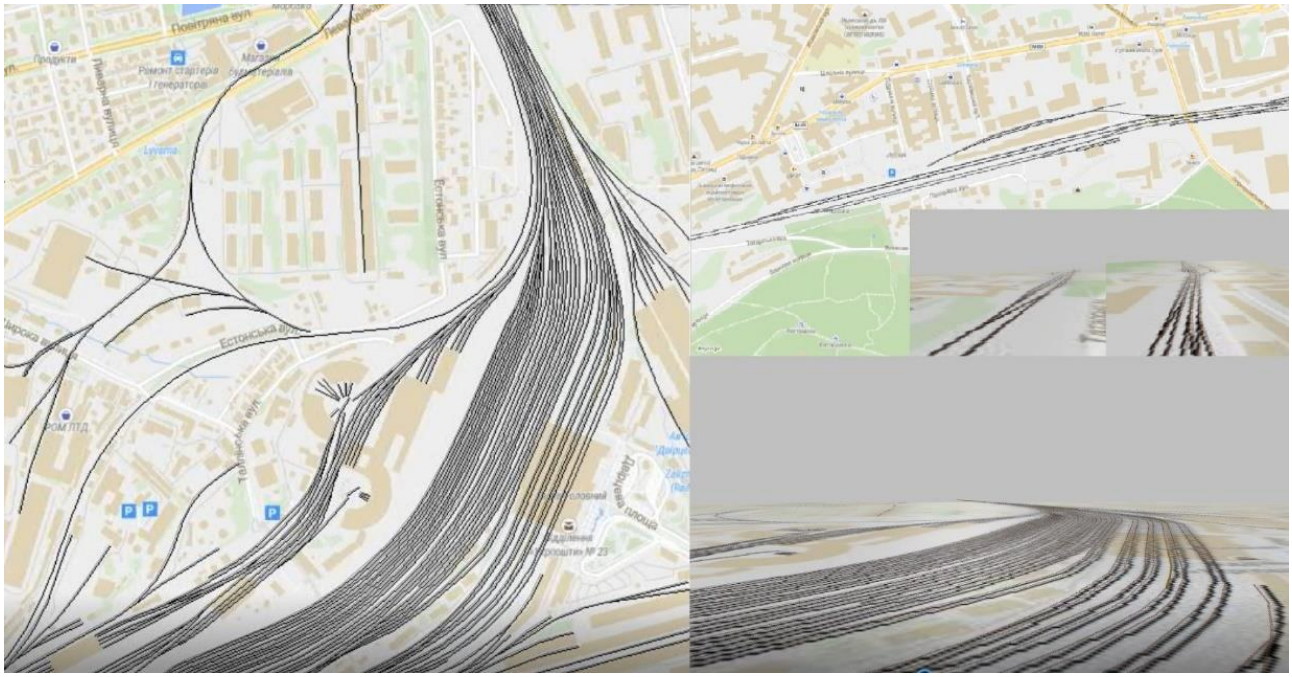
Результати дослідження та їхнє обговорення

Гіпотеза наукового дослідження, що полягає в проведенні комп'ютерного експерименту дослідження критичних параметрів управління логістичними та інфраструктурними проєктами в умовах воєнного стану полягає в розробленні оптимальних матриць логістичних проєктів, ключовими параметрами яких є вартісні показники, відстань та безпекова компонента. Що стосується умов воєнного стану, то в пріоритеті залишається саме безпекова компонента та альтернативність ланцюгів постачань. Можливі галузі прикладного застосування розроблених комп'ютерних моделей, які були ідеєю комп'ютерного експерименту – це альтернативність реалізації Чорноморської зернової ініціативи при песимістичних сценаріях, а також логістичні проєкти в секторі безпеки та оборони в умовах воєнного стану.

На рис. 1 представлені розроблені моделі логістичних проєктів засобами автомобільної та залізничної критичної інфраструктури в середовищі інтелектуального мультиагентного моделювання в рамках управління критичними параметрами інфраструктурних та логістичних проєктів на прикладі сектору безпеки та оборони в умовах війни засобами комп'ютерного експерименту.



а *б*



в

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

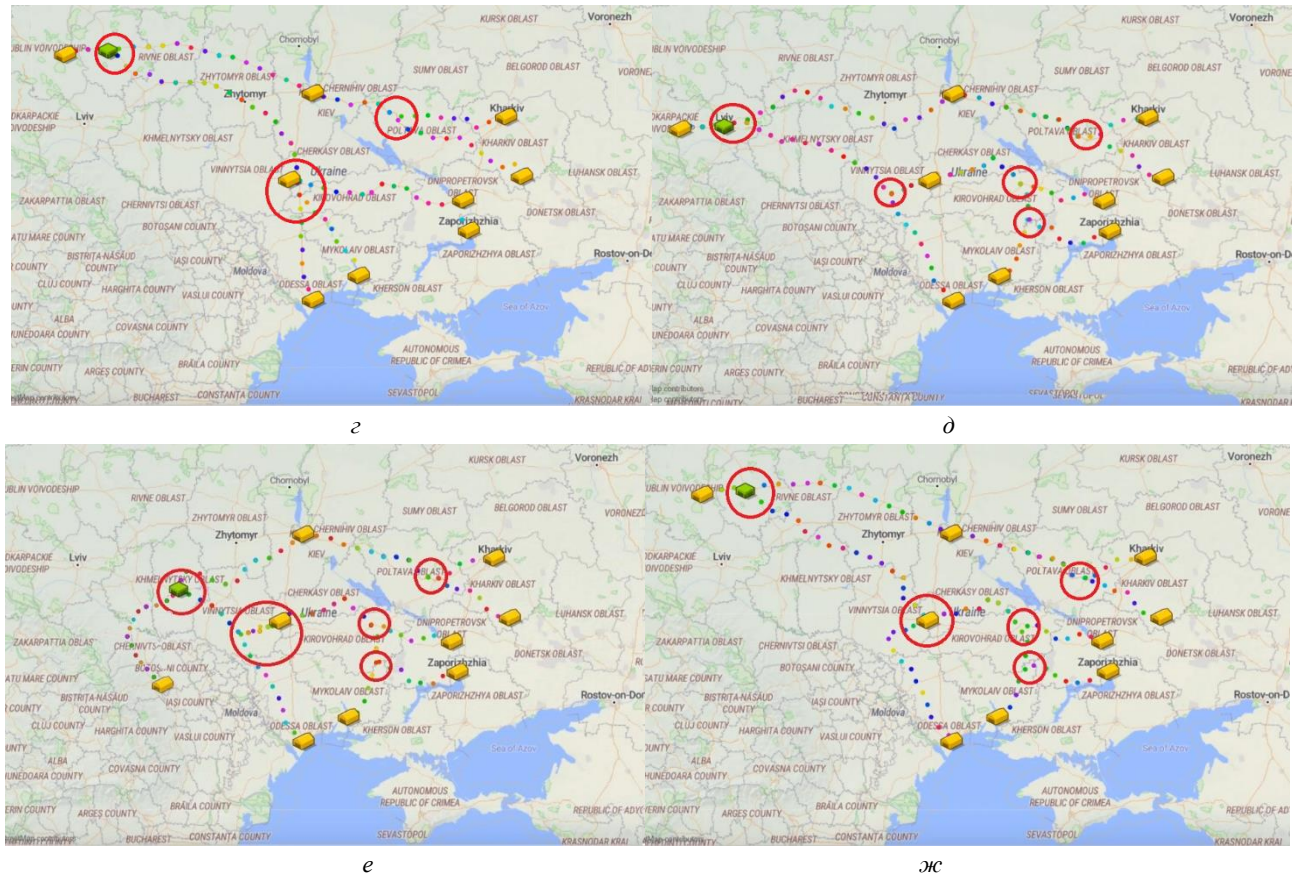


Рис. 1. Комп'ютерний експеримент управління логістичними проєктами засобами автомобільної та залізничної транспортної критичної інфраструктури в середовищі інтелектуального мультиагентного моделювання:
а – логістичний проєкт Перемишль-Львів з використанням автомобільної транспортної інфраструктури;
б – логістичний проєкт Сучава-Хмельницький з використанням автомобільної транспортної інфраструктури;
в – 3D-модель продукту реалізації логістичного проєкту Львівською залізницею;
г – логістичний проєкт Холм-Ковель з використанням автомобільної транспортної інфраструктури;
д – логістичний проєкт Перемишль-Львів з використанням залізничної транспортної інфраструктури;
е - логістичний проєкт Сучава-Хмельницький з використанням залізничної транспортної інфраструктури;
ж – логістичний проєкт Холм-Ковель з використанням залізничної транспортної інфраструктури

Вхідні параметри для комп'ютерного експерименту бралися з міжнародних та державних стандартів, що стосуються логістики [16, 17]. Оскільки прикладне застосування розроблених комп'ютерних моделей стосуватиметься як гуманітарної сфери, так і можливо у секторі безпеки та оборони, то пріоритетом була альтернативність потенційних логістичних ланцюгів на прикладі логістичних проєктів наведених на рис. 1. Технічні параметри логістичних проєктів подано в табл. 1.

Комп'ютерна модель кожного логістичного проєкту описується кортежом:

$$L = \langle S, T_x, C \rangle, \quad (1)$$

де S – відстань оцінена з урахуванням реальних масштабів і топології доріг; T_x – час у дорозі, що враховує тип перевезення, середню швидкість та можливі затримки; D – розрахунковий час перетину кордону, що враховує тип перевезення, середню швидкість та можливі затримки; C – вантажопідйомність (авто: фура ≈ 20 т; залізниця: 1 вантажний вагон ≈ 75 т (залежить від типу), 1 вантажний поїзд в середньому 50-70 вагонів.

На рис. 1 а, б, г, д, е позначені критичні точки червоними колами на картах. Це потенційні загрози безпеці (воєнні, логістичні вузли, вразливі мости та ін.) як для логістичних проєктів по Україні так і потенційних логістичних проєктів перетину кордонів. Аналізуючи якісні переваги, то автомобільна транспортна інфраструктура дає можливість оперативності та можливість доставити вантаж до конкретного об'єкта без перевантаження. Залізнична

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

транспортна інфраструктура дає вищу ефективність для великих партій, але обмежена маршрутом і станом інфраструктури. Виходячи з аналізу табл. 1 Логістичний проєкт *Перемишль–Львів* має найкращий маршрут за швидкістю та гнучкістю при використанні автомобільного транспорту (коротка відстань, швидкий час у дорозі). Залізниця підходить для великих вантажів, але має суттєві затримки на кордоні. Логістичний проєкт *Холм–Ковель* – оптимальний для автомобільних перевезень середньої відстані з хорошою гнучкістю, але з ризиками через стан інфраструктури і затримки на кордоні. Залізниця забезпечує велику вантажопідйомність, але з більшими затримками. Проєкт *Сучава–Хмельницький* – маршрут з високими ризиками затримок через відстань, стан доріг та складну логістику на залізниці. Перевага – велика вантажопідйомність залізничного транспорту, але час доставки може бути дуже довгим.

Відповідно до отриманих вихідних даних комп'ютерного експерименту сформулюємо критерії ефективності логістичних проєктів в умовах війни (табл. 2).

Таблиця 1. *Вхідні дані для комп'ютерного експерименту (дані з відкритих електронних ресурсів)*

Напрямок	Тип транспорту	Орієнтовна відстань (км)	Час у дорозі (без черг/ із технічними затримками)	Вантажо-підйомність транспортного засобу	Якісні переваги / ризики
Перемишль – Львів	Авто	~97 км	~2 год / ~2 – 6 год	~20 т (фура)	Гнучкість, затори на кордоні
	Залізниця	~95 км	~3–4 год/ ~6 – 30+ год	~75 т (вагон)	Велика пропускна здатність, складна логістика на кордоні (перестановка візків)
Холм – Ковель	Авто	~115 км	~2–2,5 год/ ~3 – 8 год	~20 т (фура)	Добра гнучкість маршруту; затримки на переході, ризики з поганою інфраструктурою
	Залізниця	~90 км	~3–6 год/ ~6 –36+ год	~75 т (вагон)	Більші обсяги перевезень, критичні вузли з низькою пропускною здатністю
Сучава – Хмельницький	Авто	~270 км	~4–5,5 год/ ~5 – 20 год	~20 т (фура)	Гнучкість маршруту, можливість обрання альтернативних КПП, довга відстань, стан доріг
	Залізниця	~290 км	~8–11,5 год/ ~24 – 72+ год	~75 т (вагон)	Висока вантажопідйомність, складний маршрут із кількома вузлами, складна логістика на кордоні (перестановка візків)

Таблиця 2. *Критерії ефективності логістичних проєктів в умовах війни*

Критерій	Автомобіль (фура)	Залізничний вантажний потяг
Середня довжина маршруту	~ 161 км	~159 км
Швидкість транспорту	Вища (швидкий рух без черг 2–5,5 год)	Нижча (рух без черг 3–11,5 год)
Середній час доставки	2–20 год з урахуванням кордону	6–72+ год з урахуванням кордону та перестановок
Вантажопідйомність одного рейсу	~20 тонн	~75 тонн (1 поїзд ≈ 50–70 вагонів)
Пропускна здатність маршруту	Обмежена кількістю фур, пунктами пропуску	Висока, але залежна від інфраструктури залізниці
Гнучкість маршруту	Висока – можна змінювати маршрут	Низька – залежність від колії, логістики станцій
Вартість транспортування (за тонну)	Вища (~1.5–2× ніж залізнична)	Нижча на великі обсяги, економія масштабу
Критичні точки	Митниці, черги, вузькі дороги та їх стан, перевантаження	Залізничні вузли, стиковки колій, зміна візків, нестача локомотивів
Якісні переваги	Гнучкість, швидкість на коротких дистанціях, можливість термінових перевезень	Обсяг, економічність, безпека
Основні ризики	ДТП, затримки, погодні умови, митний контроль, обмежена вантажопідйомність	Стиковки колій, дефіцит рухомого складу, обмежена маневровість, великі простої

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

В табл. 3 наведені результати порівняльного багатокритеріального аналізу ефективності реалізації логістичних проєктів засобами залізничної та транспортної інфраструктури в умовах війни за критеріями довжини логістичного маршруту, часу проєкту, критичних точок та ризиків.

Таблиця 3. *Порівняльний багатокритеріальний аналіз ефективності логістичних проєктів залізничною та автомобільною транспортною інфраструктурою*

Логістичний проєкт *Перемишль → Львів*		
Критерій	**Авто**	**Залізниця**
Довжина	~97 км	~95 км
Час	~2 год без черг; 2–6 год з урахуванням кордону	~3–4 год без черг; 6–30+ год з кордоном та перестановками
Критичні точки	2 (кордон, можливі затори)	3 (кордон, перестановка вагонів)
Ризики	Затори на кордоні, обмежена вантажопідйомність	Тривалі затримки на кордоні, складна логістика
Логістичний проєкт *Холм → Ковель*		
Критерій	**Авто**	**Залізниця**
Довжина	~115 км	~90 км
Час	~2–2,5 год без черг; 3–8 год з урахуванням кордону	~3–6 год без черг; 6–36+ год з кордоном і перестановками
Критичні точки	3 (кордон, інфраструктура, затримки)	4 (кордон, низька пропускна здатність вузлів)
Ризики	Затримки на переході, погана інфраструктура	Великі затримки на критичних вузлах, складність логістики
Логістичний проєкт *Сучава → Хмельницький*		
Критерій	**Авто**	**Залізниця**
Довжина	~270 км	~290 км
Час	~4–5,5 год без черг; 5–20 год з урахуванням кордону	~8–11,5 год без черг; 24–72+ год з кордоном та перестановками
Критичні точки	4 (довга відстань, кордон, стан доріг, КПП)	5 (довгий маршрут, кордон, перестановки, вузькі місця)
Ризики	Довга відстань, поганий стан доріг, затримки	Дуже тривалі затримки, складна логістика, великі простої

Використовуючи сценарний підхід змоделюємо можливі сценарії реалізації логістичних проєктів в умовах війни. Відповідно до безпекової ситуації та в залежності від того чи це гуманітарний логістичний проєкт чи в секторі безпеки та оборони, потенційних ракетних та дронових атак, руйнувань залізничної та автомобільної транспортної інфраструктури (табл. 4).

Умови війни принципово змінили підходи до управління логістичними та інфраструктурними проєктами. Так, класичними критеріями успішності логістичних проєктів були вартість та час доставки (рис. 2), в той час як зараз це безпекова ситуація.

За результатами комп'ютерного експерименту розроблено матрицю логістичних проєктів по Україні (табл. 5), враховуючи найбільші вузлові пункти транспортної інфраструктури України та на основі концептуальної параметричної моделі:

$$x = \begin{cases} \langle S, T, W, R \rangle \to \min \\ \langle V, C, P \rangle \to \max \end{cases}, \qquad (2)$$

де x – унікальний логістичний проєкт; S – відстань (км); T_x – час (год); V – швидкість (км/год); C – вантажопідйомність (т); W – вартість (грн/т); R – ризики (↑ високі, → помірні, ↓ низькі); P – переваги.

Таблиця 4. *Потенційні сценарії реалізації логістичних проєктів в умовах війни*

Сценарій	Рекомендований тип транспорту
Швидка адресна доставка	Автомобіль
Великі обсяги, менша вартість	Залізниця
Уникнення заторів/черг	Залізниця
Складні погодні умови (зима)	Залізниця
Доставка з можливістю маневрування	Автомобіль

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

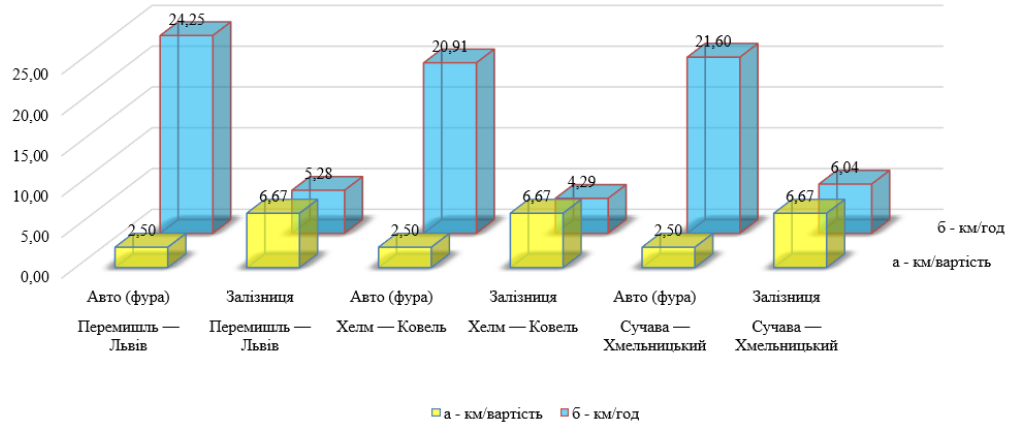


Рис. 2. Класичні критерії ефективності логістичних проєктів у довоєнний період:
а – вартість доставки за км; б – швидкість доставки за км

Таблиця 5. *Матриця логістичних проєктів по Україні*

Матриця логістичних проєктів по Україні							
Напрямок / Транспорт	**S**	**T**	**V**	**C**	**W**	**R**	**P**
Холм – Ковель → цілі по Україні (автомобільна та залізнична інфраструктура)							
Харків / Авто	1025	17	60	20	410	↑↑	Гнучкість доставки, прямі маршрути
Харків / Залізниця	1002	22	45	75	150	↑	Велика пропускна здатність, дешевше на великі відстані
Лозова / Авто	1062	18	60	20	425	↑	Door-to-door, об'їзди
Лозова / Залізниця	1081	24	45	75	162	→	Великі обсяги, нижча ціна/т
Дніпро / Авто	1107	18	60	20	443	↑	Гнучкі маршрути
Дніпро / Залізниця	1133	25	45	75	170	→	Дешевше на довгі відстані
Запоріжжя / Авто	1170	20	60	20	468	↑↑	Гнучкість, прямі поставки
Запоріжжя / Залізниця	1195	27	45	75	179	↑	Великі обсяги
Миколаїв / Авто	968	16	60	20	387	→	Прямі поставки
Миколаїв / Залізниця	988	22	45	75	148	→	Великі обсяги
Одеса / Авто	959	16	60	20	384	→	Гнучкі маршрути
Одеса / Залізниця	983	22	45	75	147	↓	Дешевше на великі обсяги
Сучава –Хмельницький → цілі по Україні (автомобільна та залізнична інфраструктура)							
Харків / Авто	1061	18	60	20	424	↑	Гнучкість, пряма доставка
Харків / Залізниця	1156	26	45	75	173	↑	Масове перевезення, дешевше
Лозова / Авто	1093	18	60	20	437	↑	Швидкість, керованість
Лозова / Залізниця	1235	27	45	75	185	→	Централізоване постачання
Дніпро / Авто	955	16	60	20	382	→	Хороша динаміка
Дніпро / Залізниця	1037	23	45	75	156	→	Регулярне масове перевезення
Запоріжжя / Авто	1040	17	60	20	416	↑	Можливість обходу
Запоріжжя / Залізн.	1100	24	45	75	165	→	Менш затратне перевезення
Миколаїв / Авто	815	14	60	20	326	↓	Гнучкість маршруту
Миколаїв / Залізн.	887	20	45	75	133	↓	Централізована доставка
Одеса / Авто	807	13	60	20	323	↓	Прямий доступ до порту
Одеса / Залізниця	881	20	45	75	132	↓	Висока ефективність, регулярність
Перемишль–Львів → Україна (автомобільна та залізнична інфраструктура)							
Харків / Авто	1127	19	60	20	451	↑↑	Гнучкий маршрут, пряме сполучення
Харків / Залізниця	1193	27	45	75	179	↑	Великі обсяги, менші витрати на великі партії
Лозова / Авто	1159	19	60	20	464	↑	Обхід критичних зон, оперативність
Лозова / Залізниця	1271	28	45	75	191	→	Масові вантажі, нижчі витрати
Дніпро / Авто	1048	17	60	20	419	→	Оптимальна відстань, надійна інфраструктура
Дніпро / Залізниця	1040	23	45	75	156	→	Стабільна логістика, дешевше на обсягах
Запоріжжя / Авто	1111	19	60	20	444	↑	Пряма доставка
Запоріжжя / Залізн.	1101	24	45	75	165	→	Обсяги, менше блокпостів
Миколаїв / Авто	908	15	60	20	363	→	Порт, хороша логістика до Одеси
Миколаїв / Залізн.	895	20	45	75	134	↓	Масові вантажі, зменшення навантаження на дороги
Одеса / Авто	899	15	60	20	360	↓	Портовий вузол, швидке реагування
Одеса / Залізниця	890	20	45	75	134	↓	Залізнична логістика з виходом на порти

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

На рис. 3 представлений порівняльний аналіз логістичних проєктів за 4 параметрами ефективності. В умовах воєнного стану на відміну від звичних підходів, на перше місце виходить безпекова компонента. Тому крім параметрів відстані, вартості та часу проєкту (рис. 3 а, б, в), що в умовах війни при реалізації проєктів в секторі безпеки та оборони, які не є критичними, необхідно пріоритезувати 4-й параметр – рівень ризиків (рис. 3г).

Нами розроблено матриці відстаней в логістичних проєктах від прикордонних вузлових станцій до ключових цілей в Україні (табл. 6).

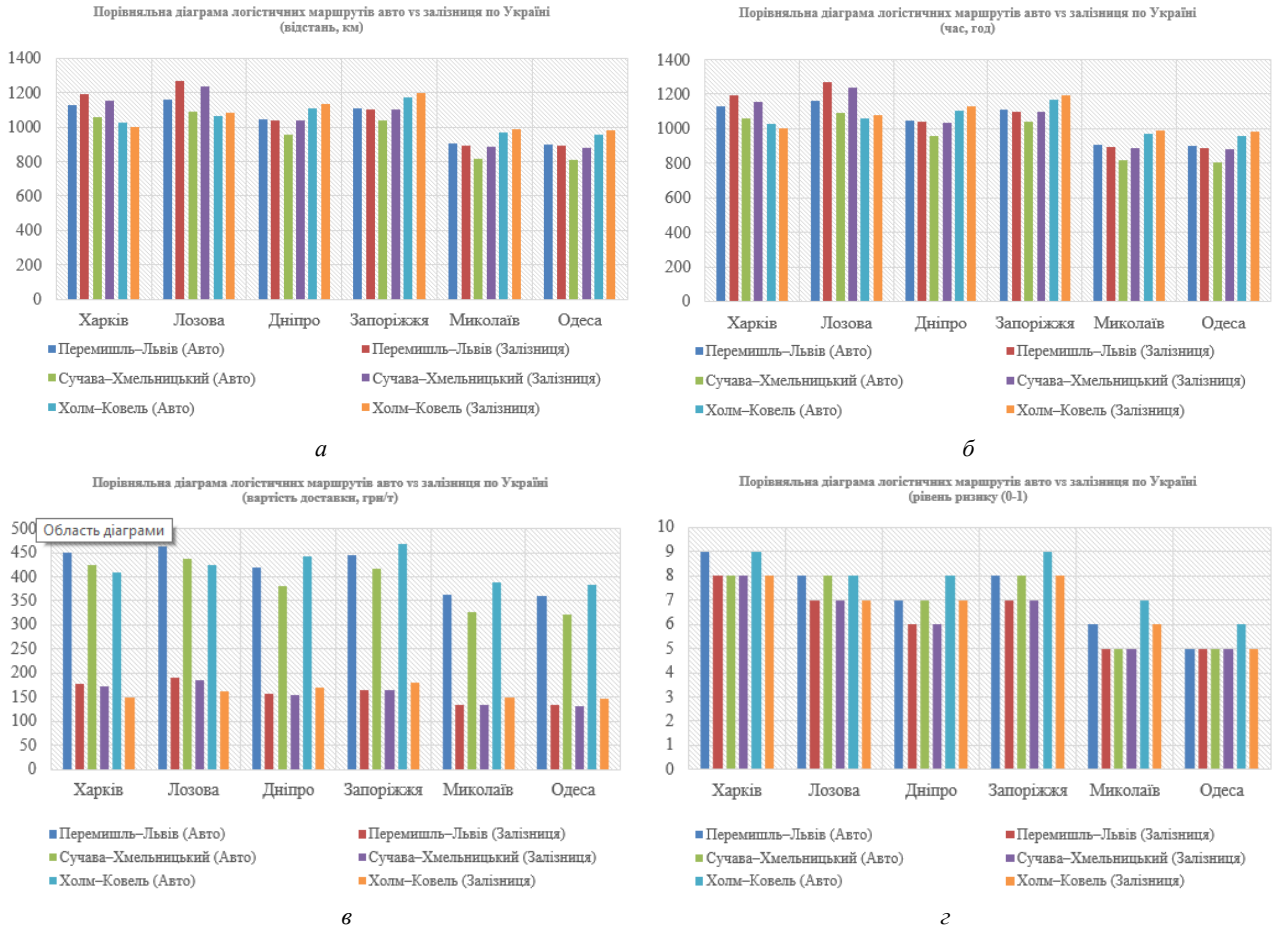


Рис. 3. Порівняльна діаграма логістичних проєктів за параметрами:
а – відстань; б – час проєкту; в – вартість проєкту; г – рівень ризиків

Таблиця 6. *Матриця відстаней в логістичних проєктах від прикордонних вузлових станцій до ключових цілей в Україні*

Напрямок, відстань, км	Харків	Лозова	Дніпро	Запоріжжя	Миколаїв	Одеса
Перемишль–Львів (Авто)	1127	1159	1048	1111	908	899
Перемишль–Львів (Залізниця)	1193	1271	1040	1101	895	890
Сучава–Хмельницький (Авто)	1061	1093	955	1040	815	807
Сучава–Хмельницький (Залізниця)	1156	1235	1037	1100	887	881
Холм–Ковель (Авто)	1025	1062	1107	1170	968	959
Хелм–Ковель (Залізниця)	1002	1081	1133	1195	988	983

На основі даних табл. 6 найкоротший маршрут до основних цілей в Україні є логістичний проєкт *Сучава–Хмельницький* автомобільною транспортною інфраструктурою. Зазвичай залізничні логістичні проєкти загалом довші на ~20–80 км через інфраструктурні особливості. Логістичний проєкт *Холм–Ковель* є найдорожчим з розрахунку відстані та витрат на транспортування до цілей в Україні (незалежно від транспорту).

Тому на основі кумуляції показників часу, вартості, ризику та швидкості (табл. 7) нами розроблено кумулятивний показник успішності логістичного проєкту в умовах воєнного стану, для визначення якого було здійснено 4 етапи.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Таблиця 7. *Кумулятивний показник успішності логістичного проєкту*

№	Маршрут	μS	μT_x	μW	μR	K_{sw}
1	Перемишль – Львів (Залізниця)	1065,00	41,67	159,75	6,33	0,717
2	Сучава – Хмельницький (Авто)	961,83	29,03	384,67	6,83	0,663
3	Сучава – Хмельницький (Залізниця)	1049,33	71,32	157,33	6,33	0,630
4	Холм – Ковель (Залізниця)	1063,67	44,64	159,33	6,83	0,575
5	Перемишль – Львів (Авто)	1042,00	21,37	416,80	7,17	0,423
6	Холм – Ковель (Авто)	1048,50	22,48	419,50	7,83	0,288

Тут μS – середнє значення відстаней логістичного проєкту по Україні (км); μT_x – середнє значення часу логістичного проєкту по Україні із врахуванням коефіцієнту затримок (год); μW – середнє значення вартості транспортування продукту логістичного проєкту по Україні (грн/т); μR – середнє значення ризику логістичного проєкту по Україні [0;10]; K_{sw} – кумулятивний показник успішності логістичного проєкту в умовах воєнного стану.

На першому етапі проведено нормалізацію $Y_{norm,j}$ мінімізувальних критеріїв $(\mu S, \mu Tx, \mu W, \mu R)$ за формулою 3.

$$Y_{norm,j} = \frac{\max(Y) - Y_j}{\max(Y) - \min(Y)}. \quad (3)$$

На другому та третьому етапах розраховані показники зваженої суми Ws_j, формула 4 та первинний індекс Ksw_{nj}, формула 5.

$$Ws_j = 1 \cdot \mu V_{norm,j} + 1 \cdot \mu Tx_{norm,j} + 1 \cdot \mu W_{norm,j} + 2 \cdot \mu R_{norm,j} \quad (4)$$

$$Ksw_{nj} = Ws_j / 5 \quad (5)$$

На четвертому етапі за формулою 6 розраховані значення кумулятивного показника успішності логістичних проєктів в умовах воєнного стану, що наведені у табл.7 та відображені на рис. 4.

$$Ksw = \frac{Ksw_{nj} - \min(Ksw_n)}{\max(Ksw_n) - \min(Ksw_n)} \quad (6)$$

Згідно даних табл. 7 та рис. 4 найефективнішим є логістичний проєкт Перемишль–Львів засобами залізничної транспортної інфраструктури через оптимальне співвідношення вартості, часу і ризиків. Найбільш збалансованим за критеріями часу, помірної вартості і середнього рівня ризиків є Сучава–Хмельницький автомобільною транспортною інфраструктурою. Найменш ефективним є Холм–Ковель автомобільною транспортною інфраструктурою через найвищі в умовах воєнного стану ризики і високу вартість при середніх показниках відстані.

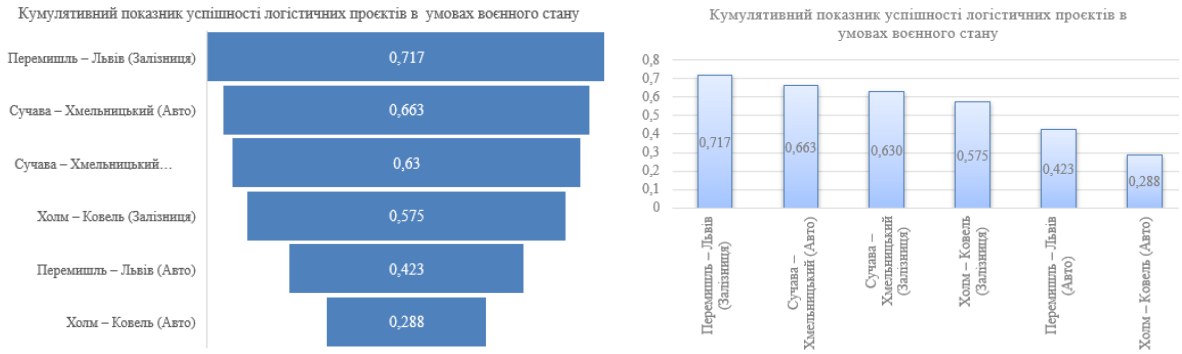


Рис. 4. Кумулятивний показник успішності логістичних проєктів в умовах воєнного стану

Висновки

В науковій статті проведено комп'ютерний експеримент з управління критичними параметрами функціонування інфраструктурних та логістичних проєктів в умовах воєнного стану з застосуванням технологій штучного інтелекту, зокрема теорії мультиагентних систем.

Отримані такі нові результати:

1. Здійснено формалізацію технічних параметрів транспортної інфраструктури та предметної області управління логістичними та інфраструктурними проєктами.

2. Створено сценарний підхід реалізації логістичних проєктів в умовах воєнного стану, що враховують параметри відстані логістичних проєктів, часу реалізації та якісних переваг.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

3. Проведено комп'ютерний експеримент з реалізації логістичних проєктів автомобільною та залізничної інфраструктурою з детальним аналізом кількісних та якісних параметрів ризик-менеджменту в умовах воєнного стану.

4. На основі комп'ютерного експерименту розроблено матриці логістичних проєктів, що враховують вартісні параметри, оцінку ризиків, технічні параметри логістичних проєктів (вантажопідйомність, швидкість, відстань).

5. Проведено моделювання критичних параметрів функціонування інфраструктурних та логістичних проєктів в умовах воєнного стану та сформовано ризики логістичних та інфраструктурних проєктів на основі аналізу критичних місць.

Напрямки подальших досліджень

Отримані наукові результати доповнюють методологію управління інфраструктурними проєктами та є перспективним напрямом наукових досліджень в секторі безпеки та оборони, зокрема:

1. Перспективним напрямом управління логістичними проєктами в умовах воєнного стану є розроблення інтелектуальних моделей, що дають змогу реалізувати проєкти у випадку песимістичного сценарію погіршення ситуації з продовольчою безпекою, зокрема як альтернативи Чорноморської зернової ініціативи засобами автомобільної та залізничної зернової ініціативи

2. Отримані результати можуть бути застосовані при плануванні логістичних проєктів постачань озброєнь Україні від союзників в умовах воєнного стану, зокрема по програмі Lend-Lease, де основними проблемами є планування маршрутів та визначення найбезпечніших та найефективніших шляхів доставки вантажів, враховуючи географічні, політичні та безпекові фактори, а також вибір транспортної інфраструктури (морська, залізнична, автомобільна та повітряна інфраструктура).

Список літератури

1. The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide) Seventh Edition / USA. Project Management Institute (PMI), 2021. 250 p.
2. The Standard for Portfolio Management. Fourth Edition / USA. Project Management Institute (PMI), 2017. 127 p.
3. P2M Bibelot (Overview of P2M Third Edition) / Japan. – Project Management Association of Japan (PMAJ), 2017. 20 p. URL: https://www.pmaj.or.jp/ENG/p2m/p2m_guide/P2M_Bibelot(All)_R3.pdf (дата звернення: 14.09.2024).
4. Bushuyev S. D., Ivko A. V. Aspects analysis of syncretic project management methodology implementation in self-managed organizations project activities. *Bulletin of Lviv State University of Life Safety*. Lviv, LSULS, Volume 29. 2024. P. 168–178. DOI: https://doi.org/https://doi.org/10.32447/20784643.29.2024.17
5. Jiang P., Zhang H. Real-time localization systems in logistics: An overview. *Logistics Research*. 2021. Volume 14(2). P. 109–124. DOI: https://doi.org/10.1007/s12159-021-00245-0
6. Bertolini M., Guido R. Warehouse and Logistics Automation: Achievements and Trends. *Automation in Industry Journal*. 2021. Volume 12(1). P. 40-56. DOI: https://doi.org/10.1016/j.aiij.2020.12.005
7. Liao S. H., Tseng M. L. Logistics management using big data. *Journal of Smart Logistics*. 2018. Volume 5(2). P. 88–102. DOI: https://doi.org/10.1016/j.jslog.2017.11.007
8. Ivko A. V. Implementation aspects analysis of syncretic methodology in the management of infrastructure restoration projects *Automobile roads and road construction*. 2024. Vol. 115, Issue 2. P. 276–289. DOI:10.32447/20784643.29.2024.17
9. Bushuyev S. D., Ivko A. V. Values spiral development method in the implementation of digitalization projects in syncretic methodology. *International Journal of Computing*, № 23(2), 2024. P. 177–186. DOI: https://doi.org/10.47839/ijc.23.4.3535
10. Flyvbjerg B., Bruzelius N., Rothengatter W. Frontmatter. In: Megaprojects and Risk: An Anatomy of Ambition. *Cambridge University Press*. 2003.
11. Flyvbjerg B. What You Should Know about Megaprojects and Why: An Overview. *Project Management Journal*. 2014. Volume 45. № 2. April-May. P. 6–19. DOI: 10.1002/pmj.21409
12. Hayat K., Hafeez M., Bilal K., Shabbir M.S. Interactive Effects of Organizational Structure and Team Work Quality on Project Success in Project Based Non Profit Organizations. *IRASD Journal of Management*. 2022. Volume 4(1), P. 84–103. DOI: https://doi.org/10.52131/jom.2022.0401.0064
13. Зачко О. Б. Інтелектуальне моделювання параметрів продукту інфраструктурного проєкту (на прикладі аеропорту "Львів"). *Східно-Європейський журнал передових технологій*. 2013. Т. 1. №-10. С. 92–94
14. Зачко О. Б. Управління безпекою складних інфраструктурних проектів в системі цивільного захисту. *Управління проектами: стан та перспективи*: матер. 10 Міжнар. наук.-практ. конф. Миколаїв: НУК. 2014. С. 91–92.
15. Зачко О. Б. Моделі, механізми та інформаційні технології портфельного управління розвитком складних регіональних систем безпеки життєдіяльності. Під заг. ред. Рака Ю.П. Монографія. Львів: Видавництво ЛДУ БЖД, 2015. 177 с.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

16. ДСТУ EN 13775-6:2022. Залізничний транспорт. Вимірювання нових і модернізованих вантажних вагонів. Частина 6. Вантажні вагони з кількох взаємно жорстко з'єднаних одиниць та зчленовані вантажні вагони.
17. ДСТУ 2609-94 "Вантажні автомобільні перевезення. Терміни та визначення".

References

1. The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition. USA. Project Management Institute (PMI). 2021. 250 p.
2. The Standard for Portfolio Management. Fourth Edition. USA. Project Management Institute (PMI), 2017. 127 p.
3. P2M Bibelot (Overview of P2M Third Edition) [Electronic resource] / Japan. Project Management Association of Japan (PMAJ), 2017. 20 p. available at: https://www.pmaj.or.jp/ENG/p2m/p2m_guide/P2M_Bibelot(All)_R3.pdf (last accessed: 2024-09-14).
4. Bushuyev, S.D., Ivko, A. V. (2024), "Aspects analysis of syncretic project management methodology implementation in self-managed organizations project activities", *Bulletin of the Lviv State University of Life Safety*. Lviv, LSULS. Vol. 29. P. 168–178. DOI: https://doi.org/10.32447/20784643.29.2024.17
5. Jiang, P., Zhang, H. (2021), "Real-time localization systems in logistics: A review", *Logistics Research*, 14(2), P. 109–124. DOI: https://doi.org/10.1007/s12159-021-00245-0
6. Bertolini, M., Guido, R. (2021), "Automation of warehouses and logistics: Achievements and trends", *Automation in Industry Journal*, 12(1), P. 40–56. DOI: https://doi.org/10.1016/j.aiij.2020.12.005
7. Liao, S. H., Tseng, M. L. (2018), "Logistics management using big data", *Journal of Smart Logistics*, 5(2), P. 88–102. DOI: https://doi.org/10.1016/j.jslog.2017.11.007
8. Ivko, A. V. (2024), "Implementation aspects analysis of syncretic methodology in the management of infrastructure restoration projects", *Motorways and Road Construction*, Issue 115. Part 2. P. 276–289. DOI:10.32447/20784643.29.2024.17
9. Bushuyev, S.D., Ivko, A. V. (2024), "Values spiral development method in the implementation of digitalization projects in syncretic methodology", *International Journal of Computing*, No. 23(2), 2024. P. 177–186. DOI: https://doi.org/10.47839/ijc.23.4.3535
10. Flyvbjerg, B, Bruzelius, N, Rothengatter, W. Frontmatter. In: Megaprojects and Risk: An Anatomy of Ambition. Cambridge University Press; 2003.
11. Bent, Flyvbjerg, (2014), "What You Should Know about Megaprojects and Why: An Overview," *Project Management Journal*, Vol. 45, No. 2, April-May. P. 6–19. DOI: 10.1002/pmj.21409
12. Hayat, K., Hafeez, M., Bilal, K., Shabbir, M.S. (2022), "Interactive Effects of Organizational Structure and Team Work Quality on Project Success in Project Based Non Profit Organizations". *IRASD Journal of Management*, 4(1), P. 84–103. DOI: https://doi.org/10.52131/jom.2022.0401.0064
13. Zachko, O.B. (2013), "Intelligent Modeling of Infrastructure Project Product Parameters (on the Example of Lviv Airport)", *Eastern-European Journal of Enterprise Technologies*. Vol. 1. No. 10. P. 92–94.
14. Zachko, O.B. (2014), "Safety management of complex infrastructure projects in the civil protection system". *Project management: state and prospects: materials. 10th International Scientific and Practical Conference*. Mykolaiv: NUK. P. 91–92.
15. Zachko, O.B. (2015), Models, mechanisms and information technologies of portfolio management of the development of complex regional life safety systems. Under the general editorship of Raka Yu.P. Monograph. Lviv: Publishing House of LSULS, 2015. 177 p.
16. DSTU EN 13775-6:2022. Railway transport. Measurement of new and modernized freight cars. Part 6. Freight wagons consisting of several rigidly connected units and articulated freight wagons.
17. DSTU 2609-94 "Freight transportation by road. Terms and definitions".

Надійшла (Received) 05.04.2025

Відомості про авторів / About the Authors

Кобилкін Дмитро Сергійович – кандидат технічних наук, доцент, Львівський державний університет безпеки життєдіяльності, учений секретар, Львів, Україна; e-mail: dmytrokobylkin@gmail.com; ORCID ID: https://orcid.org/0000-0002-2848-3572

Зачко Олег Богданович – доктор технічних наук, професор, Львівський державний університет безпеки життєдіяльності, професор кафедри права та менеджменту у сфері цивільного захисту, Заслужений діяч науки і техніки України, Львів, Україна; e-mail: zachko@ukr.net; ORCID ID: https://orcid.org/0000-0002-3208-9826

Мицько Ростислав Ігорович – аспірант, Львівський державний університет безпеки життєдіяльності, кафедра права та менеджменту у сфері цивільного захисту, Львів, Україна; e-mail: mytsko@ukr.net, ORCID ID: https://orcid.org/0009-0007-4358-4973

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

Захарчишин Сергій Васильович – аспірант, Львівський державний університет безпеки життєдіяльності, кафедра права та менеджменту у сфері цивільного захисту, Львів, Україна; e-mail: zakharchyshyn.s@viknaroff.com.ua, ORCID ID: https://orcid.org/0009-0007-0130-7345

Elmas Chetin – доктор, професор, Газі Університет, Анкара, Туреччина; e-mail: cetinelmas@hotmail.com; ORCID ID: https://orcid.org/0000-0001-9472-2327

Kobylkin Dmytro – Candidate of Technical Sciences, Associate Professor, Lviv State University of Life Safety, Academic Secretary, Lviv, Ukraine.

Zachko Oleh – Doctor of Technical Sciences (Engineering), Professor, Lviv State University of Life Safety, Professor of the Department of Law and Management in the Field of Civil Protection, Honored Worker of Science and Technology of Ukraine, Lviv, Ukraine.

Mytsko Rostyslav – PhD student, Lviv State University of Life Safety, Department of Law and Management in the Field of Civil Protection, Lviv, Ukraine.

Zakharchyshyn Serhiy – PhD student, Lviv State University of Life Safety, Department of Law and Management in the Field of Civil Protection, Lviv, Ukraine.

Elmas Chetin – Dr., Professor, Gazi University, Ankara, Turkey.

MANAGEMENT OF CRITICAL PARAMETERS OF LOGISTICS AND INFRASTRUCTURE PROJECTS BY MEANS OF COMPUTER EXPERIMENT (ON THE EXAMPLE OF THE SECURITY AND DEFENSE SECTOR IN WAR CONDITIONS)

The subject of the research in the article is the management of critical parameters of infrastructure and logistics projects using the example of the security and defense sector in wartime by means of computer experiments using modern artificial intelligence tools, in particular multi-agent systems. **The purpose** of the study is to conduct a computer experiment to study the critical parameters of the functioning of infrastructure and logistics projects based on intelligent models that apply the theory of multi-agent systems. The article solves the **following tasks**: formalization of technical parameters of transport infrastructure; formalization of the subject area of logistics and infrastructure project management, modeling of critical parameters of the functioning of infrastructure and logistics projects under martial law, risk assessment of logistics and infrastructure projects based on the analysis of critical locations. The following **methods** are used: operations research methods; computer modeling methods, artificial intelligence technologies, in particular multi-agent systems; risk theories. **Research results:** the methodology for conducting a computer experiment modeling critical parameters of the functioning of infrastructure and logistics projects under martial law was considered. The implementation of logistics projects was analyzed depending on such criteria as distance, project time, load capacity, critical zones of logistics projects and qualitative benefits from implementation. A matrix of logistics project routes under martial law was developed. A qualitative and quantitative analysis of the risks of logistics and infrastructure projects under martial law was conducted. **Conclusions:** the obtained scientific results complement the methodology of infrastructure project management and make it possible to implement logistics projects under martial law and the worsening of the food security situation, in particular as alternatives to the Black Sea Grain Initiative, as well as infrastructure and logistics projects in the security and defense sector under war conditions. The results obtained can also be applied in planning logistics projects for the supply of weapons under martial law.

Keywords: infrastructure projects; logistics projects; safety management; security and defense sector.

Бібліографічні описи / Bibliographic descriptions

Кобилкін Д.С., Зачко О.Б., Мицько Р.І., Захарчишин С.В., Elmas Chetin. Управління критичними параметрами логістичних та інфраструктурних проєктів засобами комп'ютерного експерименту (на прикладі сектору безпеки та оборони в умовах війни). *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 33–44. DOI: https://doi.org/10.30837/2522-9818.2025.3.033

Kobylkin, D., Zachko, O., Mytsko, R., Zakharchyshyn, S., Elmas Chetin. (2025), "Management of critical parameters of logistics and infrastructure projects by means of computer experiment (on the example of the security and defense sector in war conditions)", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 33–44. DOI: https://doi.org/10.30837/2522-9818.2025.3.033

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.657:004.051:004.413.5
DOI: https://doi.org/10.30837/2522-9818.2025.3.045

Д. Михневич, О. Мазурова, О. Перепичай

МЕТОД GRAPHMIGR8 МІГРАЦІЇ РЕЛЯЦІЙНИХ ДАНИХ У ГРАФОВУ МОДЕЛЬ NEO4J

У сучасному світі оброблення інформації все більшої актуальності набувають графові бази даних, що дають змогу ефективно моделювати та обробляти складні взаємозв'язки між сутностями предметних галузей. Нині основою більшості наявних програмних систем залишаються реляційні бази даних. Проте ці бази даних не завжди є кращим рішенням для програмних систем, для яких гостро постають вимоги масштабованості та високої доступності інформації. З метою підвищення продуктивності таких систем на практиці все частіше приймаються рішення щодо міграції їх реляційних баз даних в *NoSQL*, зокрема в графові бази даних. **Предметом дослідження** статті є методи міграції структурованих даних з реляційної моделі баз даних у графову модель. **Мета роботи** – підвищити ефективність міграції реляційних баз даних у графові бази даних способом розроблення продуктивного, адаптованого для системи управління базами даних *Neo4j* методу міграції та надання рекомендацій щодо ефективного впровадження методів міграції в графові бази даних. У роботі розв'язуються такі **завдання:** створення логічних моделей для реляційної та графової БД *Neo4j* в різних предметних галузях для проведення на їх основі експериментів з міграції; розроблення методу міграції реляційних даних у графову модель *Neo4j*; планування та проведення експериментального дослідження ефективності запропонованого методу порівняно з іншими методами міграції, а також розроблення рекомендацій щодо особливостей їх використання. Упроваджено такі **методи:** проєктування баз даних; міграція в графові бази даних; експериментальне оцінювання продуктивності баз даних; розроблення, основане на системах управління базами даних *MS SQL Server* 18 та *Neo4j* 5.26 і середовищі розроблення *Visual Studio* 2022. **Досягнуті результати:** запропоновано метод міграції *GraphMigr8* реляційних даних у графову модель *Neo4j*; експериментально оцінено якість методів міграції за метриками продуктивності та семантичної цілісності інформації; сформовано рекомендації щодо використання методів міграції. **Висновки:** виявлено переваги й недоліки методів міграції реляційних даних у графову модель даних, запропоновано метод міграції *GraphMigr8* реляційної БД у графову модель *Neo4j*, який продемонстрував кращу продуктивність та вищу семантичну відповідність трансформованих даних.
Ключові слова: база даних; графова модель; метод міграції; реляційна модель; СУБД; *Neo4j*.

Вступ

Сучасний стан розвитку інформаційних технологій визначається постійним зростанням обсягів та складності структурованих даних. Традиційні реляційні бази даних (БД), незважаючи на їх тривалу історію використання, демонструють певні обмеження в роботі зі складнопов'язаними даними, які є основою для роботи в проблемних сферах, що з'являються та стрімко розвиваються в наш час, а саме:

– соціальні мережі та аналіз соціальних зв'язків [1];
– системи рекомендацій;
– біоінформатика та кібербезпека;
– ігрова аналітика тощо.

Також реляційні БД мають певні проблеми із забезпеченням масштабованості [2] програмних систем. Як зазначається в роботі [3], на фоні стрімкого збільшення обсягів взаємопов'язаної інформації традиційні реляційні БД демонструють обмеження, особливо в процесі оброблення складних рекурсивних запитів. Це викликає необхідність пошуку альтернативних підходів до зберігання та оброблення інформації та, відповідно, міграції реляційних даних до нових, більш ефективних *NoSQL* моделей БД [4].

Графові БД, як представники напряму *NoSQL*, останнім часом набувають усе більшої популярності завдяки здатності ефективно подавати складні взаємозв'язки між сутностями. Графи забезпечують природний спосіб подання взаємопов'язаної інформації, що суттєво спрощує навігацію по складних структурах [5]. Відповідно до звіту *db-engines* [6] графові системи управління базами даних (СУБД) входять до топ-10 та демонструють стійку тенденцію до зростання популярності.

Отже, графові БД є основою не тільки для створення нових програмних застосунків, але й моделлю, на яку спрямовані процеси міграції реляційних БД. Сучасні методи міграції не завжди дають однозначні результати, що потребує також верифікації БД після міграції.

Отже, актуальним і практично значущим є дослідження ефективності та вдосконалення наявних методів міграції реляційних даних у графову модель,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

що дасть змогу мінімізувати втрати інформації, прискорити процес міграції даних та забезпечити семантичну цілісність перетворення.

Аналіз проблеми й наявних методів

Порівняльний аналіз продуктивності реляційних і графових БД, проведений *Vicknair* у роботі [7], демонструє значні переваги графових БД в процесі оброблення складнопов'язаних даних. Графова БД – це механізм зберігання інформації, що поєднує базові графові структури вершин і ребер із технологією збереження та мовою запитів, що оптимізовані для зберігання та швидкого отримання щільно пов'язаних даних. На відміну від інших моделей, графові БД побудовані на концепції, що відношення між сутностями є такими самими важливими, як і самі сутності [8].

Незважаючи на загальну концепцію графових БД, реалізації графової моделі в напрямі *NoSQL*-систем можуть різнитися. Відповідно до цього будуть відрізнятися і методи міграції в ці графові моделі. Отже, для *NoSQL*-систем [4], які лише поєднують в собі загальні напрями графових, документо-орієнтованих тощо БД, але не надають уніфікованих моделей БД. Методи міграції необхідно обирати чітко під певну *NoSQL* СУБД.

Нині найбільш популярною в напрямі графових СУБД є *Neo4J* [9], яка має високу масштабованість і, як особливість, забезпечує повну підтримку атрибутів для вершин і ребер графа. *Neo4J* не є мультимодельною системою, а побудована для підтримки суто графової моделі, що забезпечує їй кращу продуктивність, а це є особливо важливим у міграції складної структурованої інформації.

Загальна методологія міграції даних передбачає такі етапи: аналіз вхідної структури даних, визначення правил відтворення вхідної структури в результуючу структуру, генерація результуючої структури та перенесення інформації до неї. Необхідно нагадати, що в роботі з *NoSQL*-системами етап перетворення вхідної структури у вихідну фактично відсутній. Наприклад, в *Neo4J*, як і в будь-якій графовій моделі, відсутнє поняття "схема даних", тому під час міграції інформації з реляційної БД етап створення схеми для графової БД відсутній, тобто не створюється певна абстрактна структура даних. Натомість рядки таблиць одразу перетворюються на певні вершини або ребра графа, тобто фактично відбувається етап перенесення даних.

Основною проблемою міграції інформації між різними моделями БД є перевірка якості міграції, а саме семантичної еквівалентності БД, адже структури вхідних та вихідних даних за такої міграції не збігаються. Найбільш ефективним варіантом перевірки якості міграції в такому разі є створення та виконання еквівалентних запитів до похідної БД та результуючої. Результати виконання запитів мають збігатися.

Серед методів міграції реляційних даних у графову модель доволі популярними є *Roberto De Virgilio* [10], *Yelda Ünal* [11] тощо. Окремо необхідно виокремити групу методів, що основані на *ER*-моделі БД та за своєю сутністю є більш схожими на методи логічного проєктування графових БД на основі *ER*-моделі [12]. Яскравим представником такого напряму є метод Чжихонг Нань та Сюе Бай [13]. Проте не варто вважати такі методи повноцінними представниками методів міграції між реляційною та графовою БД.

Доволі новим і перспективним методом міграції реляційної БД в графову є *Rel2Graph* [14]. Цей метод оснований на класифікації таблиць реляційної БД на таблиці сутностей і таблиці зв'язування та їх відповідного перетворення на структури графової моделі. Так, таблиця вважається таблицею сутностей, якщо взагалі не має зовнішніх ключів, або є один чи більше ніж два, або наявні два зовнішні ключі з простим первинним ключем. Таблиця зв'язування визначається наявністю двох зовнішніх ключів, первинний ключ якої або не є простим, або складається з цих двох зовнішніх ключів. Першим кроком алгоритм визначає, які таблиці є таблицями сутностей, а які таблиці використовуються для створення зв'язків (наприклад, проміжна таблиця у зв'язку "багато до багатьох"). Далі відбувається перетворення таблиць у графові структури. Так, кожна таблиця сутностей трансформується у вершини графа. Назва таблиці стає міткою вершини, а кожен її рядок перетворюється на окрему вершину. Атрибути таблиці в цьому разі стають атрибутами вершин. Таблиці зв'язування перетворюються на ребра графа: назва таблиці стає типом ребра, її рядки – безпосередньо ребрами, а атрибути перетворюються на атрибути ребер. Додатково в графі створюються ребра типу *HAS* для відтворення зв'язків "один до одного" та "один до багатьох".

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Аналіз методу продемонстрував певні недоліки, які можуть призвести до неефективної міграції, а саме:

– таблиці, що мають два зовнішніх ключі та простий первинний ключ, визначаються як таблиці сутності, хоча дуже часто такі таблиці проєктуються саме для реалізації зв'язку "багато до багатьох", тобто мають лише три стовпці: первинний ключ (частіше за все генерується автоматично) та два зовнішніх ключі (посилання на дві таблиці, що зв'язуються);

– таблиці, що мають два зовнішніх ключі та складений первинний ключ, визначаються як таблиці зв'язування, хоча можливі таблиці сутностей, що мають два зовнішніх ключі та складений первинний ключ.

Також алгоритм приймає на вхід файл із *SQL*-скриптами для створення реляційної БД, а не під'єднується до наявної БД, витрачаючи в такий спосіб час на оброблення та парсинг текстового файлу.

Оскільки проаналізовані методи спрямовані на абстрактну міграцію даних з реляційної БД до графової без огляду на особливості тієї чи іншої графової моделі, то актуальним є адаптація наявних методів до певної СУБД, наприклад популярної *Neo4J*, удосконалення або створення нових методів для роботи саме з цією СУБД та дослідження їх ефективності.

Мета роботи й завдання

Мета статті – підвищити ефективність міграції реляційних баз даних у графові БД способом розроблення продуктивного, адаптованого для системи управління базами даних *Neo4j* методу міграції та надання рекомендацій щодо ефективного використання методів міграції в графові БД.

Дослідження потребує виконання таких завдань:

– створення логічних моделей для реляційної та графової БД *Neo4j* в різних предметних галузях для проведення на їх основі експериментів з міграції;

– розроблення методу міграції реляційних даних у графову модель *Neo4j*;

– планування та проведення експериментального дослідження ефективності запропонованого методу порівняно з іншими методами міграції та розроблення рекомендацій щодо особливостей їх упровадження.

Розв'язання завдань

А. Математичне моделювання

Для формалізації процесу міграції використано апарат теорії графів і взято до уваги реляційну БД та графову модель БД для *Neo4j*.

Розглянемо процес міграції реляційної БД D в графову БД G під керуванням *Neo4J*. Реляційну БД D можна подати у вигляді кортежу

$$D = \langle T, P, R \rangle, \quad (1)$$

де $T = \left\{ T_i \middle| i = \overline{1,n} \right\}$ – скінченна множина таблиць реляційної БД; $P = \left\{ P_i \middle| i = \overline{1,n} \right\}$ – скінченна множина атрибутів у БД, де $P_i = \left\{ P_{ij} \middle| j = \overline{1,m} \right\}$ – скінченна множина атрибутів таблиці T_i; $R \subseteq T \times T$ – множина зв'язків між таблицями.

Графова БД G під керуванням *Neo4j*, яка буде результатом міграції, може мати такий вигляд:

$$G = \langle V, E \rangle, \quad (2)$$

де V – множина вершин графа, $V = \left\{ V_l \left(PV_l \right) \middle| l = \overline{1,p} \right\}$, $V_l \left(PV_l \right)$ – скінченна множина вершин певного l-го типу, що визначається однаковою множиною атрибутів PV_l; $E \subseteq V \times V$ – множина ребер графа, $E = \left\{ E_k \left(PE_k \right) \middle| k = \overline{1,q} \right\}$, $E_k \left(PE_k \right)$ – скінченна множина ребер певного k-го типу, якій властива однакова множина атрибутів PE_k ребер.

Процес міграції можна подати як сукупність відтворень:

– $\varphi : T \to V \cup E$ – перетворення, що ставить у відповідність кожній таблиці $T_i \in T$ множину вершин $V_i \left(PV_i \right) \in V$ певного типу або множину ребер $E_i \left(PE_i \right) \in E$ певного типу;

– $\Psi : P \to PV \cup PE$ – перетворення, що ставить у відповідність кожній множині атрибутів $P_i \in P$ таблиці T_i множину PV_i атрибутів вершин або множину PE_i атрибутів ребер;

– $\Omega : R \to E$ – перетворення, що ставить у відповідність кожному зв'язку $\left(T_i, T_j \right) \in R$ множину ребер $\left(V_l, V_k \right) \in E$ певного типу.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Отже, для власного методу міграції реляційної БД у графову під керівництвом *Neo4J* необхідно визначити відтворення φ, Ψ та Ω, що задають правила перетворення таблиць, зв'язків і атрибутів реляційної БД в множини вершин і ребер певних типів графової БД та їх атрибути.

Б. Розроблення методу GraphMigr8 для міграції в Neo4J

У роботі, спираючись на метод *Rel2Graph* [14], запропоновано метод міграції *GraphMigr8* реляційної БД (1) у графову БД (2) під керівництвом *Neo4J*, що має на меті подолати недоліки попередніх методів і забезпечити більш гнучкий та ефективний процес міграції реляційних структур.

Особливістю запропонованого методу є більш однозначне й чітке виділення таблиць, що реалізують зв'язок типу "багато до багатьох". Зазвичай у сучасній практиці проєктування такі таблиці мають простий первинний ключ (сурогатний) та два зовнішніх ключі, що посилаються на відповідні таблиці, іноді з додатковим змістовним атрибутом. Альтернативний варіант структури такої таблиці зв'язування передбачає наявність двох зовнішніх ключів, що одночасно є складеним первинним ключем, також з можливим додатковим змістовним атрибутом. У методі *GraphMigr8* за наявності понад одного змістовного атрибута запропоновано розглядати такі таблиці як стрижневі, що моделюють сутності реального світу, а не зв'язки типу "багато до багатьох", та перетворювати їх під час міграції у вершини графової БД. Так, у методі *GraphMigr8* таблиця вважатиметься таблицею вершин певного типу, якщо для неї виконується одна з умов:

– зовнішні ключі відсутні;

– кількість зовнішніх ключів дорівнює одному або більше ніж двом;

– існує два зовнішніх ключі та простий первинний ключ, до того ж кількість атрибутів у таблиці перевищує чотири, тобто в наявності мінімум два змістовних атрибути;

– існує два зовнішніх ключі, що утворюють складений первинний ключ, до того ж кількість атрибутів у таблиці більше ніж три, тобто в наявності мінімум два змістовних атрибути.

Таблиця вважатиметься таблицею ребер певного типу, якщо для неї виконується одна з умов:

– існує два зовнішніх ключі та простий первинний ключ, до того ж кількість атрибутів не перевищує чотирьох, тобто є лише один змістовний атрибут;

– існує два зовнішніх ключі, що утворюють складений первинний ключ, у цьому разі кількість атрибутів не перевищує трьох, тобто може бути лише один змістовний атрибут.

На рис. 1 зображено блок-схему запропонованого методу міграції, що містить кінцеву послідовність кроків.

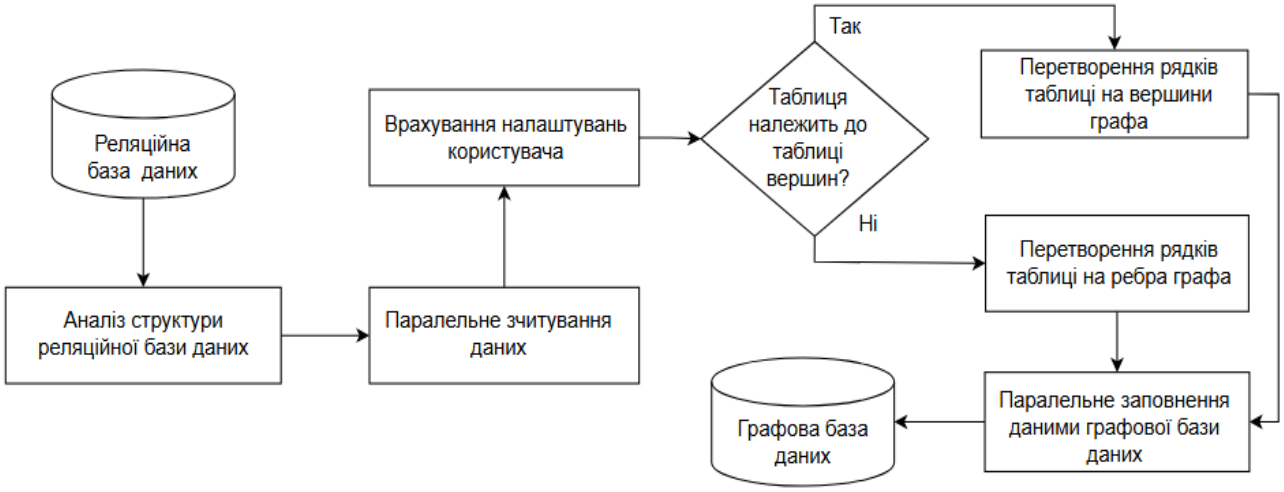


Рис. 1. Блок-схема методу міграції *GraphMigr8*

Розглянемо кроки запропонованого методу більш детально.

На кроці 1 для аналізу структури реальної реляційної БД здійснюється під'єднання безпосередньо до БД (в експериментальних дослідженнях під'єднання відбувалося до БД під керівництвом *Microsoft SQL Server*). Під час аналізу структури таблиць пропонується використовувати

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

принципи, подані в роботі *Pokorný* [15], а саме під час аналізу система збирає детальну інформацію про кожну таблицю, зокрема:

– назви таблиці та атрибутів;
– тип даних атрибутів;
– первинні ключі;
– зовнішні ключі.

Зібрана інформація зберігається в таких структурах даних, як списки, геш-мережі та словники. У цьому разі для оптимізації алгоритму використовується паралельне зчитування інформації.

На кроці 2 беруться до уваги налаштування користувача, який може задати особливості предметної галузі.

Налаштування користувача дають змогу:

– визначати правила перетворення для специфічних таблиць;

– вилучати певні таблиці або атрибути з процесу міграції.

На кроці 3 відбувається класифікація таблиць за їх призначенням і структурою відповідно до наведених вище принципів.

На кроці 4 перетворюються рядки таблиць вершин у вершини графової БД. Назва кожної таблиці вершин стає типом вершини, а кожен її рядок перетворюється на окрему вершину.

Атрибути вершин формуються на основі атрибутів вхідної таблиці. Система зберігає всю семантичну інформацію, зокрема числові, текстові та інші типи інформації.

На кроці 5 перетворюються рядки таблиць ребер у ребра графа. Назва таблиці ребер стає типом ребра, а її рядки – безпосередньо ребрами між вершинами графа. Якщо в таблиці ребер були присутні змістовні (не ключові) атрибути, то вони перетворюються на атрибути ребер.

На етапі перетворення рядків таблиць ребер у ребра графа також створюються ребра типу *HAS* для всіх зовнішніх ключів, розташованих у таблицях вершин. Кожен зовнішній ключ стає ребром типу *HAS*, яке з'єднує поточну вершину (рядок таблиці із зовнішнім ключем) з відповідною цільовою вершиною (рядок таблиці, на який посилається зовнішній ключ).

На кроці 6 відбувається паралелізація перетворення та заповнення даними графової БД, яка є ключовим елементом підвищення продуктивності алгоритму. Для реалізації паралелізації використовуються такі підходи:

– паралельне зчитування даних – система розподіляє таблиці між декількома потоками, даючи змогу одночасно обробляти різні таблиці; кожен потік незалежно виконує аналіз структури та її класифікацію;

– паралельне заповнення графової БД даними – створення вершин та ребер відбувається паралельно; це досягається способом розподілу інформації на незалежні групи, які можуть оброблятися одночасно.

Отже, запропонований метод міграції *GraphMigr8* є гнучким інструментом для перетворення реляційних даних у графову БД.

B. Планування експериментального дослідження

Для експериментального дослідження продуктивності та ефективності запропонованого методу розглянуто декілька предметних галузей, для яких є доречним використання графових БД, зокрема сфери соціальних мереж, ігрових статистик та електронної комерції.

Розглянемо підготовку експериментального дослідження на прикладі розроблення БД для соціальної мережі. Ця сфера містить поняття про такі об'єкти реального світу, як користувачі, їхні профілі, зв'язки дружби, контент і взаємодії між користувачами. На рис. 2 зображено *ER*-діаграму БД у сфері соціальної мережі, що створено за нотацією *Richard Barker* із додатковими семантичними позначками зв'язків, що є запозиченням з нотації *Peter Pin-Shan Chen*.

Для оцінювання ефективності алгоритмів міграції обрано такі метрики:

– час трансформації даних (с) – час, який витрачає алгоритм на перетворення реляційної БД у графову; вимірювання часу для БД різних розмірів дав змогу оцінити масштабованість розроблених алгоритмів;

– обсяг використаної оперативної пам'яті (МВ) – кількісна оцінка обсягу пам'яті, що споживає алгоритм під час процесу міграції; ця метрика дала змогу порівняти ресурсоємність розроблених алгоритмів;

– відсоток збігу мігрованої структури із спроєктованою розробником (%); для чистоти дослідження під час упровадження методу *GraphMigr8* етап врахування налаштувань користувача було пропущено.

Важливим показником якості міграції даних є збереження їх семантичної цілісності [16]. Для оцінювання якості та повноти отриманих графових БД застосовано метрику семантичної

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

еквівалентності запитів та їх результатів. Сутність цієї метрики полягає в порівнянні результатів, ідентичних за логікою запитів, сформованих для вхідної (реляційної) та вихідної (графової) БД.

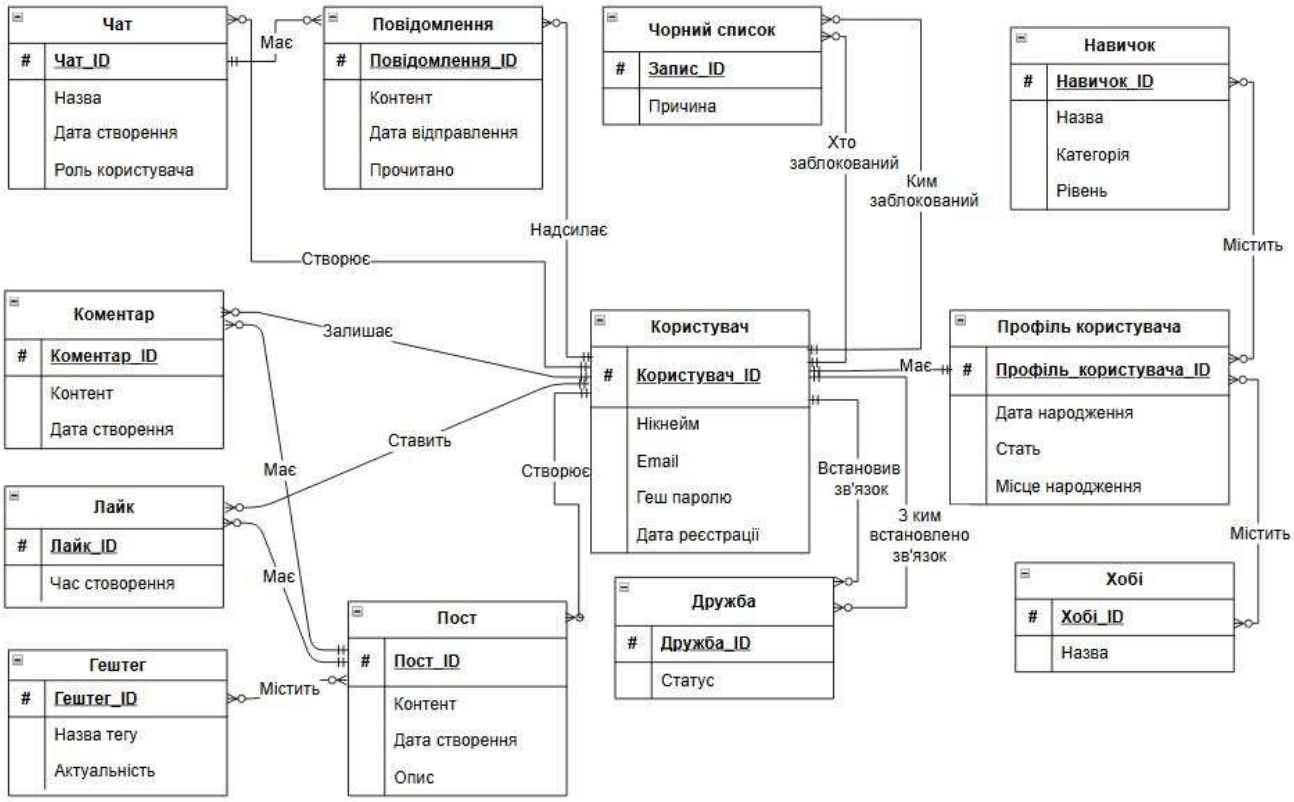


Рис. 2. *ER*-діаграма бази даних у сфері соціальної мережі

Використана для метрики технологія оцінювання передбачає такі дії:

– формування набору тестових запитів для вхідної реляційної БД;

– трансформацію цих запитів на мову *Cypher* для графової БД під керівництвом *Neo4j*;

– виконання запитів на базі СУБД *MS SQL Server ma Neo4J*, порівняння результатів тестових запитів за критерієм ідентичності кількості результуючих записів і значень за відповідними атрибутами.

Додатково для аналізу ефективності методів міграції оцінено продуктивність роботи з отриманими графовими БД способом виконання еквівалентних запитів на графових БД, створених різними алгоритмами міграції. Метриками оцінювання продуктивності роботи з графовими БД обрано:

– час виконання запитів (мс);

– завантаженість системних ресурсів (завантаженість процесора під час виконання запитів (%), обсяг споживання оперативної пам'яті (MB));

– ефективність зберігання інформації (загальний обсяг БД на диску (MB));

– кількість операцій доступу до БД, що виконуються під час виконання запиту (*DB hits*).

Результати досліджень та їх обговорення

З використанням описаних метрик порівняно два алгоритми методу *Rel2Graph* (у класичній імплементації та оптимізований алгоритм *Rel2Graph*, який використовує багатопотоковість) та розроблений алгоритм для *GraphMigr8*.

Експерименти проводилися на обчислювальній машині з такими характеристиками:

– процесор – Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30 GHz (загалом чотири ядра);

– оперативна пам'ять – 16 GB;

– 64-бітна операційна система.

Використовувалися СУБД таких версій: *MS SQL Server* 18 та *Neo4j* 5.26.

Більш детально розглянемо результати експериментів із базами даних середнього розміру, що було розроблено для більш адекватних щодо використання графових моделей сфер соціальних мереж та ігрових статистик, адже вони мають

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

найбільш зв'язану структуру. У табл. 1 подано заміри за таким важливим показником ефективності, як час міграції реляційної БД у графову БД.

Оптимізований *Rel2Graph* та *GraphMigr8* демонструють значне скорочення часу міграції для обох БД завдяки оптимізації способом багатопотоковості. Метод *GraphMigr8* має найкращі показники у зв'язку із меншою кількістю результуючих вершин і ребер, що він утворює під час міграції.

Не менш важливим є вимірювання використаних системних ресурсів під час роботи алгоритмів (табл. 2).

Таблиця 1. *Час міграції (с)*

Алгоритм	Час міграції для БД *SocialNetwork* (с)	Час міграції для БД *GameDB* (с)
Rel2Graph	213.27	202.51
Rel2Graph оптимізований	8.71	15.24
GraphMigr8	7.93	5.5

Таблиця 2. *Використання системних ресурсів*

Алгоритм	Використання оперативної пам'яті для *SocialNetwork* (МБ)	Використання оперативної пам'яті для *GameDB* (МБ)	Використання процесора для *SocialNetwork* (%)	Використання процесора для *GameDB* (%)
Rel2Graph	6.7	9.83	14.31	13.99
Rel2Graph оптимізований	15.19	11.08	3.79	2.39
GraphMigr8	11.29	6.47	2.99	3.16

Хоча алгоритм *GraphMigr8* не показує очевидної переваги з погляду застосування оперативної пам'яті, але він значно знижує навантаження на процесор, що робить його більш ефективним щодо використання системних ресурсів, а особливо для систем, де *CPU* є обмеженим ресурсом.

Проаналізуємо досягнуті результати для різних обсягів реляційних БД. Великою вважатимемо базу даних *Ecommerce*, середніми – *SocialNetwork* та *GameDB* та малими – *TravelingDb, ReportDb, SportDb*. Для середніх та малих БД наведемо усереднені значення. Отримані метрики подано в табл. 3.

Таблиця 3. *Результати міграції для БД різного обсягу*

Тип БД	Алгоритм	Час міграції (с)	Використання оперативної пам'яті (МБ)	Використання процесора (%)
Мала	*Rel2Graph*	8.6	9.6	6.5
	Rel2Graph оптимізований	4.67	2.08	3.8
	GraphMigr8	3.51	1.3	1.88
Середня	*Rel2Graph*	207.89	8.23	14.15
	Rel2Graph оптимізований	11.98	13.14	3.09
	GraphMigr8	6.715	8.88	3.08
Велика	*Rel2Graph*	10829.44	91.1	26.04
	Rel2Graph оптимізований	433.21	52.99	7.44
	GraphMigr8	427.84	42.2	7.07

Оригінальний *Rel2Graph* демонструє яскраво виражену експоненційну залежність від розміру БД. Перехід від середньої до великої БД спричиняє зростання часу міграції в понад 52 рази (з 207.89 с до 10829.44 с). Оптимізований *Rel2Graph* значно зменшує експоненційне зростання, хоча залежність усе ще не є лінійною. Алгоритм *GraphMigr8* показує найбільш наближену до лінійної залежність із незначним кутом нахилу, що є оптимальним для масштабування. На великій БД оптимізований *Rel2Graph* та *GraphMigr8* показали майже однаковий час міграції. Це пов'язано з тим, що сфера *Ecommerce* не є природною для зберігання у графовій БД. Отже, реляційна схема цієї БД містить таблиці, які обома методами визначаються як таблиці вершин, що не дає змогу задіяти переваги методу *GraphMigr8*. Незважаючи на це, розроблений *GraphMigr8* показав не гірші результати, ніж метод *Rel2Graph* [14].

Для малих БД алгоритм *GraphMigr8* використовує на 86% менше пам'яті порівняно з оригінальним *Rel2Graph*. За умови збільшення розміру БД спостерігається зростання застосованої пам'яті в усіх алгоритмах, проте *GraphMigr8* завжди демонструє кращі показники.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Оригінальний *Rel2Graph* найбільш інтенсивно використовує процесор, особливо в роботі з великими БД. Оптимізований *Rel2Graph* та *GraphMigr8* демонструють приблизно однакове навантаження на процесор для середніх і великих БД. Алгоритм *GraphMigr8* має найменше навантаження на процесор під час роботи з малими БД.

Перейдемо до розгляду експериментів щодо визначення семантичної цілісності результату міграції.

Було розроблено низку запитів мовами *SQL* та *Cypher* (офіційна мова запитів *Neo4j*) [17] та виконано їх на реляційній БД під керівництвом *MS SQL Server* та графовій БД під керівництвом *Neo4j*. Результати запитів порівнювалися автоматизовано за допомогою розробленого програмного забезпечення.

Для перевірки семантичної еквівалентності даних використовувалися серії статистичних запитів на підрахунок кількості сутностей різного типу, запитів на фільтрацію сутностей за різними атрибутами, а також запитів на перевірку зв'язності сутностей.

Наприклад, для сфери соціальної мережі застосовувалися запити, що, зокрема, містили:

– запит № 1 – пошук постів користувачів, імена яких починаються з літери "A";

– запит № 2 – пошук гештегів із релевантністю понад 0.7;

– запит № 3 – отримання середнього рівня навичок користувачів;

– запит № 4 – пошук потенційних друзів (запит до декількох зв'язаних сутностей).

Для сфери ігрової серверної системи застосовувалися запити:

– запит № 1 – пошук деталей спорядження персонажа (запит до декількох зв'язаних сутностей);

– запит № 2 – отримання інформації про предмети певного типу, що наявні в інвентарі персонажів;

– запит № 3 – пошук нагород за квести з певним атрибутом;

– запит № 4 – історія отриманих персонажем предметів.

Усі розроблені запити для обох досліджених методів повернули однакову кількість записів з ідентичними значеннями за атрибутами, що дає змогу зробити висновок про семантичну цілісність мігрованих даних.

Метрики продуктивності отриманих графових БД за результатами виконання цих запитів зведені в табл. 4. В цій серії експериментів досліджувалась лише одна реалізація методу *Rel2Graph*, адже алгоритми *Rel2Graph* та оптимізований *Rel2Graph* створюють однакову структуру графової БД, тож їх порівняння з погляду продуктивності графової БД не має сенсу. Отже, порівнювалися саме методи *Rel2Graph* та *GraphMigr8*, а також оцінювалося покращення метрик продуктивності для методу *GraphMigr8*.

Таблиця 4. *Продуктивність запитів*

База даних	**Запит**	**Метод**	**Оперативна пам'ять (байти)**	**Час (мс)**	***CPU* (%)**	***DB hits***
SocialNetwork	запит № 1	*Rel2Graph*	4472	3	7	445
		GraphMigr8	4472	2	6.3	445
		Покращення	0 %	33.3 %	10.0 %	0 %
	запит № 2	*Rel2Graph*	45000	7	10.2	5331
		GraphMigr8	40608	5	8	3257
		Покращення	9.8 %	28.6 %	21.6 %	38.9 %
	запит № 3	*Rel2Graph*	13672	5	7.3	3669
		GraphMigr8	12944	3	7.9	2512
		Покращення	5.3 %	40.0 %	-8.2 %	31.4 %
	запит № 4	*Rel2Graph*	14352	3	11.7	998
		GraphMigr8	9592	2	7.9	426
		Покращення	33.2 %	33.3 %	32.5 %	57.3 %
GameDB	запит № 1	*Rel2Graph*	2536	2	7.2	834
		GraphMigr8	2104	2	4.6	758
		Покращення	17.0 %	0 %	36.1 %	9.1 %
	запит № 2	*Rel2Graph*	8592	3	7.1	617
		GraphMigr8	8584	2	7.4	533
		Покращення	0.1 %	33.3 %	-4.2 %	13.6 %
	запит № 3	*Rel2Graph*	360	1	8.2	227
		GraphMigr8	344	1	6.2	181
		Покращення	4.4 %	0 %	24.4 %	20.3 %
	запит № 4	*Rel2Graph*	2256	2	6.2	718
		GraphMigr8	2256	1	6.8	682
		Покращення	0 %	50.0 %	-9.7 %	5.0 %

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

Унаслідок упровадження методу *GraphMigr8* спостерігається значне покращення використання оперативної пам'яті й здебільшого високі результати за часом виконання. Загалом метод *GraphMigr8* демонструє нижчу завантаженість процесора.

Детальний аналіз графових БД під час експериментів засвідчив, що однією з ключових переваг методу *GraphMigr8* є отримання більш досконалої структури графа відповідно до кількості його вершин і ребер. Кількість вузлів і ребер результуючих графових БД наведено в табл. 5 та оцінено відсоток покращення в разі використання методу *GraphMigr8*.

Таблиця 5. *Кількість вершин та ребер*

База даних	Метод	Кількість вершин	Кількість ребер
SocialNetwork	*Rel2Graph*	6744	12388
	GraphMigr8	2850	8494
	Покращення	57.7%	31.4%
GameDB	*Rel2Graph*	7885	12879
	GraphMigr8	2717	7711
	Покращення	65.5%	40.1%

Зменшення кількості вершин та ребер у процесі застосування методу міграції *GraphMigr8* безпосередньо вплинуло на розмір БД. Розміри результуючих графових БД наведено в табл. 6 та показано відсоток покращення для методу *GraphMigr8*.

Таблиця 6. *Розмір бази даних*

База даних	Метод	Розмір (MB)
SocialNetwork	*Rel2Graph*	3.35
	GraphMigr8	2.42
	Покращення	27.8 %
GameDB	*Rel2Graph*	2.46
	GraphMigr8	1.67
	Покращення	32.1 %

Останньою серією експериментів була перевірка збігу автоматично мігрованої структури до спроєктованої розробником вручну, який мав змогу взяти до уваги особливості предметної галузі. Для цього вручну спроєктовано графові БД на основі *ER*-діаграм, на базі яких було розроблено відповідні реляційні БД для обраних сфер. Приклад *ER*-діаграми та спроєктованої вручну графової БД для сфери соціальної мережі зображено на рис. 2 та 3 відповідно.

Розглянемо деякі результати цієї серії експериментів.

Подамо назви таблиць реляційних БД для кращого розуміння результату міграції:

– *TravelingDb – Achievement, Role, User, UserAchievement;*

– *SocialNetwork – BlackList, Chat, Comment, Friendship, Hashtag, HashtagPost, Hobby, Like, Message, Post, Skill, User, UserChat, UserHobby, UserProfile, UserSkill;*

– *Ecommerce – Basket, Category, Order, Product, Product_basket, Product_order, Product_property, Property, User.*

Результати експерименту наведені в табл. 7.

Результати експериментів для БД малого розміру показали, що графові моделі, мігровані за допомогою методу *GraphMigr8*, мають більш ідентичну структуру до БД, що були спроєктовані вручну. Натомість графові моделі, мігровані за допомогою методу *Rel2Graph*, менш ідентичні до спроєктованих вручну (наприклад, БД *TravelingDb*).

БД середнього розміру для сфер соціальних мереж та ігрових серверних застосунків, які визначаються високим ступенем взаємозв'язків між сутностями, є найбільш прийнятними кандидатами для використання графових БД і, відповідно, є більш показовими об'єктами для експериментальних досліджень. Як бачимо, для БД *SocialNetwork* метод *GraphMigr8* продемонстрував більш компактну графову структуру, на відміну від спроєктованої вручну. Так, таблиця *Like* була перетворена на ребра між вершинами *User* та *Post*, що привело до більш компактної та логічної структури графа. Для БД *GameDB* спроєктована вручну графова структура виявилася більш компактною. Але метод *GraphMigr8* знову продемонстрував кращі результати, ніж *Rel2Graph*.

Для БД *Ecommerce* графові структури для всіх трьох методів виявилися ідентичними, якщо не брати до уваги назви ребер. Це можна пояснити тим, що всі проміжні таблиці (таблиці зв'язування) в цій реляційній БД мають складений первинний ключ, частини якого одночасно є зовнішніми ключами, та кількість змістовних атрибутів не перевищує двох. У такому разі методи *Rel2Graph* та *GraphMigr8* створюють однакові графові структури. До того ж необхідно зазначити, що розглянутий фрагмент сфери електронної комерції майже ніколи не проєктується розробниками у вигляді графової БД. Отже, це порівняння не можна вважати показовим.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Таблиця 7. *Результати порівняння автоматично мігрованої структури із спроєктованою розробником*

База даних	Метод	Кількість вершин	Кількість ребер	Назви вершин	Назви ребер
TravelingDb	Ручне проєктування	44	112	Achievement, Role, User	Has_Achievements, Has_Role
	Rel2Graph	136	204	Achievement, Role, User, UserAchievement	HAS_Achievement _UserAchievement, HAS_Role_User, HAS_User_UserAchievement
	GraphMigr8	44	112	Achievement, Role, User	HAS_Role_User, UserAchievement
SocialNetwork	Ручне проєктування	4813	10457	Chat, Comment, Hashtag, Hobby, Like, Message, Post, Skill, User, UserProfile	HAS (чотири зв'язки), Send, Leaves, Puts, Creates, Friends_with, Has_blocked, Contains (три зв'язки)
	Rel2Graph	6744	12388	BlackList, Chat, Comment, Friendship, Hashtag, HashtagPost, Hobby, Like, Message, Post, Skill, User, UserChat, UserHobby, UserProfile, UserSkill	HAS_Chat_Message, HAS_Chat_UserChat, HAS_Hashtag_HashtagPost, HAS_Hobby_UserHobby, HAS_Post_Comment, HAS_Post_HashtagPost, HAS_Post_Like, HAS_Skill_UserSkill, HAS_UserProfile_UserHobby, HAS_UserProfile_UserSkill, HAS_User_BlackList, HAS_User_Comment, HAS_User_Friendship, HAS_User_Like, HAS_User_Message, HAS_User_Post, HAS_User_UserChat, HAS_User_UserProfile
	GraphMigr8	2850	8494	Chat, Comment, Hashtag, Hobby, Message, Post, Skill, User, UserProfile	BlackList, Friendship, HAS_Chat_Message, HAS_Post_Comment, HAS_User_Comment, HAS_User_Message, HAS_User_Post, HAS_User_UserProfile, HashtagPost, Like, UserChat, UserHobby, UserSkill
Ecommerce	Ручне проєктування	1800098	9400041	Basket, Category, Order, Product, Property, User	HAS_Order, HAS_Basket, Placed_in, HAS_Property, Ordered, HAS_Parent
	Rel2Graph	1800098	9400041	Basket, Category, Order, Product, Property, User	HAS_Category_Category, HAS_Category_Product, HAS_User_Basket, HAS_User_Order, Product_basket, Product_order, Product_property
	GraphMigr8	1800098	9400041	Basket, Category, Order, Product, Property, User	HAS_Category_Category, HAS_Category_Product, HAS_User_Basket, HAS_User_Order, Product_basket, Product_order, Product_property

Проведена серія експериментів демонструє, що БД із значною кількістю простих зв'язків (без додаткових змістовних атрибутів) отримують більший ефект від міграції до графової моделі за допомогою алгоритму *GraphMigr8*. Натомість у системах електронної комерції, де зв'язки містять багато власних змістовних атрибутів, методи *GraphMigr8* та *Rel2Graph* демонструють майже ідентичні результати. У разі коли проміжні таблиці мають простий первинний ключ та два зовнішніх ключі, алгоритм *GraphMigr8* повертатиме більш компактну графову структуру.Отже, проведені експерименти продемонстрували суттєві переваги методу *GraphMigr8* до міграції реляційних БД у графові під керівництвом *Neo4J* , якщо порівнювати з методом *Rel2Graph*, особливо для предметних галузей із високою щільністю взаємозв'язків між сутностями, де графова модель даних є природно більш відповідною для подання та оброблення інформації. Наголосимо, що дослідження підтверджують ефективність запропонованого методу *GraphMigr8* міграції реляційних БД у графові БД *Neo4J*, демонструючи його переваги за ключовими показниками продуктивності.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Висновки й перспективи подальшого дослідження

У роботі запропоновано метод міграції реляційних БД у графову БД *Neo4j* - *GraphMigr8* та експериментально перевірено його ефективність.

Одним із ключових етапів роботи стало проєктування логічних моделей предметних галузей, що дало змогу створити ефективну основу для дослідження міграції. Зокрема розроблено моделі для БД соціальних мереж та ігрових серверів, які є прикладами доцільності міграції з реляційних БД у графові.

Досягнуті результати експериментів демонструють суттєві переваги методу *GraphMigr8*, а саме: скорочення часу міграції, більш ефективне використання системних ресурсів і зниження складності отриманих графових структур. Було сформульовано рекомендації щодо впровадження розглянутих методів міграції залежно від характеру та складності вхідних показників.

Розроблений метод *GraphMigr8* має певні обмеження, оскільки він розроблявся спеціально під міграцію даних до *Neo4j*, де у ребер є можливість зберігати атрибути. Проте не всі графові моделі підтримують цю функціональність. А втім метод *GraphMigr8* без суттєвих змін підійде для міграції в такі графові СУБД, як *Amazon Neptune, OrientDb, TigerGraph*, що також підтримають атрибути у ребер.

Натомість для СУБД *Apache Giraph* або *FlockDB*, де концепція атрибутів ребер обмежена або відсутня, застосування методу *GraphMigr8* потребуватиме модифікації. У цьому разі більш ефективним буде впровадження методу *Rel2Graph*.

Надалі доцільними є додаткові експериментальні дослідження з вищезгаданими СУБД для більш точного аналізу ефективності запропонованого методу в різних технологічних середовищах. Але можна зробити прогноз, що якісні переваги методу *GraphMigr8* збережуться порівняно з методом *Rel2Graph*, хоча ефективність кількісних показників може дещо змінитися. Ці розбіжності можуть бути зумовлені особливостями реалізації відповідних СУБД, специфікою їх мов запитів, оптимізацією внутрішніх алгоритмів оброблення інформації та ефективністю роботи з різними типами структур.

Список літератури

1. Vyawahare H. Exploring the Hybrid Approach: Integrating Relational and Graph Databases for Enhanced Data Management. *Communications in Computer and Information Science*. 2024. Vol. 2235, P. 176–191. DOI: https://doi.org/10.1007/978-3-031-74701-4_13
2. Yedilkhan D., Mukasheva A., Bissengaliyeva D. Performance Analysis of Scaling NoSQL vs SQL: A Comparative Study of MongoDB, Cassandra, and PostgreSQL. *IEEE International Conference on Smart Information Systems and Technologies* (SIST). 2023. P. 479–483. DOI: https://doi.org/10.1109/SIST58284.2023.10223568
3. Ragunathan A., Bhavani K., Sasi A. Integration of NoSQL and Relational Databases for Efficient Data Management in Hybrid Cloud Architectures. *Journal of Machine and Computing*. 2025. P. 1277–1287. DOI: https://doi.org/10.53759/7669/jmc202505100
4. Kuzochkina A., Shirokopetleva M., Dudar Z. Analyzing and Comparison of NoSQL DBMS. *International Scientific-Practical Conference Problems of Infocommunications. Science and Technology* (PIC S&T). 2018. P. 560–564. DOI: https://doi.org/10.1109/INFOCOMMST.2018.8632133
5. Mhedhbi A., Deshpande A., Salihoğlu S. Modern Techniques for Querying Graph-structured Databases. *Foundations and Trends® in Databases*. 2024. №14(2). P. 72–185. DOI: https://doi.org/10.1561/1900000090
6. DB-Engines Ranking. URL: https://db-engines.com/en/ranking (дата звернення: 25.12.2024)
7. Vicknair C., Macias M., Zhao Z. A comparison of a graph database and a relational database: a data provenance perspective. *Proceedings of the 48th Annual Southeast Regional Conference*, ACM. 2010. P.1–6. DOI: https://doi.org/10.1145/1900008.1900067
8. Bonifati A., Özsu M. T., Tian Y. The Future of Graph Analytics. *Companion of the 2024 International Conference on Management of Data* (SIGMOD '24). 2024. P. 544–545. DOI: https://doi.org/10.1145/3626246.3658369
9. Neo4j, Inc. How to create a graph data model. URL: https://neo4j.com/docs/model/ (дата звернення: 01.04.2025)
10. De Virgilio R., Maccioni A., Torlone R. Model-driven design of graph databases. *International Conference on Conceptual Modeling*, Springer. 2024. P.172-185. DOI: https://doi.org/10.1007/978-3-319-12206-9_14
11. Ünal Y., Oğuztüzün H. Migration of Data from Relational Database to Graph Database. *Proceedings of the 2018 International Conference on Information Systems and Computer Science*. 2018. P. 1–5. DOI: https://doi.org/10.1145/3200842.3200852

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

12. Mazurova O., Syvolovskyi I., Syvolovska O. NOSQL database logic design methods for MONGODB and NEO4J. *Innovative technologies and scientific solutions for industries*. 2022. № 2 (20). С. 52–63. DOI: https://doi.org/10.30837/ITSSI.2022.20.052
13. Nan Z., Bai X. The study on data migration from relational database to graph database. *Journal of Physics: Conference Series*. 2019. №1345(2), P. 022061. DOI: https://doi.org/10.1088/1742-6596/1345/2/022061
14. Zhao Z., Liu W., French T. Rel2Graph: Automated Mapping From Relational Databases to a Unified Property Knowledge Graph. *Research Square* [Preprint]. 2024. DOI: https://doi.org/10.21203/rs.3.rs-5451592/v1
15. Pokorný J. Graph Databases: Their Power and Limitations. *IFIP International Conference on Computer Information Systems and Industrial Management*. 2015. P. 58–69. DOI: https://doi.org/10.1007/978-3-319-24369-6_5
16. Jouili S., Vansteenberghe V. An empirical comparison of graph databases. *Proceedings of the 2013 International Conference on Social Computing*, IEEE. 2013. P.708–715. DOI: https://doi.org/10.1109/SocialCom.2013.106
17. Neo4j, Inc. The Neo4j Cypher Manual. URL: https://neo4j.com/docs/cypher-manual/current/ (дата звернення: 08.04.2025).

References

1. Vyawahare, H. (2024), "Exploring the Hybrid Approach: Integrating Relational and Graph Databases for Enhanced Data Management", *Communications in Computer and Information Science*, No. 2235, P. 176–191. DOI: https://doi.org/10.1007/978-3-031-74701-4_13
2. Yedilkhan, D., Mukasheva, A., Bissengaliyeva, D. and Suynullayev, Y. (2023), "Performance Analysis of Scaling NoSQL vs SQL: A Comparative Study of MongoDB, Cassandra, and PostgreSQL", *IEEE International Conference on Smart Information Systems and Technologies* (SIST), P. 479–483. DOI: https://doi.org/10.1109/SIST58284.2023.10223568
3. Ragunathan, A., Bhavani, K., Sasi, A. and Kumar, S. R. (2025), "Integration of NoSQL and Relational Databases for Efficient Data Management in Hybrid Cloud Architectures", *Journal of Machine and Computing*, P. 1277–1287. DOI: https://doi.org/10.53759/7669/jmc202505100
4. Kuzochkina, A., Shirokopetleva, M., Dudar, Z., (2018), "Analyzing and Comparison of NoSQL DBMS", *International Scientific-Practical Conference Problems of Infocommunications. Science and Technology* (PIC S&T), P. 560–564. DOI: https://doi.org/10.1109/INFOCOMMST.2018.8632133
5. Mhedhbi, A., Deshpande, A. and Salihoğlu, S. (2024), "Modern Techniques for Querying Graph-structured Databases", *Foundations and Trends® in Databases*, No.14(2), P. 72–185. DOI: https://doi.org/10.1561/1900000090
6. DB-Engines Ranking, available at: https://db-engines.com/en/ranking (last accessed 25.12.2024)
7. Vicknair, C., Macias, M., Zhao, Z., Nan, X., Chen, Y. and Wilkins, D. (2010), "A comparison of a graph database and a relational database: a data provenance perspective", *Proceedings of the 48th Annual Southeast Regional Conference*, ACM, P.1–6. DOI: https://doi.org/10.1145/1900008.1900067
8. Bonifati, A., Özsu, M. T., Tian, Y., Voigt, H., Yu, W. and Zhang, W. (2024), "The Future of Graph Analytics", *Companion of the 2024 International Conference on Management of Data* (SIGMOD '24), P. 544–545. DOI: https://doi.org/10.1145/3626246.3658369
9. Neo4j, Inc. How to create a graph data model, available at: https://neo4j.com/docs/model/ (last accessed 01.04.2025).
10. De Virgilio, R., Maccioni, A. and Torlone, R. (2014), "Model-driven design of graph databases", *International Conference on Conceptual Modeling*, Springer, P.172–185. DOI: https://doi.org/10.1007/978-3-319-12206-9_14
11. Ünal, Y. and Oğuztüzün, H. (2018), "Migration of Data from Relational Database to Graph Database", *Proceedings of the 2018 International Conference on Information Systems and Computer Science*, P. 1–5. DOI: https://doi.org/10.1145/3200842.3200852
12. Mazurova, O., Syvolovskyi, I. and Syvolovska, O. (2022), "NOSQL database logic design methods for MONGODB and NEO4J", *Innovative technologies and scientific solutions for industries*, No. 2(20), P. 52–63. DOI: https://doi.org/10.30837/ITSSI.2022.20.052
13. Nan, Z. and Bai, X. (2019), "The study on data migration from relational database to graph database", *Journal of Physics: Conference Series*, No. 1345(2), P. 022061. DOI: https://doi.org/10.1088/1742-6596/1345/2/022061
14. Zhao, Z., Liu, W. and French, T. (2024), "Rel2Graph: Automated Mapping From Relational Databases to a Unified Property Knowledge Graph", *Research Square* [Preprint]. DOI: https://doi.org/10.21203/rs.3.rs-5451592/v1
15. Pokorný, J. (2015), "Graph Databases: Their Power and Limitations", *IFIP International Conference on Computer Information Systems and Industrial Management*, P. 58–69. DOI: https://doi.org/10.1007/978-3-319-24369-6_5
16. Jouili, S. and Vansteenberghe, V. (2013), "An empirical comparison of graph databases", *Proceedings of the 2013 International Conference on Social Computing*, IEEE, P.708–715. DOI: https://doi.org/10.1109/SocialCom.2013.106
17. Neo4j, Inc. The Neo4j Cypher Manual, available at: https://neo4j.com/docs/cypher-manual/current/ (last accessed 08.04.2025).

Надійшла (Received) 19.04.2025

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Відомості про авторів / About the Authors

Михневич Дмитро Костянтинович – магістр, Харківський національний університет радіоелектроніки, спеціальність 121 – Інженерія програмного забезпечення, Харків, Україна; email: dmytro.mykhnevych@nure.ua; ORCID ID: https://orcid.org/0009-0001-4854-6526

Мазурова Оксана Олексіївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, Харків, Україна; email: oksana.mazurova@nure.ua; ORCID ID: https://orcid.org/0000-0003-3715-3476

Перепичай Олег Іванович – аспірант, Харківський національний університет радіоелектроніки, спеціальність 121 – Інженерія програмного забезпечення, Харків, Україна; email: oleh.perepychai@nure.ua; ORCID ID: https://orcid.org/0009-0007-6931-7769

Mykhnevych Dmytro – Master, Kharkiv National University of Radio Electronics, Master of Specialty 121 – Software Engineering, Kharkiv, Ukraine.

Mazurova Oksana – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Software Engineering, Kharkiv, Ukraine.

Perepichai Oleg – Graduate Student, Kharkiv National University of Radio Electronics, Graduate Student of Specialty 121 – Software Engineering, Kharkiv, Ukraine.

METHOD GRAPHMIGR8 FOR MIGRATING RELATIONAL DATA TO THE NEO4J GRAPH MODEL

In the modern world of data processing, graph databases are becoming increasingly relevant, allowing for efficient modeling and processing of complex relationships between entities in subject domains. Today, relational databases remain the foundation of most existing software systems. However, relational databases are not always the best solution for software systems that face stringent requirements for scalability and high data availability. In the direction of improving the performance of such systems, decisions regarding migration of their relational databases to NoSQL, particularly to graph databases, are increasingly being made in practice. **The subject matter** of the article is methods for migrating structured data from the relational database model to the graph model. **The goal** of the work is to improve the efficiency of migrating relational databases to graph databases by developing a productive method of migration adapted for the Neo4j database management system and providing recommendations for the effective use of migration methods to graph databases. The work addresses the following **tasks**: development of logical models for relational and graph databases Neo4j in various subject areas for conducting experiments on migration based on them; development of a method for migrating relational data to the Neo4j graph model; planning and conducting experimental research on the effectiveness of the proposed method compared to other migration methods, and development of recommendations regarding the peculiarities of their use. The following **methods** are used: database design methods; migration methods to graph databases; methods for experimental evaluation of database performance; development methods based on MS SQL Server 18 and Neo4j 5.26 database management systems, Visual Studio 2022 development environment. The following **results** were obtained: the GraphMigr8 method for migrating relational data to the Neo4j graph model was proposed; experimental evaluation of the quality of migration methods according to performance metrics and semantic data integrity was conducted; recommendations for using migration methods were formed. **Conclusions:** strengths and weaknesses of existing methods for migrating relational data to the graph data model were identified, the GraphMigr8 method for migrating relational databases to the Neo4j graph model was proposed, which showed better performance and higher semantic correspondence of transformed data.

Keywords: database; graph model; migration method; relational model; DBMS; Neo4j.

Бібліографічні описи / Bibliographic descriptions

Михневич Д.К., Мазурова О.О., Перепичай О.І. Метод GRAPHMIGR8 міграції реляційних даних у графову модель NEO4J. *Сучасний стан наукових досліджень та технологій в промисловості.* 2025. № 3 (33). С. 45–57. DOI: https://doi.org/10.30837/2522-9818.2025.3.045

Mykhnevych, D., Mazurova, O., Perepichai, O. (2025), "Method GRAPHMIGR8 for migrating relational data to the NEO4J graph model", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 45–57. DOI: https://doi.org/10.30837/2522-9818.2025.3.045

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

УДК 004.85 DOI: https://doi.org/10.30837/2522-9818.2025.3.058

А. НОВАКОВСЬКИЙ, І. ЯЛОВЕГА

ВИЗНАЧЕННЯ РОЛЕЙ АГЕНТІВ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ (LLM) У МОДЕЛІ ДИЗАЙН-ПРОЦЕСУ

Предметом дослідження є здатності та обмеження, які демонструють великі мовні моделі (LLM) при їх впровадженні в інтелектуальні, технічні та творчі процеси, зокрема в дизайн. **Мета** роботи – визначити місце і ролі агентів великих мовних моделей в дизайні та розробити відповідну модель дизайн-процесу, доповненого LLM-агентами. У статті ставляться наступні **завдання**: (1) провести огляд сучасних публікацій щодо підходів та методів оцінки здатностей великих мовних моделей, зокрема, при виконанні творчих, технічних та дизайнерських задач; (2) провести аналіз сучасних підходів та методів взаємодії з LLM; (3) розробити модель, що визначає місце та ролі LLM-агентів в дизайні у взаємодії з дизайн-командою, зовнішнім середовищем та дизайн-артефактами. Під час проведення дослідження використані такі **методи**: порівняльно-історичний та ретроспективний аналіз змісту технічних, економічних, філософських, лінгвістичних наукових та методичних досліджень для формування цілісного бачення поточного стану розвитку великих мовних моделей та підходів до взаємодії з ними; структурно-логічний аналіз для формування моделі дизайн-процесу, доповненого LLM-агентами. Досягнуто наступні **результати**: визначено дев'ять основних груп здатностей великих мовних моделей; визначені основні сучасні патерни взаємодії з великими мовними моделями; розроблено модель дизайну як ітеративного процесу збагачення знань, що дає змогу описати взаємодію людей та агентів великих мовних моделей між собою, з зовнішнім середовищем та з дизайн артефактами. У **висновках** підкреслена новизна таких здатностей LLM в родині технологій штучного інтелекту, як розуміння та генерація тексту природною мовою та мовами програмування, багатомовність, володіння загальними та галузевими знанням, здатність до міркування, агентність. Додатково актуалізовано необхідність концептуального осмислення місця та ролі великих мовних моделей в творчих процесах. Розроблена структурна модель, що представляє дизайн як ітеративний процес збагачення знань та дозволяє визначити ролі агентів великих мовних моделей у взаємодії з дизайн-командою, артефактами та зовнішнім середовищем.

Ключові слова: великі мовні моделі; LLM агенти; штучний інтелект; людино-комп'ютерна взаємодія; інновації в дизайн.

Вступ

Технології штучного інтелекту (ШІ) суттєво впливають на економіку, наукові дослідження, освіту та інші галузі. За даними звіту про індекс штучного інтелекту за 2024 рік (AI Index Report 2024), що випускається раз на рік спеціалізованим департаментом університету Стенфорду, станом на квітень 2024 року такі генеративні моделі, як GPT-4, Gemini та Claude 3, вже перевершували рівень людських показників за численними бенчмарками (benchmarks) у виконанні інтелектуальних завдань низької та середньої складності [1], хоча варто зауважити, що "рівень людських показників" є умовним та визначається специфікою кожного з бенчмарків.

У вересні 2024 року компанія OpenAI представила нову модель o1, яка розроблялась з фокусом на виконання задач, що потребують міркування (reasoning) та демонструє значний прогрес у розв'язанні складних інтелектуальних завдань, таких як задачі олімпіадного рівня з математики (AIME2024) та програмування (Codeforces), наукові питання рівня PhD з різних дисциплін (GPQA Dimond) [2]. В грудні 2024 року компанія OpenAI анонсувала наступну версію з моделей цього сімейства – o3 [3]. o1 та o3 значно перевершують показники попередніх моделей в числених завданнях, що потребують комплексного міркування, включаючи задачі з програмування, логіки, квантової фізики. Хоча, як демонструє звіт про індекс штучного інтелекту за 2025 рік, здатності штучного інтелекту продовжили стрімкий розвиток, моделі все ще демонструють не сталі результати в задачах, що потребують складного міркування [4].

Наукова спільнота та комерційні організації приділяють сьогодні значну увагу дослідженню генеративного штучного інтелекту, зокрема великих мовних моделей, та впровадженню цієї технології в дослідницькі, творчі, інноваційні, виробничі та управлінські процеси. Технологія вже знайшла своє практичне застосування в низці індустрій

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

в різних регіонах світу та демонструє значний потенціал. Згідно опитування, проведеного міжнародною консалтинговою компанією McKinsey & Company, в 2024 році 78% організацій використовують ШІ хоча в одній з бізнес-функцій [4]. Обсяг приватних інвестицій в генеративний штучний інтелект збільшився в 9 разів в 2023 році в порівняння з 2022 до 25 млрд. доларів [1] та досяг рівня в 34 млрд. доларів в 2024 році [4].

Сьогодні ще неможливо однозначно спрогнозувати масштаби змін, які очікують на суспільство під впливом розвитку ШІ. Існує ймовірність, що технологія зіткнеться з внутрішніми чи зовнішніми обмеженнями, і її вплив залишиться помірним. Однак, є очевидним, що нові генеративні моделі, зокрема великі мовні моделі (LLM), відкривають принципово нові можливості для впровадження ШІ в інтелектуальну та креативну працю, і технологія продовжує стрімко розвиватися.

Аналіз останніх досліджень і публікацій

Методологія оцінки здатностей штучного інтелекту є складним міждисциплінарним напрямом, що динамічно розвивається. Дослідники створюють нові та збагачують вже існуючі бенчмарки (benchmark) – "стандартизовані тести або набори задач, що використовується для оцінювання та порівняння продуктивності, ефективності чи якості різних систем, алгоритмів або моделей" [5].

У контексті штучного інтелекту бенчмарки зазвичай складаються з трьох компонентів [6]:

1) *завдання* (специфікація конкретної проблеми), виконання якого оцінюється;

2) *набори даних*, призначені для виконання завдання та перевірки результатів;

3) *метрики* – показники, які дозволяють кількісно охарактеризувати результати роботи моделі при виконанні завдання.

За даними AI Index Report 2024, у дев'яти бенчмарках показники провідних генеративних моделей ШІ перевищили або впритул наблизилися до показників виконання цих завдань людиною, хоча варто зауважити, що "людський рівень" є умовним та визначається специфікою кожного з бенчмарків:

– класифікація зображень (ImageNet Top-5);

– використання здорового глузду при аргументації на основі зображень (Visual Commonsense Reasoning, VCR);

– виявлення логічних зв'язків між фразами (aNLI);

– розуміння англійської мови (SuperGLUE);

– розуміння тексту базового рівня (SQuAD 1.1);

– розуміння тексту середнього рівня складності (SQuAD 2.0);

– багатозадачне розуміння мови (MMLU);

– відповіді на запитання за зображеннями (Visual Question Answering, VQA);

– математичні задачі олімпіадного рівня (MATH).

Таким чином, існуючі бенчмарки все частіше стають недостатньо інформативними, оскільки результати сучасних генеративних моделей наближаються до їх граничних значень чи перевищують середні показники, досягнуті людьми. Протягом 2023 року з'явилася низка нових, більш складних бенчмарків для оцінки певних здатностей ШІ, серед яких:

– SWE-bench – виправлення помилок в коді;

– HEIM – генерація зображень з дотриманням складних інструкцій;

– MMMU – мультимодальна логічна аргументація;

– MoCa – аргументація в моральній площині;

– AgentBench – агентність;

– HaluEval – розпізнавання галюцинацій та генерація фактологічно коректного тексту.

Важливим трендом є те, що методи оцінки моделей ШІ почали зміщуватися від автоматизованих бенчмарків, на кшталт ImageNet або SQuAD, в бік рейтингів на основі людських оцінок, таких як Chatbot Arena Leaderboard [1]. Цей тренд відображає зростаючу складність здатностей генеративних моделей, що оцінюються, а також важливість врахування етичної складової, точок зору різноманітних стейкхолдерів та громадської думки в оцінюванні прогресу ШІ.

Описаний вище тренд швидкого вичерпання інформативності бенчмарків, зберігся та навіть посилився протягом 2024 року. Великі мовні моделі суттєво покращили свої результати за бенчмарками, що з'явились в 2023 році (MMMU, GPQA, SWE-bench) – наприклад, результати провідних моделей покращились з 4.4% в 2023 році до 71.7% в 2024 за бенчмарком SWE-bench. Протягом року дослідники запропонували низку більш складних та комплексних бенчмарків (Humanity's Last Exam, FrontierMath, BigCodeBench) та нові підходи до розробки бенчмарків (BenchBuilder, WildBench) [4].

Здатності та поточні обмеження великих мовних моделей (LLM). До останнього часу

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

розвиток штучного інтелекту в галузі обробки природної мови (Natural Language Processing, NLP) зосереджувався на створенні вузькоспеціалізованих моделей для виконання конкретних завдань, таких як переклад, аналіз емоційного забарвлення тексту (sentiment analysis) чи визначення наміру користувача (intent detection) [7].

Поява великих мовних моделей (LLM), таких як моделі сімейства GPT, призвела до якісного стрибка в здатностях штучного інтелекту в галузі обробки природної мови. Основним завданням, для виконання якого були розроблені великі мовні моделі на основі архітектури трансформерів, є прогнозування найбільш ймовірного наступного фрагменту (токена) у тексті природною мовою. Крім досягнення визначного рівня у виконанні цього завдання (генерації тексту природною мовою), LLM демонструють широкий спектр додаткових емерджентних здібностей, які виходять за межі їхньої початкової мети [6, 8].

В своєму дослідженні Чанг та коллеги [7] виділяють декілька груп здатностей великих мовних моделей, дослідження яких є актуальним та активним сьогодні.

1) *Розуміння тексту природною мовою (Natural language understanding, NLU)* – ця група включає такі здатності великих мовних моделей як аналіз емоційного забарвлення, класифікація тексту, виявлення логічних зв'язків між фразами, семантичний аналіз, розуміння соціального контексту.

2) *Генерація тексту природною мовою (Natural language generation, NLG)* – до цієї групи входять такі здатності LLM як генерація підсумків, ведення діалогів, генерація відповідей на питання, робота зі стилістикою та форматуванням речень.

3) *Міркування (Reasoning)* є емерджентною здатністю великих мовних моделей, яка ще не є сталою та знаходиться в процесі розвитку. Сьогодні LLM до певної міри можуть розпізнавати логічні зв'язки в тексті, робити логічно коректні висновки та поєднувати їх в ланцюжки послідовних аргументів. Здебільшого здатність великих мовних моделей до міркування оцінюють в чотирьох площинах: математичне міркування, міркування здорового глузду, логічне міркування та галузево-специфічне міркування. Цікавим окремим випадком міркування є здатність великих мовних моделей до планування – вони досить стало можуть створювати та виконувати прості інструкції, але показники суттєво знижуються при виконанні складного, багатоетапного планування.

4) *Багатомовні завдання (Multilingual tasks)* – ця група включає в себе переклад та інші здатності, пов'язані з оперуванням багатьма мовами, наприклад, крос-мовна класифікація чи порівняльний аналіз понять у різних мовах. Дослідження показують, що LLM добре працюють з англійською та іншими поширеними мовами, але продуктивність може знижуватися при використанні більш рідкісних мов.

5) *Фактологічність (Factuality)* відображає здатність LLM генерувати твердження, що відповідають реальним фактам. Значна увага в цій області приділяється дослідженню галюцинування великих мовних моделей – випадків, коли модель вигадує факти, що не відповідають реальності.

6) *Робастність, етичність, неупередженість та довірчість (Robustness, Ethics, Bias, Trustworthiness)* відображає стійкість моделі до несподіваних або шкідливих вхідних даних, відсутність токсичності та упередженості та відповідність етичним нормам. Наприклад, дослідження показують, що LLM можуть засвоювати стереотипи, що містяться в навчальних даних, або генерувати шкідливий контент.

7) *Галузеві знання (Social and Natural Science, Engineering, Medicine)* є необхідною передумовою для вирішення великими мовними моделями специфічних задач: соціальні знання необхідні, наприклад для аналізу політичної риторики, оцінки емоцій або психологічних тестів; природничі, математичні та інженерні знання – для розв'язання, зокрема математичних задач чи задач з генерації або перевірки програмного коду; медичні знання – для аналізу клінічних кейсів.

8) *Агентність (Agent Applications)* виражається в здатності моделей інтерпретувати дані, отримані з зовнішнього середовища, приймати узгоджені рішення, планувати, оперувати інструментами (наприклад, через API або веб-інтерфейс), взаємодіяти з іншими інформаційними системами та моделями ШІ.

9) *Інші здатності* моделей можуть включати такі вузькоспеціалізовані сценарії, як виконання різноманітних завдань в галузі освіти (наприклад, автоматичне оцінювання студентських робіт), пошук й рекомендації, тестування особистісних характеристик людини тощо.

Чанг та коллеги роблять висновки, що на момент публікації звіту великі мовні моделі демонструють стабільно високі результати у таких задачах як

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

розуміння природної мови, генерація зв'язного та точного тексту з урахуванням контексту, переклад. Сталими є також такі здатності окремих LLM, як арифметичне та логічне міркування, а також міркування, що потребує розуміння часових залежностей. Окремими перспективними напрямками є математичне міркування та міркування на основі структурованих даних.

В той же час великі мовні моделі здебільшого демонструють обмежені та несталі результати в таких задачах, як абстрактне міркування, розуміння деталей складного контексту, визначення семантичної подібності, виявлення логічних зв'язків. Продуктивність LLM може суттєво знижуватися при роботі з рідкісними мовами, вузько-специфічними сценаріями, візуально-модальними завданнями та при необхідності враховувати людські протиріччя. Обмеженнями великих мовних моделей є також схильність до галюцинування, формування упереджень та небажаної поведінки під час навчання, токсичність, вразливість до несподіваних або шкідливих вхідних даних та неспроможність швидко оновлювати знання в реальному часі.

Більшість дослідників сходиться у висновках, що вже на поточному етапі розвитку великі мовні моделі можуть значно допомогти людині у виконанні інтелектуальних та творчих завдань, які потребують обробки мультимодальної інформації [9,10]. Разом з тим сьогодні продовжуються дискусії щодо креативних здатностей великих мовних моделей – наскільки LLM здатні генерувати принципово новий контент та якими бенчмарками це можна виміряти кількісно [11,12]. Крім того, концептуальні моделі взаємодії людини з великими мовними моделями все ще знаходяться на стадії формування та потребують додаткового осмислення та досліджень [13,14].

Великі мовні моделі в дизайні та креативних задачах. Застосування генеративного ШІ у дизайні має численні переваги, проте постають і суттєві обмеження [10]. Такі інструменти, як ChatGPT та DALL·E, здатні швидко генерувати значну кількість альтернативних ідей "без втоми", що притаманна людині, та вільно оперувати загальновідомими концептами. Водночас результати роботи цих моделей не завжди є точними та достовірними: інколи вони пропонують недостовірні факти, а якість створюваних рішень, як правило, не перевершує рівень досвідчених фахівців. Крім того, здатність генеративного ШІ розв'язувати складні аналітичні задачі все ще обмежена, а процес взаємодії вимагає спеціалізованих навичок і додаткових зусиль, що може ускладнювати творчий потік і знижувати рівень занурення дизайнера в проблему. Додаткові труднощі виникають з передачею контексту (через обмеження контекстного вікна моделі та складність чіткого формулювання сутнісних нюансів), а також із ризиком звуження бачення дизайнера через надто детальні або нав'язані системою рішення.

Дослідження Чжібін Чжоу [9] за участю 30 дизайнерів показало, що (1) застосування LLM (наприклад, ChatGPT) може прискорити процес концептуального дизайну без суттєвої втрати якості результату; (2) хоча дизайнери схильні вважати LLM менш ефективним та надійним дизайн-партнером ніж людину, проте розглядають їх як потенційного помічника під своїм керівництвом.

Одним з напрямків кількісних досліджень креативних здатностей великих мовних моделей є адаптації тесту Торренса, що є визнаним інструментом для оцінки креативного мислення людини [11, 12]. Класична версія тесту оцінює рівень креативність за чотирма параметрами: (1) продуктивність (Fluency) – кількість ідей або рішень, які здатний згенерувати учасник; (2) гнучкість (Flexibility) – різноманітність ідей, варіативність підходів та здатність переключатися між різними концепціями; (3) оригінальність (Originality) – рівень новизни та унікальності запропонованих ідей; (4) розробленість (Elaboration) – ступінь деталізації та опрацьованості ідей, здатність розвинути їх до завершеної концепції.

Базуючись на тесті Торренса, Жао та коллеги [12] пропонують адаптовану методику для оцінки рівня креативності LLM, виділяючи сім груп креативних завдань: незвичайне застосування; наслідки; а що коли; ситуаційне завдання; звична проблема; поліпшення; уявні історії.

За допомогою GPT-4 дослідники згенерували набір із 700 задач (по 100 в кожній з груп), що використали для оцінки креативності шести LLM моделей за чотирма критеріями з оригінального тесту Торренса: продуктивність, гнучкість, оригінальність, розробленість. Дослідники прийшли до висновків, що великі мовні моделі демонструють найвищі результати за критерієм розробленості (elaboration) та найнижчі – за критерієм оригінальності (originality). Важливо зазначити, що результати виконання задач великими мовними моделями оцінювались

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

за допомогою GPT-4, що ставить під деякий сумнів отримані результати.

Також базуючись на Тесті креативного мислення Торренса (ТКМТ), Чакрабарті, Тухін та колеги [11] пропонують Тест креативного письма Торренса (ТКПТ). Дослідники залучили десять креативних авторів-експертів, за допомогою яких провели оцінку 48 історій, створених або професійними письменниками, або LLM. Дослідження демонструє, що історії, згенеровані LLM, в 3-10 разів рідше успішно проходять тест ТКПТ, ніж тексти, написані професіоналами. Крім цього, дослідники оцінили, наскільки ефективною може бути оцінка текстів за ТКПТ за допомогою LLM – виявилось, що результати оцінки жодної з моделей LLM не корелюють з експертними оцінками.

Розвиток креативних здатностей великих мовних моделей спонукає до концептуального переосмислення ролі ШІ у творчій діяльності та відкриває нові питання про співіснування та співтворчість людини і машини. У практичній площині машина стає здатною не тільки виконувати інтелектуальні завдання, які до цього вважалися "виключно людськими", але й набуває певного рівня агентності – стає здатною сприймати та аналізувати середовище, планувати та виконувати дії. З погляду етики це ставить нові питання щодо майбутнього співіснування та співтворчості людини й машини. Виникають та розвиваються такі концепти, як постантропоцентричний дизайн, співтворчість (людини та машини), агентність машини тощо. Ставляться питання пошуку синергії між сильними аспектами людського та машинного інтелекту для створення майбутнього дизайну та інновацій [13].

В роботі [10] було запропоновано структурну модель дизайн-мислення (рис. 1), яка дає змогу до певної міри декомпозувати складний творчий процес на нелінійну ітеративну послідовність кроків, на кожному з яких виконуються специфічні техніки. Результати кожного з кроків зберігаються у вигляді відповідних артефактів, збагачуючи обсяг і якість накопичених знань щодо проблеми та ідей її розв'язання. Розроблена структурна модель закладає основу для системного впровадження технологій генеративного штучного інтелекту в процеси дизайну.

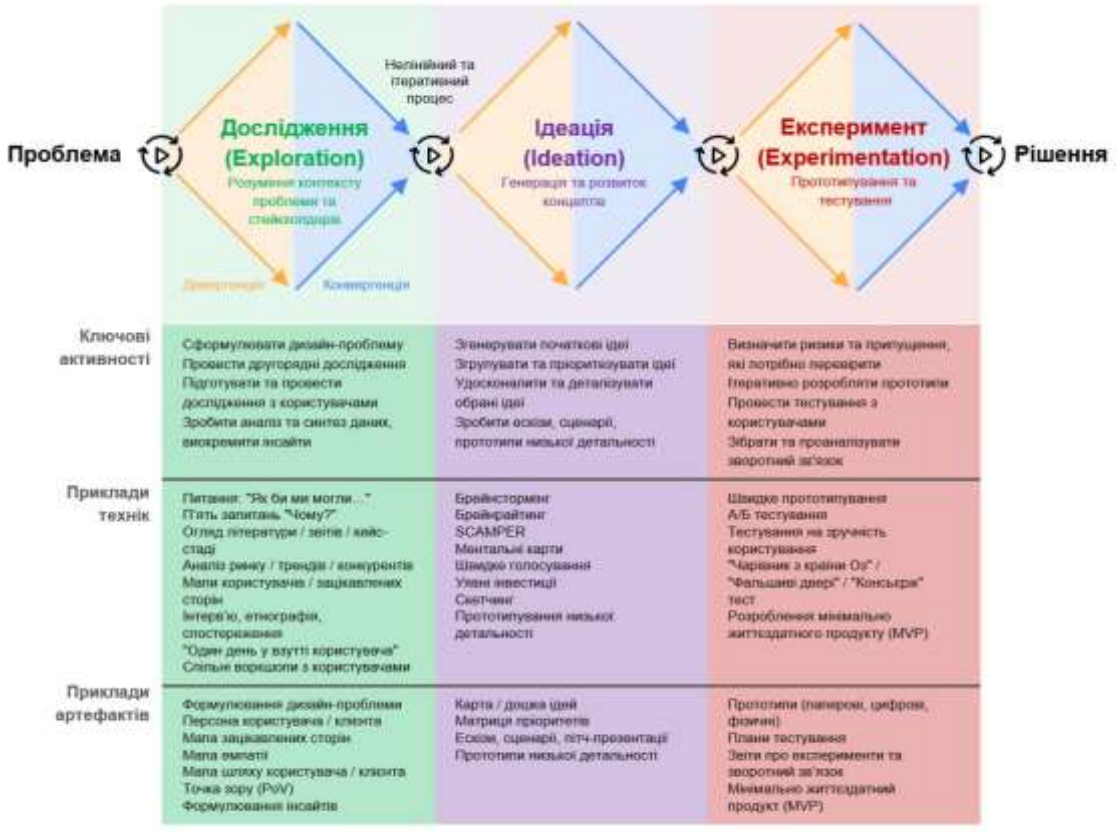


Рис. 1. Структурна модель процесу дизайн-мислення

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Катя Торинг зі співавторами [14] представляють концептуальний підхід до осмислення місця генеративних моделей в дизайні, базуючись на концепціях *дизайн-знання (design knowledge)* та *еволюційної креативності (evolutionary creativity)*.

Модель дизайн-знання складається з чотирьох рівнів: (A) знання втілені в артефактах, (B) неявні знання на рівні дизайн-інтуїції, (C) явні знання виражені мовою дизайну (D) теорії дизайну. Дизайнери взаємодіють з дизайн-знаннями на кожному з рівнів та створюють нові знання на певному рівні, базуючись на знаннях з інших рівнів – наприклад, екстерналізують елементи дизайн інтуїції у вигляді дизайн-технік чи осмислюють нові концепції та розробляють на основі них нові теорії дизайну. Катя Торинг зі співавторами приходять до висновку, що штучний інтелект, зокрема великі мовні та мультимодальні моделі, можуть взаємодіяти з дизайн-знанням та створювати нове на перших трьох рівнях: (A) взаємодіяти з фізичними та цифровими артефактами (ескізами, текстами, прототипами) за допомогою, наприклад, комп'ютерного зору чи керуючи 3Д принтером; (B) зберігати в неявному вигляді в параметрах нейронної мережі дизайн-патерни чи зв'язки між дизайн-концепціями; (C) створювати, адаптувати та виконувати явно сформульовані інструкції. Разом з тим дослідники вважають, що станом на сьогодні ШІ не здатний збагачувати четвертий рівень дизайн знання (D), а саме формувати нові теорії дизайну.

Згідно з концепцією еволюційної креативності дизайн-процесс складається з циклів варіацій – створення численних варіантів через мутацію та рекомбінацію ідей і відбору найпридатніших рішень з урахуванням контексту та вимог. Катя Торинг зі співавторами вважають, що у співпраці генеративних моделей та людини ШІ має спеціалізуватись на варіаціях – швидко генерувати та деталізувати нові концепції, а дизайнери – на оцінці та відборі.

Підсумовуючи зазначимо, що дослідники здебільшого сходяться в якісних висновках – в тому, що взаємодія людини та ШІ може суттєво прискорити процес дизайну в першу чергу за рахунок здатності ШІ швидко генерувати та детально пропрацьовувати значну кількість ідей та концепцій. Разом з тим концептуальні моделі такої взаємодії та кількісні бенчмарки для оцінки креативних здатностей ШІ все ще знаходяться на стадії формування та потребують додаткового осмислення та дослідження.

Мета роботи, основні завдання та методи

Мета роботи – визначити місце і ролі агентів великих мовних моделей в дизайні та розробити відповідну модель дизайн-процесу, доповненого LLM-агентами.

Для досягнення поставленої мети виконані наступні завдання:

– проведено огляд сучасних публікацій щодо підходів та методів оцінки здатностей великих мовних моделей, зокрема, при виконанні творчих та дизайнерських задач;

– проведено аналіз сучасних підходів та методів взаємодії з великими мовними моделями, зокрема в контексті дизайну;

– розроблено структурну модель, що визначає місце та ролі штучних агентів в дизайні у взаємодії з дизайн-командою, зовнішнім середовищем та дизайн-артефактами.

Використані наступні методи:

– порівняльно-історичний та ретроспективний аналіз змісту технічних, економічних, філософських, лінгвістичних наукових та методичних досліджень для формування цілісного бачення поточного стану розвитку великих мовних моделей та підходів до взаємодії з технологією;

– структурно-логічний аналіз для формування структурної моделі дизайн-процесу, доповненого агентами великих мовних моделей.

Результати досліджень та їх обговорення

Серед основних підходів до взаємодії з великими мовними моделями можна виділити такі.

Чат-інтерфейс. Чат є найбільш розповсюдженим (але не єдиним) інтерфейсом взаємодії людини з технологіями генеративного штучного інтелекту. Великі мовні та мультимодальні моделі демонструють вражаючі здатності до сприйняття інструкцій природною мовою з урахуванням значного за обсягом контексту, що дозволяє реалізовувати складні комунікаційні та обчислювальні сценарії. Зазвичай, користувач формулює запит до моделі у вигляді промпту (prompt) – повідомлення з інструкціями, прикладами, деталями контексту (рис. 2) та у відповідь отримує результат виконання (completion).

У сучасній практиці промптингу зазвичай виділяють кілька типових складових, включення

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

яких у запит допомагає генеративним моделям штучного інтелекту точніше та повніше виконати задачу з урахуванням очікувань.

– *Інструкція* є основною частиною промпту, та містить формулювання задачі, яку необхідно виконати. Інструкція може не бути явно прописана, а випливати з наведених прикладів.

– *Приклади* є поширеною складовою промптів оскільки в більшості випадків інструкції з кількома прикладами (few-shot prompts) допомагають значно покращити результати роботи моделі.

– *Контекст* містить додаткову змістовну інформацію щодо обставин, що зумовлюють виконання задачі – це допомагає отримати результат з урахуванням нюансів середовища, в якому, наприклад, планується використовувати результати.

– *Обмеження* включаються в промпт з метою звуження діапазону можливих відповідей моделі.

– *Мета* допомагає спрямувати генерацію специфічним чином без надання детальних інструкцій чи обмежень.

– *Роль* може бути включена в промпт для визначення того, від імені кого має діяти модель, здійснюючи генерацію. Це дає можливість скеровувати стиль, тон, логіку відповіді відповідно до очікуваної ролі – наприклад, експерта, консультанта чи викладача.

– *Вимоги щодо форматування* включаються в промпт, якщо необхідно отримати результат в специфічному вигляді – це можуть бути (1) вимоги щодо форматування звичайного тексту – наприклад, бажана кількість абзаців чи наявність/відсутність списків, або (2) вимоги щодо подання відповіді в специфічному форматі – наприклад, у вигляді програмного коду або в JSON форматі.

– *Елементи форматування промту*: хорошою практикою є використання розмітки – від простих розділювачів (кавиць, списків) до більш складних, таких як мови розмітки Markdown чи HTML – це допомагає моделі краще розпізнати структуру та змістовну ієрархію запиту.

Звернемо увагу, що взаємодія з LLM може відбуватись, як безпосередньо через чат, де споживачем результатів роботи LLM зазвичай є людина (як, наприклад, в таких застосунках, як ChatGPT), так і через API виклики, де споживачем результату роботи LLM зазвичай є інша комп'ютерна система.

Ланцюжки промптів. Реалізація більш складних сценаріїв за допомогою великих мовних моделей може потребувати декомпозиції комплексної задачі на послідовну серію більш простих підзадач. Це зазвичай актуально, коли обсяг вхідних даних перевищує розмір контекстного вікна моделі або коли складність обчислень більша, ніж модель здатна виконати за один виклик. У таких випадках поширеною практикою є використання ланцюжків промптів (prompt chaining), де результати виконання кожного попереднього промпта впливають на наступні. Такий підхід дає змогу розподілити обробку інформації, зберігаючи цілісність сценарію, що в результаті дає більш точне та стале рішення.

На рис. 2 наведено схему перетворення даних LLM-застосунком "Аналізатор коментарів" з використанням ланцюжкового підходу. Послідовність з трьох промптів обробляє масив користувацьких відгуків, а саме: (1) виділяє ключові теми, що присутні в кожному окремому коментарі; (2) групує їх за категоріями; (3) генерує текстовий підсумок за кожною з категорій. Додаток допомагає структурувати інформацію, присутню в коментарях, спрощуючи подальшу роботу дизайнера або продакт менеджера з больовими точками та потребами користувачів сервісу або продукту. Використання такого додатку в дизайн-процесі є доцільним у випадку, коли обсяг коментарів первищує той, який модель здатна обробити, зберігаючи необхідну точність обчислень.

Варто зауважити, що аналізатор коментарів є прикладом додатку, в якому взаємодія користувача та великої мовної моделі відбувається не обов'язково через чат-інтерфейс, а може бути реалізована, наприклад, через завантаження файлу з коментарями через веб-інтерфейс та отримання результату на електронну пошту після опрацювання.

Для роботи з ланцюжками промптів використовують такі спеціальні фреймворки, як LangChain або Semantic Kernel, що допомагають впорядкувати логіку переходів між промптами та зберігати проміжні результати. Крім ланцюжків послідовних запитів, спеціалізовані фреймворки дають змогу будувати й складніші структури для взаємодії з великими мовними моделями, зокрема ієрархічні або розгалужені сценарії. Завдяки цьому можна комбінувати кілька потоків чи рівнів обробки даних в межах одного процесу, що робить взаємодію з LLM більш гнучкою та забезпечує вищу контрольованість кінцевого результату.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

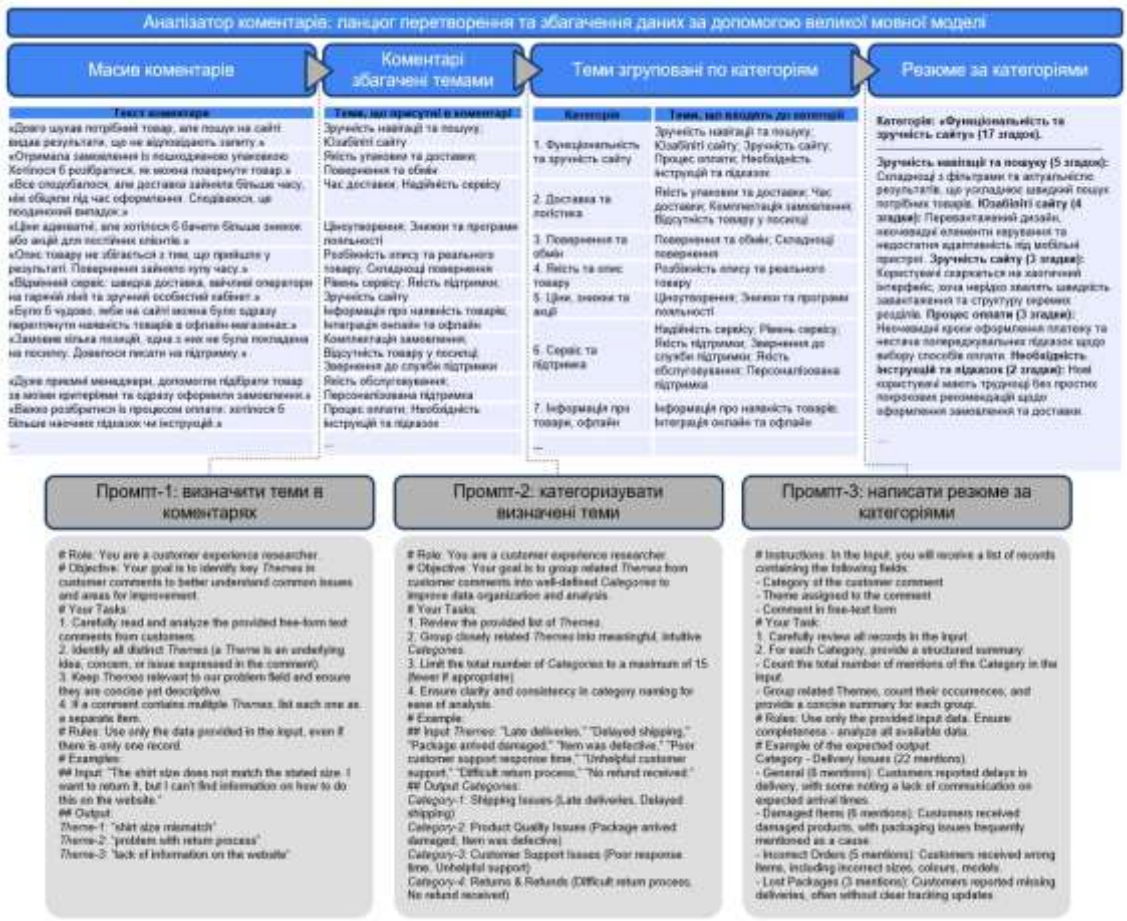


Рис. 2. Схема перетворень даних LLM-додатком "Аналізатор коментарів"

LLM-Агенти. Агенти великих мовних моделей (LLM-Agents) – новий напрямок в розробці програмного забезпечення, що привертає значну увагу з боку наукової та бізнес спільнот [15]. Використання архітектури, в основі якої лежать агенти, дає змогу вирішувати проблеми високої складності, які важко або неможливо подати у вигляді повністю детермінованих алгоритмів. Даний напрямок знаходиться на стадії становлення, осмислення концепцій та формування методологічної бази. Агент зазвичай розглядається як концептуальний, функціональний або архітектурний компонент, який включає механізми для:

– інтерпретації стану середовища та намірів користувача;

– міркування та планування наступних кроків;

– виконання цих кроків – наприклад, генерації відповіді на запит користувача або API викликів до інших агентів та інформаційних систем.

Сі з колегами [15], аналізуючи розвиток концепту від філософських основ до галузі ШІ, дають наступне визначення: "штучний інтелектуальний агент – це штучна сутність, що сприймає навколишнє середовище, приймає рішення та виконує дії". Основні властивості агентів: (1) автономність – здатність діяти та приймати рішення самостійно; (2) реактивність – здатність вчасно реагувати на зміни в оточенні; (3) проактивність – здатність формувати цілі та ініціювати дії; (4) соціальність – здатність взаємодіяти з людьми та іншими агентами, що базується на певній спільній мові.

Дослідники пропонують модель агенту (рис. 3), що складається з трьох основних модулів: сенсорний модуль, модуль "мозок" та модуль дій. Сенсорний модуль можна співставити с органами чуттів людини – таким як очі чи вуха. Він отримує сигнали з зовнішнього середовища (запити від користувачів, кодифіковану інформацію, данні про об’єкти, запити та відповіді від інших агентів та інформаційних систем) та перетворює мультимодальну інформацію у зрозуміле агенту представлення. "Мозок" агенту виступає центром управління: обробляє одержану

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

інформацію, забезпечує міркування, планування, прийняття рішень, керує операціями зі збереження знань, що можуть бути корисними в майбутньому. Модуль дій, що можна співставити з людськими кінцівками, виконує дії за допомогою інструментів та справляє вплив на середовище агенту. Така структура дозволяє агенту взаємодіяти з оточенням та отримувати зворотний зв'язок.

Сьогодні LLM-Агенти вже активно використовуються для вирішення задач в таких галузях, як наукові дослідження, робототехніка, соціальні науки, економіка, медицина, розробка програмного забезпечення [16,17]. Яскравим прикладом є платформа Cursor, де LLM-агент у реальному часі допомагає інженеру-розробнику безпосередньо в інтегрованому середовищі розробки (IDE): генерує код і тести; знаходить помилки, визначає їх причини, пропонує виправлення.

Можна виділити три архітектурні патерни в розробці програмного забезпечення на основі LLM-агентів, що є розповсюдженими сьогодні: один агент, взаємодія між кількома агентами, колаборативна взаємодія між агентом та людиною (рис. 4).

Рис. 3. Три-модульна модель агенту ШІ

Рис. 4. Архітектурні патерни з використанням LLM-агентів

Так, прикладом сценарію, побудованого на базі одного агента, може бути чат-бот «Перекладач технічних специфікацій», що допомагає інженерам-розробникам з розподілених команд працювати зі специфікаціями тією мовою, яка є для них зручнішою – агент робить переклад на визначену розробником мову, враховуючи специфічний словник термінів проекту. Ще одним прикладом простого чат-бота може бути «Генератор юніт-тестів», що отримує фрагмент коду та опис граничних випадків і на цій основі генерує юніт-тести.

Взаємодія трьох агентів може бути проілюстрована у застосунку «Асистент спостереження за серверними журналами», що отримує в якості вхідних даних архів логів сервера за останні 24 години. *Агент-категоризатор* виявляє похибки та аномалії в сирих даних з сервера та кластеризує схожі проблеми. *Агент-аналізатор* формує гіпотези щодо можливих причин для кожного з кластерів проблем, аналізуючи кодову базу, конфігураційні файли тощо. *Агент-пріоритезатор* визначає рівень критичності кожної проблеми, створює тікети в застосунку Jira для високопріоритетних інцидентів і сповіщає чергового інженера. У цьому сценарії кожен агент працює за власними інструкціями та може надсилати запити іншим агентам. Наприклад, Агент-пріоритезатор може повернути гіпотезу Агенту-аналізатору, якщо обґрунтування недостатнє для визначення пріоритету.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Прикладом, що ілюструє взаємодію агента з людиною, є застосунок «Асистент документування кодової бази», в якому LLM-агент працює під наглядом інженера-розробника – сканує кодову базу проєкту й підтримує коментарі та проєктну документацію в актуальному й повному стані. Агент виконує три основні завдання: (1) додає відсутні аннотації в коді до кожної функції, класу тощо; (2) переписує наявні коментарі, щоб вони відповідали корпоративному стилю та були узгоджені з рештою кодової бази; (3) оновлює та впорядковує проектну документацію. Після кожної операції агент представляє проміжний результат на розгляд інженеру, який надає зворотний зв'язок: *підтвердити* та перейти до наступного етапу або *повернути* задачу агенту для корекції. Такий ітеративний підхід із залученням людини є ефективним для виконання складних задач, що потребують високого рівня акуратності чи відповідності певним нормам (наприклад, безпековим або етичним).

LLM-агенти вже сьогодні демонструють високий рівень інтегрованості та значну ефективність в задачах з розробки програмного забезпечення: від перекладу технічних специфікацій, генерації юніт-тестів та аналізу логів до написання складного коду, визначення причин помилок та їх виправлення. Враховуючи складну та нелінійну природу дизайн-процесу, LLM-агенти мають великий потенціал для доповнення дизайнерських команд.

Дизайн як ітеративний процес збагачення знань: структурна модель

Розглянемо дизайн як циклічний процес обміну інформацією та артефактами (ідеями, концепціями та прототипами) між дизайн-командою, доповненою агентами великих мовних моделей, і зовнішнім середовищем (рис. 5).

(A): Дизайн-команда отримує інформацію з зовнішнього середовища через активну емпатію; спостереження за користувачами, системою та середовищем; збір наявної кодифікованої інформації, такої як інструкції, журнали, записи чи звіти.

(B): Зібрані дані опрацьовується, збагачуючи індивідуальні знання учасників команди, зокрема їхній досвід, інтуїцію та уявлення, та спільні знання команди, що представлені дизайн-артефактами. Артефакти не лише акумулюють знання, але й слугують основою для узгодження колективної роботи та комунікації всередині команди, зі стейкхолдерами та з LLM-агентами. Процес взаємного збагачення індивідуальних знань учасників команди і артефактів є неперервним, двостороннім і забезпечує основу для ітеративної роботи над вирішенням проблеми.

(C): На етапах, коли артефакти (ідеї, концепції, прототипи) досягають певного рівня зрілості, вони повертаються у зовнішнє середовище для тестування й перевірки їхнього впливу на систему. Такі експерименти дають змогу оцінити, наскільки розроблені рішення відповідають реальним потребам користувачів і як саме вони впливають на вихідну проблему.

(A – початок нового циклу): Результати тестування дають команді додаткову інформацію, яка збагачує розуміння команди про контекст, потреби користувачів і характер проблеми та дозволяє вдосконалювати розроблені рішення.

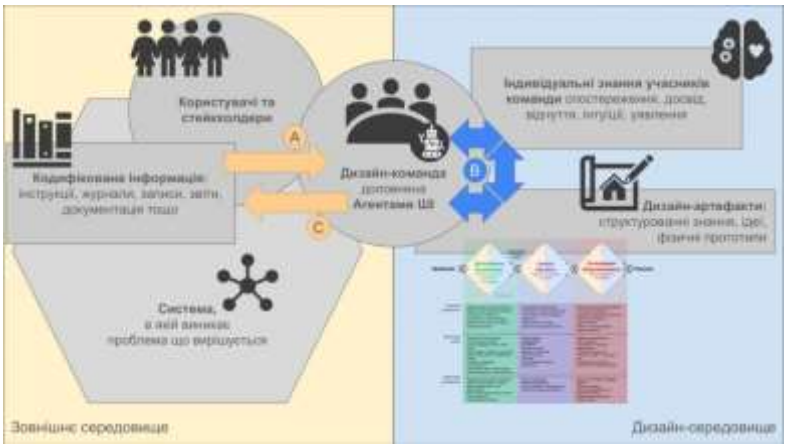


Рис. 5. Модель дизайну як ітеративного процесу збагачення знань

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Таким чином, процес є ітеративним та циклічним: кожен етап збагачує команду новими знаннями, які використовуються для уточнення й вдосконалення рішень. Різні фази (збір даних, аналіз, генерування ідей, побудова прототипів, тестування) часто відбуваються не по черзі, а накладаються одна на одну, забезпечуючи можливість швидко реагувати на нову інформацію та мінливий контекст. Суттєвою рисою такого підходу є постійне перетворення "сирих" даних на осмислену інформацію, а згодом і на знання, котре збагачує розуміння проблеми та можливих шляхів її вирішення.

Важливо зауважити, що у середовищах, де присутній соціокультурний контекст (а не суто технічний чи природний), дизайнерам необхідно збагачувати створювані ними артефакти не лише функціональними характеристиками, а й сюжетним та емоційним наповненням. Проілюструємо це на прикладі розробки корпоративного чат-боту. З функціональної точки зору чат-бот є інструментом для відповіді на поширені питання та виконання нескладних задач за дорученням співробітника. Роль, в якій виступає чат-бот в корпоративному культурному середовищі, в певних межах визначає пов’язаний з ним сюжет, персонаж, тон комунікації та емоційну складову. Наприклад, якщо бот виступає в ролі "помічника співробітників", як персонаж для нього може бути обраний Джин, що був помічником Аладіна у відповідній казці, з відповідними сюжетними, комунікаційними та емоційними особливостями. Таким чином, дизайн виходить за суто функціональні рамки, інтегруючи бота в корпоративне соціокультурне середовище.

Запропонована модель дизайну як трансформації знань містить ключові ідеї п’яти класичних підходів до концептуалізації дизайну [18], кожен з яких розглядає дизайн під певним кутом: (1) створення артефактів (Саймон, 1969); (2) рефлексивна практика (Шен, 1983); (3) діяльність, спрямована на вирішення проблем (Б’юкенен, 1992); (4) спосіб міркування / осмислення речей (Лоусон, 2006; Крос, 2006); (5) створення сенсів (Кріппендорфф, 2006).

Потенційна роль і місце штучних інтелектуальних агентів в процесі дизайну. Розгляд дизайну як процесу ітеративної трансформації знання відкриває новий кут зору для осмислення ролі ШІ агентів у співпраці з дизайн-командою. Вони здатні суттєво прискорити окремі етапи дизайну – від збору й аналізу інформації до генерації й деталізації ідей, проте мають і низку обмежень, які варто враховувати.

1. Взаємодія з об’єктами зовнішнього та дизайн-середовищ. Сучасні великі мовні моделі можуть автономно або напівавтономно збирати різнорідні дані з цифрових джерел і допомагати в їх аналізі. Наприклад, алгоритм може сортувати та класифікувати відгуки користувачів або визначати "вузькі місця" в процесі взаємодії з сайтом. ШІ може створювати нові структуровані інформаційні дизайн-артефакти або інтегрувати знахідки в існуючі (наприклад, узагальнити користувацькі вимоги у форматі user stories, оновити онлайн-документацію чи прототипи у Figma). Здатність систем на основі ШІ безпосередньо взаємодіяти з фізичними об’єктами на даний момент є менш сталою, але розвивається бурхливими темпами в таких галузях, як робототехніка, автономне керування автомобілями, комп’ютерний зір, 3D друк. Взаємодія з фізичними об'єктами в багатьох дизайн-процесах все ще вимагає суттєвої участі людини – наприклад, побудова фізичних прототипів чи тестування у польових умовах.

2. Співпраця з людиною. Чат-інтерфейс, що сьогодні є найбільш поширеним способом взаємодії людини з великими мовними та мультимодальними моделями, дозволяє реалізовувати складні та гнучкі колаборативні сценарії. Водночас такий формат взаємодії може створювати труднощі з передачею повного контексту складних проектів або невисловлених інтуїтивних нюансів, якими часто керується дизайнер. Додатковою суттєвою проблемою є обмежена здатність моделей сприймати та розуміти внутрішній світ людини (почуття, емоції, інтуіції), міжособистісні відносини та соціальний контекст, які є значною складовою в комунікації між людьми та можуть бути критичними, наприклад, в UX-дизайні. Це вимагає від команди відповідального комбінування висновків ШІ з власним аналізом, побудованим на емпатії. Дані обмеження стосуються, як взаємодії ШІ з дизайн-командою так і з користувачами та стейкхолдерами. Варто зауважити, що робота з мовною чи мультимодальною моделлю передбачає наявність специфічних навичок в дизайнерів в формулюванні запитів до моделі (prompt engineering).

3. Автономність. Для того, щоб автономно діяти в складному та динамічному дизайн-процесі важливо, щоб агенти могли (1) орієнтуватись в загальній структурі та зв’язках між елементами зовнішнього та

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

дизайн-середовищ; (2) утримувати загальну картину (зум-аут) та (3) заглиблюватись в конкретні деталі (зум-ин). Великі мовні моделі не завжди стало справляються з такою роботою, проте сучасні ШІ агенти вже здатні генерувати та підтримувати складні інформаційні структури (програмний код, великі текстові документи тощо), а також планувати й виконувати певні дії за допомогою зовнішніх інструментів. Важливо зазначити, що незважаючи на певний рівень автономності ШІ агентів, контроль та відповідальність має залишатись на людині – одним з поширених патернів взаємодії ШІ агентів та людини в складних сценаріях є обов'язкове підтвердження з боку людини перед виконанням агентом значущих дій (наприклад, перед оплатою рахунків, оновленням кодової бази або публікацією результатів), що слугує запобіжником від помилкових кроків ШІ агента.

4. Креативність. Здатність ШІ до оригінальної творчості є дискусійною, однак у низці досліджень відзначається, що моделі здатні генерувати ідеї, котрі дизайнери можуть використати як стартові варіанти для подальшого доопрацювання. Швидке формування великої кількості альтернативних ідей або стилістичних рішень (наприклад, у дизайні інтерфейсів) може значно пришвидшити роботу. Водночас фінальні рішення мають обов'язково оцінюватися з огляду на реальні потреби користувачів, соціальний контекст і якісні критерії.

5. Етичність та безпека. Інтеграція ШІ агентів потребує дотримання принципів приватності даних, запобігання дискримінаційним упередженням, прозорості дій агентів. Якщо вхідні дані містять викривлення або нерепрезентативні вибірки (наприклад, здебільшого від молодих користувачів), модель може "мимоволі" транслювати й підсилювати упередження щодо інших груп. У соціально чутливих проектах (зокрема, в охороні здоров'я, освіті, роботі з людьми з особливими потребами) варто особливо ретельно стежити, аби ШІ не пропонував рішення, що порушують етичні норми або права певних категорій користувачів. Прозорість – це критичний фактор для довіри з боку команди та стейкхолдерів, тому варто чітко розуміти, звідки ШІ отримав дані, як він формував рекомендації чи рішення.

Підсумовуючи викладене вище, зазначимо, що потенціал ШІ агентів у процесі дизайну визначається їх здатністю доповнювати та підсилювати роботу людей, залишаючи останнє слово людині – насамперед у питаннях емпатії, розумінні складних соціокультурних контекстів і відповідальності за прийняті рішення. Ефективність ШІ зростає там, де необхідно швидко обробляти великі обсяги даних або генерувати численні варіанти рішень, у той час як команда критично оцінює та доопрацьовує ці пропозиції. Відповідно, успішна інтеграція ШІ у дизайн-процес вимагає ітеративного залучення моделей на різних етапах, чіткого розподілу ролей, а також спеціальних компетенцій, які дозволяють дизайнерам і стейкхолдерам ефективно співпрацювати з інтелектуальними агентами.

Висновки та напрями майбутніх досліджень

Великі мовні моделі демонструють такі принципово нові здатності в родині технологій штучного інтелекту, як розуміння та генерація тексту природною мовою та мовами програмування, багатомовність, володіння загальними та галузевими знанням, здатність до міркування, агентність. Хоча не всі ці здатності є однаково добре розвинутими та сталими сьогодні, наукові установи та комерційні організації інвестують значний обяг зусиль та ресурсів в розробку наступних поколінь великих мовних моделей, методологій оцінки та вдосконалення підходів у взаємодії з ними. Одним з найбільш перспективних напрямків сьогодні є розробка штучних інтелектуальних LLM-агентів.

Більшість дослідників сходиться у висновках, що вже на поточному етапі розвитку великі мовні моделі можуть значно допомогти людині у виконанні інтелектуальних та творчих завдань, які потребують обробки мультимодальної інформації. Разом з тим сьогодні продовжуються дискусії щодо креативних здатностей великих мовних моделей – наскільки LLM здатні генерувати принципово новий контент та якими бенчмарками це можна виміряти кількісно. Крім того концептуальні моделі взаємодії людини з великими мовними моделями все ще знаходяться на стадії формування та потребують додаткового осмислення та досліджень.

Дана робота пропонує експлоративну модель, що представляє дизайн як ітеративний процес збагачення знань та дозволяє визначити місце та роль інтелектуальних агентів великих мовних моделей у взаємодії з дизайн-командою, артефактами та зовнішнім середовищем. Такий кут зору задає п'ять напрямків для подальших досліджень потенціалу великих мовних моделей в творчих

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

та дизайн процесах: (1) взаємодія LLM-агентів з об'єктами зовнішнього та дизайн-середовищ, (2) співпраця з людиною, (3) автономність, (4) креативність, (5) етичність та безпека.

Список літератури

1. Perrault R., Clark J. Artificial Intelligence Index Report 2024. 2024. URL: https://hai.stanford.edu/ai-index/2024-ai-index-report
2. Learning to Reason with LLMs. *OpenAI*. 2024. URL: https://openai.com/index/learning-to-reason-with-llms/
3. Today, we shared evals for an early version of the next model in our o-model reasoning series: OpenAI o3. *OpenAI Twitter Account*. URL: https://x.com/OpenAI/status/1870186518230511844
4. Perrault R. et al. Artificial Intelligence Index Report 2025. 2025. URL: https://hai.stanford.edu/ai-index/2025-ai-index-report
5. Chumachenko D. et al. "Glossary of Terms in the Field of Artificial Intelligence. *Ministry of Digital Transformation of Ukraine*. 2024. 37 p.
6. Raji, I. et al. AI and the everything in the whole wide world benchmark. *arXiv preprint arXiv:2111.15366*. 2021. DOI: 10.48550/arXiv.2111.15366
7. Chang Y. et al. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*. 2024. No. 15.3. P. 1–45. DOI: 10.1145/3641289
8. Schulhoff S. et al. The Prompt Report: A Systematic Survey of Prompting Techniques. *arXiv preprint arXiv:2406.06608*. 2024. DOI:10.48550/arXiv.2406.06608
9. Zhou Z. et al. Examining how the large language models impact the conceptual design with human designers: A comparative case study. *International Journal of Human–Computer Interaction*. 2024. P. 1–17. DOI: 10.1080/10447318.2024.2370635
10. Novakovskyi A., Yaloveha I. Implementation of generative artificial intelligence technologies in creative activities: development of a structural model of design thinking. *Innovative Technologies and Scientific Solutions for Industries*. 2024. No. 2(28). P. 108–120. DOI: 10.30837/2522-9818.2024.2.108
11. Chakrabarty T. et al. Art or artifice? large language models and the false promise of creativity. *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024. DOI: 10.1145/3613904.3642731
12. Zhao Y. et al. Assessing and understanding creativity in large language models. *arXiv preprint arXiv:2401.12491*. 2024. DOI: 10.48550/arXiv.2401.12491
13. Tholander J., Jonsson M. Design ideation with ai-sketching, thinking, and talking with Generative Machine Learning Models. *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2023. P. 1930–1940. DOI: 10.1145/3563657.3596014
14. Thoring K., Huettemann S., Mueller R. M. The augmented designer: a research agenda for generative AI-enabled design. *Proceedings of the Design Society*. 2023. No. 3. P. 3345–3354. DOI:10.1017/pds.2023.335
15. Xi Z. et al. The rise and potential of large language model based agents: A survey. *Science China Information Sciences*. 2025. No. 68.2**.** 121101. DOI: 10.1007/s11432-024-4222-0
16. Hou X. et al. Large language models for software engineering: A systematic literature review. *ACM Transactions on Software Engineering and Methodology*. 2024. No. 33.8. P. 1–79. DOI: 10.1145/3695988
17. Guo T. et al. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*. 2024. DOI: 10.48550/arXiv.2402.01680
18. Johansson-Sköldberg U., Woodilla J., Çetinkaya M. Design thinking: Past, present, and possible futures. *Creativity and innovation management*. 2013. No. 2. P. 121–146. DOI: 10.1111/caim.12023

References

1. Perrault R., Clark J. (2024), "Artificial Intelligence Index Report 2024", available at: https://hai.stanford.edu/ai-index/2024-ai-index-report
2. Learning to Reason with LLMs, *OpenAI*. 2024, available at: https://openai.com/index/learning-to-reason-with-llms/

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

3. "Today, we shared evals for an early version of the next model in our o-model reasoning series: OpenAI o3", *OpenAI Twitter Account*, available at: https://x.com/OpenAI/status/1870186518230511844
4. Perrault R. et al (2025), "Artificial Intelligence Index Report 2025", available at: https://hai.stanford.edu/ai-index/2025-ai-index-report
5. Chumachenko D. et al. (2024), "Glossary of Terms in the Field of Artificial Intelligence", *Ministry of Digital Transformation of Ukraine*, 37 p.
6. Raji, I. et al. (2021), "AI and the everything in the whole wide world benchmark", *arXiv preprint arXiv:2111.15366*. DOI: 10.48550/arXiv.2111.15366
7. Chang Y. et al. (2024), "A survey on evaluation of large language models", *ACM Transactions on Intelligent Systems and Technology*, No. 15.3, P. 1–45. DOI: 10.1145/3641289
8. Schulhoff S. et al. (2024), "The Prompt Report: A Systematic Survey of Prompting Techniques", *arXiv preprint arXiv:2406.06608*. DOI:10.48550/arXiv.2406.06608
9. Zhou Z. et al. (2024), "Examining how the large language models impact the conceptual design with human designers: A comparative case study", *International Journal of Human–Computer Interaction*, P. 1–17. DOI: 10.1080/10447318.2024.2370635
10. Novakovskyi A., Yaloveha I. (2024), "Implementation of generative artificial intelligence technologies in creative activities: development of a structural model of design thinking", *Innovative Technologies and Scientific Solutions for Industries*, No. 2(28), P. 108–120. DOI: 10.30837/2522-9818.2024.2.108
11. Chakrabarty T. et al. (2024), "Art or artifice? large language models and the false promise of creativity", *Proceedings of the CHI Conference on Human Factors in Computing Systems*. P. 1–34. DOI: 10.1145/3613904.3642731
12. Zhao Y. et al. (2024), "Assessing and understanding creativity in large language models", *arXiv preprint arXiv:2401.12491*. DOI: 10.48550/arXiv.2401.12491
13. Tholander J., Jonsson M. (2023), "Design ideation with ai-sketching, thinking, and talking with Generative Machine Learning Models", *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, P. 1930–1940. DOI: 10.1145/3563657.3596014
14. Thoring K., Huettemann S., Mueller R. M. (2023), "The augmented designer: a research agenda for generative AI-enabled design", *Proceedings of the Design Society*, No. 3, P. 3345–3354. DOI:10.1017/pds.2023.335
15. Xi Z. et al. (2025), "The rise and potential of large language model based agents: A survey", *Science China Information Sciences*, No. 68.2**,** 121101 p. DOI: 10.1007/s11432-024-4222-0
16. Hou X. et al. (2024), "Large language models for software engineering: A systematic literature review", *ACM Transactions on Software Engineering and Methodology*, No. 33.8, P. 1–79. DOI: 10.1145/3695988
17. Guo T. et al. (2024), "Large language model based multi-agents: A survey of progress and challenges", *arXiv preprint arXiv:2402.01680*. DOI: 10.48550/arXiv.2402.01680
18. Johansson-Sköldberg U., Woodilla J., Çetinkaya M. (2013), "Design thinking: Past, present, and possible futures", *Creativity and innovation management*, No. 2, P. 121–146. DOI: 10.1111/caim.12023

Надійшла (Received) 30.04.2025

Відомості про авторів / About the Authors

Новаковський Антон Валерійович – аспірант, Харківський національний університет радіоелектроніки, кафедра прикладної математики, Харків, Україна; e-mail: anton.novakovskyi@nure.ua; ORCID ID: https://orcid.org/0009-0007-6129-5374

Яловега Ірина Георгіївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри прикладної математики; заступник директора (керівника) навчально-наукового інституту міжнародних відносин ХНЕУ ім. Семена Кузнеця, доцент кафедри економіко-математичного моделювання, Харків, Україна; e-mail: iryna.ialoveha@nure.ua; ORCID ID: https://orcid.org/0000-0002-2486-1812

Novakovskyi Anton – Postgraduate, Kharkiv National University of Radio Electronics, Department of Applied Mathematics, Kharkiv, Ukraine.

Yaloveha Iryna – PhD (Engineering Sciences), Associate Professor at the Department of Applied Mathematics, Kharkiv National University of Radio Electronics; Deputy Director of the Educational and Scientific Institute of International Relations, Simon Kuznets Kharkiv National University of Economics; Associate Professor at the Department of Economic and Mathematical Modeling, Kharkiv, Ukraine.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

DEFINING THE ROLES OF LARGE LANGUAGE MODELS (LLM) AGENTS IN THE MODEL OF DESIGN

The **subject** of the study is the capabilities and limitations that large language models (LLM) demonstrate when they are implemented in intellectual, technical and creative processes, in particular in design. The **goal** of the work is to determine the place and roles of LLM-agents in design and to develop an appropriate model of the design process augmented by LLM-agents. The article sets the following **tasks**: (1) to conduct a review of modern publications on approaches and methods for assessing the capabilities of large language models, in particular, when performing creative, technical and design tasks; (2) to conduct an analysis of modern approaches and methods of interaction with LLM; (3) to develop a model that defines the place and roles of LLM agents in design in interaction with the design team, the external environment and design artifacts. The following **methods** were used during the study: comparative-historical and retrospective analysis of the content of technical, economic, philosophical, linguistic scientific and methodological research to form a holistic vision of the current state o f development of large language models and approaches to interaction with them; structural-logical analysis for the formation of a model of the design process, augmented by LLM agents. The following **results** were achieved: nine main groups of abilities of large language models were identified; the main modern patterns of interaction with large language models were identified; a model of the design was developed as an iterative process of knowledge enrichment, which allows describing the interaction of people and LLM-agents with each other, with the external environment and with design artifacts. The **conclusions** emphasize the novelty of such abilities of LLM in the family of artificial intelligence technologies as understanding and generating text in natural and programming languages, multilingualism, possession of general and industry knowledge, the ability to reason, and agency. Additionally, the need for conceptual understanding of the place and role of large language models in creative processes is highlighted. A structural model has been developed that represents design as an iterative process of knowledge enrichment and allows to define the roles of LLM-agents in interaction with the design team, artifacts, and the external environment.

Keywords: large language models; LLM agents; artificial intelligence; human-computer interaction; design innovations.

Бібліографічні описи / Bibliographic descriptions

Новаковський А. В., Яловега І. Г. Визначення ролей агентів великих мовних моделей (LLM) у моделі дизайн-процесу. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 58–72. DOI: https://doi.org/10.30837/2522-9818.2025.3.058

Novakovskyi, A., Yaloveha, I. (2025), "Defining the roles of large language models (LLM) agents in the model of design", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 58–72. DOI: https://doi.org/10.30837/2522-9818.2025.3.058

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.8+551.58
DOI: https://doi.org/10.30837/2522-9818.2025.3.073

Т. Петренко, А. Задорожний

ПЛАТФОРМА ДЛЯ ІНТЕГРАЦІЇ ІНСТРУМЕНТІВ І СЕРВІСІВ ОБРОБЛЕННЯ МЕТЕОДАНИХ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ

Предметом дослідження є інструменти, сервіси та платформи забезпечення прогнозування локальних метеоумов. Процес прогнозування метеоумов за певною геолокацією доволі складний. Джерелами помилок прогнозування є об'єктивні причини, які є наслідками складності метеопроцесів, що взагалі існували завжди, а також суттєвих кліматичних змін через глобальне потепління. Використання моделей машинного та глибокого навчання (*Machine Learning and Deep Learning*, ML&DL) разом з уточненням результатів класичних фізичних моделей атмосфери – важливий крок підвищення точності моделей прогнозування. Моделі для прогнозування метеоумов стають усе більше гібридними, а інформація, що застосовується для навчання ML&DL-моделей, – усе більш різноманітною та має різні джерела походження. Для трансформації структурованих, неструктурованих та напівструктурованих метеоданих і прогнозування метеоумов використовуються потужні й не завжди безкоштовні середовища провідних розробників. **Мета роботи** – аналіз можливостей наявних платформ використання ML&DL-моделей для прогнозування метеоумов і створення платформи для прогнозування метеоумов, яка має гібридну полегшену архітектуру (*Hybrid LightWeight Architecture*, HLWA). Платформа на основі HLWA виконує такі **завдання**: розподілення етапів оброблення метеоданих між різними постачальниками інструментів і сервісів із хмарних середовищ, але водночас дає змогу інтегрувати ресурси та інструменти оброблення на одній платформі. Розгортання інструментів і сервісів підготовки метеоданих і прогнозування метеоумов у роботі пропонується на сервері *AWS Lightsail* з використанням *Node-RED, MongoDB* та *AWS SageMaker AI*. У статті впроваджено **методи** декомпозиції процесів прогнозування метеоумов. **Результатом роботи** є створення моделі платформи у вигляді UML-діаграми компонентів з уточненням властивостей кожного компонента платформи та інтерфейсів. **Висновком статті** є твердження, що застосування пропонованої платформи для дослідження гібридних моделей прогнозування метеоумов на основі ML&DL-моделей є зручним, економічним і перспективним рішенням.

Ключові слова: модель платформи прогнозування метеоумов; штучний інтелект; сервіси хмарних середовищ.

Вступ

Створення сучасних моделей, підходів і технологій прогнозування метеоумов на основі штучного інтелекту речей (*Artificial Internet of Things*, AIoT) посилило розвиток гібридних моделей прогнозування метеоумов і платформ щодо науки про дані та машинного навчання для метеопрогнозу (*Data science and machine learning platforms*, DSML platforms). Нові підходи до прогнозування метеоумов, побудовані на ML&DL-моделях, довели спроможність покращити процес і результати прогнозування. Поєднання моделей відбувається на різних рівнях і додає різні властивості гібридним моделям.

Використання гібридних моделей стимулює створення платформ прогнозування, що дають змогу:

1) інтегрувати процеси отримання інформації для прогнозування метеоумов з різних глобальних і локальних джерел, зокрема з процесами отримання даних з інших платформ, які також підтримують генерацію синтетичних даних;

2) забезпечувати підготовку до застосування метеоданих відповідно до вимог ML&DL-моделей;

3) зберігати інформацію для прогнозування метеоумов і його результати в сучасних базах даних;

4) навчати й тестувати ML&DL-моделі із використанням метрик оцінювання результативності моделей;

5) забезпечувати безпосередньо процес прогнозування метеоумов;

6) створювати та виконувати сценарії, що дають змогу зменшувати розмірність ML&DL-моделей з метою створення малих лінгвістичних моделей (*Small Language Models*, SLMs), що підтримують оброблення гетерогенних метеоданих на рівні граничних обчислень;

7) забезпечувати відповідні внутрішні інтерфейси між компонентами системи й зовнішні інтерфейси між платформою прогнозування та джерелами інформації, а також з користувачами системи;

8) зменшити накладні витрати для прогнозування метеоумов за певною геолокацією.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Визначення архітектури платформи прогнозування метеоумов потребує комплексного підходу та є актуальним. У цій роботі запропоновано модель створення платформи підготовки інформації та прогнозування на основі гібридної полегшеної архітектури HLWA, яка забезпечує розподілення етапів оброблення метеоданих між різними постачальниками інструментів і сервісів, але водночас дає змогу інтегрувати ресурси та інструменти оброблення на одній платформі. Розгортання інструментів і сервісів підготовки метеоданих і прогнозування метеоумов запропоновано на сервері *AWS Lightsail* з використанням *Node-RED, MongoDB Atlas* та *AWS SageMaker AI*. У статті обгрунтовано нову модель інтеграції ресурсів та інструментів прогнозування метеоумов, що забезпечує компактність і економічність створеної платформи прогнозування.

Аналіз останніх досліджень і публікацій

Прогнозування метеоумов є нелегким завданням [1–4] через труднощі природних процесів формування метеоумов. З метою зменшення складності модельного уявлення використовується декомпозиція процесів формування метеоумов, яка розподіляє процеси на різні рівні їх просторового уявлення, іншими словами, горизонтального та вертикального розподілення. Для просторового уявлення процесів упроваджено термін "просторова роздільна здатність", що задає параметри куба атмосфери (км^3), для якого значення властивостей метеоумов є однаковими в межах цього куба загальної тривимірної сітки охоплення прогнозу.

Тривимірна сітка охоплення прогнозу є фактично нерівномірною для атмосфери над різними територіями планети Земля. Крім цього, просторова роздільна здатність для різних моделей прогнозування метеоумов є неоднаковою. Для вертикального уявлення зазвичай використовують до 64 рівнів вертикального розподілення, починаючи з поверхні землі. Але, наприклад, для моделі ERA5 значення вертикального розподілення дорівнює до 137 рівнів щодо поверхні землі до 0.01 hPa (приблизно 80 км), а площа горизонтального розподілення дорівнює 31 км^2 [1].

Моделі чисельного прогнозування погоди (*Numerical Weather Prediction*, NWP), до яких належить ERA5, побудовані на фізичних рівняннях динаміки шарів атмосфери та гідросфери [1, 3, 4]. NWP використовує складні математичні моделі, що описують динаміку процесів атмосфери, океанів і поверхневих процесів, але розрахунки за NWP є надзвичайно ресурсомісткими [5].

Для декомпозиції систем прогнозування метеоумов розглядається також часова розподіленість метеоумов за допомогою застосування різних термінів прогнозування. Виокремлюють постійне прогнозування метеоумов (тобто на наступні години на сьогодні й на завтра на основі погодинних метеоданих за сьогодні), короткочасне прогнозування (на 3–5 діб), середньострокове прогнозування (на тиждень або 10–14 діб) та довгострокове прогнозування метеоумов (на місяць і пізніше) [6]. Системи прогнозування метеоумов надають доволі точні постійні та короткочасні прогнози (до 90%), але для середньострокового та довгострокового прогнозування якість результату знижується (до 80 та 50%) [6].

Розвиток AIoT й супутникового зв’язку забезпечив наявність значних обсягів метеоданих. Приблизно 85% інформації для глобальних моделей клімату отримують зараз саме із систем супутникового зв’язку [7]. Також збільшилась кількість та якість отримання метеоданих із локальних метеостанцій. Розвиток апаратного забезпечення хмарних середовищ дає змогу інтегрувати не тільки спостереження, а й результати метеопрогнозів, добутих за різними сучасними моделями прогнозування. Інтеграція джерел метеоданих і моделей прогнозування метеоумов стає найбільш актуальною проблемою, розв’язання якої дослідники вбачають в об’єднанні підходів NWP та ML&DL-моделей [8, 9].

Застосування ML&DL покращує якість прогнозу [10–15], але успішність ML&DL у прогнозуванні потребує наявності значної кількості різних так званих історичних показників про минулі погодні умови [1, 3, 4].

Історичні метеодані відіграють важливу роль у тренуванні та вдосконаленні ML&DL-моделей, оскільки вони дають змогу навчити систему розпізнавати певні тенденції (патерни) в метеоумовах, що можуть виникати за різних кліматичних сценаріїв. Для навчання ML&DL-моделей використовують різні за технологією формування дані. Застосування метеоданих повторного аналізу (*Reanalysis Weather Data*, RWD) допомагає акумулювати інформацію з різних джерел та аналізувати спостереження, які були недоступні в режимі реального часу [1, 3, 4].

Отримання історичних метеоданих з глобальних метеорологічних баз даних, зокрема Європейського центру середньострокових прогнозів погоди *(European*

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

Centre for Medium-Range Weather Forecasts, ECMWF) [16], Національного управління океанічних і атмосферних досліджень США (*National Oceanic and Atmospheric Administration*, NOAA) [17] та з API сайтів метеоресурсів [18], дає змогу сформувати необхідну кількість інформації, що разом з локальними метеоданими відповідної геолокації спроможні забезпечити навчання й тестування ML&DL-моделей.

Метеодані мають бути зібрані, оброблені та структуровані до навчання ML&DL-моделей. Кліматичний стан планети динамічно змінюється та реагує на виклики, що спричиняють техногенні катастрофи. Війна в Україні не тільки призвела до загибелі значної кількості населення та суттєво вплинула на життя мільйонів людей, але й змінила метеоумови й умови отримання впевнених показників із локальних метеостанцій, а також на окремих територіях спричинила знищення метеостанцій [19]. Масштабність таких змін на фоні мінливості клімату всієї планети також змінює уяву про поняття аномалії в метеоданих, а тому до етапів їх оброблення, таких як очищення, перетворення й нормалізація, для дотримання сумісності інформації з ML&DL-моделями додаються також етапи аналізу додаткових показників. Крім того, інформація має зберігатися в спеціалізованих базах даних, що забезпечують швидкий доступ до інформації різного типу й уможливлюють постійне оновлення для покращення моделей.

Успішність використання ML&DL-моделей залежить також від обраної DSML-платформи [20]. Успішними є, наприклад, платформи таких провідних компаній, як *Microsoft, Google, Amazon* та *IBM* [21]. Набули поширення DSML-платформи *Databricks Unified Analytics Platform, KNIME Analytics platform, TIBCO Software, Alteryx Analytics, SAS, H2O.ai* та *DataRobot* [21]. Популярні інструменти, зокрема *Scikit-learn, PyTorch, TensorFlow, Weka, KNIME, Colab* тощо, також відіграють важливу роль у дослідженнях ML&DL-моделей [21].

Наявність у провідних компаній своїх моделей прогнозування метеоумов покращує інтеграцію ресурсів для прогнозування. Спеціалізовані платформи для прогнозування метеоумов створюють і використовують сотні різних постачальників метеопрогнозу для кінцевих користувачів мобільних застосунків [14]. Важливість прогнозування локальних метеоумов для підприємств різних галузей (агропромисловість, транспорт, охорона здоров'я тощо) стимулює створення інтегрованих платформ, що беруть до уваги властивості певного напряму застосування метеопрогнозів і використовують моделі AIoT для оброблення метеоданих та прогнозування метеоумов [14, 15].

Якість прогнозу метеоумов залежить від розуміння складності формування метеоумов, обрання або створення відповідних баз метеоданих, моделей, платформ та інструментів прогнозування метеоумов.

Створення платформи для прогнозування метеоумов потребує з'ясування місця розміщення платформи у відповідному хмарному середовищі, визначення компонентів платформи та ресурсів для використання компонентів, а також визначення інтерфейсів між компонентами для злагодженої роботи платформи.

У роботі запропоновано гібридну полегшену архітектуру платформи прогнозування метеоумов, що дає змогу проводити експерименти та інтегрує можливості AIoT, сучасних ML&DL-моделей та документо-орієнтованих баз даних. Для розміщення платформи обрано сервер *AWS Lightsail* [22]; *Node-RED* як інтеграційне середовище [23]; *MongoDB Atlas* для збереження історичних і прогнозованих метеоданих [24] та *AWS SageMaker AI* як сервіс для розроблення, навчання та розгортання ML&DL-моделей [25].

Мета роботи й завдання

Інтенсивні дослідження ML&DL-моделей, що дають змогу прогнозувати метеоумови [5, 7–15], а також створювати різноманітні бази метеоданих [16–18], підтверджує актуальність експериментів з метою покращення метеопрогнозів. Але потужні середовища прогнозування та великі бази метеоданих здебільшого налаштовані на попит з боку бізнес-користувачів, а безкоштовний режим використання сервісів та інформації є обмеженим.

Безоплатні бази метеоданих забезпечують зацікавленість до них з боку освітньо-наукової спільноти як до інформаційного ресурсу для навчання та дослідження. Якість і формати зберігання метеоданих у різних базах можуть бути неоднаковими. Тому одним із важливих питань прогнозування метеоумов є розуміння джерел і властивостей інформації, що використовується для метеопрогнозування. Первинні метеодані отримують унаслідок вимірювання показників метеоумов, таких як температура, вологість, тиск, швидкість вітру, ультрафіолет тощо. Метеоінформацію збирають за допомогою наземних метеорологічних станцій, радарів, літаків і супутників.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

У бази даних метеопоказники потрапляють уже в RWD-вигляді, але наявність різних баз із різними джерелами метеоданих призводить до необхідності повторного оброблення вже обробленої інформації для подальшого її використання в ML&DL-моделях. Для навчання, тестування та валідації ML&DL-моделей обсяги інформації мають бути достатніми для виявлення тенденцій у даних. Показники, зібрані з відкритих метеоресурсів за допомогою програмного інтерфейсу застосунку (*Application Programming Interface*, API), вимагають спеціального оброблення, оскільки формати даних і методи доступу змінюються, що ускладнює інтеграцію та стабільність метеопоказників. Проблемою є об'єднання різної метеорологічної інформації в єдину базу, здатну адаптуватися до нових джерел без втрати якості даних.

Метою роботи є аналіз можливостей наявних платформ використання ML&DL-моделей для прогнозування метеоумов і створення платформи для прогнозування метеоумов, яка має гібридну полегшену архітектуру, здатна збирати різноманітні метеопоказники з різних метеорологічних API-ресурсів і з локальної наземної метеостанції, зберігати метеопоказники в документо-орієнтованій базі даних; аналізувати метеоінформацію та прогнозувати метеоумови відповідно до певної геолокації за допомогою ML&DL-моделей.

Матеріали й методи

Виконання аналізу наявних DSML-платформ дало змогу виокремити властивості, які є важливими щодо платформи прогнозування метеоумов для певної геолокації:

1) залучення структурованої та неструктурованої інформації (текст, зображення, відео, аудіо та геопросторові показники) з безкоштовних баз метеоданих і файлових сховищ, які можуть бути розташовані локально або в хмарі, а також отримання інформації від локальної метеостанції;

2) розміщення отриманих метеоданих у сучасній документо-орієнтованій базі, що дає змогу зберігати гетерогенні метеопоказники;

3) використання, створення та оцінювання за відповідними метриками гібридних ML&DL-моделей і за допомогою популярних інструментів їх досліджень переважно в безкоштовних хмарних середовищах;

4) розгортання, розміщення та обслуговування компонентів платформи на сервері, який забезпечує захищений доступ і віддалене керування компонентами платформи;

5) забезпечення інтерфейсу з низьким вмістом коду, придатного для неекспертів у сфері оброблення інформації, але експертів із прогнозування метеоумов;

6) забезпечення кодового інтерфейсу для науковців з оброблення інформації для доступу до даних, їх підготовки й розміщення в базі, розроблення ML&DL-моделей та публікації;

7) управління життєвим циклом ML&DL-моделей після розгортання для перенавчання та адаптації моделей;

8) створення передумов використання на проєктованій платформі SLM, що підтримує оброблення гетерогенних метеопоказників на рівні граничних обчислень.

У роботі запропоновано формування платформи на основі HLWA-архітектури для підготовки даних та прогнозування. Ця платформа забезпечує важливі властивості платформи прогнозування метеоумов для певної геолокації. Вагомим компонентом запропонованої платформи є *Node-RED*, що інтегрує інструменти керування компонентами платформи для збирання, попереднього оброблення, збереження інформації та прогнозування на основі ML&DL-моделей.

Платформа побудована на сервері *AWS Lightsail*, що втілює архітектуру захищеного хостингу та забезпечує доступ до віртуальних машин і розгортання компонентів платформи. *AWS Lightsail* є приватним віртуальним сервером (*Virtual Private Server*, VPS), що уможливлює просте розгортання віртуальних машин, стабільність і масштабованість за умови доступної ціни. Використання *AWS Lightsail* потребує реєстрації (рис. 1). VPS має переваги щодо спільного хостингу, оскільки дає змогу ефективніше контролювати, налаштовувати, масштабувати ресурси та забезпечує більш надійну безпеку. *Lightsail* працює на базі *Amazon Web Service* та підтримує різноманітні конфігурації для задоволення потреб оброблення метеоданих. У запропонованій платформі прогнозування метеоумов сервер працює під управлінням ОС *Ubuntu* з 1 GB оперативної пам'яті та 40 GB SSD. Для безпечного доступу до сервера *Lightsail* використовується мережевий протокол SSH (*Secure Shell*), що дає змогу адміністратору виконувати команди з віддаленого комп'ютера. Протокол SSH забезпечує захист інформації, а завдяки клієнтській програмі *Putty* з віддаленого комп'ютера можна керувати, наприклад, установленням необхідних компонентів, таких як *Node-RED*.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

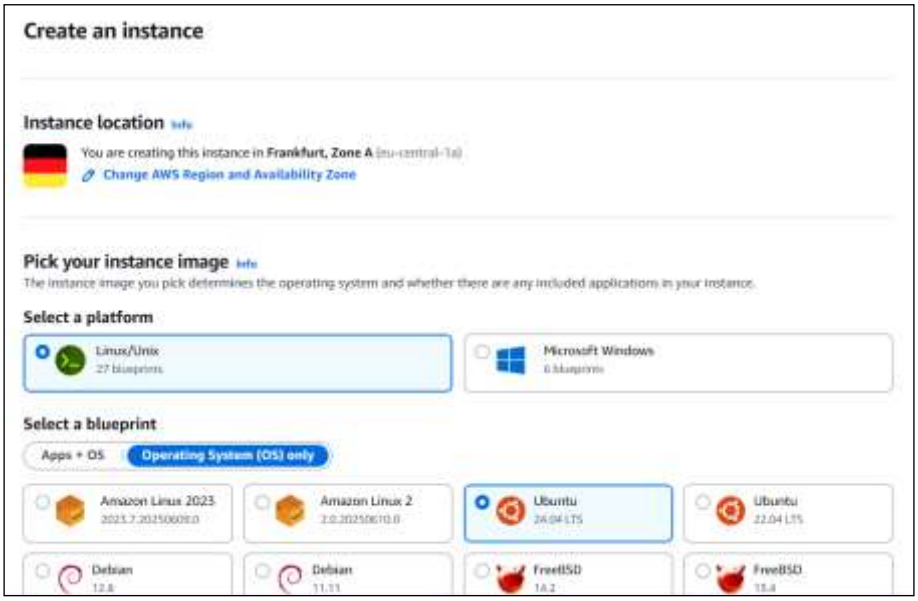


Рис. 1. Етап створення та налаштування віртуального сервера *AWS Lightsail*

Відповідно до HLWA-архітектури платформи, запропонованій у цій роботі, передбачено два основних компоненти, що розгортаються на сервері *AWS Lightsail*.

Компонент *Reanalysis Weather Data Retrieval* реалізовано як програму *Python*, яка автоматично запускається щоденно за визначеним часом доби планувальником завдань *crontab*, що є системним ресурсом ОС *Ubuntu*, та забезпечує налаштування режиму отримання метеопоказників із відкритих API-метеоресурсів.

Наприклад, якщо необхідно порівняти результати прогнозів метеопоказників з відкритих API-метеоресурсів з оновленими результатами вимірювання метеопоказників з тих самих відкритих API-метеоресурсів за попередню годину, можна налаштувати виклик програми *Python* для запуску двічі:

1) о 13:00 UTC запускається програма на отримання прогнозованих метеопоказників по всіх АР-ресурсах на наступну годину для певної геолокації (з параметром *forecast* для виклику запитів);

2) о 14:00 UTC запускається програма на отримання вже уточнених (оновлених) метеопоказників по всіх API-ресурсах на попередню годину для певної геолокації (з параметром *current*).

У цьому разі налаштований файл конфігурації *crontab* для періодичного запуску програми *Python* на сервері *AWS Lightsail* має вигляд, зображений на рис. 2.

```
00·13·*·*·*·/home/ubuntu/weather-forecasting/.venv/bin/python3·/home/ubuntu/weather-
forecasting/main.py·--task·forecast·>>·/home/ubuntu/weather-forecasting/main.log·2>&1¶
00·14·*·*·*·/home/ubuntu/weather-forecasting/.venv/bin/python3·/home/ubuntu/weather-
forecasting/main.py·--task·current·>>·/home/ubuntu/weather-forecasting/main.log·2>&1¶
```

Рис. 2. Текст файлу *crontab*

Компонент *Reanalysis Weather Data Retrieval* опитує відкриті API-метеоресурси з метою відбору необхідних для певного прогнозування історичних RWD-метеоданих та відбору значень метеопараметрів, які метеоресурси пропонують як результати прогнозування. Відкриті метеоресурси попередньо проаналізовано та обрано. Основні метеопоказники, такі як температура, вологість, швидкість вітру, ультрафіолетовий індекс і тиск, отримуються асинхронно з використанням бібліотеки *asyncio*. Зібрані відомості передаються за допомогою бібліотеки *paho-mqtt* через протокол MQTT у режимі реального часу до другого компонента *Node-RED*, що розгортається на сервері *AWS Lightsail*.

Node-RED є платформою з відкритим кодом і дає змогу інтегрувати метеодані з різних джерел, а також створювати потоки оброблення метеоданих за допомогою зручного графічного інтерфейсу. *Node-RED* налаштовується на сервері *AWS Lightsail* у глобальному файлі *settings.js*,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

де також створюються облікові записи з різними правами доступу.

Для безперебійної роботи *Node-RED* використовується *tmux*, термінальний мультиплексор, який дає змогу паралельно запускати декілька процесів на одному сервері. Це забезпечує можливість збереження та продовження роботи в разі розриву з'єднання із сервером.

Основними компонентами *Node-RED* є вузли (*Nodes*), що забезпечують функціональність окремих етапів оброблення метеоданих. Вузли згруповані за категоріями, зокрема *common, function, network, storage* тощо, що полегшує використання. Наприклад, вузли категорії *network* дають змогу отримувати інформацію із зовнішніх джерел за допомогою HTTP-запитів або MQTT-протоколу.

Для контролю процесу оброблення використовуються *Node-RED* вузли *debug*, що забезпечують перевірку стану метеоданих на різних етапах. Результати прогнозування та метрики якості прогнозу користувач системи може отримати на панелі *Dashboard.*

Вузли в *Node-RED* забезпечують з'єднання між такими компонентами запропонованої платформи:

1) *Reanalysis Weather Data Retrieval*, що отримує значення метеопараметрів з API відкритих джерел метеоданих;

2) *Local Observation Weather Data*, який отримує числові значення метеопараметрів з локальної метеостанції та за протоколом MQTT передає компоненту *Weather Data Preprocessing* для підготовки інформації до використання компонентом оброблення метеоданих;

3) *Weather Data Preprocessing*, що забезпечує оброблення метеопараметрів (очищення інформації, її нормалізацію, агрегацію та аналіз її додаткових властивостей), передачу метеоданих до компонента *MongoDB Atlas*, передачу параметрів налаштування ML&DL-моделей у компонент ML&DL для прогнозування метеоумов та отримання результатів прогнозування й значень метрик для відтворення результатів користувачеві;

4) *MongoDB Atlas*, який дає змогу зберігати метеоінформацію різних типів у документо-орієнтованій базі даних;

5) ML&DL (*AWS SageMaker AI*), що виконує навчання, тестування та валідацію ML&DL-моделей прогнозування метеоумов.

Безпосередньо в *Node-RED* реалізовано у вигляді потоків компонент *Weather Data Preprocessing, function*-вузли якого сформовано на *JavaScript*. Фрагменти налаштованих потоків у *Node-RED* зображено на рис. 3–5, а призначення вузлів описано в табл. 1.

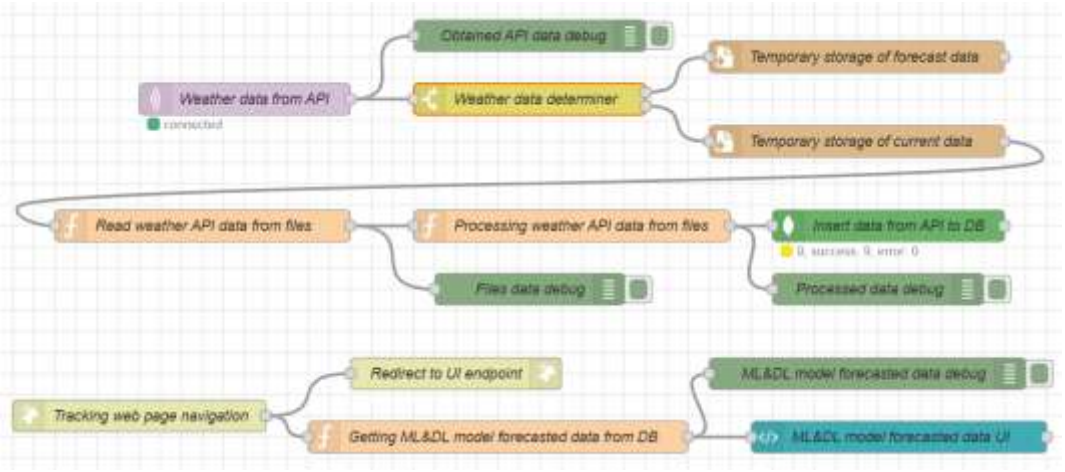


Рис. 3. Потік *Node-RED* оброблення інформації з API-метеоресурсів

Рис. 4. Потік *Node-RED* оброблення метеоданих із локальної метеостанції

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Рис. 5. Потік *Node-RED* налаштування та передачі гіперпараметрів моделі прогнозування до компонента ML&DL

Таблиця 1. *Опис вузлів Node-RED як компонента запропонованої платформи*

Вузол *Node-RED*	**Призначення вузла**
Weather data from API	MQTT-брокер, що отримує метеопоказники з API-ресурсів.
Obtained API data debug	Перегляд метеопоказників із зовнішніх API-ресурсів.
Weather data determiner	Аналіз метеопоказників і визначення властивостей інформації для подальшого оброблення.
Temporary storage of forecast data	Тимчасове збереження метеопоказників прогнозу з API-ресурсів у файлі.
Temporary storage of current data	Тимчасове збереження метеопоказників поточного стану із API-ресурсів у файлі.
Read weather API data from files	Інтеграція метеопоказників в єдиному *JavaScript*-об'єкті.
Files data debug	Перегляд об'єднаних метеопоказників із файлів.
Processing weather API data from files	Попереднє оброблення об'єднаних метеопоказників.
Processed data debug	Перегляд оброблених метеоданих.
Insert data from API to DB	Збереження оброблених метеопоказників з API в базі даних *MongoDB Atlas*.
Tracking web page navigation	Відстеження переходів користувачів на Web-сторінку.
Redirect to UI endpoint	Відповідь користувачеві та перенаправлення користувача на Web-сторінку огляду результатів прогнозування та метрик моделі.
Getting ML&DL model forecasted data from DB	Отримання останніх спрогнозованих метеопоказників із *MongoDB Atlas*.
ML&DL model forecasted data debug	Перегляд отриманих спрогнозованих метеопоказників із MongoDB Atlas.
ML&DL model forecasted data UI	Опис структури Web-сторінки відтворення результатів прогнозування метеопоказників із *MongoDB Atlas*.
Obtained local weather data	MQTT-брокер, що отримує метеопоказники з локальної метеостанції.
Obtained local weather data debug	Перевірка метеопоказників з локальної метеостанції.
Local weather data processing	Попереднє оброблення метеопоказників з локальної метеостанції.
Insert weather local data to DB	Збереження метеопоказників з локальної метеостанції в базу даних *MongoDB Atlas*.
Function starts every hour	Запуск роботи функції генерації параметрів щогодини для роботи ML&DL-моделі.
Generation of parameters for ML&DL model	Генерація параметрів для запуску ML&DL-моделей.
ML&DL model parameters debug	Перегляд інформації, отриманої у відповідь на POST-запит.
Send request for ML&DL triggering	POST HTTP-запит для запуску ML&DL-моделі з параметрами.

Одним із ключових компонентів платформи є база даних *MongoDB Atlas*. Завдяки своїй гнучкій схемі зберігання *MongoDB Atlas* дає змогу зберігати неструктуровану або напівструктуровану інформацію, отриману з різноманітних API без необхідності попередньої трансформації в жорстко визначений формат. Такий підхід спрощує оброблення даних, які можуть мати різні структури залежно від джерела. *MongoDB Atlas* підтримує динамічне оновлення схеми, що є надійним інструментом для об'єднання інформації з різних джерел, навіть якщо формат вихідних даних змінюється з часом. Крім того, *MongoDB Atlas* забезпечує високопродуктивний доступ до інформації завдяки вбудованій індексації та можливостям горизонтального масштабування, що допомагає зберігати великі обсяги метеопоказників і швидко обробляти запити навіть у режимі реального часу.

MongoDB є зручною базою даних для зберігання метеоданих, адже структура JSON-документів бази дає змогу масштабувати систему та адаптувати формат інформації. Для з'єднання з *MongoDB Atlas* в *Node-RED* налаштовано спеціальний вузол, що автоматично записує зібрані дані в *MongoDB Atlas* – хмарний сервіс, який підтримує реплікацію та моніторинг інформації в реальному часі.

Оброблена інформація надходить з *Node-RED* у форматі JSON-документів та зберігається в базі даних *MongoDB* під назвою *weather-forecasting*, а саме в колекцію *historic-data* (рис. 6).

Колекція в *MongoDB Atlas* є набором документів, що мають схожу структуру та зберігаються в межах однієї бази даних. Колекція подібна до таблиці в реляційних базах даних, проте документи в одній колекції можуть розрізнятися за своєю структурою.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

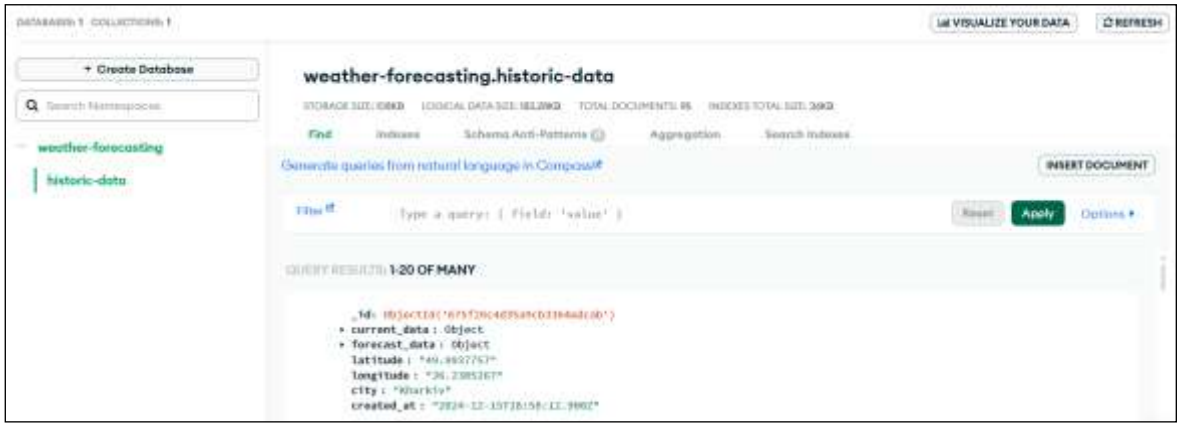


Рис. 6. Збереження метеопоказників у колекції *MongoDB Atlas*

MongoDB Atlas має свої обмеження, але вони не є суттєвими для використання бази з метою збереження метеопоказників.

Можливості безкоштовного збереження метеоінформації у *MongoDB Atlas* обмежені 512 МБ, але її збереження є оптимізованим, а запити до *MongoDB Atlas* можуть використовувати вбудовані функції, що дають змогу обробляти метеодані з огляду на час і відсутні значення. Обмеження на вхідний та вихідний трафік (10 GB вхідного та 10 GB вихідного трафіку за 7-денний період) можуть бути зняті внаслідок переходу до платних кластерів у зв'язку з доступністю вертикального масштабування. Перевагою використання *MongoDB Atlas* є також можливість вибору хмарного провайдера, наприклад AWS.

Одним із потужних сервісів *Amazon Web Services* для розроблення, навчання та розгортання ML&DL-моделей є хмарна платформа *AWS SageMaker AI*. Сервіс надає різноманітні інструменти та гнучке середовище для роботи з ML&DL-моделями, які можуть бути використані з метою прогнозування метеоумов для певної геолокації. *AWS SageMaker AI* підтримує автоматизоване масштабування, оптимізацію гіперпараметрів та інтеграцію з потоковими метеоданими, що робить його ефективним рішенням для експериментів у сфері AIoT та аналізу кліматичних змін.

Використання як компонента ML&DL-сервісу *AWS SageMaker AI* дає змогу спростити процеси створення, навчання та розгортання моделей прогнозування. Запит до історичних метеоданих у *MongoDB Atlas* з компонента ML&DL є прямим.

Сервіс *AWS SageMaker AI* допомагає обрати та налаштувати певні екземпляри (*instances*) програмних і апаратних інструментів та кількість таких екземплярів; надає різні конфігурації досягнення мети формування моделі (*endpoint configuration*); дає змогу обрати й налаштувати процес отримання висновків (*inferences*), які забезпечують виконання прогнозування метеоумов за обраними ML&DL-моделями. Кожна з можливостей *AWS SageMaker AI* має певні обмеження як за ресурсами, так і термінами безкоштовного використання. Наприклад, можливе застосування безоплатного рівня *AWS SageMaker AI* протягом перших двох місяців з моменту початку роботи з *AWS SageMaker AI*. Але щодо таких екземплярів програмних інструментів, як блокноти (*Notebooks*), наразі також надається обмежена (відповідно до регіону) кількість годин застосування екземплярів (*instances*) апаратних ресурсів за місяць. Наприклад, *ml.t3.medium* означає використання двох віртуальних CPU, 4 GiB пам'яті, збереження тільки в *Amazon Elastic Block Store* (*Amazon EBS*) і продуктивність мережі до 5 Gigabit. Після двох місяців або в разі перевищення лімітів усі ресурси необхідно оплачувати за стандартними тарифами AWS, наприклад, оплата *ml.t3.medium* може становити приблизно $0.06 за годину.

Для збереження інформації та моделей *AWS SageMaker AI* інтегрується з хмарним сховищем об'єктів AWS S3. Інструменти *SageMaker AI* можуть використовувати як локальні сховища на час своєї роботи, або для тривалого збереження інформації застосовувати S3. AWS пропонує безкоштовний рівень використання S3 з параметрами 5 GB стандартного сховища, 20000 запитів GET, 2000 запитів PUT, COPY, POST або LIST на місяць. Ліміти діють упродовж 12 місяців з моменту реєстрації нового облікового запису AWS.

Підтримка життєвого циклу ML&DL-моделей прогнозування метеоумов передбачає етапи збирання

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

та підготовки метеоданих, обрання або створення власної моделі з подальшим тренуванням моделі, оцінювання та вдосконалення моделі, і, нарешті, її розгортання для прогнозування. Незважаючи на використання в запропонованій платформі компонентів від різних постачальників, що може ускладнювати процес керування життєвим циклом ML&DL-моделей, платформа на основі HLWA-архітектури має переваги для виконання експериментів з ML&DL-моделями для прогнозування метеоумов.

Модель описаної платформи у вигляді UML-діаграми демонструє автономність компонентів платформи (рис. 7) та їх взаємодію у хмарному середовищі за допомогою різних інтерфейсів (табл. 2).

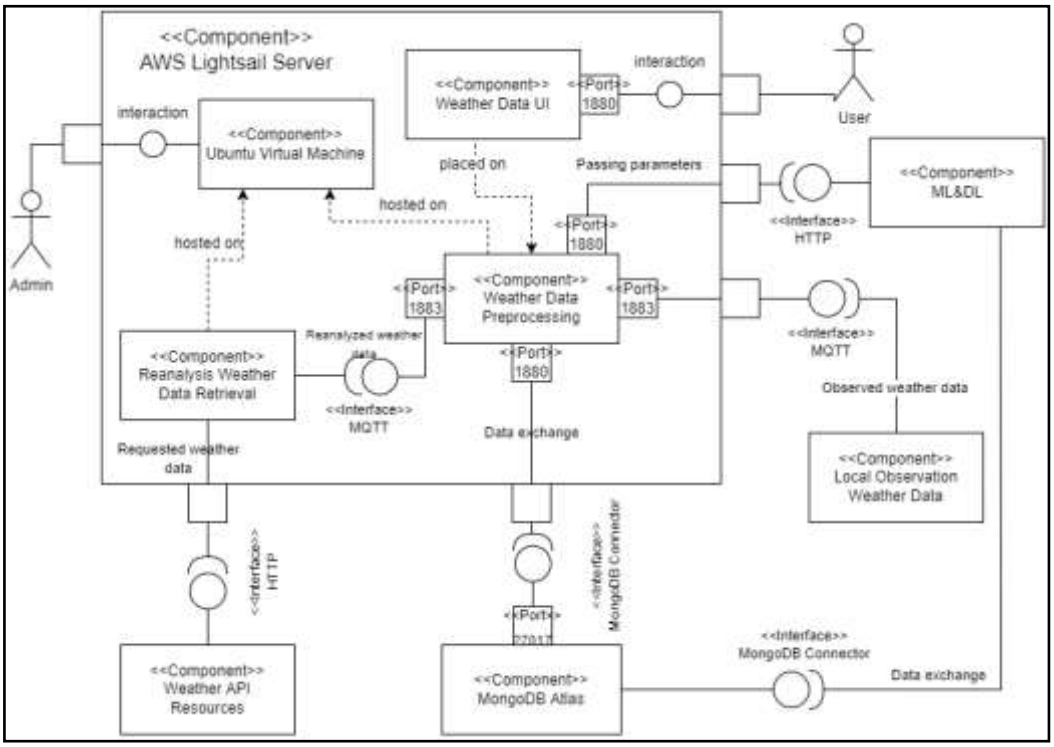


Рис. 7. UML-діаграма компонентів запропонованої платформи на основі HLWA-архітектури

Таблиця 2. *Опис інтерфейсів між компонентами платформи на основі HLWA-архітектури*

Інтерфейс	Опис
1	2

<<Interface>> HTTP (<<Component>> Reanalysis Weather Data Retrieval → <<Component>> Weather API Resources)	Запит метеопоказників з обраних API-метеоресурсів за методом GET HTTP:
	async def fetch_weather_data(scraper, url):
	  try:
	  response = scraper.get(url)
	  if response.status_code == 200:
	  return response.json()
	  else:
	  print(f"Unexpected status code from {url}: {response.status_code}")
	  send_email(
	  "Weather API Failed",
	  f"Failed to retrieve data from the API: {url}.\nDetails: {response}"
	  )
	  return None
	  except Exception as e:
	  print(f"Error fetching data from {url}: {e}")
	  send_email(
	  "Weather API Failed",
	  f"Failed to retrieve data from the API: {url}.\nDetails: {e}"
	  )
	  return None

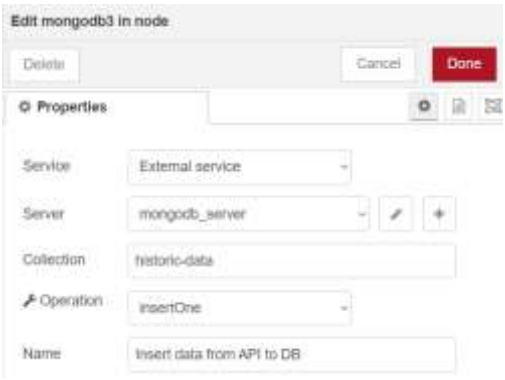
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Продовження таблиці 2

1	2

<<Interface>> MQTT (<<Component>> Reanalysis Weather Data Retrieval → <<Component>> Weather Data Preprocessing)	Передача метеопоказників з API-метеоресурсів до *Node-RED* за допомогою MQTT-брокера:
	mqtt_client = mqtt.Client()
	mqtt_broker = "localhost"
	mqtt_topic = "weather_data_topic"
	
	def publish_to_mqtt(client, topic, data):
	try:
	client.publish(topic, json.dumps(data))
	except Exception as e:
	print(f"Error publishing to MQTT: {e}")

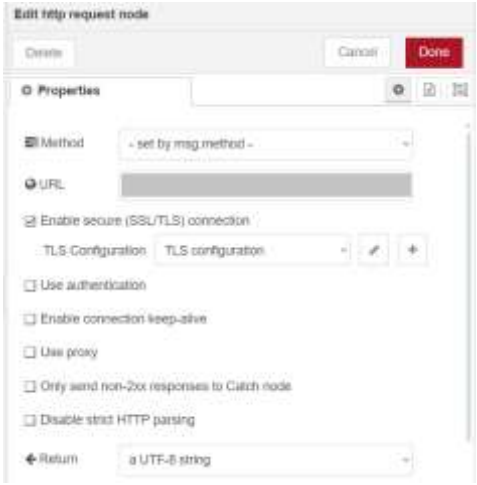
<<Interface>> MQTT (<<Component>> Local Observation Weather Data → <<Component>> Weather Data Preprocessing)	Передача метеопоказників з локальної метеостанції до *Node-RED* за допомогою MQTT-брокера (приклад): String payload = "{\"temperature\": " + String(temperature) + ", \"humidity\": " + String(humidity) + "}"; client.publish(mqtt_topic, payload.c_str());

<<Interface>> MongoDB Connector (<<Component>> Weather Data Preprocessing → <<Component>> MongoDB Atlas)	Обмін даними між компонентами *Node-RED* і *MongoDB Atlas* – передача підготовленого об'єкта метеопоказників у вузлі *Processing of all received data* до вузла *Insert data from API to DB*:
	msg.payload = {
	current_data: sanitizeKeys(currents),
	forecast_data: sanitizeKeys(forecasts),
	latitude: currentObject.latitude,
	longitude: currentObject.longitude,
	city: currentObject.city,
	created_at: todayFullData
	};
	
	Налаштування вузла *Insert data from API to DB* у *Node-RED*
	 

<<Interface>> MongoDB Connector (<<Component>> ML&DL → <<Component>> MongoDB Atlas)	Обмін даними між компонентами ML&DL (*SageMaker*) та *MongoDB Atlas*:
	MONGO_URI = "mongodb+srv://username:password@cluster-url"
	DATABASE_NAME = "database"
	COLLECTION_NAME = "collection"
	
	def get_data_from_mongo():
	client = pymongo.MongoClient(MONGO_URI)
	db = client[DATABASE_NAME]
	collection = db[COLLECTION_NAME]
	# Fetch data from MongoDB
	data = pd.DataFrame(list(collection.find({})))
	# Drop MongoDB object ID field if exists
	if "_id" in data.columns:
	data = data.drop("_id", axis=1)
	return data

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Продовження таблиці 2

1	2
<<Interface>> HTTP (<<Component>> Weather Data Preprocessing → <<Component>> ML&DL)	Передача параметрів з компонента *Node-RED* до компонента ML&DL (*SageMaker*). Використання ML&DL відбувається щогодини. Передача підготовленого об'єкта у вузлі *Generation of parameters for ML&DL* до вузла *Send request with parameters*: msg.method = "POST"; msg.url = "Amazon API url"; msg.payload = JSON.stringify({}); return msg; Налаштування вузла *Send request with parameters* в *Node-RED* для відправлення HTTP-запиту на *API endpoint* 

Результати дослідження

Розвиток платформ щодо науки про дані та машинне навчання (*DSML platforms*) стимулює створення предметно-орієнтованих платформ для різних напрямів використання ML&DL-моделей. У роботі вперше запропоновано модель платформи автоматичної підготовки метеоінформації та прогнозування метеоумов на основі HLWA-архітектури. HLWA-архітектура втілює якості, необхідні для проведення дослідницьких експериментів з різними інструментами аналізу метеоданих. Запропонована платформа містить компоненти, що дають змогу отримувати метеопоказники будь-яких типів з різних джерел, зберігати в документо-орієнтованій базі даних і прогнозувати метеоумови за допомогою ML&DL-моделей. Платформа спроможна отримувати метеопоказники температури, вологості, тиску, швидкості та напрямку вітру, атмосферних явищ та інших метеопараметрів для конкретної геолокації за допомогою API-метеоресурсів відкритого доступу, а також метеоінформацію від локальної метеостанції. Інтеграція різних за типом і джерелами отримання метеопоказників, їх збереження у створеній базі даних *MongoDB Atlas*, подальша підготовка метеоінформації для використання ML&DL-моделей щодо прогнозування та, зрештою, візуалізація результатів прогнозування виконуються на одній платформі.

Запропонована платформа, створена на захищеному сервері *AWS Lightsail*. *Node-RED*, що розгортається на сервері *AWS Lightsail*, дає змогу інтегрувати в платформу результати роботи інструментів отримання та збереження метеоданих, а також результатів інструментів прогнозування. Інструменти оброблення метеопоказників здебільшого розміщуються в хмарах постачальників інструментів, що забезпечує компактність практично безкоштовного розміщення інформації за етапами її оброблення. Запропонована платформа розв'язує проблеми, що виникають під час прогнозування метеоумов за допомогою сучасних ML&DL-моделей – обмеженість в інтеграції різних джерел метеоданих та їх оброблення для прогнозування локальних метеоумов; складність створення єдиного інструменту для ефективного збереження та доступу

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

до історичних метеопоказників у масштабованих базах даних, таких як *MongoDB Atlas*, та автоматизації налаштування регулярних завдань для збирання та оброблення метеоданих з огляду на часові пояси та різні частоти оновлення інформації.

Для формування платформи в цій роботі використано мову програмування *Python* та *Python*-бібліотеки для роботи з інформацією. У *Node-RED* для формування вузлів застосовано *Node-RED*-бібліотеки й мову програмування *JavaScript*. Визначення певних інтерфейсів для взаємодії між компонентами платформи забезпечило можливість формування складних сценаріїв дослідження метеоданих з використанням ML&DL-моделей.

Висновки й перспективи подальшого дослідження

У роботі запропоновано платформу на основі HLWA-архітектури для підготовки метеоданих і прогнозування метеоумов. Платформа забезпечує автоматичний збір метеопоказників із безкоштовних API-метеоресурсів і локальної метеостанції за певним розкладом; оброблення метеоданих з метою подальшого використання їх для прогнозування за допомогою ML&DL-моделей; зберігання метеопоказників у документо-орієнтованій базі даних, та саме прогнозування локальних метеоумов. Платформа підтримує всі необхідні етапи для виконання прогнозування метеоумов і є економічною, масштабованою та зручною для проведення експериментів з різними ML&DL-моделями та комбінаціями моделей. Застосування *Node-RED* як інтеграційного середовища дає змогу виконувати потоки оброблення метеопоказників, залучаючи до цього процесу різні компоненти платформи, і постійно мати доступ до візуалізації результатів кожного етапу оброблення метеоданих. Суттєвою перевагою платформи є можливість використання як компонентів платформи системи моніторингу метеопоказників засобами локальної метеостанції, застосування бази даних *MongoDB Atlas*, що забезпечує надійне сховище з можливістю масштабування та підтримкою JSON-документів, а також упровадження *AWS SageMaker AI*, що виконує навчання, тестування та валідацію ML&DL-моделей прогнозування метеоумов. Завдяки можливостям *AWS Lightsail* платформа є гнучкою і може бути легко адаптована для розширення та використання як нових джерел метеоінформації, так і нових компонентів її оброблення. Перспективним є долучення компонента з SLMs.

Подальший розвиток платформи також передбачає створення різних сценаріїв застосування платформи та уточнення характеристик ML&DL-моделей прогнозування метеоумов, дослідження з якими проводились автономно в *AWS SageMaker AI*.

Список літератури

1. Hersbach H. ERA5 atmospheric reanalysis. URL: https://climatedataguide.ucar.edu/climate-data/era5-atmospheric-reanalysis (дата звернення: 1.04.2025)
2. Beucler T. et al. Next-Generation Earth System Models: Towards Reliable Hybrid Models for Weather and Climate Applications. *Atmospheric and Oceanic Physics*. 2023. DOI: https://doi.org/10.48550/arXiv.2311.13691
3. Lopez N. Reanalysis Q&As. URL: https://climate.copernicus.eu/reanalysis-qas#:~:text=Reanalysis%2C%20on%20the%20other%20hand,where%20weather%20observations%20are%20missing (дата звернення: 1.04.2025)
4. Hersbach H. et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. 2020. No. 146 (730), P. 1999–2049. DOI: https://doi.org/10.1002/qj.3803
5. Bauer P. What if? Numerical weather prediction at the crossroads. *Journal of the European Meteorological Society*. 2024. No. 1. DOI: https://doi.org/10.1016/j.jemets.2024.100002
6. Redmon M. How Accurate Are Weather Forecasts? 5 Reasons for Inaccuracy. URL: https://tempest.earth/resources/how-accurate-are-weather-forecasts/#:~:text=It's%20accurate%20about%2090%25%20of,term%20planning%20and%20decision%2Dmaking (дата звернення: 19.01.2025)
7. Waqas M. et al. Artificial intelligence and numerical weather prediction models: A technical survey. *Natural Hazards Research*. 2024. DOI: https://doi.org/10.1016/j.nhres.2024.11.004
8. Willard J. et al. Integrating Physics-Based Modeling with Machine Learning: A Survey. *Computational Physics.* 2022. DOI: https://doi.org/10.48550/arXiv.2003.04919
9. Chen L. et al. Machine Learning Methods in Weather and Climate Applications: A Survey. *Applied Sciences*. 2023. No. 13. 12019 p. DOI: https://doi.org/10.3390/app132112019

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

10. The Ultimate Guide to Weather Forecast Models in 2025. URL: https://climavision.com/resources/the-ultimate-guide-to-weather-forecast-models/ (дата звернення: 1.04.2025)
11. Lam R. et al. Learning skillful medium-range global weather forecasting. *Science*. 2023. No. 382 (6677). P. 1416–1421. DOI: https://doi.org/10.1126/science.adi233
12. Zhang H. et al. Machine Learning Methods for Weather Forecasting: A Survey. *Atmosphere*. 2025. No. 16(1). 82 p. DOI: https://doi.org/10.3390/atmos16010082
13. Kochkov D. et al. Neural General Circulation Models for Weather and Climate. *Nature*. 2024. No. 632. P. 1060–1066. DOI: https://doi.org/10.1038/s41586-024-07744-y
14. Martorana F. et al. A new tool to process forecast meteorological data for atmospheric pollution dispersion simulations of accident scenarios: A Sicily-based case study. *Journal of Sustainable Development of Energy, Water and Environment Systems*. 2021. No. 9(3). 1080377 p. DOI: https://doi.org/10.13044/j.sdewes.d8.0377
15. Das S., Nayak P. Integration of IoT- AI powered local weather forecasting: A Game-Changer for Agriculture. *Other Computer Science*, 14 p. 2025. DOI: https://doi.org/10.48550/arXiv.2501.14754
16. European Centre for Medium-Range Weather Forecasts. URL: https://www.ecmwf.int/en/about (дата звернення: 1.04.2025)
17. National Oceanic and Atmospheric Administration. URL: https://www.noaa.gov/about-our-agency (дата звернення: 1.04.2025)
18. Top Websites Ranking. URL: https://www.similarweb.com/top-websites/science-and-education/weather/ (дата звернення: 1.04.2025)
19. Balabukh V. et al. Possible contamination of Ukraine and neighboring countries by Cs-137 due to a hypothetical accident at the Zaporizhzhia NPP as a consequence of the Russian aggression. *European Geosciences Union General Assembly 2024 (EGU24)*, Vienna, Austria, 2024. URL: https://meetingorganizer.copernicus.org/EGU24/EGU24-10247.html (дата звернення: 1.04.2025)
20. Jaffri A. et al. Magic Quadrant for Data Science and Machine Learning Platforms. 2024. URL: https://www.gartner.com/doc/reprints?id=1-2HIEGHWG&ct=240508&st=sb&__hstc=&__hssc=&hsCtaTracking=b9f5a528-32c7-4b6e-8dce-c3d23844ba2f%7C08736e49-9973-4af1-b53e-3ba4c12dd1ff (дата звернення: 1.04.2025)
21. Patel B. 10 Best Machine Learning Platforms in 2025. 2025. URL: https://www.spaceotechnologies.com/blog/machine-learning-platforms/ (дата звернення: 1.04.2025)
22. Amazon's Free Virtual Cloud Server. URL: https://aws.amazon.com/free/compute/lightsail/ (дата звернення: 1.04.2025)
23. Demme M. Node-RED in a Unified Namespace Architecture. 2024. URL: https://flowfuse.com/blog/2024/02/node-red-unified-namespace-architecture/ (дата звернення: 1.04.2025)
24. Structured vs Unstructured Data: An Overview. URL: https://www.mongodb.com/resources/basics/unstructured-data/structured-vs-unstructured (дата звернення: 1.04.2025)
25. Amazon SageMaker AI. Guide to getting set up with Amazon SageMaker AI. URL: https://docs.aws.amazon.com/sagemaker/latest/dg/gs.html?icmpid=docs_sagemaker_lp/index.html (дата звернення: 1.04.2025)

References

1. Hersbach, H. "ERA5 atmospheric reanalysis", available at: https://climatedataguide.ucar.edu/climate-data/era5-atmospheric-reanalysis (last accessed: 1.04.25)
2. Beucler, T., Koch, E., Kotlarski, S., Leutwyler, D., Michel, A., Koh, J. (2023), "Next-Generation Earth System Models: Towards Reliable Hybrid Models for Weather and Climate Applications". *Atmospheric and Oceanic Physics* DOI: https://doi.org/10.48550/arXiv.2311.13691
3. Lopez, N. "Reanalysis Q&As", available at: https://climate.copernicus.eu/reanalysis-qas#:~:text=Reanalysis%2C%20on%20the%20other%20hand,where%20weather%20observations%20are%20missing (last accessed: 1.04.25)
4. Hersbach, H. et al. (2020), "The ERA5 global reanalysis", *Qquarterly Journal of the Royal Meteorological Society*, No. 146 (730), P. 1999-2049. DOI: https://doi.org/10.1002/qj.3803
5. Bauer, P., (2024), "What if? Numerical weather prediction at the crossroads", *Journal of the European Meteorological Society*, No.1. DOI: https://doi.org/10.1016/j.jemets.2024.100002
6. Redmon, M. "How Accurate Are Weather Forecasts? 5 Reasons For Inaccuracy", available at: https://tempest.earth/resources/how-accurate-are-weather-forecasts/#:~:text=It's%20accurate%20about%2090%25%20of,term%20planning%20and%20decision%2Dmaking (last accessed: 19.01.25)
7. Waqas, M., Humphries, U., Chueasa, B., Wangwongchai, A. (2024), "Artificial intelligence and numerical weather prediction models: A technical survey", *Natural Hazards Research*. DOI: https://doi.org/10.1016/j.nhres.2024.11.004

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

8. Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V. (2022), "Integrating Physics-Based Modeling with Machine Learning: A Survey". *Computational Physics.* DOI: https://doi.org/10.48550/arXiv.2003.04919
9. Chen, L., Han, B., Wang, X., Zhao, J.; Yang, W.; Yang, Z. (2023), "Machine Learning Methods in Weather and Climate Applications: A Survey", *Applied Sciences*, No. 13, 12019 p. DOI: https://doi.org/10.3390/app132112019
10. "The Ultimate Guide to Weather Forecast Models in 2025", available at: https://climavision.com/resources/the-ultimate-guide-to-weather-forecast-models/ (last accessed: 1.04.25)
11. Lam, R. et al. (2023), "Learning skillful medium-range global weather forecasting", *Science*, No. 382 (6677), P. 1416–1421. DOI: https://doi.org/10.1126/science.adi233
12. Zhang, H., Liu, Y., Zhang, C., Li, N. (2025), "Machine Learning Methods for Weather Forecasting: A Survey", *Atmosphere*, No. 16(1), 82 p. DOI: https://doi.org/10.3390/atmos16010082
13. Kochkov, D. et al. (2024), "Neural General Circulation Models for Weather and Climate", *Nature*, No. 632, P. 1060–1066. DOI: https://doi.org/10.1038/s41586-024-07744-y
14. Martorana, F., Giardina, M., Buffa, P., Beccali, M., Zammuto, C. (2021), "A new tool to process forecast meteorological data for atmospheric pollution dispersion simulations of accident scenarios: A Sicily-based case study", *Journal of Sustainable Development of Energy, Water and Environment Systems*, No. 9(3), 1080377 p. DOI: https://doi.org/10.13044/j.sdewes.d8.0377
15. Das, S., Nayak, P. (2025), "Integration of IoT- AI powered local weather forecasting: A Game-Changer for Agriculture". *Other Computer Science*, 14 p. DOI: https://doi.org/10.48550/arXiv.2501.14754
16. "European Centre for Medium-Range Weather Forecasts", available at: https://www.ecmwf.int/en/about (last accessed: 1.04.2025)
17. "National Oceanic and Atmospheric Administration", available at: https://www.noaa.gov/about-our-agency (last accessed: 1.04.2025)
18. "Top Websites Ranking", available at: https://www.similarweb.com/top-websites/science-and-education/weather/ (last accessed: 1.04.25)
19. Balabukh, V., Skrynyk, O., Sergiy Bubin, S., Laptev, G. (2024), "Possible contamination of Ukraine and neighboring countries by Cs-137 due to a hypothetical accident at the Zaporizhzhia NPP as a consequence of the Russian aggression", *European Geosciences Union General Assembly 2024 (EGU24)*, Vienna, Austria, Apr. 14–19, available at: https://meetingorganizer.copernicus.org/EGU24/EGU24-10247.html (last accessed: 1.04.2025)
20. Jaffri, A., Popa, A., Krensky, P., Hare, J., Bhati, R., Hassanlou, M., Zhang, T. (2024), "Magic Quadrant for Data Science and Machine Learning Platforms", available at: https://www.gartner.com/doc/reprints?id=1-2HIEGHWG&ct=240508&st=sb&__hstc=&__hssc=&hsCtaTracking=b9f5a528-32c7-4b6e-8dce-c3d23844ba2f%7C08736e49-9973-4af1-b53e-3ba4c12dd1ff (last accessed: 1.04.25)
21. Patel, B. (2025), "10 Best Machine Learning Platforms in 2025", available at: https://www.spaceotechnologies.com/blog/machine-learning-platforms/ (last accessed: 1.04.25)
22. "Amazon's Free Virtual Cloud Server", available at: https://aws.amazon.com/free/compute/lightsail/ (last accessed: 1.04.25)
23. Demme, M., (2024), "Node-RED in a Unified Namespace Architecture", available at: https://flowfuse.com/blog/2024/02/node-red-unified-namespace-architecture/ (last accessed: 1.04.25)
24. "Structured vs Unstructured Data: An Overview", available at: https://www.mongodb.com/resources/basics/unstructured-data/structured-vs-unstructured (Last accessed: 1.04.2025)
25. "Amazon SageMaker AI. Guide to getting set up with Amazon SageMaker AI", available at: https://docs.aws.amazon.com/sagemaker/latest/dg/gs.html?icmpid=docs_sagemaker_lp/index.html (last accessed: 1.04.25)

Надійшла (Received) 05.04.2025

Відомості про авторів / About the Authors

Петренко Тетяна Григорівна – кандидат технічних наук, доцент, Український державний університет залізничного транспорту, доцент кафедри інформаційних технологій, Харків, Україна; e-mail: petrenko_tg@kart.edu.ua, ORCID ID: http://orcid.org/0000-0001-6305-7918

Задорожний Антон Юрійович – аспірант, Український державний університет залізничного транспорту, кафедра інформаційних технологій, Харків, Україна; e-mail: zadorojniy85@kart.edu.ua, ORCID ID: https://orcid.org/0009-0000-5044-6068

Petrenko Tetyana – PhD, associate professor, department of information technology, Ukrainian State University of Railway Transport, Kharkiv, Ukraine.

Zadorozhnyi Anton – PhD student, Ukrainian State University of Railway Transport, department of information technology, Kharkiv, Ukraine.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

PLATFORM FOR INTEGRATION OF METEODATES PROCESSING TOOLS AND SERVICES USING ARTIFICIAL INTELLIGENCE

The **subject** of the study is tools, services and platforms for forecasting local weather conditions. The process of forecasting weather conditions for a specific geolocation is quite complex. The sources of forecasting errors are objective reasons that are consequences of the complexity of weather processes, which have always existed, as well as significant climate changes due to global warming. The use of Machine Learning and Deep Learning (ML&DL) models, together with the refinement of the results of classical physical models of the atmosphere, is an important step in increasing the accuracy of forecasting models. Models for forecasting weather conditions are increasingly becoming hybrid, and the data used to train ML&DL models is increasingly diverse and has different sources of origin. Powerful and not always free environments from leading developers are used to transform structured, unstructured and semi-structured weather data and forecast weather conditions. **The purpose of the work** is to analyze the capabilities of existing platforms for using ML&DL models for weather forecasting and to create a platform for weather forecasting that has a hybrid lightweight architecture (Hybrid LightWeight Architecture, HLWA). The HLWA-based platform solves such **problems** as distributing the stages of weather data processing between different providers of tools and services from cloud environments, but at the same time allows integrating resources and processing tools on a single platform. The deployment of tools and services for preparing weather data and forecasting in the work is proposed on the AWS Lightsail server using Node-RED, MongoDB and AWS SageMaker AI. The article uses **methods** for decomposition of weather forecasting processes. The **results** of the research are the creation of a platform model in the form of a UML component diagram with clarification of the properties of each platform component and interfaces. The **conclusion** of the article is the statement that using the proposed platform for studying hybrid weather forecasting models based on ML&DL models is a convenient, economical and promising solution.
Keywords: weather forecasting platform model; artificial intelligence; cloud services.

Бібліографічні описи / Bibliographic descriptions

Петренко Т. Г., Задорожний А. Ю. Платформа для інтеграції інструментів і сервісів оброблення метеоданих засобами штучного інтелекту. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 73–87. DOI: https://doi.org/10.30837/2522-9818.2025.3.073

Petrenko, T., Zadorozhnyi, A. (2025), "Platform for integration of meteodates processing tools and services using artificial intelligence", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 73–87. DOI: https://doi.org/10.30837/2522-9818.2025.3.073

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

UDC 004.932.2 DOI: https://doi.org/10.30837/2522-9818.2025.3.088

YU. POHULIAIEV, K. SMELYAKOV

METHOD FOR COMPARING FINGERPRINTS BASED ON A PREVIOUS ALIGNMENT MODEL

The subject of this article is the development and analysis of a new method for comparing fingerprints that uses the geometric Euclidean characteristics of minutiae for biometric identification. The research focuses on minutiae—special points of interruption or bifurcation of papillary lines—as key biometric features, contrasting them with traditional global and local descriptors such as SIFT, HOG, or LPQ. The **purpose** of the **article** is to develop a robust and efficient fingerprint comparison method that uses Euclidean geometric characteristics of minutiae and pre-alignment to improve the accuracy of biometric identification without relying on machine learning. **The research task** was performed in three stages: first, studying the theoretical provisions of minutia-based descriptors and their invariance to affine transformations (shift, rotation, scaling); second, development of a model using a shift vector and distance functions for matching minutiae; third, experimental verification of the model, determination of optimal parameters, and evaluation on a standard data set. **Methods** include theoretical analysis and experimental evaluation. The theoretical basis establishes the stability of alignment to shifts. The descriptor is formed through minutiae coordinates, distance functions, and alignment optimization. An image processing algorithm with filtering and minutiae analysis is used to extract features. **The results** are achieved through experimental verification on the FVC2000 (DB1_B) dataset and demonstrate high performance, as evaluated by classification metrics and execution time. **Conclusions** indicate the theoretical and practical achievements of the research. The model demonstrates theoretical and practical resistance to Euclidean shifts, with advantages for processing prints from different scanners. Experiments confirm the effectiveness of shift detection, achieving a high Van Rijseberg score (0.735), although dependence on positive matches requires additional filtering of false positives. The method is workable and can be applied in practice.

Keywords: dactyloscopy; fingerprints; preliminary alignment; minutiae; biometric identification; affine transformations; tensor representation; machine learning; FVC2000; binary classification metrics.

Introduction

Dactyloscopy, which emerged as a science in the second half of the 19th century, remains relevant in the 21st century thanks to the uniqueness of papillary lines and the ease of obtaining prints. The study of the structure of fingerprints and the discovery of the exceptional properties and shapes of papillary patterns were carried out by British pioneers: Sir William James Herschel, a colonial administration official in Bengal; Henry Folds, a Scottish doctor in Japan; and Sir Francis Galton, a British researcher. Their work laid the foundation for the use of dactyloscopic methods in identification and forensics, ensuring significant competitiveness compared to the system proposed by French researcher and lawyer Alphonse Bertillon [1]. The greater resistance of fingerprints to change, combined with the ease of taking them compared to the Bertillonage procedure, gave fingerprinting an advantage in the 20th century, especially with the progress of computer technology.

One of the key milestones in the development of fingerprinting was the transition of criminalistics to automated fingerprint identification systems (AFIS) during the 1970s and 1980s, primarily in Western Europe and the United States. The evolution of computer technology, databases, and the ability to automatically search for image fragments without human intervention significantly accelerated the processing of biometric information and reduced storage resources—cumbersome fingerprint card indexes were replaced by electronic storage. The second important milestone was the improvement of computer vision tools and machine learning methods, which made it possible to study the details of papillary patterns in greater depth and expanded the variety of approaches to fingerprint matching. This has ensured the competitiveness of fingerprinting in the 21st century compared to new biometric methods such as facial or iris recognition.

Despite significant progress in fingerprinting in the 20th century, it faces a number of challenges that remain partially or completely unresolved even in the 21st century. Fingerprinting methods are sensitive to the quality of the prints, which depends on finger pressure, skin condition, or scanner performance. Issues of distortion (shifting, scaling, inversion, or rotation of prints) remain partially unresolved. Modern fingerprinting systems must meet requirements for quality and speed of results.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

These features and contradictions confirm the relevance of developing new methods for recognizing and comparing fingerprints. This study proposes a fingerprint comparison method based on the geometric Euclidean characteristics of minutiae and preliminary alignment for effective biometric identification with minimized distortion without the use of machine learning.

Analysis of the problem and existing methods

Recent works describe new problems and challenges in fingerprint processing and matching. For example, *Machado et al*. (2025) consider aspects of working with newborn fingerprints [2], while *Raj et al.* (2020) investigates the nature of latent fingerprints and their application [3]. Affine transformations, such as scaling, translation, and rotation, can significantly affect the quality of the system. *Guan et al.* (2025) investigated the problem of matching and aligning partial prints [4], while *Wang et al.* (2025) proposed a model to overcome these limitations, which includes not only the patterns of all fingers, but also the patterns of blood vessels (veins) [5]. This largely determines the divergence of methods for applying fingerprints and their processing. Equally important is the aspect of biometric ADIS performance, especially given the increasing prevalence of fingerprint scanners in consumer electronic devices. *Ravulakollu et al.* (2025) proposed a fast *few-shot* learning method for establishing and recognizing latent fingerprints [6].

The challenges outlined above for modern fingerprinting systems determine the variety of approaches used in current research. Structurally, they can be divided into the following groups: methods that use minutiae (or Galton points); methods based on the structure of fingerprint ridges; various models of fingerprint alignment (both with and without the introduction of machine learning methods); methods of comparing prints using machine learning methods, as well as other methods and models intended for limited use.

Methods based on minutiae are among the key ones in modern approaches to fingerprint comparison. Minutes (or Galton points) are specific points on a fingerprint that mark the beginning or end of one or more papillary lines. The most important points for study are the points where the papillary pattern is interrupted and the points where it splits. Their advantage is the uniqueness of the aggregate set and the relative ease of recognition, but minutiae are very sensitive to image artifacts and interference. *Maltoni et al*. (2009) systematized approaches to minutiae extraction, which have become the standard in fingerprinting [7]. *Ibragimov et al.* (2025) summarized pore detection methods (level 3) for latent and partial prints, proposing a basic detector and comparing it on the PolyU-HRF, L3-SF, and IITI-HRF datasets [8]. *Mani et al.* (2024) developed a two-stage authentication method that uses contour, edge, and line extraction with Euclidean distance and mutual information matching, achieving high accuracy [9].

Crest-based methods analyze the entire papillary pattern, unlike minutiae, which are only a unique structure of pattern intersections. On the one hand, this complicates processing, since it is necessary to extract specific lines rather than points. On the other hand, the pattern structure is more unique than a set of points and less prone to artifacts. Gu et al. (2006) proposed a method of combining ridges and minutiae, improving performance on FVC2002 [13]. *Li et al.* (2021) introduced a ridge pattern analysis algorithm that is robust to low image quality [14].

The next group of methods is alignment and correction of impressions. They are necessary for preliminary operations to extract ridges and minutiae most accurately. *Chen et al.* (2006) applied RANSAC for alignment based on coordinates and minutiae angles [10]. *Jain et al.* (2007) used orientation fields for coarse alignment at high resolution [11]. *Guan et al.* (2024) proposed a multi-task CNN transformer network for joint face verification and partial fingerprint alignment, improving accuracy on the NIST SD14 and FVC datasets [4]. *Dias et al.* (2025) developed a super-resolution method for newborn prints using *Vision Transformer*, improving noise-free details with a 10% increase in accuracy [2]. *Zhang et al.* (2022) implemented an optimized alignment algorithm that reduces computational costs [12].

Machine learning methods are presented in a significant number of studies. Their main advantage is their resistance to image artifacts and transformations, which makes these methods universal for working with images. The disadvantage is the high-performance requirements necessary for working with virtually any artificial intelligence model, which makes the development of high-performance machine learning-based models a priority. *Wang et al.* (2023) proposed a localized deep representation (LDRF) that combines alignment and feature extraction [15]. *Liu et al.* (2021) applied transformers for an interpretable fixed-length fingerprint representation [16]. *Kim et al.* (2022) developed hybrid neural networks that improve matching

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

in noisy environments [17]. *Yang et al.* (2021) proposed a deep learning model for complex patterns that is resistant to deformations [18]. *Park et al.* (2020) implemented an algorithm that minimizes the impact of low image quality [19]. *Ravulakollu et al*. (2025) applied *few-shot* learning with a prototype network based on DenseNet121 for latent fingerprint recognition, achieving an accuracy of 91.66% on the IIIT-D dataset [6].

The last group of methods has a narrow purpose. For the most part, they solve a specific problem, but the proposed approaches and solutions may also be of interest in the general context of matching and identification. *Zhao et al.* (2025) developed a technique for latent prints that improves identification at crime scenes [20]. *Xu et al.* (2025) proposed a method for partial prints based on local correlations [21]. *Lee et al.* (2024) implemented a secure matching protocol with encrypted information [22]. *Choi et al.* (2023) improved encryption protocols for biometric systems [23]. Wu et al. (2025) proposed a method for removing interfering minutiae from latent fingerprints [24]. *Han et al.* (2024) optimized computations for ADIS [25]. *Kim et al.* (2023) presented architectures for improving accuracy [26]. *Park et al.* (2023) developed deep learning for complex deformations [27]. *Lim et al.* (2020) implemented adaptive filtering for preprocessing [28]. *Sathya et al.* (2020) improved latent prints with linear feature extraction, achieving an accuracy of over 80% [3]. *Wulandari et al.* (2021) developed an identification system with edge detection and GLCM, achieving an accuracy of 83% [29]. *Patel et al.* (2021) proposed a technique for high-noise fingerprints [30]. *Wang et al.* (2024) created a contactless multimodal system for four-finger fingerprints and veins with the ACFNet network, achieving an EER of 0.07% [5]. *Yu et al.* (2025) implemented 3D fingerprint unfolding with B-spline for 2D recognition, with an EER ranging from 0.2072% to 1.50% [31]. *Calderón-Calderón et al.* (2025) proposed a method for improving the quality of prints for low-resolution scanners, increasing image clarity and facilitating feature extraction, which is particularly useful for inexpensive consumer devices [32].

A comparative analysis of the groups of methods is presented in Table 1.

Table 1. Comparison of fingerprint matching methods

Method	Features	Noise resistance	Accuracy, EER (%)	Computational complexity	Data nets
Minutiae	Galton points	low	1–2	low	FVC, PolyU-HRF
Rows	papillary patterns	medium	0.5–1	medium	FVC2002
Alignment	coordinates, orientation	medium	0.3–1	high	NIST SD14, FVC
Machine learning	deep features	high	0.4–1	very high	IIIT-D, NIST

Analysis of research confirms the relevance of developing new methods for comparing fingerprints using minutiae features. Machine learning methods are often cumbersome and unsuitable for simple application scenarios. However, it is necessary to take into account the problems outlined by the authors, in particular image artifacts during minutiae recognition, the need for preliminary alignment, etc. The proposed pre-alignment method is based on Euclidean characteristics of minutiae, minimizes the impact of affine transformations, and optimizes computational complexity, making it suitable for real-time applications.

Purpose and objectives of the research

The purpose of this research is to develop a robust and efficient fingerprint comparison method that uses Euclidean geometric features of minutiae and pre-alignment to improve the accuracy of biometric identification without relying on machine learning, ensuring computational efficiency and suitability for real-time operation in the presence of noise, artifacts, and affine transformations.

Research objectives:

- to analyze existing fingerprint processing methods, evaluate their robustness to noise, artifacts, and affine transformations such as shift, rotation, and scaling;

- to develop a mathematical model for fingerprint comparison based on pre-alignment, relying on tensor representation of images, shift vector, and minutiae matching using distance function and tolerance ratio;

- to implement a software solution for practical application of the model using image processing tools and high-performance computing;

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

- to perform experimental verification of the model on the standard FVC2000 (DB1_B) dataset using binary classification metrics and measure the execution time;

to determine the optimal parameters to ensure a balance between the accuracy and performance of the method.

Fingerprint comparison method

Within the scope of the outlined research objectives and tasks, we propose to consider a method of comparing fingerprints based on a pre-alignment model. Earlier, we pointed to the problem of pre-alignment as one of the most important stages of biometric fingerprint matching. To this end, we present a mathematical model of the method.

Let there be a vector space *V*:

$$\mathrm{V} \cong \mathbb{R}^3, \tag{1}$$

where $\mathbb{R}$ – field of real numbers.

Let's introduce the basis of the vector space *V*:

$$\begin{gathered} \forall \mathrm{i} \in [1,3] \cap \mathbb{Z} : \exists \mathbf{e}_\mathrm{i} \in \mathrm{V} \\ \mathbf{e}_1 = (1,0,0) \\ \mathbf{e}_2 = (0,1,0) \\ \mathbf{e}_3 = (0,0,1) \end{gathered}, \tag{2}$$

where $\boldsymbol{e}_i$ – basis vectors.

Then we define a pixel as a tensor of type (1, 0). Here and below, the notation $a^i b_i$ refers to Einstein's notation for tensors.

$$\forall \mathrm{p}^\mathrm{i} \in [0,255] \cap \mathbb{Z} : \mathbf{p} = \mathrm{p}^\mathrm{i} \mathbf{e}_\mathrm{i}, \quad \mathbf{p} \in \mathrm{V}, \tag{3}$$

where p^i – integer pixel components; *p* – first-order contravariant tensor.

Let's introduce vector spaces *H* і *W*:

$$\begin{gathered} \mathrm{H} = \mathrm{V}\left\{\mathbf{f}_\mathrm{m} \middle| \forall \mathrm{m} \in [1,\mathrm{M}] \cap \mathrm{Z}\right\}, \quad \mathbf{f}_\mathrm{m} \in \mathrm{H} \\ \mathrm{W} = \mathrm{V}\left\{\mathbf{f}_\mathrm{n} \middle| \forall \mathrm{n} \in [1,\mathrm{N}] \cap \mathrm{Z}\right\}, \quad \mathbf{f}_\mathrm{n} \in \mathrm{W} \end{gathered}, \tag{4}$$

where *V* – linear span of vector space elements; $\boldsymbol{f}_m$ і $\boldsymbol{f}_n$ – basis space vectors; *M* і *N* – space dimensions (width and height).

Then we define the image as a tensor of type (3, 0):

$$\mathbf{P} = \mathrm{P}^{\mathrm{mni}} \left(\mathbf{f}_\mathrm{m} \otimes \mathbf{f}_\mathrm{n} \otimes \mathbf{e}_\mathrm{i}\right), \quad \mathbf{P} \in \mathrm{H} \otimes \mathrm{W} \otimes \mathrm{V}, \tag{5}$$

where P^{mni} – image components; $\boldsymbol{P}$ – third-order contravariant tensor.

Within specific values of *m* and *n*, the image and pixel components are related:

$$\forall \mathrm{m}, \mathrm{n}, \mathrm{i} : \mathrm{P}^{\mathrm{mni}} = \mathrm{p}^\mathrm{i}. \tag{6}$$

Since our method is based on the characteristics of minutiae, we will briefly describe their extraction (since we are working with a ready-made minutia detector). Here and further, we will understand minutia as a tuple of values from a mixed vector space:

$$\begin{gathered} \mathrm{M} = \mathbb{R}^2 \times \mathrm{S}^1 \times \mathrm{T} \\ \mathrm{S}^1 = \mathbb{R} / 2\pi\mathrm{k} \quad \forall \mathrm{k} \in \mathbb{Z} \\ \mathrm{T} = \{0,1\} \\ \mathrm{m} = (\mathrm{x}, \mathrm{y}, \alpha, \mathrm{t}) \quad \forall \mathrm{m} \in \mathrm{M} \end{gathered}, \tag{7}$$

where *M* – minutiae space; S^1 – factor set denoting the single circle of minutiae orientation angles; *T* – set of minutiae types (bifurcation or interruption); *m* – minutia itself is a set of four values.

Then the minutiae detector

$$\Phi : \mathbf{P} \to \mathrm{M}, \tag{8}$$

where Φ – reproducing the image in minutiae space.

Since the pre-alignment model is available to us and works with a set of minutiae for an image, we will also describe it:

$$\mathrm{A} : 2^\mathrm{M} \times 2^\mathrm{M} \to \mathbb{R}^2, \tag{9}$$

where *A* – reproducing the space of minutiae sets into a vector of image shifts relative to each other, 2^M – a set of all subsets.

Let us assume that, as a result of applying operators *F* and *A*, we obtain a shift vector *s* for two images $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$. Then, the first step of the procedure will be to align the second image with the first one:

$$\begin{gathered} \sigma_\mathrm{s} : \mathrm{M} \to \mathrm{M} \\ \sigma_\mathrm{s}(\mathrm{m}) = (\mathrm{x} + \Delta\mathrm{x}, \mathrm{y} + \Delta\mathrm{y}, \alpha, \mathrm{t}) \end{gathered}, \tag{10}$$

where σ_s is the image shift operator; Δx and Δy are components of vector *s*.

We assume that the shift model should correctly describe the same fingerprint under different affine transformations. However, this should imply the equivalence of two fingerprints. Therefore, we formulate the working hypothesis as follows:

$$\begin{gathered} \forall \mathbf{P}_1, \mathbf{P}_2 \in \mathbf{P} : \\ \exists \mathrm{s} \in \mathbb{R}^2 : \mathrm{s} = \mathrm{A}\left(\Phi(\mathbf{P}_1), \Phi(\mathbf{P}_2)\right) \\ \Phi(\mathbf{P}_1) \sim \sigma_\mathrm{s}\left(\Phi(\mathbf{P}_2)\right) \leftrightarrow \mathbf{P}_1 = \mathbf{P}_2 \end{gathered}, \tag{11}$$

where ~ is the equivalence relation between sets of minutiae, the criterion for which will be given below.

Therefore, we assume that if two sets of minutiae are equivalent after applying the calculated shift, then we are dealing with the same fingerprint. Given the structure of the minutiae tuple, we propose the following criterion for the equivalence of sets:

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

$$\begin{gathered}\forall M_1, M_2 \subset M: \\ \exists L \subset M: L = \begin{cases} M_1, |M_1| \le |M_2| \\ M_2, |M_1| > |M_2| \end{cases} \\ \exists G \subset M: G = \left((M_1 \cup M_2) \setminus L\right) \cup \left(M_1 \cap M_2\right), \\ \exists C \in \mathbb{N}, C = \left|\left\{m \in L \middle| \exists m_2 \in G: m \simeq m_2\right\}\right| \\ \frac{C}{|L|} \ge C_{threshold} \leftrightarrow M_1 \sim M_2\end{gathered} \quad (12)$$

where L is the smaller of the sets according to the power criterion; G is the larger of the sets according to the power criterion; C is the number of minutiae that satisfy the tolerance ratio; $\simeq$ is the tolerance ratio on the set of minutiae; $C_{threshold}$ is the threshold value for the number of tolerant minutiae.

On the space of minutiae M, we define the distance function and the order relation:

$$\begin{gathered}\forall m_1 = (x_1, y_1, \alpha_1, t_1), m_2 = (x_2, y_2, \alpha_2, t_2) \in M: \\ d(m_1, m_2) = \left(|x_1 - x_2|, |y_1 - y_2|, \min_{k \in \mathbb{Z}} |\alpha_1 - \alpha_2 + 2\pi k|, t_1 \odot t_2\right) \in M \\ S^1 = \mathbb{R}/2\pi\mathbb{Z} = \left\{e^{i\theta} \middle| \theta \in \mathbb{R} \wedge e^{i\theta} = e^{i(\theta + 2\pi k)}\ \forall k \in \mathbb{Z}\right\} \\ \exists \theta_0 = 0, \varphi(\alpha) = (\theta - \theta_0) \bmod 2\pi: \alpha_1 \le \alpha_2 \leftrightarrow \varphi(\alpha_1) \le \varphi(\alpha_2) \\ (x_1 \le x_2) \wedge (y_1 \le y_2) \wedge (\alpha_1 \le \alpha_2) \wedge (t_1 \vee \neg t_2) \leftrightarrow m_1 \le m_2\end{gathered} \quad (13)$$

where d is the distance function; $\odot$ is the XNOR operator; θ_0 is the starting reference angle; φ is the function that reduces the angle on the single circle to the range [0, 2π).

Then we set the tolerance ratio in the minutiae space:

$$\forall m_1, m_2 \in M: \exists \Delta m \in M: d(m_1, m_2) \le \Delta m \leftrightarrow m_1 \simeq m_2, \quad (14)$$

where Δm is the vector of permissible deviation during minutiae comparison.

It is easy to see that the definition given in formula (14) does indeed correspond to the tolerance ratio. It is reflexive:

$$\begin{gathered}\forall m, \Delta m \in M: \\ d(m, m) = \left(\inf \mathbb{R}_{+0}, \inf \mathbb{R}_{+0}, \inf [0, 2\pi), \inf \{0,1\}\right), \\ \forall Q, q \in Q: \inf Q \le q \to d(m, m) \le \Delta m \to m \simeq m\end{gathered} \quad (15)$$

symmetrical:

$$\begin{gathered}\forall m_1, m_2, \Delta m \in M: \\ |x| = |-x| \to d(m_1, m_2) = d(m_2, m_1) \\ d(m_1, m_2) \le \Delta m \to d(m_2, m_1) \le \Delta m \\ (m_1 \simeq m_2) \leftrightarrow (m_2 \simeq m_1)\end{gathered}. \quad (16)$$

At the same time, the relationship is not transitive:

$$\begin{gathered}\forall m_1, m_2, m_3, \Delta m \in M: \\ \left(d(m_1, m_2) \le \Delta m \wedge d(m_2, m_3) \le \Delta m\right) \nrightarrow \left(d(m_1, m_3) \le \Delta m\right). \\ \left((m_1 \simeq m_2) \wedge (m_2 \simeq m_3)\right) \nleftrightarrow (m_1 \simeq m_3)\end{gathered} \quad (17)$$

The procedure proposed in formula (12) does not guarantee optimality under the condition of a shift search model found. Moreover, a situation may easily arise when a threshold value that is not selected accurately enough will cut off the correct pair of minutiae sets. Therefore, it is necessary to verify the optimality of the result.

We set the reproduction that takes into account specific pairs of minutiae matched in formula (12):

$$\begin{gathered}S: M \times M \to J, \quad J \subset \mathbb{R}^2: |J| = \frac{C}{|L|} \\ S(M_1, M_2) = \left\{m_1 \in L \middle| \exists! m_2 \in G: (m_1 \simeq m_2) \wedge \left(O(m_1, G) = m_2\right)\right\} \to \{\delta(m_1, m_2)\}\end{gathered}, \quad (18)$$

where O is the optimal matching search function; δ is the Euclidean distance function between minutiae.

The description of the optimal matching search function should begin with the introduction of the minutiae similarity metric. Above, in formula (14), the tolerance ratio is specified, which allows us to distinguish between previously matching minutiae and non-matching minutiae. However, in the process of completely enumerating pairs among two sets of minutiae, we cannot guarantee that each minutia of the first set corresponds bijectively to only one minutia of the second set. Then we will define the minutiae similarity function:

$$\begin{gathered}\forall m_1, m_2, \Delta m \in M: \\ \exists M_d \subset \mathbb{R} \times [0, \pi] \times T, \gamma: M \to M_d \\ \gamma(m) = \left(\sqrt{x^2 + y^2}, \varphi(\alpha), t\right) = (\rho, \beta, t) \\ \forall m_d \in M_d: \varepsilon(m_d) = \left(\frac{1}{\sqrt{x^2 + y^2}}, \frac{1}{\alpha}, t\right) \\ \forall m_{d1}, m_{d2} \in M_d\ \exists \eta: M_d \times M_d \to \mathbb{R}_{+0} \cup \{+\infty\} \\ \eta(m_{d1}, m_{d2}) = \rho_1 \cdot \rho_2 + \beta_1 \cdot \beta_2 + (\neg t_1 \wedge t_2) \\ \lambda(m_1, m_2) = \begin{cases} \eta\left(\gamma\left(d(m_1, m_2)\right), \gamma\left(\varepsilon(\Delta m)\right)\right), m_1 \simeq m_2 \\ +\infty, m_1 \not\simeq m_2 \end{cases}\end{gathered} \quad (19)$$

where M_d is the minutiae space with the transformation of coordinate components into distance; γ is the function that converts coordinate components and angle; ε is the function that returns a vector of values for normalization; η is the function of scalar multiplication of minutiae; λ is the similarity function.

Given the definitions in formulas (13) and (14), we are interested in minimizing the value of the function, since this will mean the most accurate match of two minutiae in the case of $\lambda = 0$. The worst match of minutiae will be described by the value $\lambda = 1$, and in the case of a mismatch, the function will have a value of positive infinity.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Then the function for finding the optimal match looks like this:

$$\forall M_1, M_2 \subset M, m_1 \in M_1 : \\ O(m_1, M_2) = \arg\min_{m_2 \in M_2} \lambda(m_1, m_2). \quad (20)$$

Then we set a recursive description of the result search function:

$$R(M_1, M_2) = \max_{B \in \{S(M_1, \sigma_s(M_2))\} \cup \{S(M_1, \sigma_\delta(M_2)) | \forall \delta \in S(M_1, \sigma_s(M_2))\}} |B|, \quad (21)$$

where $R(M_1, M_2) : M \times M \to \mathbb{N} \cup \{0\}$

where *B* is all possible shift options obtained based on the results of matching minutiae pairs; *s* is the initial shift parameter found using the *A* ratio (see formula (11)).

Then the result of the method is the following ratio:

$$\forall M_1, M_2 \subset M : \\ \frac{R(M_1, M_2)}{|L|} \geq C_{threshold} \leftrightarrow M_1 \sim M_2. \quad (22)$$

Therefore, within the framework of the working hypothesis, we assume that if the number of optimal minutiae matches in relation to the power of the smaller of the two minutiae sets exceeds a certain threshold value, we consider these two minutiae sets (and, therefore, the two fingerprints) to be equivalent. The proposed comparison scheme takes into account both possible affine transformations of images (using a pre-alignment model) and various detection situations (differences in artifacts, image quality, etc.).

Program implementation

The proposed mathematical model for comparing fingerprints, which uses preliminary alignment and minutiae analysis, can be implemented as a multi-module software system. The software provides the phases described in formulas (1)– (22), which facilitates image processing, feature extraction, and comparison, taking into account affine transformations and quality discrepancies.

The following technologies have been implemented for this purpose: Python and OpenCV for image processing, C# for high-performance computing, and ASP.NET for module integration and user interface interaction. This contributes to the efficiency and flexibility required for real-time applications [33].

Below is the module structure.

1. The image preprocessor converts the input scan images of fingerprints, presented as third-order tensors ***P*** (formula (5)), into a form suitable for analysis. Noise filtering, brightness normalization, and contrast enhancement of papillary lines are performed using OpenCV algorithms such as Gaussian smoothing and adaptive binarization. This minimizes the impact of artifacts related to skin quality, finger or scanner pressure, preparing images P_1 and P_2 for further detection. An example of the detector and preprocessor in action is shown in Figures 1 and 2.

Fig. 1. First scan image of fingerprint

Fig. 2. Second scan image of fingerprint

2. The minutiae detector implements the reproduction of Φ (formula (8)), extracting minutiae – special points of prints, such as interruptions and bifurcations of papillary lines. Each minutia is represented by a tuple *m* (formula (7)), where *x*, *y* are coordinates; α is the orientation angle; *t* is the type. An open detector based on skeletonization and local analysis algorithms from OpenCV is used. The result is sets of minutiae for images $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, shown in Figures 3 and 4 for a pair from the FVC2000 (DB_1B) dataset.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

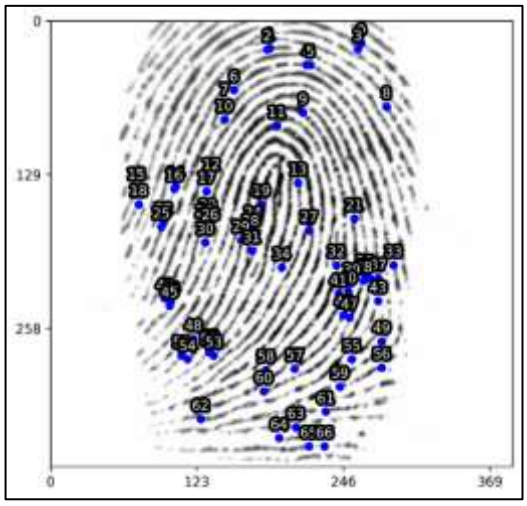


Fig. 3. Results of detecting the first scan image of the fingerprint

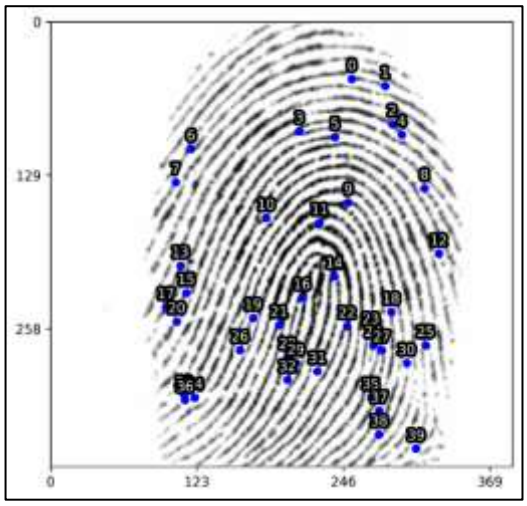


Fig. 4. Results of detecting the second scan image of the fingerprint

3. The alignment model implements the mapping A (formula (9)), calculating the shift vector s for aligning the images. Based on the minutiae sets, the coordinates and orientations are analyzed, and the optimal shift is determined. The Python implementation uses NumPy for efficient computations. The operator σ_s (formula (10)) is applied to image P_2, correcting its position relative to P_1. This compensates for the affine distortions observed in Figures 1 and 2.

4. The fingerprint matching method is implemented as a separate module and works with matching indices, calculating the distance function d (formula (13)) and similarity λ (formula (19)) for pairs of minutiae from sets L and G. The tolerance ratio $\simeq$ (formula (14)) cuts off mismatched pairs, taking into account the permissible deviations $\gamma(\Delta m)$. The function λ minimizes deviations in coordinates, angles, and minutiae types. Implementation in C# ensures high performance, and integration with Python via API allows the use of numerical libraries for tensor operations. The equivalence criterion (formula (12)) and the search for the optimal match O (formula (20)) are also implemented. Based on the shift vector s and a recursive approach (formula (21)), pairs of matching minutiae are determined by minimizing λ. The ratio (formula (22)) compares the number of optimal matches C with the threshold value $C_{threshold}$, confirming the equivalence of the prints. For a pair of images from FVC2000 (DB_1B) (see Figs. 1 and 2), the similarity score was 0.62 (in the range of 0–1), which preliminarily confirms the satisfactory performance of the method. The module is implemented on ASP.NET for integration with databases and the interface.

Hardware implementation

To ensure reliable results and efficient processing, all experiments were performed on a high-performance computing system. A workstation with an Intel Core i9-13900K processor (24 cores, 32 threads, base clock speed of 3.0 GHz, maximum clock speed of 5.8 GHz) and 64 GB of DDR5 RAM running at 5,200 MHz formed the basic hardware configuration. A 2 TB NVMe SSD was used for fast data transfer and access. To ensure compatibility with the software frameworks used in the experiments, the system was configured to run on Windows 11 Pro Build 26100.

Planning and description of experiments

The experiments in this study used the *Fingerprint Verification Competition* (FVC2000) DB1_B dataset, which is recognized as the gold standard for evaluating fingerprint recognition systems. Launched during the first fingerprint verification competition, the FVC2000 dataset is highly regarded in the field of biometric research for its consistent data collection methods and diverse fingerprint samples. The DB1_B subset was specifically selected for this study because it is small in size and suitable for evaluating algorithm productivity under controlled conditions.

The FVC2000 DB1_B dataset contains fingerprints captured using an inexpensive KeyTronic optical sensor with a resolution of 500 dpi. This dataset contains fingerprints from 10 individuals, each of whom provided eight prints from the same finger, for a total of 80 fingerprint images. The images are stored in 8-bit

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

grayscale format with a size of 300 × 300 pixels. The collection covers numerous fingerprint attributes, including varying degrees of line clarity, orientation, and inherent variations resulting from finger position or pressure.

The fingerprint scans in DB1_B were obtained under semi-controlled conditions, without strict regulations regarding finger placement or pressure, resulting in internal heterogeneity in the dataset. The optical sensor in DB1_B generates images that are characterized by moderate noise and sporadic artifacts, such as spots or incomplete ridge patterns, which complicates feature extraction and matching algorithms. These characteristics make DB1_B particularly suitable for evaluating the robustness of fingerprint recognition technologies to practical anomalies.

The dataset is publicly available and complies with the FVC2000 standard, which ensures consistency in data format and evaluation criteria across different research studies. Each photograph is given a unique identifier that indicates the person number and fingerprint number (e.g., "101_1.tif" for the first fingerprint of the first subject), which improves information management and research reproducibility.

The main objective of the experiments is to determine the optimal value of the tuple of minutiae threshold deviations $\gamma(\Delta m)$, the threshold value $C_{threshold}$, and to evaluate the performance of the method as a whole. To do this, we will use single-factor analysis methods and evaluate using binary classification metrics—assuming that each pair of images belongs to either the "match" class or the "mismatch" class. Accordingly, by calculating the derivative metrics of binary classification, we will be able to evaluate the performance of the method as a whole. We will conduct an experiment in an "all against all" format, i.e., each image in the dataset will be compared with all the others. With a dataset size of 80 images, we will perform 6,400 comparisons and obtain 6,400 results of the method's performance.

We propose the following metrics to study the productivity of the proposed method: primary metrics of binary classification (true positives, true negatives, false positives, false negatives), as well as derivative metrics such as: false positive rate (FPR), Van Rijsbergen's skewed measure with a coefficient of β = 0.5 (F_β), Matthews correlation coefficient, which is a special case of Pearson's correlation coefficient (PCC), balanced accuracy measure (*Balanced Accuracy*), as well as the method's runtime. This focus on balanced metrics and metrics for evaluating the contribution of false positive results is due to the research area—from the point of view of the authentication task, a false positive result is less desirable than a false negative, since an unauthorized access error is significantly worse than an access error in general.

A detailed description of the primary metrics is proposed to be considered as follows:

$$\begin{gathered}\forall \mathbf{P}_1, \mathbf{P}_2 \in \mathrm{P}: \\ \mathrm{R}_{12} = \left(\frac{\mathrm{R}(\mathrm{M}_1, \mathrm{M}_2)}{|\mathrm{L}|} \geq \mathrm{C}_{\text{threshold}} \right) \\ \mathrm{TP} = \mathrm{R}_{12} \wedge (\mathbf{P}_1 \sim \mathbf{P}_2) \\ \mathrm{FP} = \mathrm{R}_{12} \wedge (\mathbf{P}_1 \not\sim \mathbf{P}_2) \\ \mathrm{FN} = \neg \mathrm{R}_{12} \wedge (\mathbf{P}_1 \sim \mathbf{P}_2) \\ \mathrm{TN} = \neg \mathrm{R}_{12} \wedge (\mathbf{P}_1 \not\sim \mathbf{P}_2)\end{gathered}, \quad (23)$$

where R_{12} is the comparison result for two images according to formula (22); *TP* is the measure of true positive results; *FP* is the measure of false positive results; *TN* is the measure of true negative results; *FN* is the measure of false negative results.

It is proposed to consider the derived metrics as follows:

$$\begin{gathered}\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \\ \mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \\ \mathrm{F}_\beta = \frac{(1+\beta^2)\cdot \mathrm{P}\cdot \mathrm{R}}{\beta^2 \cdot \mathrm{P} + \mathrm{R}} \in [0,1] \\ \mathrm{MCC} = \frac{\mathrm{TP}\cdot \mathrm{TN} - \mathrm{FP}\cdot \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})\cdot(\mathrm{TP}+\mathrm{FN})\cdot(\mathrm{TN}+\mathrm{FP})\cdot(\mathrm{TN}+\mathrm{FN})}} \in [-1,1] \\ \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \in [0,1] \\ \mathrm{BAC} = \frac{\mathrm{R}+1-\mathrm{FPR}}{2} \in [0,1]\end{gathered}, \quad (24)$$

where *P* is the positive prediction rate; *R* is the completeness measure; F_β is the shifted Van Rijsenberg measure; *MCC* is the Matthews correlation coefficient; *FPR* is the false positive prediction rate; *BAC* is the balanced accuracy.

Results

The first experiment was designed to determine the optimal threshold distance Δd. Table 2 shows the results of the experiment to study the distance component of the permissible deviation of minutiae.

For clarity, we will also illustrate the results obtained using graphs (see Figs. 5 and 6).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Table 2. Method productivity metrics depending on the permissible deviation of the distance component

Testing №	Δd	F_β	MCC	FPR	BAC	Time (ms)
1	3	0.437	0.35	0	0.567	29155
2	6	0.536	0.415	0	0.594	42245
3	9	0.613	0.471	0	0.62	66942
4	12	0.689	0.535	0.0007	0.657	69515
5	15	0.735	0.586	0.003	0.697	111973
6	18	0.706	0.575	0.01	0.714	153514
7	21	0.635	0.539	0.024	0.729	141679
8	24	0.567	0.506	0.043	0.744	146207
9	27	0.499	0.47	0.068	0.753	179326
10	30	0.426	0.417	0.101	0.749	190190

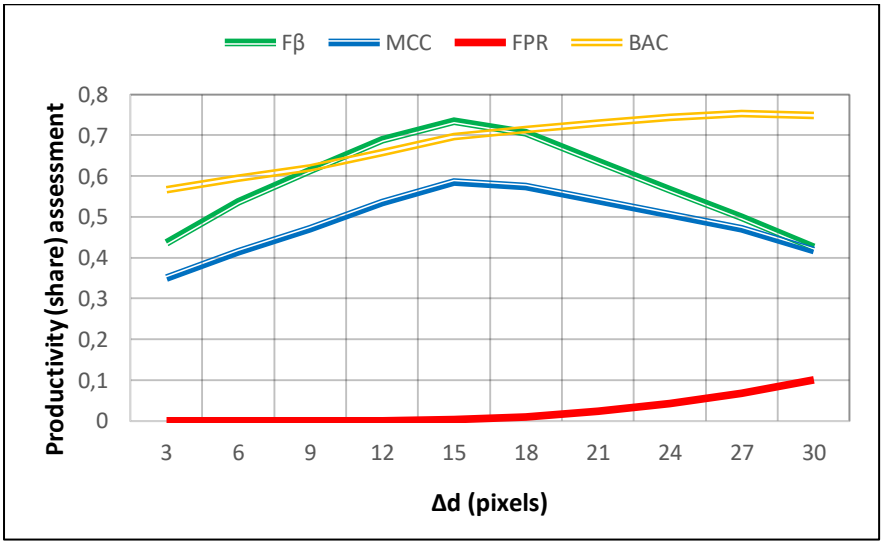


Fig. 5. Graphs of fluctuations in experimental values from experiment No. 1

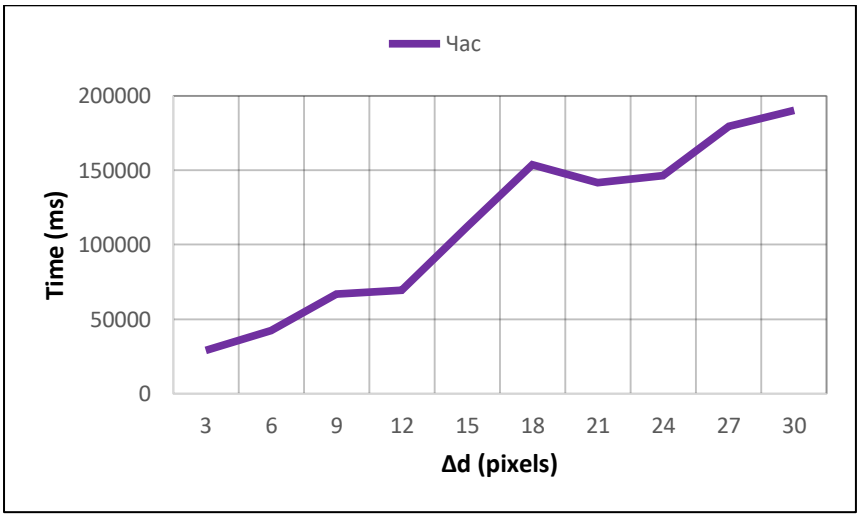


Fig. 6. Graph showing the dependence of the method's operating time on the distance limit

Several important conclusions can be drawn from the first experiment. First, the overall performance of the model has been preliminarily confirmed – a significant Van Rijsbergen measure result has been obtained (0.735 in the best case), and this performance is also illustrated by other metrics – the highest correlation

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

coefficient and balanced accuracy measure. Second, the naive assumption about finding the optimal boundary distance has been experimentally confirmed—the graph line resembles a Gaussian and has an optimal value approximately in the middle of the range of definition. Third, the experiment demonstrates an increase in computational complexity in accordance with an increase in the boundary distance. This follows directly from the features of the procedure presented in formula (21). Since we are trying to find the optimal result by adding additional shifts, the number of these additional iterations directly depends on how many positive identification results we return. Interestingly, the main increase in time costs occurs in the middle of the range of definition, while before and after it we see either a more gradual increase or even a decrease in execution time (as shown in sections (18)– (21)). Fourth, the only metric that deviates from the trend is the balanced accuracy BAC (it begins to decline significantly later than the optimal threshold distance). However, there is an explanation for this: the accuracy metric takes into account the contribution of both false positive growth and true positive growth. Depending on how we see the growth of the false positive rate, the balanced accuracy metric will increase to the extent that the growth of the true positive rate exceeds the growth of the false positive rate. Both of these increases directly result from the fact that the number of false negatives decreases as the threshold value increases, thereby increasing the true positive and false positive results.

The second experiment was designed to determine the optimal threshold angle $\Delta\alpha$. Table 3 shows the results of the experiment to study the angle component of the permissible deviation of minutiae.

For clarity, we will also illustrate the results obtained using graphs (see Figs. 7 and 8).

Table 3. Productivity metrics of the method depending on the permissible deviation of the angle component

Testing №	Δα	F_β	MCC	FPR	BAC	Time (ms)
1	3	0.475	0.374	0	0.577	47858
2	6	0.609	0.468	0	0.619	67279
3	9	0.693	0.54	0.002	0.666	86475
4	12	0.735	0.586	0.003	0.697	131448
5	15	0.732	0.589	0.005	0.705	98835
6	18	0.691	0.566	0.012	0.714	134568
7	21	0.658	0.55	0.019	0.723	190983
8	24	0.625	0.531	0.026	0.728	193243
9	27	0.602	0.515	0.03	0.727	155286
10	30	0.564	0.491	0.039	0.727	165305

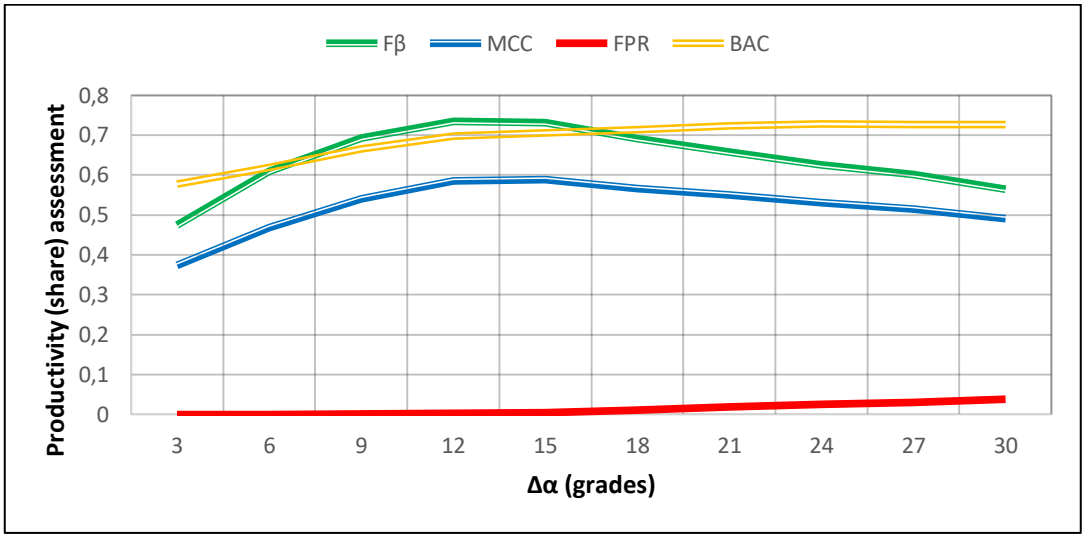


Fig. 7. Graphs of fluctuations in experimental values from experiment No. 2

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

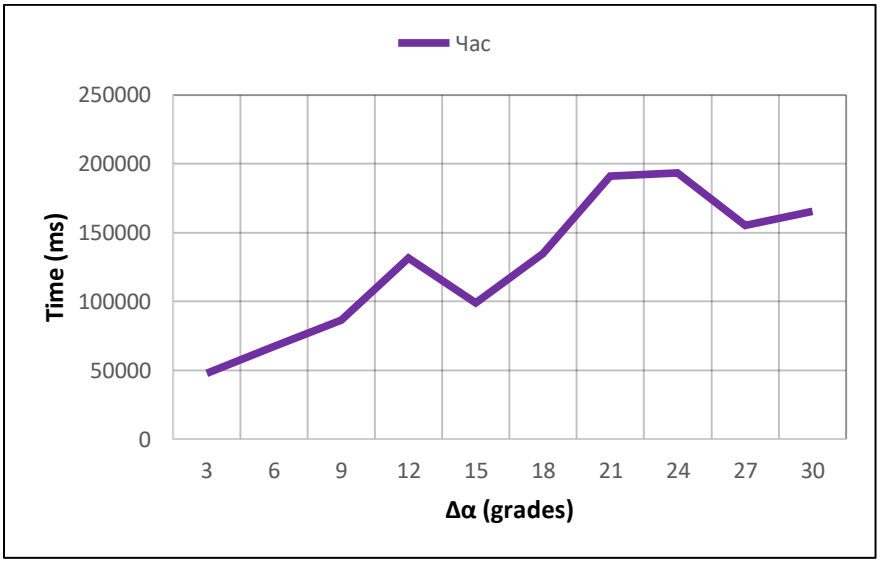


Fig. 8. Graph showing the dependence of the method's operating time on the threshold angle value

The second experiment largely confirms the results of the first – the graphs and trends are approximately the same. The optimal angle is closer to the center of the range of definition, and abnormal behavior in the method's execution time was noted during the experiment. This confirms the nonlinear dependence of the method's complexity on the thresholds and the number of images being processed.

The third experiment was designed to find the optimal value for minutiae type matching. Table 4 shows the results of the experiment studying the minutiae type component of the minutiae tolerance vector**.**

Table 4. Productivity metrics of the method depending on the permissible deviation of the minutiae component

Testing №	Δt	F_β	MCC	FPR	BAC	Time (ms)
1	0	0.733	0.583	0.003	0.695	103078
2	1	0.735	0.586	0.003	0.697	105674

The third experiment confirms the need to include minutiae type comparison in the method—under other equal conditions (such as false positive rate and execution time), the main indicators are better. That is, it is necessary to add minutiae type to the analysis.

Thus, the optimal value of the deviation vector component is $\gamma(\Delta m)$ = (15, 12, 1).

The fourth experiment was designed to find the optimal threshold value $C_{threshold}$. Table 5 shows the results of the experiment to find the optimal threshold value.

Table 5. Productivity metrics of the method depending on the threshold value

Testing №	$C_{threshold}$	F_β	MCC	FPR	BAC
1	1	0.417	0.338	0	0.5625
2	0.9	0.43	0.346	0	0.566
3	0.8	0.463	0.366	0	0.57
4	0.7	0.53	0.411	0	0.592
5	0.6	0.65	0.5	0.0003	0.637
6	0.5	0.735	0.586	0.003	0.697
7	0.4	0.652	0.556	0.023	0.737
8	0.3	0.465	0.455	0.088	0.764
9	0.2	0.225	0.258	0.377	0.712
10	0.1	0.136	0.107	0.864	0.559

For clarity, we will also illustrate the results obtained using a graph (see Fig. 9).

Therefore, the optimal threshold value is $C_{threshold}$ = 0.5 or 50% of correct minutiae from the total number. Any lower threshold value begins to lead to an overly sharp disproportionate increase in the false positive rate.

Empirical studies confirm the effectiveness of the proposed method and demonstrate the correctness of the proposed working hypothesis regarding the equivalence of fingerprints. The result achieved is high enough for the practical application of such a model.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

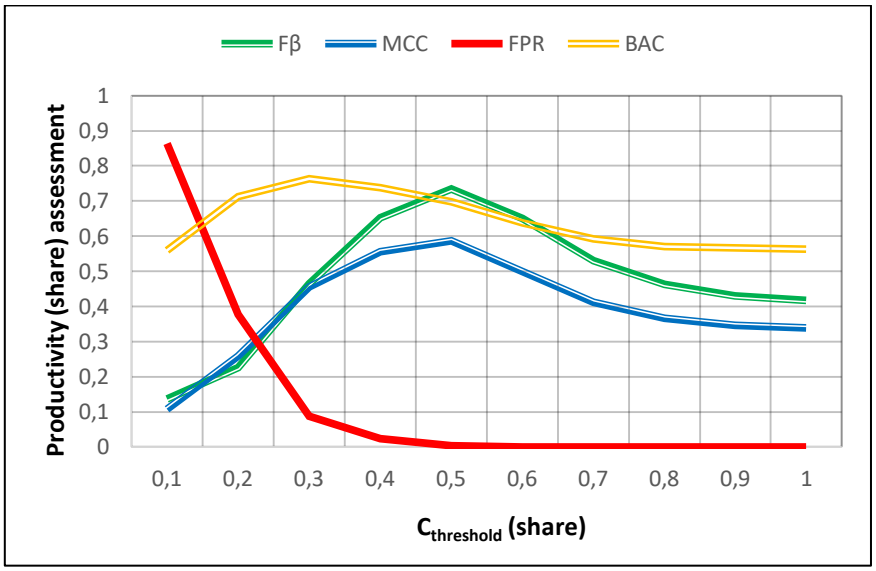


Fig. 9. Graphs of fluctuations in experimental values from experiment No. 4

Conclusions

Within the scope of the study, an original method for comparing fingerprints was proposed, based on a model of preliminary alignment and analysis of minutiae (characteristic points of papillary lines), which compensates for affine distortions by calculating a shift vector and determines the equivalence of prints through tolerance ratios and optimized minutiae matching. The mathematical model represents fingerprint images as third-order tensors in vector space, describes minutiae as tuples in mixed space considering coordinates, orientation angles, and types (breaks or bifurcations), and evaluates the equivalence of minutiae sets using a distance function, tolerance ratio, and minimization of deviations during matching.

The method is implemented as a multi-module system using Python and OpenCV for image processing, C# for high-performance computing, and ASP.NET for module integration on the FVC2000 (DB_1B) dataset. Experimental testing was performed on the specified set, which contains 80 images from 10 subjects with a resolution of 500 dpi, under conditions that simulate real scenarios with noise and artifacts, using 6400 comparisons in an “all against all” format, resulting in the following optimal parameters: vector of permissible deviation of minutiae ($\gamma(\Delta m)$ = (15, 12, 1)) for distance, angle, and type, as well as threshold value ($C_{threshold}$ = 0.5). The maximum Van Rijsbergen measure is 0.73, confirming the method's performance, while metrics such as the Matthews correlation coefficient, balanced accuracy, and false positive rate show satisfactory results.

The proposed method is characterized by resistance to affine distortions and artifacts, as well as low computational complexity, which ensures its suitability for real-time tasks and prospects in biometric systems. It is also worth noting that, despite the potential for complete resistance to affine transformations, this study focuses only on Euclidean shifts. Potential improvements may include accounting for the rotation angle of two images and the scaling effect.

References

1. Sengoopta, C. (2004), *Imprint of the Raj: how fingerprinting was born in colonial India.* London, Pan, 234 p.
2. Machado, J. H. P., Koop, B. D. O., Filipak, M., Barbosa, M. A. C., Oliva, J. T. Southier, L. F. P. (2025), "A Super-Resolution Approach for Image Resizing of Infant Fingerprints With Vision Transformers". *IEEE Access*, Vol. 13, P. 67718-67728. DOI: 10.1109/ACCESS.2025.3561206
3. Raj, J. M., Rakshitha, S., Priya, S. S., Vaishnavi, S. & Sivaranjani, A. (2020), "Latent Fingerprint Enhancement for Investigation". *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, IEEE, P. 644-648. DOI: 10.1109/ICACCS48705.2020.9074191
4. Guan, X., Pan, Z., Feng, J. & Zhou, J. (2025), "Joint Identity Verification and Pose Alignment for Partial Fingerprints". *IEEE Transactions on Information Forensics and Security*, Vol. 20, P. 249-263. DOI: 10.1109/TIFS.2024.3516566

5. Wang, S., Shen, Y. Yang, W. (2025), "Touchless Finger Vein and Fingerprint Verification via Exploiting Attention-Based Cross-Domain Fusion". *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 35, No. 4, P. 3426-3437. DOI: 10.1109/TCSVT.2024.3504270
6. Ravulakollu, K. K., Reddy, G. S. N., Jadhav, M., Batti, V., Vupputuri, V. K. Polishetty, R. K. (2025), "Limited Training Approach To Model Latent Fingerprint Data For Time-Constrained Solutions". *2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS)*, Bengaluru, India, IEEE, P. 90-95. DOI: 10.1109/COMSNETS63942.2025.10885690
7. Maltoni, D., Maio, D., Jain, A.K. & Prabhakar, S. (2009), "Handbook of Fingerprint Recognition". *2nd ed. London, Springer*, 496 p. DOI: 10.1007/978-1-84882-254-2
8. Ibragimov, A., Ferreira, F. Santos, G. (2025), "Fingerprint Pore Detection: A Survey and Comparative Analysis". *IEEE Transactions on Biometrics, Behavior, and Identity Science*, Vol. 7, No. 1. DOI: 10.1109/TBIOM.2025.3560655
9. Mani, K., Kumar, R., Singh, P. Sharma, A. (2024), "Two-Step Fingerprint Authentication Using Contour and Edge Matching". *2024 International Conference on Advanced Computer Information Technologies (ACIT)*, Rzeszow, Poland, IEEE, P. 123-128. DOI: 10.1109/ACIT62805.2024.10877179
10. Chen, L., Moon, Y.S. Lee, H.K. (2006), "Efficient Alignment of Fingerprint Images". *2006 IEEE International Conference on Image Processing*, Atlanta, GA, USA, IEEE, P. 169-172. available at: https://ieeexplore.ieee.org/document/1698755 (last accessed 01.05.2025).
11. Jain, A.K., Feng, J. Nandakumar, K. (2007), "Orientation Field Alignment for Fingerprint Matching". *Advances in Biometrics: International Conference, ICB 2007*, Seoul, South Korea, Springer, P. 745-754. DOI: 10.1007/978-3-540-74549-5_75
12. Zhang, Y., Li, X. & Wang, H. (2022), "Optimized Fingerprint Alignment for Large-Scale Databases". *2022 IEEE International Conference on Systems, Signals and Image Processing (IWSSIP)*, Sofia, Bulgaria, IEEE, P. 1-4. DOI: 10.1109/SILCON55242.2022.10028828
13. Gu, J., Zhou, J. Zhang, C. (2006), "Fingerprint Matching Using Ridges". *Pattern Recognition*, Vol. 39, No. 11, P. 2131-2140. DOI: 10.1016/j.patcog.2006.04.015
14. Li, X., Zhang, Y. Chen, L. (2021), "Ridge Pattern Analysis for Low-Quality Fingerprint Images". *2021 International Conference on Smart Cities and Energy Efficiency (ICSCEE)*, Istanbul, Turkey, IEEE, P. 1-6. DOI: 10.1109/ICSCEE50312.2021.9497996
15. Wang, H., Li, Y. Zhang, X. (2023), "Localized Deep Representation for Fingerprint Matching". *arXiv preprint arXiv:2311.18576v2*, 15 p. available at: https://arxiv.org/html/2311.18576v2 (accessed 01.05.2025)
16. Liu, Z., Zhang, Q. Wang, S. (2021), "Transformer-Based Fingerprint Representation for Matching". *IEEE Transactions on Information Forensics and Security*, Vol. 16, P. 4867-4878. DOI: 10.1109/TIFS.2021.3134867
17. Kim, S., Park, J. Lee, H. (2022), "Hybrid Neural Networks for Fingerprint Matching in Noisy Conditions". *2022 International Conference on Future Trends in Intelligent Computing (ICFTIC)*, Shanghai, China, IEEE, P. 45-50. DOI: 10.1109/ICFTIC57696.2022.10075139
18. Yang, J., Chen, X. Zhang, L. (2021), "Deep Learning for Complex Fingerprint Patterns". *IEEE Transactions on Information Forensics and Security*, Vol. 16, P. 3921-3932. DOI: 10.1109/TIFS.2021.3139219
19. Park, T., Kim, Y. Lee, S. (2020), "Algorithm for Low-Quality Fingerprint Image Matching". *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Niteroi, Brazil, IEEE, P. 123-128. DOI: 10.1109/ICSSIT48917.2020.921420
20. Zhao, Q., Li, J. Wang, X. (2025), "Latent Fingerprint Processing for Crime Scene Identification". *2025 IEEE Workshop on Applications of Computer Vision (WACV Workshops)*, Tucson, AZ, USA, IEEE, P. 56-61. DOI: 10.1109/WACVW65960.2025.00159
21. Xu, L., Zhang, Y. Chen, H. (2025), "Local Correlation-Based Matching for Partial Fingerprints". *IEEE Access*, Vol. 13, P. 12345-12354. DOI: 10.1109/ACCESS.2025.3555311
22. Lee, H., Kim, S. Park, J. (2024), "Secure Protocol for Fingerprint Matching with Encrypted Data". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 46, No. 8, P. 5678-5690. DOI: 10.1109/TPAMI.2024.3486179
23. Choi, Y., Lee, H. Kim, D. (2023), "Enhanced Encryption Protocols for Biometric Systems". *2023 International Conference on Biometrics and Signal Processing (BIOSIG)*, Darmstadt, Germany, IEEE, P. 89-94. DOI: 10.1109/BIOSIG58226.2023.10345974
24. Wu, X., Zhang, L. Chen, Y. (2025), "Obstructive Minutiae Removal for Latent Fingerprints". *IEEE Sensors Letters*, Vol. 9, No. 3, P. 1-4. DOI: 10.1109/LSENS.2025.3550874
25. Han, C., Li, X. Wang, Y. (2024), "Optimized Computations for Automated Fingerprint Identification Systems". *2024 International Conference on Systems and Electronics (ICSES)*, Singapore, IEEE, P. 34-39. DOI: 10.1109/ICSES63445.2024.10763052
26. Kim, D., Park, J. Lee, S. (2023), "New Architectures for Fingerprint Recognition Accuracy". *2023 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Taipei, Taiwan, IEEE, P. 123-128. DOI: 10.1109/APSIPAASC58517.2023.10317455
27. Park, J., Kim, Y., Lee, H. (2023), "Deep Learning for Complex Fingerprint Deformations". *2023 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, Hangzhou, China, IEEE, P. 56-61. DOI: 10.1109/AICAS57966.2023.10168628
28. Lim, S., Chen, X., Zhang, Y. (2020), "Adaptive Filtering for Fingerprint Preprocessing". *2020 IEEE International Conference on Advances in Science, Engineering and Technology (ASET)*, Dubai, UAE, IEEE, P. 78-83. DOI: 10.1109/ASET48392.2020.9118193

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

29. Wulandari, R., Santoso, A., Nugroho, H. (2021), "Fingerprint Identification Using Edge Detection and GLCM Analysis". *2021 International Conference on Computer Science (ICCS)*, Jakarta, Indonesia, IEEE, P. 123-128. DOI: 10.1109/ICITech50181.2021.9590134
30. Patel, R., Sharma, S. & Kumar, V. (2021), "Technique for Processing Noisy Fingerprints with Extreme Distortions". *2021 International Conference on Computer Systems (ICCS)*, Bangalore, India, IEEE, P. 45-50. DOI: 10.1109/ICCS54944.2021.00068
31. Yu, T., Zhang, X. & Li, Y. (2025), 3D Fingerprint Unwrapping Using B-Spline for 2D Recognition. *IEEE Open Journal of the Computer Society*, Vol. 6, P. 123-134. DOI: 10.1109/OJCS.2025.3559975
32. Calderón-Calderón, M.D., Medina-Pérez, M.A. & Monroy, R. (2025), "Fingerprint Quality Enhancement for Low-Resolution Scanners. *IEEE Access*, Vol. 13, P. 56789-56798. DOI: 10.1109/ACCESS.2025.3527071
33. Pohuliaiev, Y. (2025), "Euclidean fingerprint descriptor and comparator model software". available at: https://drive.google.com/file/d/1bRD2yqc7Zg6NSBpGYvtAVv-hiJrekJJh/view?usp=sharing (last accessed: 01.05.2025)

Надійшла (Received) 03.06.2025

Відомості про авторів / About the Authors

Pohuliaiev Yurii – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Software Engineering, Kharkiv, Ukraine; e-mail: yurii.pohuliaiev@nure.ua; ORCID ID: 0009-0005-5883-1661

Smelyakov Kirill – Doctor of Sciences (Engineering), Professor, Head at the Department of Software Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; e-mail: kyrylo.smelyakov@nure.ua; ORCID ID: 0000-0001-9938-5489

Погуляєв Юрій Сергійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна.

Смеляков Кирило Сергійович – доктор технічних наук, професор, завідувач кафедри програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

МЕТОД ПОРІВНЯННЯ ВІДБИТКІВ ПАЛЬЦІВ, ЗАСНОВАНИЙ НА МОДЕЛІ ПОПЕРЕДНЬОГО ВИРІВНЮВАННЯ

Предметом статті є розробка та аналіз нового методу порівняння відбитків пальців, що використовує геометричні евклідові характеристики мінуцій для біометричної ідентифікації. Дослідження зосереджується на мінуціях — специфічних точках переривання або роздвоєння папілярних ліній — як ключових біометричних ознаках, протиставляючи їх традиційним глобальним і локальним дескрипторам, таким як SIFT, HOG чи LPQ. **Мета** полягає у розробці стійкого й ефективного методу порівняння відбитків пальців, що використовує геометричні евклідові характеристики мінуцій і попереднє вирівнювання, для підвищення точності біометричної ідентифікації без залежності від машинного навчання. **Завдання** дослідження триетапні: по-перше, дослідити теоретичні основи дескрипторів на основі мінуцій та їхню інваріантність до афінних перетворень (зсув, обертання, масштабування); по-друге, розробити модель із використанням вектора зсуву та функцій відстані для зіставлення мінуцій; по-третє, експериментально перевірити модель, визначивши оптимальні параметри та оцінивши її на стандартному наборі даних. **Методи** охоплюють теоретичний аналіз і експериментальну оцінку. Теоретична основа встановлює стійкість вирівнювання до зсувів. Дескриптор формується через координати мінуцій, функції відстані та оптимізацію зіставлення. Для вилучення ознак застосовано алгоритм обробки зображень із фільтрацією й аналізом мінуцій. **Результати** отримані з допомогою експериментальної перевірки на наборі FVC2000 (DB1_B) демонструють високу продуктивність, яка оцінюється метриками класифікації та часом виконання. **Висновки** підкреслюють теоретичні та практичні надбання дослідження. Модель демонструє теоретичну й практичну стійкість до евклідових зсувів, пропонуючи переваги для обробки відбитків із різних сканерів. Експерименти підтверджують ефективність виявлення зсувів, досягаючи високої міри Ван Різбергена (0.735), хоча залежність від позитивних збігів вимагає додаткової фільтрації помилково позитивних результатів. Метод є робочим і може бути застосований практично.

Ключові слова: дактилоскопія, відбитки пальців, попереднє вирівнювання, мінуції, біометрична ідентифікація, афінні перетворення, тензорне представлення, машинне навчання, FVC2000, метрики бінарної класифікації..

Бібліографічні описи / Bibliographic descriptions

Погуляєв Ю.С., Смеляков К.С. Метод порівняння відбитків пальців, заснований на моделі попереднього вирівнювання. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 88–101. DOI: https://doi.org/10.30837/2522-9818.2025.3.088

Pohuliaiev, Y., Smelyakov, K. (2025), "Method for comparing fingerprints based on a previous alignment model", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 88–101. DOI: https://doi.org/10.30837/2522-9818.2025.3.088

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

УДК 004.94 DOI: https://doi.org/10.30837/2522-9818.2025.3.102

В. Поддубний, О. Сєвєрінов

МОДЕЛЬ АВТЕНТИФІКАЦІЇ КОРИСТУВАЧІВ З ВИКОРИСТАННЯМ НУЛЬОВОГО ВОДЯНОГО ЗНАКУ НА ОСНОВІ RGB-ЗОБРАЖЕНЬ

Предметом дослідження методи та схеми автентифікації користувачів в інформаційно-комунікаційних мережах. **Мета роботи** – розробка моделі автентифікації користувачів в інформаційно-комунікаційних системах з використанням методу нульового водяного знаку на основі RGB-зображень **завдання:** розгляд основних проблем під час інтеграції систем контролю доступом; розробка моделі автентифікації користувачів з використанням нульового водяного знаку; дослідження потенційних характеристик та висування пропозицій/рекомендацій щодо її використання для автентифікації, побудова графічних схем, що відображають процеси та процедури Для виконання окреслених завдань використовувалися такі **методи**: методи моделювання, методи формального опису **Досягнуті результати:** запропонована вдосконалена модель автентифікації користувачів заснована на алгоритмі нульового водяного знаку, що являється альтернативою існуючим схемам автентифікації в інформаційно комунікаційних системах, досліджені її потенційні характеристики, що підтверджують перспективність розробки, побудовані графічні представлення процесів та процедур **Висновки** під час роботи було коротко розглянуто основні проблеми з якими стикаються власники інформаційно-комунікаційних систем при інтеграції систем автентифікації, запропоновано вдосконалену модель автентифікації за допомогою нульового водяного знаку, в якості нульового водяного знаку використовувався метод заснований на RGB-зображеннях з використанням перетворення K-means та DWT, в якості другого фактору використовується пароль. Представлено три режими роботи: реєстрація, автентифікація та зміна ключових даних для автентифікації. Така модель є альтернативою існуючим схемам та можуть потенційно зменшити вартість автентифікації, кількість помилок першого та другого роду у порівняні з аналогічними схемами автентифікації, підвищити зручність автентифікації в інформаційно-комунікаційних системах. Також це розширює множину вибору схем автентифікації що мають власники систем при їх інтеграції.

Ключові слова: цифровий водяний знак; автентифікація; зображення; модель; пароль.

Вступ

Зі зростанням ролі інформаційних технологій та їх активним упровадженням у різні сфери діяльності посилюється необхідність використання засобів автентифікації користувачів, зокрема під час отримання доступу до приміщень, технічних засобів чи інформації в інформаційно-комунікаційних системах. Традиційно такі засоби ґрунтуються на паролях, магнітних картках, біометричних даних або QR-кодах. Одним із шляхів підвищення рівня захисту є поєднання зазначених методів у межах багатофакторної автентифікації [1].

Разом з тим, такі системи характеризуються значними витратами, оскільки вимагають застосування спеціалізованого обладнання (наприклад, зчитувачів відбитків пальців чи інших біометричних сенсорів). Використання карток або кодів також створює додаткові ризики: наявність фізичного носія сигналізує потенційному зловмиснику про його важливість, що стимулює спроби викрадення, копіювання чи підміни даних.

Альтернативним підходом є застосування нульових водяних знаків для побудови систем автентифікації. Ключова перевага цього методу полягає в тому, що критичні дані не змінюються та залишаються невидимими для зловмисника, оскільки він не має змоги визначити їхню наявність. Крім того, реалізація подібних рішень є відносно маловартісною, адже водяний знак може ґрунтуватися на різних типах інформації, для опрацювання якої не потрібне спеціальне обладнання [2, 3].

Основна проблематика застосування нульових водяних знаків у багатофакторній автентифікації полягає у новизні технології та недостатній кількості досліджень у цій сфері. Водночас перспективи її використання видаються обнадійливими, особливо з огляду на інтеграцію з сучасними IoT-засобами. Таким чином виникає необхідність в дослідженні та розробці нових методів автентифікації заснованих на нульовому водяному знакові.

Аналіз проблеми

Автентифікація поділяється за кількістю факторів авторизації (фактор – набір доказів які пред’являє користувач для підтвердження особи, так вона може бути однофакторною та багатофакторною.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Для забезпечення належного рівня безпеки рекомендується використовувати мінімум двофакторну автентифікацію [4, 5].

Під час побудови систем автентифікації власники систем зустрічаються з наступними проблемами:

- вибору фактору;
- вибору схеми автентифікації.
- вартості;
- безпеки;
- зручності;
- завадозахищеності.

Проблема вибору фактору полягає у визначені які фактори слід використовувати (знання чогось, володіння чимсь, характеристики об'єкту). Різні фактори мають різні характеристики вартості, надійності, зручності ітд. Власнику системи слід визначитись які фактори слід використовувати, як їх поєднувати, інтегрувати в систему тощо.

Проблема вибору схеми автентифікації заключається у виборі порядку поєднання та впливу факторів, так існують послідовні, паралельні та послідовно паралельні схеми автентифікації. Власнику системи слід визначайсь в якому порядку використовувати фактори автентифікації та як обчислювати кінцевий результат автентифікації.

Проблема вартості полягає у зменшені показників вартості кожної авторизації, вартості обслуговування системи автентифікації, вартості додаткових матеріалів та засобів для автентифікації кожного користувача.

Проблема безпеки закладається у забезпеченні достатнього рівня достовірності авторизації. Під безпекою розуміється характеристика яка є сукупністю вірогідностей помилок першого та другого роду, можливостей злому системи зловмисником, можливістю викрадення даних, їх підробкою.

Проблема зручності полягає у максимізації суб'єктивних параметрів зручності авторизації (складності дій, швидкості, необхідності запам'ятовувати дані тощо).

Проблема завадозахищеності з'являється в системах в яких рівень шуму вищий ніж 0, вона полягає у мінімізації помилок першого та другого роду при автентифікації у зв'язку з виникненням помилок в обробці чи введенні наборів даних.

Тому при інтеграції багатофакторної автентифікації власники систем зазвичай не розробляють власні моделі а використовують наявні, у зв'язку з чим виникає необхідність в розробці готових моделей для мінімізації та вирішення наведених вище проблем та створені множини вибору в залежності від пріоритетів власників систем.

Аналіз літератури

Схеми автентифікації за допомогою нульових водяних знаків вже відомі. Так наприклад в роботі [6] наводиться огляд поточного стану досліджень у сфері автентифікації зображень, зокрема з використанням методів нульового водяного знаку. В роботі [7] представлений метод є забезпечення автентичності медичних зображень з використанням схеми НВЗ, що зберігає оригінальність зображення (без вбудовування змін до пікселів). В роботі [8] наводиться приклад використання нульових водяних знаків разом з QR-кодами. Автори пропонують схему автентифікації медичних зображень, яка поєднує НВЗ та QR-код. В роботі[9] Автори пропонують гібридний метод, який поєднує цифровий підпис та НВЗ для забезпечення автентичності, цілісності, виявлення маніпуляцій і захисту прав автора для чутливих текстових документів.

Як можна побачити кількість наявних схем автентифікації з використанням НВЗ досить велика, проте дані схеми не підходять для автентифікації користувачів в інформаційно-комунікаційних системах. Схеми автентифікації що використовують в якості контейнеру текст не можуть бути використані для автентифікації користувачів в інформаційно-комунікаційних системах так як користувачу незручно вводити великі текстові дані, а при введенні тексту з носія краще використовувати дані криптографічних перетворень. Дані методи підходять для машинної автентифікації або для підтвердження авторства.

Тому виникає необхідність у вдосконаленні існуючих моделей автентифікації користувачів саме для інформаційно-комунікаційних систем, де в якості одного з факторів використовується цифрове зображення.

Модель багатофакторної автентифікації на основі методу нульового водяного знаку

Проведений аналіз показав, що для вирішення наведених проблем запроваджуються нові методи

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

багатофакторної автентифікації, серед яких розповсюдженості набули методи, засновані на використанні OTP-кодів, адаптивна та поведінкова автентифікація, WebAuthn та інші. Досить новою є автентифікація за допомогою нульових водяних знаків. На даний момент запропоновано декілька таких методів [10]. Один з достатньо простих в використанні є запропонований авторами в роботах [11, 12] метод багатофакторної автентифікації на основі пароля та RGB-зображення.

Метод формування водяного знаку детально наведений в роботі [13]. Проте відсутні моделі використання даного та аналогічних методів для автентифікації користувачів в інформаційно-комунікаційних системах.

Розглянемо загальну модель використання нульового водяного знаку для автентифікації. В такій моделі основні дії користувача це:

– реєстрація;
– автентифікація;
– зміна ключових даних.

Дії користувача виконуються послідовно, тобто спочатку реєстрація, автентифікація за необхідності зміна ключових даних. Зміна ключових даних без автентифікаціїх неможлива, в свою чергу автентифікація неможлива без першочергової реєстрації.

Для системи автентифікації за допомогою нульових водяних знаків обрана типова схема для багатьох систем та засобів які здійснюють процедуру автентифікації. Проте виникає необхідність в деталізації процесів, ключових даних, параметрів та інших даних при використанні нульового водяного знаку як алгоритму підтвердження правильності автентифікації.

Запропонована модель автентифікації за допомогою нульових водяних знаків зображена на рис.1.

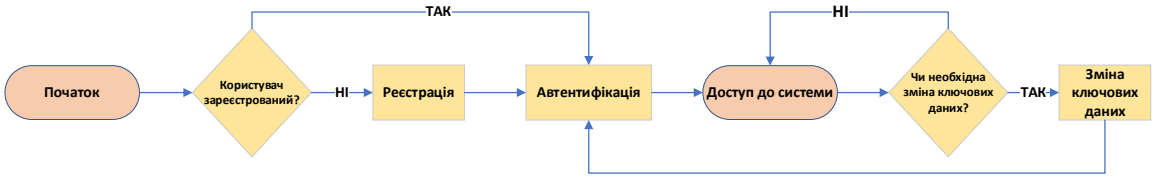


Рис. 1. Модель автентифікації

В роботі [12] розглянуті основні компоненти роботи типової моделі автентифікації з використанням нульового водяного знаку. Тому виникає необхідність в деталізації процедур кожного окремого процесу. Метод формування водяного знаку та розширення ключа не буде розглядатися детально, оскільки він описаний в роботі [13].

В такій системі в якості ключових даних виступають:

- пароль обраний користувачем;
- цифрове RGB зображення.

У якості підтвердження автентифікації використовується результат порівняння сформованого нульового водяного знаку з еталонною копією.

Процес реєстрації користувача з використанням запропонованого методу водяного знаку полягає в послідовному виконанні наступних кроків:

- крок 1. Користувач обирає ключем довільне зображення O (або генерує його засобами системи, в якій інтегрована схема, в тому числі в якості генератора зображень може використовуватися штучний інтелект):

$$O \in \{0, \dots, 255\}^{H*W*3}, \quad (1)$$

де $H, W \in \mathbb{N}$ – висота та ширина вхідного зображення. Де $\mathbb{N} \in \{1, 2, 3 \dots \infty\}$;

- крок 2. Користувач обирає випадковий пароль доступу $PASS$, в залежності від політики паролів організації

$$PASS \in \{0, \dots, 255\}^{L}, \quad (2)$$

де L – кількість символів в паролі;

- крок 3. Користувач вводить в систему (за допомогою засобів зчитування та цифровізації IT) обране зображення O та пароль $PASS$

$$O \rightarrow IT, \ PASS \rightarrow IT; \quad (3)$$

- крок 4. Під час реєстрації генерується нульовий водяний знак W, де вхідні дані – це ключ $PASS$ (або його розгорнутий варіант) та обране зображення O:

$$W = F(O, PASS); \quad (4)$$

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

- крок 5. Сформований нульовий водяний знак зберігається в базі даних системи як еталонна копія

$$W \rightarrow DB. \quad (5)$$

Процес реєстрації користувача в системі з використанням запропонованого методу водяного знаку представлений на рис. 2.

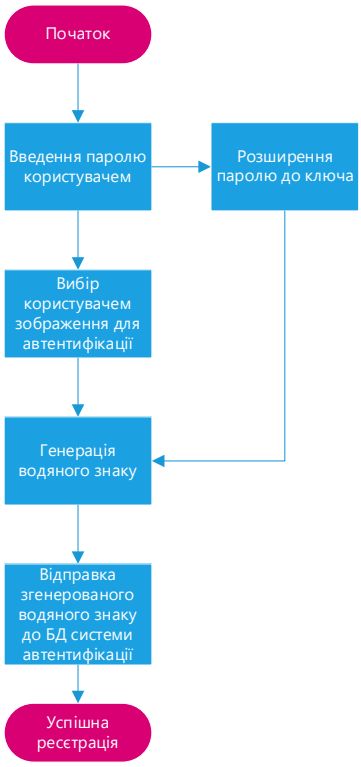


Рис. 2. Процес реєстрації користувача в системі

Для процедури автентифікації на основі запропонованого методу водяного знаку пропонується обрати стандартну послідовність дій, що складається з наступних кроків:

- крок 1. Користувач вводить в систему IT власний пароль $PASS$ та засобами системи зчитує зображення O, яке він обрав як ключ, аналогічно (3);
- крок 2. Генерується нульовий водяний знак, де вхідні дані – це ключ (або його розгорнутий варіант) та обране зображення:

$$W' = F(O, PASS); \quad (6)$$

- крок 3. Модуль автентифікації порівнює згенерований нульовий водяний знак з еталонним нульовим водяним знаком W з бази даних за допомогою функції порівння S, повертаючи певне значення подібності x:

$$x = S(W', W); \quad (7)$$

- крок 4. Система порівнює коефіцієнт схожості [14] водяного знаку з еталонним нульовим знаком (x) зі встановленим значенням допустимої похибки (y); якщо x < y, то користувачеві блокується доступ до системи, а якщо x ≥ y, то користувачеві надається доступ до системи.

На рис 3. наведений процес автентифікації користувача за допомогою запропонованої моделі.

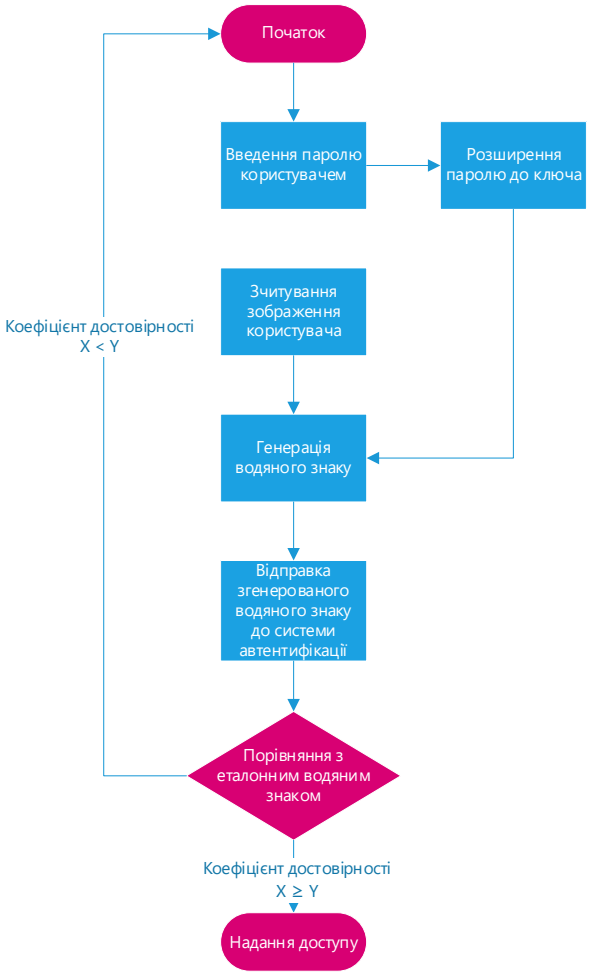


Рис. 3. Процес автентифікації за допомогою запропонованої моделі

Для зміни ключових даних при використанні запропонованого методу водяного знаку пропонується здійснити типові кроки:

– крок 1. Користувач повинен пройти успішну автентифікацію, для надання йому повноважень щодо зміни ключових даних:

$$Auth \Leftrightarrow x \geq y, \quad (8)$$

– крок 2. Користувач після успішної автентифікації повинен заново ввести нові уключові дані;

$$O' \rightarrow IT, \; PASS' \rightarrow IT, \quad (9)$$

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

– крок 3. Метод формування нульового водяного знаку здійснює повторне формування водяного знаку за новими введеними ключовими даними

$$W' = F(O', PASS'), \quad (10)$$

– крок 4. Новосформований водяний знак записується в базу даних системи автентифікації

$$W' \rightarrow DB, \quad (11)$$

На рис 4. Наведений процес зміни даних автентифікації користувача за допомогою запропонованої моделі.

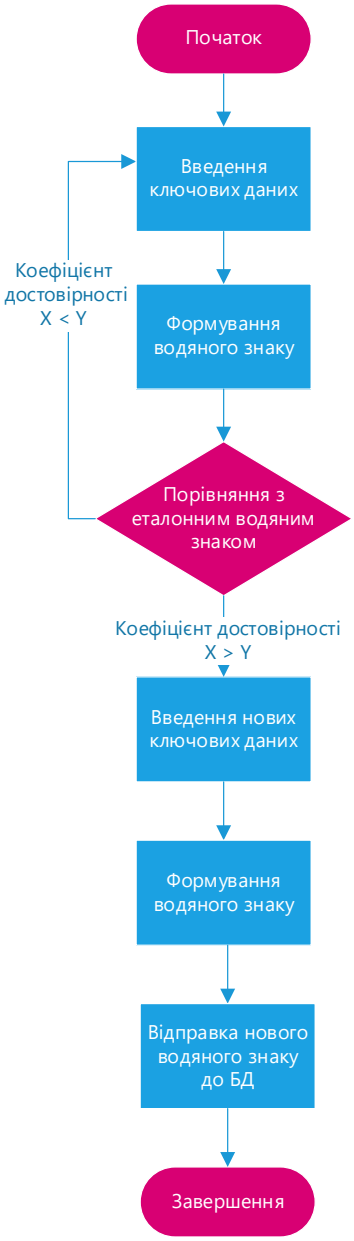


Рис 4. Процес зміни даних користувача за допомогою запропонованої моделі

За умови застосуванні IoT-пристроїв така модель може контролювати доступ до приміщень або певної інформації. Засобом зчитування зображення можуть бути відеокамери, для перетворення цифрового водяного знаку можна використовувати мікроконтролери. Після обчислення перетворення цифрового водяного знаку передається результат на сервер доступу організації, що перевіряє:

1) коректність введеної інформації;

2) чи має користувач з такими даними доступ до певного ресурсу.

Оскільки обробка зображення складніший процес ніж формування та перевірка геш-значення, то рекомендується додатково зберігати геш-значення від паролю $PASS$ в базі даних системи автентифікації DB

$$H(PASS) \rightarrow DB \quad (12)$$

Під час автентифікації перевіряти геш-значення H' від введеного пароля $PASS`$ та між еталонним значенням H перед формуванням водяного знаку. Це може пришвидшити процедуру автентифікації відкидаючи одразу помилки в паролі.

В запропонованому варіанті процедура реєстрації та автентифікації не містить додаткових ідентифікаторів користувача для авторизації (прав доступу, ролей ітд), для використання рольової моделі пропонується використовувати ідентифікатор у вигляді логіну при записі в базу даних. Таким чином формується комбінація логін-водяний знак. Для невеликих систем, або систем де не потрібна адміністративний розподіл, може здійснюватися перевірка лише на наявність схожого водяного знаку в базі даних.

Потенційні характеристики запропонованої моделі

Розглянемо дану модель с точки зору потенційного вирішення вищезазначених проблем.

Проблема вибору фактору. Дана модель ґрунтується на двох факторах, це знання чогось – пароль, та володіння чимось – зображення. Тобто вона пропонує вже готове рішення поєднання факторів автентифікації.

Проблема вибору схеми автентифікації. Модель являється послідовною, оскільки фактори в ній застосовуються один за одним. Кінцевим результатом застосування є водяний знак, що перевіряється схемою автентифікації. Модель пропонує готову схему поєднання факторів.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

Проблема вартості. Розглянемо загальну вартість моделі автентифікації з факторами «занання чогось» та «володіння чимсь»:

$$C = x * z + K_з + K_в + (A_з + A_в) * x, \quad (13)$$

де $A_з$ – вартість адміністрування системи знання (створення облікових записів, зміна паролів) у розрахунку на одного користувача;

$A_в$ – вартість адміністрування системи володіння (створення облікових записів, зміна засобів, видача та контроль) у розрахунку на одного користувача;

x – кількість користувачів;

z – вартість одного засобу для користувача;

$K_з$ – вартість комплексу засобів захисту для роботи з знанням (засоби вводу);

$K_в$ – вартість комплексу засобів захисту для роботи з інформацією (засоби зчитування то обробки інформації).[15]

Запропонована модель потенційно знижує z – вартість засобу доступу (носій з інформацією, магнітна картка, ключ тощо), оскільки збереження та розповсюдження цифрового зображення не потребує багатьох ресурсів, це може бути аркуш паперу, особистий телефон співробітника чи зовнішнє інтернет посилання.

Проблема безпеки. Розглянемо потенційну надійність моделі автентифікації. Математична складність перебору пароля:

$$S(pass) = y^c, \quad (14)$$

де *c*- кількість символів в паролі; *y* – розмір алфавіту (діапазон символів які може набувати пароль).

Потенційні варіації зображення:

$$S(image) = 256^{M*M*3} = 2^{24M^2}, \quad (15)$$

де 256 – значення яскравості кожного біту; M – розмір зображення; 3 – кількість спектрів (R,G,B).

Дані характеристики зменшуються, оскільки самий метод водяного знаку містить колізії та толерантний до утворення шумів та викривлень, тому маючи початкові високі показники надійності сам модель формування водяного знаку потребує додаткового вивчення з метою оцінки його криптостійкості.

Проблема зручності. З точки зору користувача "зручність" це перш за все простота використання. Модель використовує широко розповсюджені паролі які є зрозумілим та розповсюдженим методом автентифікації , в якості другого фактору використовується цифрове зображення, яке також зрозуміле більшості користувачам. Навіть користувачі які не мають досвіду роботи з інформаційно-комунікаційними системами можуть зрозуміти принцип використання зображення для автентифікації. Тим самим дана модель має високу зручність та гнучкість при використанні її для автентифікації[16][17].

Проблема завадозахищеності. Дана проблема виникає в мережах з певним рівнем шуму. Запропонована модель має певну толерантність (як показано в роботі [13] водяний знак може бути розпізнаний з мінімальною кількістю помилок при рівні шуму13 PSTR[18]). Модель стійка до шумів при:

– зчитування зображення в систему автентифікації;

– при передачі сформованого водяного знаку для перевірки з еталонним водяним знаком.

Тобто не обов'язково $W' = W$ та $O' = O$, якщо система налаштована відповідним чином.

Отже запропонована модель є досить практичною з точки вирішення основних проблем при вибору моделей автентифікації, та пропонує власникам систем додаткову гнучкість при побудові систем автентифікації. Проте деякі аспекти моделі (такі як криптографічна стійкість, практична зручність) потребують додаткового вивчення та дослідження.

Висновок

В роботі запропонована модель багатофакторної автентифікації за допомогою методу нульового водяного знаку заснованого на RGB-зображеннях. Дана розроблена модель є альтернативою існуючим моделям та потенційно зменшити вартість автентифікації, кількість помилок першого та другого роду у порівняні з аналогічними схемами автентифікації, підвищити зручність автентифікації в інформаційно-комунікаційних системах. Також це розширює множину вибору схем автентифікації що мають власники систем. Така модель потенційно вирішує проблеми що виникають при побудові систем автентифікації, проте потребує дослідження з точки зору крипостійкості методу формування нульового знаку та практичної зручності.

Перевагою такої моделі є її стійкість до виникнення шумів при введенні ключових даних чи при передачі результату автентифікації. Додатковими перевагами є: використання двох факторів автентифікації, висока зрозумілість користувачам.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Тому хоч запропонована модель є перспективною, пропонуються наступні вектори дослідження:

– дослідження криптографічної стійкості методу нульового водяного знаку, та схеми автентифікації в цілому;

– практична реалізація моделі з метою дослідження швидкодії та вартості ресурсів необхідних для автентифікації;

– практична реалізація моделі з метою тестування зручності на вибірці користувачів.

Список літератури

1. Lal N. A., Prasad S., Farik M. A Review of Authentication Methods. *International Journal of Scientific & Technology Research*. 2016. Vol. 5, No. 11. P. 246–249. URL: https://www.researchgate.net/publication/311514269_A_Review_Of_Authentication_Methods
2. Arévalo-Ancona R. E., Cedillo-Hernandez M., Ramirez-Rodriguez A. E., Nakano-Miyatake M., Perez-Meana H. Secure Medical Image Authentication Using Zero-Watermarking Based on Deep Learning Context Encoder. *Computación y Sistemas*. 2024. Vol. 28, No. 1. P. 199–210. DOI: 10.13053/cys-28-1-4898
3. Seenivasagam V., Velumani R. A QR Code Based Zero-Watermarking Scheme for Authentication of Medical Images in Teleradiology Cloud. *Computational and Mathematical Methods in Medicine*. 2013. Article ID 516465. 10 p. DOI: 10.1155/2013/516465
4. Chapple M. *Implementing Access Control Systems*. Indiana (USA): Wiley, 2020. P. 110–135.
5. Айко О. В., Василенко О. Д., Степаненко В. М. Системи автоматизації контролю доступу на об’єкти. *Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених*. 2023. P. 106-108. URL: https://ela.kpi.ua/server/api/core/bitstreams/e4de99d9-a572-452d-8194-3ba4a39f5bb4/content;
6. Qi X., Liu Y. Cloud Model Based Zero-Watermarking Algorithm for Authentication of Text Document. *Proceedings of the 2013 Ninth International Conference on Computational Intelligence and Security* (Emeishan, China, 2013). 2013. P. 712–715. DOI: 10.1109/CIS.2013.15
7. Hasan B. M. S., Ameen S. Y., Hasan O. M. Image Authentication Based on Watermarking Approach: Review. *Asian Journal of Research in Computer Science And Information Technology*. 2021. Vol. 9, No. 3. P. 34-51. DOI: 10.9734/AJRCOS/2021/v9i330224
8. Arévalo-Ancona, R. E., Cedillo-Hernández, M., Ramírez-Rodríguez, A. E., Nakano-Miyatake, M., Pérez-Meana, H. Secure Medical Image Authentication Using Zero-Watermarking Based on Deep Learning Context Encoder. *Computación y Sistemas*. 2024. Vol. 28, No. 1 P. 199-210. DOI: 10.13053/CyS-28-1-4898
9. Seenivasagam V., Velumani R. A QR Code Based Zero-Watermarking Scheme for Authentication of Medical Images in Teleradiology Cloud. *Computational and Mathematical Methods in Medicine*. 2013. Vol. 2013, Article ID 516465, 16 p. DOI: 10.1155/2013/516465
10. Tayan O., Kabir M. N., Alginahi Y. M. A Hybrid Digital-Signature and Zero-Watermarking Approach for Authentication and Protection of Sensitive Electronic Documents. *The Scientific World Journal*. 2014. Article ID 514652. 10 p. DOI: 10.1155/2014/514652
11. Поддубний В., Сєвєрінов О., Непокритов Д. Дослідження ефективності алгоритмів оброблення зображень у схемах нульового водяних знаків. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 1(31). С. 102–114. DOI: 10.30837/2522-9818.2025.1.102
12. Поддубний В. О., Сєвєрінов О. В., Гвоздьов Р. Ю. Використання нульових водяних знаків для підтвердження авторства зображень та багатофакторної автентифікації. *Радіотехніка*. 2024. Вип. 218. С. 35–43. DOI: 10.30837/rt.2024.3.218.02
13. Poddubnyi V., Sievierinov O. Method of User Authentication in Information and Communication Systems Using Zero Watermark. *Innovative Technologies and Scientific Solutions for Industries*. 2025. № 2(32). P. 69–78. DOI: 10.30837/2522-9818.2025.2.069
14. Pal A., Mondal B., Bhattacharyya N., Ghosh S. Similarity in Fuzzy Systems. *Journal of Uncertainty Analysis and Applications*. 2014. Vol. 2. Article 18. DOI: 10.1186/s40467-014-0018-0
15. Altinkemer K., Wang T. Cost and Benefit Analysis of Authentication Systems. *Decision Support Systems*. 2011. Vol. 51, No. 3. P. 394–404. DOI: 10.1016/j.dss.2011.01.005
16. Phillips T. Usability and Security of Different Authentication Methods for an Electronic Health Records System. *Proceedings of the 14th International Conference on Health Informatics (HEALTHINF)*. 2021. 8 p. URL: https://www.academia.edu/125169716/Usability_and_Security_of_Different_Authentication_Methods_for_an_Electronic_Health_Records_System;

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

17. Kruzikova A., Knapova L., Smahel D., Dedkova L., Matyas V. Usable and Secure? User Perception of Four Authentication Methods for Mobile Banking. *Computers & Security*. 2022. Vol. 115. Article 102603. DOI: 10.1016/j.cose.2022.102603
18. Nadipally M. Optimization of Methods for Image-Texture Segmentation Using Ant Colony Optimization. In: *Intelligent Data-Centric Systems: Intelligent Data Analysis for Biomedical Applications*. Academic Press, 2019. Ch. 2. P. 21–47. DOI: 10.1016/B978-0-12-815553-0.00002-1

References

1. Lal, N.A., Prasad, S. Farik, M. (2016), "A review of authentication methods", *International Journal of Scientific & Technology Research*, Vol. 5, No. 11, P. 246–249. available at: https://www.researchgate.net/publication/311514269_A_Review_Of_Authentication_Methods
2. Arévalo-Ancona, R.E., Cedillo-Hernandez, M., Ramirez-Rodriguez, A.E., Nakano-Miyatake, M. Perez-Meana, H. (2024), "Secure medical image authentication using zero-watermarking based on deep learning context encoder", *Computación y Sistemas*, Vol. 28, No. 1, P. 199–210. DOI:10.13053/cys-28-1-4898
3. Seenivasagam, V. & Velumani, R. (2013), "A QR code based zero-watermarking scheme for authentication of medical images in teleradiology cloud", *Computational and Mathematical Methods in Medicine*, 2013, Article ID 516465, 10 p. DOI:10.1155/2013/516465
4. Chapple, M. (2020), *Implementing Access Control Systems*, Indiana (USA): Wiley, P. 110–135.
5. Aiko, O.V., Vasylenko, O.D. Stepanenko, V.M. (2023), "Systems for access control automation at facilities" ["Systemy avtomatyzatsii kontroliu dostupu na ob'yekty"], *Vseukrainska naukovo-praktychna konferentsiia studentiv, aspirantiv ta molodykh vchenykh*. P. 106-108. available at: https://ela.kpi.ua/server/api/core/bitstreams/e4de99d9-a572-452d-8194-3ba4a39f5bb4/content
6. Qi, X. Liu, Y. (2013), "Cloud model based zero-watermarking algorithm for authentication of text document", *Proceedings of the 2013 Ninth International Conference on Computational Intelligence and Security* (Emeishan, China), P. 712–715. DOI:10.1109/CIS.2013.15
7. Hasan, B.M.S., Ameen, S.Y. and Hasan, O.M. (2021), "Image Authentication Based on Watermarking Approach: Review", *Asian Journal of Research in Computer Science and Information Technology*, № 9(3), P. 34–51. DOI:10.9734/AJRCOS/2021/v9i330224
8. Arévalo-Ancona, R.E., Cedillo-Hernández, M., Ramírez-Rodríguez, A.E., Nakano-Miyatake, M. and Pérez-Meana, H. (2024), "Secure Medical Image Authentication Using Zero-Watermarking Based on Deep Learning Context Encoder", *Computación y Sistemas*, 28(1), P. 199–210. DOI:10.13053/CyS-28-1-4898
9. Seenivasagam, V. and Velumani, R. (2013), "A QR Code Based Zero-Watermarking Scheme for Authentication of Medical Images in Teleradiology Cloud", *Computational and Mathematical Methods in Medicine*, Article ID 516465, 16 p. DOI:10.1155/2013/516465
10. Tayan, O., Kabir, M.N. Alginahi, Y.M. (2014), "A hybrid digital-signature and zero-watermarking approach for authentication and protection of sensitive electronic documents", *The Scientific World Journal*, Article ID 514652, 10 p. DOI:10.1155/2014/514652
11. Poddubnyi, V., Sievierinov, O. & Nepokrytov, D. (2025), "Research of the effectiveness of image processing algorithms in zero-watermarking schemes", ["Doslidzhennia efektyvnosti alhorytmiv obroblennia zobrazhen u skhemakh nulovoho vodianoho znaka"], *Suchasnyi stan naukovykh doslidzhen ta tekhnolohii v promyslovosti*, No. 1(31), P. 102–114. DOI:10.30837/2522-9818.2025.1.102
12. Poddubnyi, V.O., Sievierinov, O.V. & Hvozdiov, R.Yu. (2024), "Using zero watermarks for image copyright protection and multifactor authentication" ["Vykorystannia nul′ovykh vodianykh znakiv dlia pidtverdzhennia avtorstva zobrazhen′ ta bahatofaktornoï avtentyfikatsii"], *Radiotekhnika*, Issue 218, P. 35–43. DOI:10.30837/rt.2024.3.218.02
13. Poddubnyi, V. Sievierinov, O. (2025), "Method of user authentication in information and communication systems using zero watermark", *Innovative Technologies and Scientific Solutions for Industries*, No. 2(32), P. 69–78. DOI:10.30837/2522-9818.2025.2.069
14. Pal, A., Mondal, B., Bhattacharyya, N. Ghosh, S. (2014), "Similarity in fuzzy systems", *Journal of Uncertainty Analysis and Applications*, Vol. 2, Article 18. DOI:10.1186/s40467-014-0018-0
15. Altinkemer, K. Wang, T. (2011), "Cost and benefit analysis of authentication systems", *Decision Support Systems*, Vol. 51, No. 3, P. 394–404. DOI: 10.1016/j.dss.2011.01.005
16. Phillips, T. (2021), "Usability and security of different authentication methods for an electronic health records system", *Proceedings of the 14th International Conference on Health Informatics (HEALTHINF)*. 8 p. available at: https://www.academia.edu/125169716/Usability_and_Security_of_Different_Authentication_Methods_for_an_Electronic_Health_Records_System

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

17. Kruzikova, A., Knapova, L., Smahel, D., Dedkova, L. Matyas, V. (2022), "Usable and secure? User perception of four authentication methods for mobile banking", *Computers & Security*, Vol. 115, Article 102603. DOI: 10.1016/j.cose.2022.102603
18. Nadipally, M. (2019), "Optimization of methods for image-texture segmentation using ant colony optimization". In: *Intelligent Data-Centric Systems: Intelligent Data Analysis for Biomedical Applications*. Academic Press, Ch. 2, P. 21–47. DOI: 10.1016/B978-0-12-815553-0.00002-1

Надійшла (Received) 05.04.2025

Відомості про авторів / About the Authors

Поддубний Вадим Олександрович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій, Харків, Україна; e-mail: vadym.poddubnyi@nure.ua; ORCID ID: https://orcid.org/0000-0002-4380-491X

Сєвєрінов Олександр Васильович – кандидат технічних наук, Харківський національний університет радіоелектроніки, доцент кафедри безпеки інформаційних технологій, Харків, Україна; e-mail: oleksandr.sievierinov@nure.ua; ORCID ID: https://orcid.org/0000-0002-6327-6405

Poddubnyi Vadym – Kharkiv National University of Radio Electronic, PhD Student at the Department of Information Technology Security, Kharkiv, Ukraine.

Sievierinov Oleksandr – PhD (Engineering Sciences), Kharkiv National University of Radio Electronic, Associate Professor at the Department of Information Technology Security, Kharkiv, Ukraine.

USER AUTHENTICATION MODEL USING ZERO WATERMARKING BASED ON RGB IMAGES

The subject of the research is the methods and schemes of user authentication in information and communication networks. **The aim of the work** is to develop a user authentication model in information and communication systems using the zero-watermarking method based on RGB images. **Objectives**: to examine the main issues arising during the integration of access control systems; to develop a user authentication model using zero-watermarking; to study potential characteristics and put forward proposals/recommendations for its application in authentication; to construct graphical schemes that illustrate processes and procedures. **Methods used**: modeling methods and formal description methods. **Results achieved**: an improved user authentication model based on the zero-watermarking algorithm has been proposed as an alternative to existing authentication schemes in information and communication systems; its potential characteristics have been studied, confirming the prospects of the development; graphical representations of processes and procedures have been constructed. **Conclusions**: in the course of the work, the main problems faced by owners of information and communication systems during the integration of authentication systems were briefly reviewed; an improved authentication model using zero-watermarking was proposed. The zero-watermarking method is based on RGB images employing K-means clustering and DWT, while a password is used as the second factor. Three operating modes are presented: registration, authentication, and modification of key authentication data. This model serves as an alternative to existing schemes and can potentially reduce the cost of authentication, lower Type I and Type II error rates compared to similar authentication schemes, and improve the convenience of authentication in information and communication systems. It also expands the range of authentication schemes available to system owners during integration.

Keywords: digital watermarking; authentication; image; model; password.

Бібліографічні описи / Bibliographic descriptions

Поддубний В.О., Сєвєрінов О.В. Модель автентифікації користувачів з використанням нульового водяного знаку на основі RGB-зображень. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 102–110. DOI: https://doi.org/10.30837/2522-9818.2025.3.102

Poddubnyi, V., Sievierinov, O. (2025), "User authentication model using zero watermarking based on RGB images", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 102–110. DOI: https://doi.org/10.30837/2522-9818.2025.3.102

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 005.963.1:629.7.-027.3 DOI: https://doi.org/10.30837/2522-9818.2025.3.111

Ю. Полупан, О. Малєєва

МОДЕЛІ ВИБОРУ РАЦІОНАЛЬНОЇ СТРУКТУРИ ПОСТАЧАННЯ КОМПЛЕКТОВАННЯ ДЛЯ ВИРОБНИЦТВА БПЛА В УМОВАХ РИЗИКУ, ОБМЕЖЕНИХ ЛОГІСТИЧНИХ ВИТРАТ І ЧАСУ

Предметом дослідження є логістичні структури для забезпечення постачання комплектування для виробництва БПЛА. **Мета статті** – підвищення ефективності постачання ресурсів для виробництва БПЛА внаслідок вибору раціональної логістичної структури. Окреслено такі **завдання**: аналіз особливостей формування логістичної структури постачання деталей для виробництва БПЛА; формування множини варіантів логістичних структур постачання та критеріїв для їх оцінювання; розроблення математичних моделей багатокритеріального вибору логістичної структури. Упроваджено такі **методи:** аналіз та узагальнення інформації, структурний варіантний синтез, метод основного критерію. **Досягнуті результати.** Для різних класів БПЛА узагальнено основні види комплектувальних елементів. Аналіз можливих постачальників комплектування продемонстрував варіантність вибору, а звідси й множину та різнорідність імовірних структур постачання. Формування та аналіз основних видів логістичних структур дав змогу визначити низку критеріїв для оцінення їх недоліків і переваг. Порівняння різних логістичних структур підприємства розкриває їх особливості в контексті ключових критеріїв, таких як мінімальний ризик, мінімальні витрати, швидкість постачання, масштабованість і гнучкість у кризових ситуаціях. Сформовано дерево вибору логістичної структури з огляду на стратегічні цілі підприємства (зниження вартості, часу доправлення вантажу або логістичних ризиків). Сформульовано узагальнену задачу багатокритеріального вибору логістичної структури із застосуванням методу основного критерію. Розглянуто три варіанти формалізованої моделі задачі вибору раціональної структури. **Висновки:** результати дослідження можуть бути використані для вдосконалення логістичних стратегій у сфері виробництва БПЛА, зокрема щодо зниження часу доправлення (відповідно до визначеного ритму виробництва) й мінімізації ризиків в умовах воєнного періоду. Напрямом подальших досліджень є розроблення моделей оптимізації на комбінованих структурах постачання та збуту.

Ключові слова: логістична структура; вибір постачальників; виробництво БПЛА; маршрути постачання; ризики; ланцюги постачання; багатокритеріальний вибір; мінімізація витрат; метод основного критерію.

Вступ

У сучасних умовах зростає потреба у високотехнологічних військових рішеннях, здатних забезпечити стратегічну перевагу на полі бою. Одним із ключових напрямів розвитку оборонних технологій є безпілотні літальні апарати (БПЛА), які відіграють важливу роль у розвідуванні, спостереженні, коригуванні артилерійського вогню та нанесенні ударів по ворожих позиціях. Саме тому вдосконалення дронів і створення їх нових моделей є критично важливим завданням для оборонної промисловості.

Поточні військові конфлікти демонструють, що потреба у БПЛА значно зростає, зокрема у компактних і мобільних моделях, які можуть працювати в складних умовах бойових дій. Значна увага приділяється дронам-камікадзе, що здатні самостійно знаходити та знищувати цілі, а також багатофункціональним БПЛА, які поєднують завдання розвідки та ударного застосування [1].

Однак створення та модернізація безпілотників неможливі без ефективної логістики. Логістичні процеси охоплюють постачання комплектування, організацію виробництва, тестування та доправлення готових дронів до місць їх застосування. В умовах війни або підвищеного військового навантаження важливо налагодити безперебійні логістичні канали постачання, зокрема для компонентів високоточної електроніки, двигунів, систем зв'язку та навігації. Крім того, необхідно зважати на можливість оперативного обслуговування та ремонту БПЛА, що вимагає створення розгалуженої системи технічної підтримки та сервісного обслуговування [2].

Отже, удосконалення виробництва БПЛА та оптимізація логістичних процесів є ключовими завданнями для підприємств, що працюють у сфері оборонних технологій. Це не лише сприяє підвищенню боєздатності армії, а й забезпечує стратегічну незалежність у виробництві критично важливих систем. Дослідження, спрямовані на формування ефективних методів постачання з використанням інформаційних технологій для управління логістикою процесів виробництва БПЛА, є актуальним завданням.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Аналіз публікацій та постановка завдань

Існує значна кількість публікацій, присвячених питанням логістичного забезпечення виробничого підприємства. Так, у статті [3] проаналізовано теоретичні аспекти управління витратами в процесі логістичної діяльності підприємства. Узагальнено основні типи логістичних рішень, критерії їх формування та основні методи прийняття. Запропоновано підходи до управління витратами під час прийняття рішень, що сприяють зниженню витрат і створенню конкурентних переваг підприємства. Автори статті [4] проаналізували фактори, що впливають на витрати логістично-постачальницької діяльності підприємств. Виокремлено та систематизовано внутрішні та зовнішні чинники, що визначають ефективність логістичних процесів, таких як стратегія постачання, складське господарство, конкуренція та інфраструктура. У статті [4] досліджено фактори, що впливають на прийняття логістичних рішень у ланцюгу постачання. Особливу увагу приділено класифікації логістичних витрат та їх впливу на ефективність процесів постачання.

У студіях [5–7] розкрито сутність оптимізації основних логістичних бізнес-процесів на підприємстві. Проаналізовано еталонну модель оптимізації системи управління бізнес-процесами, зокрема використання методів контролю на етапі оцінювання результатів оптимізації. Розглянуто необхідність адаптації стратегічного маркетингового управління до умов бізнес-процесів ланцюга постачання товарів повсякденного попиту. Особливу увагу зосереджено на інтеграції маркетингу й логістики для підвищення ефективності обслуговування клієнтів, управління попитом і скорочення витрат [8, 9].

У роботі [10] запропоновано алгоритм управління асортиментом і товарними запасами в умовах ринкової нестабільності. Упроваджено методи статистичного аналізу, прогнозування, машинного навчання [11, 12].

Стаття [13] присвячена інноваціям у логістичному менеджменті, зокрема методам проєктування логістичних систем на основі процесно-матричної організаційної структури для оптимізації матеріальних, інформаційних і фінансових потоків. У роботі [14] проаналізовано вплив сталого управління ланцюгами постачання на динамічні можливості та ефективність підприємств, з особливою увагою до компаній у країнах, що розвиваються.

У праці [15] досліджено роль блокчейну та інформаційно-комунікаційних технологій у ланцюгу постачання. Автори виявляють ключові чинники впровадження блокчейну за допомогою рейтинг-кон'юнктурного аналізу, що дає змогу розробити ефективну архітектуру для інтеграції та сталого розвитку ланцюга постачання. Розв'язано завдання управління ризиками в ланцюгах постачання з визначенням основних ризикових подій і стратегій мінімізації загроз через організаційні підходи [16, 17]. Стаття [18] присвячена аналізу змін у логістичних ланцюгах постачання внаслідок війни та євроінтеграції. Розглянуто виклики, цифрову трансформацію та адаптацію маршрутів, а також запропоновано способи вдосконалення за допомогою аналітики й міжнародної співпраці.

Автори роботи [19] проаналізували мобільні логістичні та моніторингові системи, що використовують рої БПЛА. Крім того, досліджено основні виклики впровадження таких систем та перспективи їх розвитку для покращення ефективності логістичних операцій і моніторингу в динамічних умовах.

Отже, значну кількість досліджень присвячено підвищенню ефективності управління логістичними процесами, зокрема в умовах невизначеності та ризиків. Але водночас недостатньо уваги приділено саме формуванню логістичних структур, зокрема логістичних ланцюгів постачання ресурсів для виробництва продукції подвійного призначення в умовах воєнного стану. Не взято до уваги значний рівень невизначеності в плануванні логістичних систем виробництва у воєнний період.

Мета статті – підвищення ефективності постачання ресурсів для виробництва БПЛА за допомогою вибору раціональної структури з огляду на невизначеність на етапі попереднього планування логістичних завдань.

Для досягнення мети необхідно розв'язати такі завдання:

1) проаналізувати особливості формування логістичної структури постачання деталей для виробництва БПЛА;

2) сформувати множину варіантів логістичних структур постачання та критеріїв для їх оцінювання;

3) розробити математичні моделі багатокритеріального вибору логістичної структури.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Матеріали й методи

На етапі попереднього планування логістики постачання та збуту необхідно обрати структуру з огляду на низку критеріїв. Отже, виникає завдання побудови моделі багатокритеріального оцінювання відповідної ситуації прийняття рішення [20]. Формалізація процедури оцінювання варіантів логістичних структур часто пов'язана з необхідністю переходу від якісних лінгвістичних змінних до певної метрики заданого набору окремих критеріїв варіантів.

З формального погляду вибір найкращого варіанта логістичної структури з допустимого набору є завданням прийняття рішень з визначення найкращого з них $x^{\circ} \in X$, де X – допустима множина варіантів. Кожне рішення $x_i \in X$, $i = \overline{1,n}$ визначається кортежем різнорідних часткових критеріїв

$$K = \langle k_j(x) \rangle,\ j = \overline{1,m}. \quad (1)$$

Тоді оптимальним буде рішення

$$x^{\circ} = \arg \underset{x \in X}{extr} \langle k_j(x) \rangle,\ \forall j = \overline{1,m}. \quad (2)$$

Труднощі прийняття рішення про вибір найкращого варіанта в багатокритеріальних задачах виникають у ситуаціях, коли один частковий критерій не може бути покращений без погіршення хоча б одного іншого критерію. Ця ситуація ототожнюється із так званою областю компромісів (область Парето). Розв'язання компромісної задачі може здійснюватися запровадженням деякого додаткового правила (принципу оптимальності), що дає змогу прийняти рішення про вибір єдиного найкращого варіанта.

Для вибору пріоритетного варіанта з множини структур, що подані на третьому рівні дерева (рис. 8), можна застосувати метод основного критерію.

Цей метод базується на виділенні одного з критеріїв (який є переважним в умовах певного підприємства) та переведення всіх інших критеріїв в обмеження. Для цього аналізуються конкретні особливості багатокритеріальної задачі, з множини часткових критеріїв обирається один – найважливіший. Для кожного з інших часткових критеріїв призначається граничне значення, нижче за яке він не може опускатися. Отже, всі часткові критерії, крім одного, перетворюються на обмеження, що додатково звужують множину допустимих рішень *X*. Тоді вихідна багатокритеріальна задача (1) перетворюється на однокритеріальну:

$$x^{\circ} = \arg \underset{x \in X}{extr}\, k^*(x),\ k_i(x) \geq (\leq) k_{iHГ}(x),\ i = \overline{1,n-1}, \quad (3)$$

де $k^*(x)$ – оптимізаційний скалярний критерій;

$k_{iHГ}$ – найгірші допустимі значення часткових критеріїв-обмежень;

знак "≥" використовується для критеріїв, які необхідно максимізувати, а знак "≤" – мінімізувати.

У процесі реалізації розглянутого методу важливо звертати особливу увагу на те, щоб припустима множина рішень, яка визначається частковими критеріями-обмеженнями, не виявилася порожньою.

Результати дослідження

1. Аналіз видів комплектування для виробництва БПЛА

Безпілотні літальні апарати можна розрізняти за різними параметрами, зокрема за призначенням, дальністю польоту, рівнем автономності та технічними характеристиками. Основні класи БПЛА [21]:

– розвідувальні – призначені для збирання інформації, спостереження та коригування артилерії;

– ударні – оснащені зброєю та здатні завдавати високоточних ударів по ворогу;

– тактичні – використовуються на оперативному рівні для підтримки військових підрозділів безпосередньо на полі бою;

– стратегічні – великі апарати, які можуть здійснювати далекі польоти та виконувати завдання високої складності.

Усі класи БПЛА містять численні комплектування, кожне з яких відіграє важливу роль у забезпеченні ефективності, безпеки та стабільності роботи дронів [22]. Але певні класи мають інноваційні компоненти, призначені для забезпечення (удосконалення) виконання особливих функцій – від руху і маневреності до збирання та передачі інформації, необхідної для виконання військових місій.

Основою будь-якого БПЛА є рама й корпус. Рама забезпечує структурну цілісність апарата й утримує всі інші елементи. Для виготовлення рам використовуються матеріали, що поєднують легкість і міцність, а також стійкість до механічних пошкоджень. Найбільш поширеними є карбонові волокна, алюмінієві сплави й пластикові композити. Карбонові матеріали дають високу міцність за умови мінімальної ваги, алюмінієві сплави додають корозійну стійкість, а пластикові композити роблять конструкцію більш маневреною та знижують її вартість. В Україні компанії, зокрема *Drone Frames*

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

та *Vyriy Drone*, виготовляють рами та корпуси для БПЛА, постачаючи їх як для комерційних, так і для військових потреб.

Наступний важливий елемент БПЛА – це двигуни та пропелери, що забезпечують основну рухову силу для підйому та переміщення апарата. Для багатороторних дронів (мультикоптерів) кожен ротор оснащений окремим двигуном, що уможливлює високу маневреність і стабільність польоту. У дронів із фіксованими крилами двигуни забезпечують швидкість і підйомну силу. Для живлення двигунів використовуються безщіткові електричні мотори, яким властиві висока ефективність і низький рівень шуму. В Україні компанії, зокрема *Escadrone* та *AIR 3F*, спеціалізуються на виготовленні двигунів і пропелерів для різних типів дронів, забезпечуючи високу якість та надійність своїх компонентів.

Акумулятори є критично важливими для гарантування тривалості польоту БПЛА. Від ефективності акумуляторів залежить здатність апарата виконувати свої функції протягом заданого часу. Найпоширенішими є літій-полімерні (LiPo) та літій-іонні (Li-ion) акумулятори. LiPo-акумулятори мають високу енергетичну щільність і здатні швидко заряджатися, що робить їх ідеальними для дронів. Літій-іонним акумуляторам властивий значний термін служби, а також вони більш безпечні, хоча важчі та мають меншу ємність для однакового об'єму. Для підвищення безпеки акумулятори зазвичай оснащують системами захисту від перезаряджень та перегріву. В Україні компанії *PAWELL BATTERY* та *BTRY Energy* постачають літій-полімерні та літій-іонні акумулятори, що відповідають вимогам високих стандартів безпеки та ефективності.

Щоб забезпечити стабільність польоту та контроль над рухами БПЛА, використовують системи управління та навігації. Польотний контролер (FC) координує роботу двигунів і датчиків, що дає змогу дрону залишатися стабільним у польоті. Гіроскопи визначають орієнтацію апарата в просторі, акселератори вимірюють прискорення, а барометри – висоту польоту. GPS-модулі допомагають точно позиціювати БПЛА та здійснювати автономні польоти. Вітчизняні компанії *USMT* та *SkyLab UA* постачають високоточні навігаційні системи, польотні контролери та GPS-модулі для БПЛА, що забезпечують їх стабільність і точність у роботі.

Системи зв'язку є критичними для ефективного керування БПЛА. Вони забезпечують двосторонній зв'язок між оператором і дроном, що дає змогу здійснювати передачу команд та отримувати зворотний зв'язок від апарата. З метою передачі інформації можуть використовуватися радіочастотні системи (RF), Wi-Fi або супутникові канали зв'язку. Для дронів, що працюють на великих відстанях або в умовах поганого зв'язку, часто застосовуються супутникові канали. В Україні компанії *Dronarnia* та *3DTech* розробляють і постачають радіочастотні та супутникові системи для передачі інформації на значні відстані.

Сенсори та камери на борту БПЛА виконують важливу роль під час збирання інформації та моніторингу місцевості. Вони можуть бути оптичними, інфрачервоними (для нічного бачення) або термальними (для вимірювання температури). Також з метою створення точних тривимірних карт місцевості використовуються лідарні сенсори. Камери та сенсори дають змогу БПЛА виконувати різноманітні завдання – від спостереження до наукових досліджень. В Україні компанії *UAV in UA* і *Modelistam* постачають високоякісні сенсори та камери для різних завдань – від військових до комерційних.

Крім того, важливими є системи захисту та додаткові функції БПЛА, які забезпечують безпеку та надійність операцій. До таких систем належать антивірусні програми для захисту від хакерських атак, а також резервні системи для аварійного приземлення дрону в разі збоїв або втрати зв'язку. У нашій країні компанії *Kvertus* та *DracoTech* розробляють такі системи, забезпечуючи безпеку та стабільність роботи БПЛА навіть у складних умовах.

Фірми, що постачають ці компоненти, зазвичай спеціалізуються на одному або кількох категоріях, але також можуть пропонувати додаткові продукти для комплексних рішень. Нижче подано таблицю, що відтворює основні компоненти БПЛА, зокрема рами, двигуни, акумулятори, системи управління, камери, сенсори та системи захисту.

Також важливо зауважити, що компанії можуть розширювати свій асортимент, додаючи нові компоненти, що дає змогу їм бути більш гнучкими в задоволенні потреб клієнтів. Результати аналізу подано в табл. 1. Зазначимо, що за більш детального аналізу асортименту перелічених компаній можуть бути додані й інші продукти для БПЛА. Однак, як видно з таблиці, для забезпечення виробництва комплектування для БПЛА постає завдання вибору множини постачальників.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

Таблиця 1. *Постачальники й типи комплектування для БПЛА*

Постачальник	Рами та корпуси	Двигуни та пропелери	Акумулятори	Системи управління	Системи зв'язку	Камери та сенсори	Системи захисту
Drone Frames	+	+	+	+	+	+	+
VYRIY Drone	+	+	+	+	+	+	+
Escadrone	+	+	+	+	+	+	+
AIR 3F	+	+	+	+	+	+	+
PAWELL BATTERY	–	–	+	–	–	–	–
BTRY.Energy	–	–	+	–	–	–	–
USMT	–	–	–	+	+	+	+
SkyLab UA	+	+	+	+	+	+	+
Dronarnia	+	+	+	+	+	+	+
3DTech	+	+	+	+	+	+	+
UAV in UA	+	+	+	+	+	+	+
Modelistam	+	+	+	+	+	+	+
Kvertus	+	+	+	+	+	+	+
DracoTech	–	–	–	–	–	–	+

У виборі постачальників необхідно брати до уваги множину ймовірних маршрутів постачання. Потрібно оцінити варіанти широкої мережі постачальників (з вузькою спеціалізацією) або співробітництва тільки з окремими постачальниками (із значним асортиментом комплектування та позитивним досвідом співпраці). Другий варіант може бути переважним щодо зменшення логістичних шляхів. Але він має недолік – для проєктів модернізації БПЛА може виникнути потреба в інноваційних елементах, які здатні забезпечити лише окремі постачальники.

В умовах активних бойових дій постачання компонентів для БПЛА від закордонних постачальників може здаватися менш ризикованим (якщо порівнювати з місцевим виробництвом) завдяки стабільності політичної ситуації в багатьох країнах та розвинутій інфраструктурі. Однак важливо зауважити, що терміни постачання таких компонентів можуть бути значно тривалішими через митні процедури, міжнародні санкції та транспортні обмеження. Це створює ризики затримок, які можуть негативно вплинути на оперативність постачання, що критично важливо в умовах, де кожен момент є вирішальним. Крім того, закордонні постачальники можуть стати ненадійними через непередбачувані фактори, наприклад зміни в політичній або економічній ситуації, введення нових санкцій чи порушення стабільності міжнародних торгових шляхів. Навіть добре зарекомендовані компанії можуть зіткнутися з труднощами у виконанні зобов'язань, що підвищує ризики вчасного постачання критичних компонентів.

Тому, незважаючи на потенційну надійність імпортних компонентів, варто максимально орієнтуватися на місцевих виробників. Це дасть змогу знизити залежність від зовнішніх факторів, забезпечуючи оперативність, гнучкість і надійність у виконанні військових завдань.

2. Варіанти логістичних структур постачання та критерії їх оцінювання

У процесі формування логістичних каналів постачання ресурсів і збуту продукції можна розглядати різні структури взаємодії учасників логістичних процесів, кожна з яких має свої особливості, переваги та виклики. Одна з найпростіших структур – це модель організації постачання, що передбачає прямий зв'язок між одним постачальником та одним замовником без посередників або складських приміщень (рис. 1). У цій моделі постачальник безпосередньо постачає товар або ресурси замовнику, що дає змогу значно знизити витрати, оскільки немає необхідності в додаткових етапах логістики, таких як склади або логістичні посередники [23]. Така структура особливо ефективна для малих підприємств або компаній, що мають обмежену кількість замовників і постачальників. Вона дає змогу краще контролювати постачальні процеси, оскільки всі етапи обмежуються лише двома учасниками. Однак така модель має і певні ризики. Залежність від одного постачальника й одного замовника може стати суттєвим недоліком у разі виникнення проблем на будь-якому з етапів – від виробництва до доправлення. Крім того, така структура є обмеженою за масштабами, оскільки для більш складних і масштабних операцій необхідно мати більшу кількість учасників у ланцюгу постачання, що знижує гнучкість і швидкість реагування.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Рис. 1. Пряма структура постачання

Іншою моделлю є мережа з кількома постачальниками та одним замовником, де останній отримує ресурси від постачальників (рис. 2). Модель з кількома постачальниками та одним замовником зменшує ризики перебоїв у постачанні завдяки диверсифікації джерел ресурсів. Це підвищує надійність постачання, знижує залежність від одного постачальника й дає змогу обирати більш вигідні умови співпраці, що може знизити витрати й підвищити конкурентоспроможність [24].

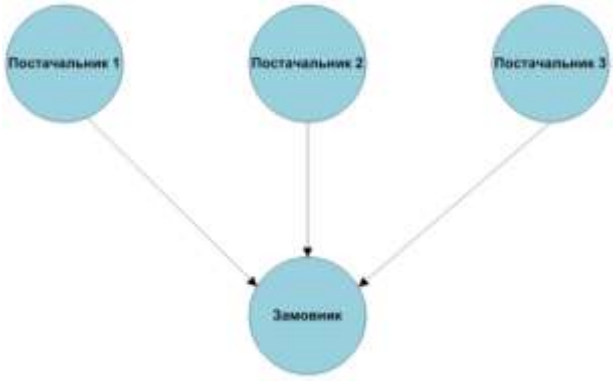


Рис. 2. Модель з кількома постачальниками та одним замовником

Однак управління такою мережею вимагає значних зусиль для координації постачальників, забезпечення ефективного обміну інформацією, узгодження графіків і контролю виконання угод. Крім того, складна логістика, необхідність моніторингу якості та вчасності постачань можуть ускладнити управління запасами й транспортуванням. Отже, хоча ця модель підвищує гнучкість і знижує ризики, вона потребує ретельного управління для забезпечення ефективності та мінімізації потенційних недоліків.

Мережа з кількома постачальниками та кількома замовниками створює складніші ланцюги постачання, що вимагає ретельного управління та координації. У такій структурі часто виникає потреба в складських приміщеннях або спеціальних послугах для підтримки постачань і вчасного розподілу ресурсів. Це допомагає зберігати необхідні запаси та забезпечувати безперервність постачання, особливо в умовах непередбачуваних змін попиту або пропозиції (рис. 3).

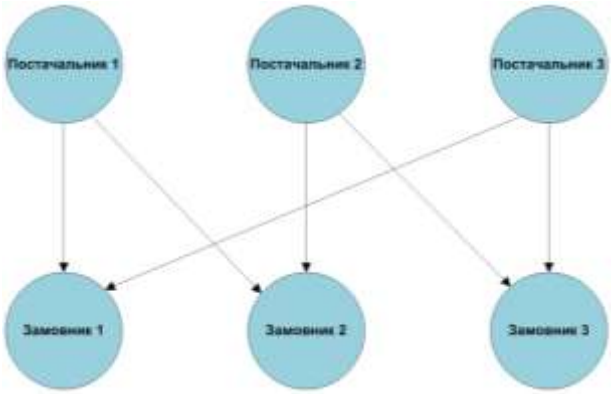


Рис. 3. Мережа з кількома постачальниками та кількома замовниками

З одного боку, така структура забезпечує більшу гнучкість і знижує залежність від окремих постачальників, оскільки замовник може обирати з кількох джерел постачання. Це дає змогу адаптуватися до змін на ринку та зменшити ризики, пов'язані з перебоями в постачанні [25]. Однак, з іншого боку, така мережа потребує більше ресурсів для координації та контролю, оскільки необхідно управляти взаємодією між численними учасниками, що може бути складним завданням.

Крім того, зростання кількості учасників у ланцюзі постачання може збільшити кількість можливих ризиків, пов'язаних з багатофакторним управлінням. Наприклад, можуть виникнути проблеми з комунікацією, невчасним виконанням зобов'язань або якістю постачання, що, ймовірно, вплине на загальну ефективність ланцюга постачання.

У сучасних ланцюгах постачання широко застосовуються моделі, що поєднують використання посередників, таких як логістичні оператори та дистриб'ютори, зі складськими приміщеннями для зберігання товарів. Такі структури сприяють ефективному управлінню логістичними процесами, оптимізації витрат і забезпеченню безперебійного постачання продукції (рис. 4).

Посередники відіграють важливу роль у координації процесів доправлення та розподілу товарів між постачальниками та замовниками. Використання логістичних компаній дає змогу підприємствам скоротити витрати на транспортування та управління запасами, оскільки посередники володіють розвиненою інфраструктурою та технологіями, що підвищують ефективність перевезень і складування [26]. Крім того, централізоване управління запасами допомагає уникнути нестачі

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

чи надлишку продукції, що позитивно впливає на стабільність постачань [27].

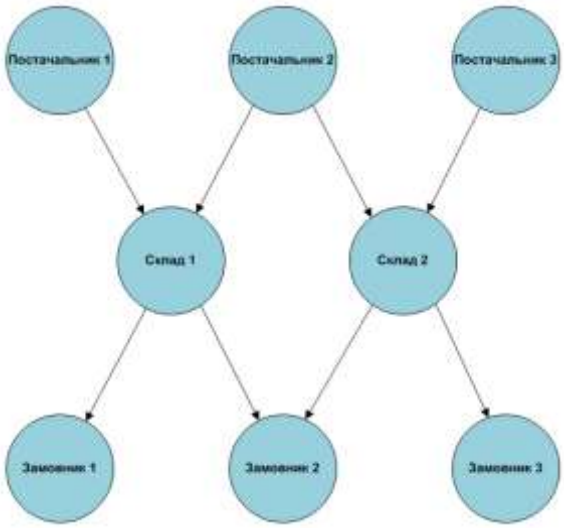


Рис. 4. Мережа з кількома постачальниками та кількома замовниками зі складами

Водночас логістичні мережі часто містять складські приміщення, що виконують функцію буфера між постачальниками та кінцевими споживачами. Завдяки цьому можна швидко реагувати на зміни в попиті та забезпечувати вчасне доправлення продукції без необхідності чекати нових партій товару від виробника. Це особливо важливо в умовах коливань ринку та непередбачуваних затримок у постачанні.

Попри численні переваги, така модель має і певні недоліки. Залучення посередників передбачає додаткові фінансові витрати, які можуть бути суттєвими, особливо в разі значних обсягів операцій. Крім того, передача частини логістичних функцій зовнішнім компаніям може зменшити рівень контролю над постачанням, що іноді призводить до затримок або інших операційних ризиків. Використання складських приміщень також супроводжується витратами на їх оренду, утримання персоналу, обслуговування запасів і впровадження сучасних інформаційних систем для управління ними.

Отже, поєднання посередників та складських приміщень у логістичних мережах дає змогу підприємствам підвищити ефективність постачань і зменшити витрати, проте вимагає ретельного аналізу всіх супутніх ризиків та витрат. Оптимальний вибір моделі постачання залежить від особливостей діяльності компанії, характеристик ринку та стратегічних цілей бізнесу.

У разі, коли в мережі постачання присутній посередник, тобто зовнішній логістичний оператор, його роль полягає в оптимізації процесів доправлення та розподілу товарів між постачальниками та замовниками (рис. 5). Зовнішні логістичні оператори мають доступ до більш ефективних транспортних і складських ресурсів, що дає змогу знижувати витрати на транспортування, зберігання та інші логістичні послуги. Вони також можуть централізовано управляти запасами, що забезпечує вчасність постачань і зменшує витрати на зберігання товарів.

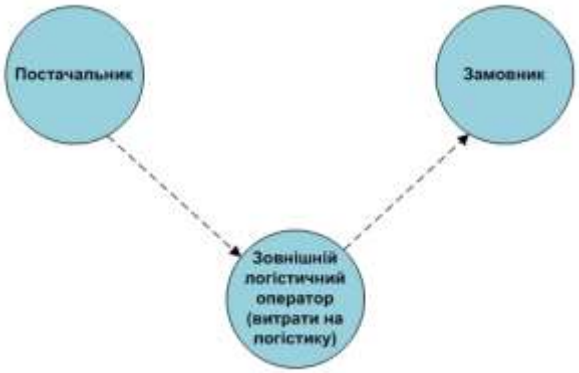


Рис. 5. Логістична структура із залученням зовнішніх логістичних операторів

Залучення таких операторів допомагає підприємствам знизити свої внутрішні витрати, оскільки вони не змушені утримувати власну інфраструктуру для транспортування та зберігання товарів або вести логістичну діяльність самостійно. Проте ця модель має недоліки. Присутність посередника в ланцюгу постачання може призвести до додаткових витрат, адже оператори отримують оплату за свої послуги. Це особливо відчутно за умови великих обсягів операцій, що може знижувати загальну ефективність системи. Крім того, відсутність прямого контролю над процесами спричиняє підвищення ризиків, пов’язаних з ненадійністю оператора або затримками в доправленнях.

Отже, хоча зовнішні логістичні оператори можуть значно оптимізувати витрати на логістику та покращити управління запасами, підприємства мають ретельно оцінювати вигоди від таких послуг порівняно з потенційними витратами та ризиками, що можуть виникнути під час використання цієї моделі постачання.

Ще більш складною є структура, що містить інтегровані інформаційні технології, такі як ERP (*Enterprise Resource Planning*), TMS (*Transportation*

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Management Systems) та WMS (*Warehouse Management Systems*) [28]. Використання цих технологій дає змогу автоматизувати чимало процесів управління постачаннями, що знижує витрати й покращує контроль за всіма етапами логістики (рис. 6). Наприклад, ERP-системи дають змогу інтегрувати всі бізнес-процеси компанії, зокрема планування, закупівлі, облік, та розподіл ресурсів, а TMS допомагає управляти процесами транспортування, моніторити вантажі та оптимізувати маршрути. WMS забезпечує ефективне управління складськими запасами, що зменшує ризик нестачі товарів чи надлишку на складах.

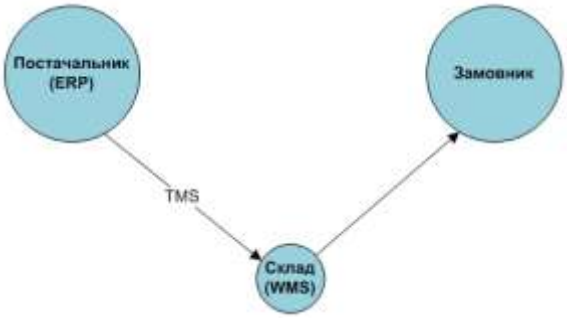


Рис. 6. Логістична структура з інтегрованими технологіями (ERP, TMS та WMS)

Однак упровадження таких складних і дорогих систем потребує значних капіталовкладень і ресурсів для підтримки та інтеграції в наявну інфраструктуру компанії. Це може бути серйозною перешкодою для малих і середніх підприємств, що мають обмежений бюджет для таких технологічних інвестицій. Однак для великих компаній, що працюють на глобальних ринках, така інтеграція може стати ключем до зниження витрат і підвищення ефективності логістики.

Альтернативні маршрути транспортування, що передбачають використання різних видів транспорту або зміну маршрутів залежно від умов, є важливим складником ефективного управління логістикою (рис. 7). Така модель дає змогу значно збільшити гнучкість, оскільки забезпечує можливість швидко адаптуватися до непередбачуваних ситуацій, таких як затори, погодні умови або технічні несправності [29]. Використання кількох видів транспорту, наприклад, поєднання залізничного, автотранспортного й морського транспорту, сприяє забезпеченню безперервності постачання, навіть якщо один із маршрутів стає недоступним. Це також допомагає знижувати ризики, пов’язані з перебоями в постачанні або затримками.

Рис. 7. Логістична структура з альтернативними маршрутами транспортування

Однак така система потребує високого рівня планування та координації. Необхідно постійно моніторити ситуацію на всіх етапах маршруту, вчасно коригувати логістичні плани та взаємодіяти з різними транспортними операторами. Для цього потрібні додаткові ресурси та інформаційні системи, що дають змогу ефективно управляти комплексними ланцюгами постачання. Упровадження альтернативних маршрутів може бути складним і витратним, оскільки кожен новий маршрут і вид транспорту потребує окремого аналізу витрат, ризиків і часу доправлення.

Проте для великих компаній, що працюють на глобальних ринках, ця модель є важливим інструментом для підтримки конкурентоспроможності, забезпечення вчасності постачань і зменшення загальних витрат на транспортування. Модель дає змогу не тільки реагувати на поточні проблеми, а й стратегічно планувати постачання, максимально використовуючи доступні транспортні можливості.

У формуванні логістичних каналів постачання та збуту продукції важливо обрати оптимальну модель, що знижує витрати та підвищує гнучкість. Моделі прямого постачання ефективні для малих підприємств, але обмежені в масштабах. Мережі з кількома постачальниками знижують ризики, але потребують складного управління. Структури з посередниками та складськими приміщеннями забезпечують гнучкість і зниження витрат, але вимагають додаткових витрат на управління. Для більшості підприємств оптимальним є поєднання

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

кількох постачальників з посередниками, що забезпечує баланс витрат і ефективності.

У табл. 2 подано порівняння різних логістичних структур підприємства, розкрито їх основні характеристики й переваги в контексті ключових критеріїв, таких як мінімальний ризик, мінімальні витрати, швидкість постачання, масштабованість і гнучкість у кризових ситуаціях ("+" означає, що цей критерій є перевагою логістичної структури, "–" – критерій є недоліком структури). Пряма структура постачання, яка є найбільш простою та дешевою, забезпечує швидкість постачання завдяки мінімальній кількості етапів, але її основний недолік полягає у високому ризику через залежність від одного постачальника чи замовника. Для зменшення загроз можна застосувати мережу з кількома постачальниками та одним замовником, яка додає гнучкість і знижує ймовірність перебоїв. Однак така структура потребує додаткових витрат на координацію та управління.

Таблиця 2. *Критерії оцінювання логістичних структур*

Види логістичної структури	**Ризик**	**Вартість**	**Швидкість**	**Масштабованість**	**Гнучкість**
1. Пряма структура постачання	–	+	+	–	–
2. Мережа з кількома постачальниками та одним замовником	+	–	+	+	+
3. Мережа з кількома постачальниками та кількома замовниками	+	–	–	+	+
4. Модель постачання з посередником (зовнішній логістичний оператор)	–	–	+	–	–
5. Логістична структура із складськими приміщеннями	+	–	+	+	+
6. Логістична структура з інтегрованими технологіями (ERP, TMS, WMS)	+	–	+	+	+
7. Альтернативні маршрути транспортування	+	–	+	–	+

Якщо ж у мережі є кілька постачальників і кілька замовників, це дає змогу значно підвищити масштабованість і адаптивність бізнесу, проте таке управління може стати складним і вартісним. Моделі постачання з посередниками сприяють оптимізації логістичних процесів і зменшенню навантаження на компанію, але вони створюють додаткові витрати й ризики залежності від надійності посередника. У разі мережі без посередників компанія має повний контроль над процесами, що знижує ризики затримок, але потребує складної координації, що може уповільнювати постачання.

Логістична структура, яка містить складські приміщення, допомагає зберігати запаси, що сприяє стабільності постачання, однак вимагає значних витрат на утримання складів та управління запасами. Залучення зовнішніх логістичних операторів дає змогу оптимізувати транспортування та знизити витрати, але це створює залежність від сторонніх компаній і може виникнути проблема з якістю послуг. Інтеграція сучасних технологій, таких як ERP, TMS чи WMS, здатна автоматизувати процеси, прискорити роботу й зменшити помилки, але впровадження таких систем пов’язане з високими витратами.

Застосування альтернативних маршрутів транспортування дає змогу швидко реагувати на проблеми, знижувати ризик затримок і забезпечувати більшу гнучкість. Однак підтримка цих альтернативних шляхів спричиняє додаткові витрати, що може зробити їх менш вигідними в довгостроковій перспективі.

Зазначимо, що табл. 2 містить узагальнені оцінки. Для кожного окремого підприємства її зміст може бути уточений або доповнений.

3. Математичні моделі вибору логістичної структури з огляду на стратегічні цілі підприємства

У розробленні ефективної логістичної стратегії для підприємства важливо не лише організувати постачання та збут продукції, але й обрати структуру логістичних каналів, що найкраще відповідає цілям компанії. Як випливає з наведеного вище аналізу, оптимальний вибір структури залежить від низки критеріїв, серед яких мінімізація витрат, зниження ризиків та забезпечення високої швидкості постачання. Завдання полягає у створенні чітких правил вибору структури, що забезпечать максимальний ефект під заданий набір умов.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Мінімізація ризиків для виконання замовлень з виробництва БПЛА в умовах військових дій є критично важливим фактором, оскільки ризики можуть бути зумовлені залежністю від одного постачальника, перебоями в постачаннях, затримками через посередників або проблемами в управлінні запасами. Більша кількість етапів у логістичному процесі підвищує ймовірність відмов. Пряма структура постачання має менше учасників, що робить її ефективною за швидкістю, але більш вразливою до перебоїв. Натомість диверсифікована мережа з кількома постачальниками та замовниками дає змогу скоротити ризики, хоча й ускладнює управління.

Мінімізація витрат є не менш важливим критерієм, адже вартість організації логістики залежить від обраної моделі. Пряма модель є найменш витратною, оскільки не потребує посередників чи складських приміщень, але підвищує залежність від окремих контрагентів. Більш складні моделі з кількома постачальниками чи збутовими каналами знижують ризики, але потребують додаткових витрат на координацію. Залучення посередників або складських приміщень може оптимізувати транспортування та забезпечити наявність запасів, проте супроводжується додатковими витратами.

Швидкість постачання та збуту є особливо важливою для підприємств, що працюють у динамічному конкурентному середовищі або мають обмеження за часом. Пряма структура є найбільш ефективною щодо швидкості, оскільки мінімізує кількість етапів у логістичному ланцюгу, скорочуючи час доправлення. Водночас використання складських приміщень чи зовнішніх посередників може сприяти утриманню запасів і оптимізації транспортування, але спричиняє складність у координації.

Вибір структури логістичного каналу має базуватися на пріоритетах підприємства. Якщо основна мета – зниження витрат за мінімальних ризиків, тоді доцільним вибором може бути мережа з кількома постачальниками та одним замовником або безпосередній зв’язок без посередників. Це допомагає уникнути зайвих витрат на координацію, хоча потребує ефективного управління логістичними процесами.

Якщо підприємство прагне до гнучкості та масштабованості, оптимальним варіантом є мережа з кількома постачальниками й замовниками. Така модель сприяє диверсифікації ризиків, проте потребує більшої координації та фінансових ресурсів. Вона ефективна для компаній з великими обсягами постачання та потребою швидкого реагування на зміни ринку.

Для забезпечення максимальної швидкості постачання та збуту найкращою є пряма структура постачання. Це дає змогу мінімізувати час доправлення, проте підвищує залежність від ключових контрагентів. У разі необхідності зменшення навантаження на логістичну інфраструктуру можна розглянути варіант із залученням посередників, що оптимізує транспортування та знижує витрати на утримання складів. Водночас необхідно зважати на можливу залежність від зовнішніх операторів.

Компаніям, що прагнуть забезпечити максимальний контроль над логістикою та мінімізувати ризики затримок, доцільно впроваджувати інтегровані технології, такі як ERP, TMS чи WMS. Вони дають змогу автоматизувати управління запасами, транспортуванням і координацією, зменшуючи ймовірність помилок. Проте їх упровадження потребує значних фінансових і ресурсних вкладень.

Загалом, вибір логістичної структури залежить від стратегічних цілей підприємства: зниження ризиків, мінімізації витрат, швидкості постачання. Рішення необхідно приймати, зважаючи на ринкові умови та здатність компанії адаптуватися до змін. Узагальнена процедура вибору подана у вигляді дерева на рис. 8.

Зазначимо, що критерії масштабованості та гнучкості управління (див. табл. 1) не є стратегічними цілями виробництва. Вони розглядаються як умови, тобто є обмеженнями в постановці багатокритеріальної задачі.

Зазначимо, що на рис. 8 стратегічні цілі підприємства (другий рівень дерева) є основними критеріями, за якими обирається множна допустимих структур третього рівня. Для вибору переважної структури з визначеної множини можна застосовувати низку додаткових обмежень, якими є умови замовника для постачання продукції, а також виробничі обмеження підприємства.

Подамо варіанти формалізованої задачі вибору раціональної структури. Запропонуємо такі познаки: s_i – варіанти видів логістичних структур, $s_i \in S, i = \overline{1,n}$; $W(s_i)$ – оцінка логістичних витрат у структурі s_i; $T(s_i)$ – оцінка часу доправлення в структурі s_i; $R(s_i)$ – оцінка ризику доправлення в структурі s_i.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Рис. 8. Дерево вибору логістичної структури з огляду на стратегічні цілі підприємства

1. Стратегічна мета підприємства – мінімізація логістичних витрат, тоді правило вибору переважного варіанта:

$$s^o = \arg\min_{s_i \in S} W(s_i),\ T(s^o) \le T^*,\ R(s^o) \le R^*, \qquad (4)$$

де T^* – максимально допустимий час доправлення;

R^* – максимально допустимий ризик доправлення.

2. Стратегічна мета підприємства – мінімізація часу доправлення, тоді правило вибору переважного варіанта:

$$s^o = \arg\min_{s_i \in S} T(s_i),\ W(s^o) \le W^*,\ R(s^o) \le R^*, \qquad (5)$$

де W^* – максимально допустимі логістичні витрати.

3. Стратегічна мета підприємства – мінімізація ризиків доправлення, тоді правило вибору переважного варіанта:

$$s^o = \arg\min_{s_i \in S} R(s_i),\ W(s^o) \le W^*,\ T(s^o) \le T^*. \qquad (6)$$

На рівні попереднього планування логістичної структури немає змоги застосувати кількісні оцінки логістичних параметрів (часу, витрат, ризиків) і формалізоване подання задачі. Тому переважна структура обирається неформалізовано способом послідовного вибору умов і відхилення варіантів, які їх не задовольняють.

Висновки

Під час дослідження розглянуто основні методи й технології оптимізації логістичних процесів на прикладі постачання комплектування для БПЛА. Особливу увагу приділено вибору постачальників, управлінню маршрутами постачання та оцінюванню ризиків, які є критичними для забезпечення ефективності логістичних операцій у цій сфері.

Для різних класів БПЛА узагальнено основні види комплектувальних елементів, кожен з яких відіграє важливу роль у забезпеченні їх ефективності, безпеки та стабільності роботи. Аналіз можливих постачальників комплектування продемонстрував варіантність множини постачальників, а звідси й множину та різнорідність можливих структур постачання. Формування й аналіз основних видів логістичних структур дали змогу сформувати низку критеріїв для оцінювання їх недоліків і переваг. Сформовано дерево вибору логістичної структури з огляду на стратегічні цілі підприємства (зменшення витрат, часу та ризиків доправлення). Сформульовано узагальнену задачу багатокритеріального вибору логістичної структури. Наведено три варіанти математичних моделей для застосування методу основного критерію відповідно до стратегічних цілей підприємства.

Результати дослідження продемонстрували значний потенціал для вдосконалення логістичних стратегій у сфері безпілотних систем. Упровадження запропонованих моделей сприятиме вдосконаленню логістичних стратегій у сфері виробництва БПЛА щодо зниження часу доправлення (відповідно до визначеного ритму виробництва) та мінімізації ризиків в умовах воєнного періоду.

Напрямом подальших досліджень є розроблення моделей оптимізації на комбінованих структурах постачання та збуту.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Список літератури

1. Чирак, М. Шляхи підвищення бойових спроможностей військових частин за рахунок оптимізації застосування ударних БПЛА. *Повітряна міць України*, № 1(6), 2024, С. 99–104. DOI: https://doi.org/10.33099/2786-7714-2024-1-6-99-104
2. Корсунов, С. І., Волков, А. Ф., Оборонов, М. І., Орєхов, С. В., Гуртовенко, В. В., Федченко, С. І. Трансформація завдань безпілотної авіації: від створення до застосування у воєнних конфліктах сучасності. *Наука і техніка Повітряних Сил ЗСУ*, № 3(44), 2021, С. 66–81. DOI: https://doi.org/10.30748/nitps.2021.44.08
3. Muha, R. An Overview of the Problematic Issues in Logistics Cost Management. *Pomorstvo*, Vol. 33, No. 1, 2019, P. 102–109. DOI: https://doi.org/10.31217/p.33.1.11
4. Skerlic, S. Factors that influence logistics decision making in the supply chain of the automotive industry. *Transport Problems*, Vol. 15, No. 3, 2020, P. 117–126. DOI: https://doi.org/10.21307/tp-2020-038
5. Завадська, О., Місюкевич, В., Сисоєв, В. В. Оптимізація ланцюга постачань у комерційній логістиці: вплив на ефективність та прибутковість. *Вісник Хмельницького національного університету*, 322(5), 2023. С. 234–241. DOI: https://doi.org/10.31891/2307-5740-2023-322-5-39
6. Fedorovich, O., Pronchakov, Y., Leshchenko, Y., Yelizieva, A. Using of the component method in logistics of supplies of high-tech production components. *Advanced Information Systems*, No. 5(3), 2021, P. 40–45. DOI: https://doi.org/10.20998/2522-9052.2021.3.06
7. Mahat, D., Niranjan, K., Naidu, C. S. K. V. R., Babu, S. B. G. T., Kumar, M. S., N. L. AI-Driven Optimization of Supply Chain and Logistics in Mechanical Engineering, *10th IEEE UPCON, Gautam Buddha Nagar*, India, 2023, P. 1611–1616. DOI: https://doi.org/10.1109/UPCON59197.2023.10434905
8. Крикавський, Є. В., Якимишин, Л. Я. Компліментарність стратегій маркетингу та логістики в ланцюгу поставок товарів повсякденного попиту. *Маркетинг і цифрові технології,* Т. 2, № 1, 2018. С. 21–32. DOI: https://doi.org/10.15276/mdt.2.1.2018.2
9. Rybakov, D. S. A process model of a logistics system as a basis for optimisation programme implementation. *Int. J. of Logistics Research and Applications*, Vol. 21, No. 1, 2018, P. 72–93. DOI: https://doi.org/10.1080/13675567.2017.1361910
10. Ковтун, В. В., Ворона, М. В., Михелєв, І. Л. Алгоритм управління асортиментом та прогнозування рівня товарних запасів. *Таврійський науковий вісник. Серія: Технічні науки*, № 4, 2024. С. 75–91. DOI: https://doi.org/10.32782/tnv-tech.2024.4.7
11. Onukwulu, E. C., Agho, M. O., Eyo-Udo, N. L. Developing a framework for AI-driven optimization of supply chains in energy sector. *Global Journal of Advanced Research and Reviews*, Vol. 1, No. 2, 2023, P. 82–101. DOI: https://doi.org/10.58175/gjarr.2023.1.2.0064
12. Ahmed, A., Rahman, S., Islam, M., Chowdhury, F., Badhan, I. A. Challenges and Opportunities in Implementing Machine Learning For Healthcare Supply Chain Optimization. *Int. Journal of Business and Management Sciences*, Vol. 3, No. 7, 2023, P. 6–31. DOI: https://doi.org/10.55640/ijbms-03-07-02
13. Cherchata, A., Popovychenko, I., Andrusiv, U., Shkuropatskyi, O. Innovations in logistics management as a direction for improving logistics activities. *Management Systems in Production Engineering*, Vol. 30, No. 1, 2022, P. 9–17. DOI: https://doi.org/10.2478/mspe-2022-0002
14. Hong, J., Zhang, Y., Ding, M. Sustainable supply chain management practices, dynamic capabilities, and enterprise performance. *J. of Cleaner Production*, Vol. 172(2). 2017, P. 3508-3519. DOI: https://doi.org/10.1016/j.jclepro.2017.06.093
15. Samant, S., Dey, K. Blockchain technology adoption, architecture, and sustainable agri-food supply chains. *J. of Cleaner Production*, Vol. 284, 2021, 124731. DOI: https://doi.org/10.1016/j.jclepro.2020.124731
16. Olson, D. L., Wu, D. Enterprise risk management in supply chains. In: *Enterprise Risk Management Models*, 2023, P. 1–14. DOI: https://doi.org/10.1007/978-3-662-68038-4_1.
17. Hashemi, S. R., Arasteh, A., Paydar, M. M. Risk Management of Disruption and Sustainable Development of Supply Chains. *Interdisciplinary J. of Management Studies*, No. 16(1), 2023, P. 277–297. DOI: https://doi.org/10.22059/ijms.2022.329830.674732
18. Завербний, А. С., Ломага, Ю. Р. Проблеми та перспективи формування логістичних ланцюгів постачання у воєнний період. *Економіка та суспільство*, № 45. 2022. DOI: https://doi.org/10.32782/2524-0072/2022-45-23
19. Крюченков, О. І., Морозова, О. І., Нікітіна, Т. С. Мобільні логістичні та моніторингові системи на базі роїв БПЛА. *Системи управління, навігації та зв'язку*, № 4, 2024, С. 96–102. DOI: https://doi.org/10.26906/SUNZ.2024.4.096
20. Артюх, Р. В., Малєєва, О. В. Структурні та параметричні моделі вибору та оцінки варіантів технологічного розвитку соціотехнічної системи. *Математичні моделі та новітні технології управління економічними та технічними системами*, 2017. С. 302–314. URL: https://openarchive.nure.ua/server/api/core/bitstreams/b55000b9-9782-4244-b1be-9930e7419da8/content

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

21. Fahlstrom, P. G., Gleason, T. J., Sadraey, M. H. *Introduction to UAV Systems*. Wiley, 2022. URL: https://gnindia.dronacharya.info/ME/Common-Subjects-8th-Sem/Downloads/FUNDAMENTALS-OF-DRONE-TECHNOLOGY/Books/FUNDAMENTALS-OF-DRONE-TECHNOLOGY-text-book-4.pdf
22. Vasić, Z., Maksimović, S., Georgijević, D. Applied Integrated Design in Composite UAV Development. *Applied Composite Materials*, Vol. 25, 2018, P. 221–236. DOI: https://doi.org/10.1007/s10443-017-9611-y
23. Chiang, A.-H., Huang, M.-Y. Demand-pull vs supply-push strategy: org. structure, supply chain integration and response. *J. of Manufacturing Technology Management*, Vol. 32, No. 8, 2021, P. 1493–1514. DOI: https://doi.org/10.1108/JMTM-08-2020-0324
24. Мороз, М., Загорянський, В., Гайкова, Т., Кузєв, І. Використання методів дослідження операцій для оптимізації перевезень. *Вісник НТУ "ХПІ",* № 1(11), 2022, С. 44–50. DOI: https://doi.org/10.20998/2413-4295.2022.01.07
25. Ламберт, Д. М., Купер, М. С. Supply Chain Management: Implementation Issues and Research Opportunities. *Int. J. of Logistics Management*, 2000, Vol. 11, No. 2, P. 1–20. DOI: https://doi.org/10.1108/09574099810805807
26. Гірна, О. Б., Кобилюх, О. Я. До питання функціонування ланцюга постачання в аспекті аутсорсингу. *Економічний простір*, 2023, № 187, С. 91–96. DOI: https://doi.org/10.32782/2224-6282/187-15
27. Загурський, О. М. Аналіз ефективності транспортних процесів у ланцюгах постачань. *Machinery & Energetics*, Vol. 9, No. 4, 2018, С. 43–48. DOI: https://doi.org/10.31548/me2018.04.043
28. Polusmiak, Y., Pavliuk, T., Kosach, I. Project management in the construction industry: logistics optimisation. *Management and Entrepreneurship Trends of Development*, Vol. 4(30), 2024, P. 69–78. DOI: https://doi.org/10.26661/2522-1566/2024-4/30-06
29. Горпенко, Д. Р., Болтьонков, В. О. Порівняльний аналіз методів розв'язання оперативних задач транспортної логістики. *Інформатика та мат. методи в моделюванні*, Т. 13, № 1–2, 2023, С. 34–47. DOI: https://doi.org/10.15276/imms.v13.no1-2.34

References

1. Chyrak, M. (2024), "Ways to improve the combat capabilities of military units through optimizing the use of strike UAVs". *Povitrjana Mic Ukrajiny*, no. 1(6), P. 99–104. DOI: https://doi.org/10.33099/2786-7714-2024-1-6-99-104
2. Korsunov, S. I., Volkov, A. F., Oboronov, M. I., Oriekhov, S. V., Hurtovenko, V. V., Fedchenko, S. I. (2021), "Transformation of unmanned aviation tasks: from creation to application in modern armed conflicts". *Science and Technology of the Air Force of the Armed Forces of Ukraine*, No. 3(44), P. 66–81. DOI: https://doi.org/10.30748/nitps.2021.44.08
3. Muha, R. (2019),"An overview of the problematic issues in logistics cost management". *Pomorstvo*, Vol. 33, No. 1, P. 102–109. DOI: https://doi.org/10.31217/p.33.1.11
4. Skerlic, S. (2020), "Factors that influence logistics decision making in the supply chain of the automotive industry". *Transport Problems*, Vol. 15, No. 3, P. 117–126. DOI: https://doi.org/10.21307/tp-2020-038
5. Zavadska, O., Misyukevych, V., Sysoiev, V. V. (2023)."Optimization of the supply chain in commercial logistics: impact on efficiency and profitability". *Bulletin of the Khmelnytskyi National University*. 322(5), C. 234-241. DOI: https://doi.org/10.31891/2307-5740-2023-322-5-39
6. Fedorovich, O., Pronchakov, Y., Leshchenko, Y., Yelizieva, A. (2021), "Using of the component method in logistics of supplies of high-tech production components". *Advanced Information Systems*, No. 5(3), P. 40–45. DOI: https://doi.org/10.20998/2522-9052.2021.3.06
7. Mahat, D., Niranjan, K., Naidu, C. S. K. V. R., Babu, S. B. G. T., Kumar, M. S., N. L. (2023), "AI-driven optimization of supply chain and logistics in mechanical engineering". *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Gautam Buddha Nagar, India, P. 1611–1616. DOI: https://doi.org/10.1109/UPCON59197.2023.10434905.
8. Krykavskyi, Ye. V., Yakymyshyn, L. Ya. (2018), "Complementarity of marketing and logistics strategies in the supply chain of consumer goods". *Marketing and Digital Technologies*, Vol. 2, No. 1, P. 21–32. DOI: https://doi.org/10.15276/mdt.2.1.2018.2
9. Rybakov, D. S. (2018), "A process model of a logistics system as a basis for optimisation programme implementation". *International Journal of Logistics Research and Applications*, Vol. 21, No. 1, P. 72–93. DOI: https://doi.org/10.1080/13675567.2017.1361910
10. Kovtun, V. V., Vorona, M. V., Mikhelev, I. L. (2024), "Algorithm for assortment management and forecasting of stock levels". *Tavriya Scientific Bulletin. Series: Technical Sciences*, No. 4. P. 375–91. DOI: https://doi.org/10.32782/tnv-tech.2024.4.7
11. Onukwulu, E. C., Agho, M. O., Eyo-Udo, N. L. (2023), "Developing a framework for AI-driven optimization of supply chains in energy sector". *Global Journal of Advanced Research and Reviews*, Vol. 1, No. 2, P. 82–101. DOI: https://doi.org/10.58175/gjarr.2023.1.2.0064

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

12. Ahmed, A., Rahman, S., Islam, M., Chowdhury, F., Badhan, I. A. (2023), "Challenges and opportunities in implementing machine learning for healthcare supply chain optimization: A data-driven examination". *International Journal of Business and Management Sciences*, Vol. 3, No. 7, P. 6–31. P. 6–31. DOI: https://doi.org/10.55640/ijbms-03-07-02
13. Cherchata, A., Popovychenko, I., Andrusiv, U., Shkuropatskyi, O. (2022), "Innovations in logistics management as a direction for improving the logistics activities of enterprises". *Management Systems in Production Engineering*, Vol. 30, No. 1, P. 9–17. DOI: https://doi.org/10.2478/mspe-2022-0002
14. Hong, J., Zhang, Y., Ding, M. (2017), "Sustainable supply chain management practices, supply chain dynamic capabilities, and enterprise performance". *Journal of Cleaner Production*, Vol. 172, No. 2. P. 3508–3519. DOI: https://doi.org/10.1016/j.jclepro.2017.06.093
15. Samant, S., Dey, K. (2021),"Blockchain technology adoption, architecture, and sustainable agri-food supply chains". *Journal of Cleaner Production*, Vol. 284, 124731. DOI: https://doi.org/10.1016/j.jclepro.2020.124731
16. Olson, D. L., Wu, D. (2023), "Enterprise risk management in supply chains". In: *Enterprise Risk Management Models*, P. 1–14. DOI: 10.1007/978-3-662-68038-4_1
17. Hashemi, S. R., Arasteh, A., Paydar, M. M. (2023), "Risk management of disruption and sustainable development of supply chains". *Interdisciplinary Journal of Management Studies*, Vol. 16, No. 1, P. 277–297. DOI: https://doi.org/10.22059/ijms.2022.329830.674732
18. Zaverbnyi, A. S., Lomaha, Y. R. (2022), "Problems and prospects of forming supply chains during wartime under the conditions of European integration". *Economy and Society*, No. 45. DOI: https://doi.org/10.32782/2524-0072/2022-45-23
19. Kryuchenkov, O. I., Morozova, O. I., Nikitina, T. S. (2024), "Mobile logistics and monitoring systems based on UAV swarms: challenges and development directions". *Control, Navigation and Communication Systems*, No. 4, P. 96–102. DOI: https://doi.org/10.26906/SUNZ.2024.4.096
20. Artiukh, R. V., Malyeyeva, O. V. (2017), "Structural and parametric models for selection and evaluation of technological development options of socio-technical systems". In: *Mathematical Models and Modern Technologies for Managing Economic and Technical Systems*. P. 302–314. available at: https://openarchive.nure.ua/server/api/core/bitstreams/b55000b9-9782-4244-b1be-9930e7419da8/content
21. Fahlstrom, P. G., Gleason, T. J., Sadraey, M. H. (2022), Introduction to UAV Systems. Wiley, available at: https://gnindia.dronacharya.info/ME/Common-Subjects-8th-Sem/Downloads/FUNDAMENTALS-OF-DRONE-TECHNOLOGY/Books/FUNDAMENTALS-OF-DRONE-TECHNOLOGY-text-book-4.pdf
22. Vasić, Z., Maksimović, S., Georgijević, D. (2018) "Applied integrated design in composite UAV development". *Applied Composite Materials*, Vol. 25, P. 221–236. DOI: https://doi.org/10.1007/s10443-017-9611-y
23. Chiang, A.-H., Huang, M.-Y. "Demand-pull vs supply-push strategy: the effects of organizational structure on supply chain integration and response capabilities". *Journal of Manufacturing Technology Management*, 2021, Vol. 32, No. 8, P. 1493–1514. DOI: https://doi.org/10.1108/JMTM-08-2020-0324
24. Moroz, M., Zahoryanskyi, V., Haikova, T., Kuzev, I. (2022), "Using operations research methods to optimize bulk cargo transportation in the agro-industrial complex". *Bulletin of the National Technical University "KhPI". Series: New Solutions in Modern Technologies*, No. 1(11), P. 44–50. DOI: https://doi.org/10.20998/2413-4295.2022.01.07
25. Lambert, D. M., Cooper, M. S. (2000), "Supply chain management: Implementation issues and research opportunities". *International Journal of Logistics Management*, Vol. 11, No. 2, P. 1–20. DOI: https://doi.org/10.1108/09574099810805807
26. Hirna, O. B., Kobyliukh, O. Ya. (2023), "On the functioning of the supply chain in the context of outsourcing and logistics operators". *Economic Space: Collected Scientific Works*, No. 187, P. 91–96. DOI: https://doi.org/10.32782/2224-6282/187-15
27. Zahurskyi, O. M. (2018), "Analysis of transport process efficiency in supply chains". *Machinery Energetics. Journal of Rural Production Research*, Vol. 9, No. 4, P. 43–48. DOI: 10.31548/me2018.04.043
28. Polusmiak, Yu. I., Pavliuk, T. S., Kosach, I. (2024), "Project management in the construction industry: identifying logistics problems and ways to optimise them". *Management and Entrepreneurship Trends of Development*, Vol. 4, No. 30, P. 69–78. DOI: https://doi.org/10.26661/2522-1566/2024-4/30-06
29. Horpenko, D. R., Boltonkov, V. O. (2023), "Comparative analysis of methods for solving operational problems in transport logistics". *Informatics and Mathematical Methods in Simulation*, Vol. 13, No. 1–2, P. 34–47. DOI: https://doi.org/10.15276/imms.v13.no1-2.34

Надійшла (Received) 03.01.2025

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Відомості про авторів / About the Authors

Полупан Юрій Володимирович – аспірант, Національний аерокосмічний університет, кафедра комп'ютерних наук та інформаційних технологій, Харків, Україна; e-mail: yuriypolupan6@gmail.com; ORCID ID: 0009-0009-3030-8448

Малєєва Ольга Володимирівна – доктор технічних наук, професор, Національний аерокосмічний університет, професор кафедри комп'ютерних наук та інформаційних технологій, Харків, Україна; e-mail: o.maleyeva@khai.edu; ORCID ID: 0000-0002-9336-4182

Polupan Yuriy – PhD Student, National Aerospace University "Kharkiv Aviation Institute", Department of Computer Science and Information Technologies, Kharkiv, Ukraine.

Malyeyeva Olga – Doctor of Sciences (Engineering), Professor, National Aerospace University "Kharkiv Aviation Institute", Professor at the Department of Computer Science and Information Technologies, Kharkiv, Ukraine.

MODELS FOR SELECTING A RATIONAL SUPPLY STRUCTURE OF COMPONENTS FOR UAV MANUFACTURING UNDER RISK CONDITIONS, CONSTRAINED LOGISTICS COSTS AND TIME LIMITATIONS

The subject of this paper is logistics structures used to ensure the supply of components for the production of unmanned aerial vehicles (UAVs). **The goal** of the article is to improve the efficiency of supply of components for UAV production through the selection of a rational logistics structure. The following **tasks** are addressed: to analyze the peculiarities of the formation of the logistic structure of supply of components of the UAV production; to form a set of logistics supply structures options and criteria for their evaluation; and to developement mathematical models for the multicriteria selection of a logistics structure. **The methods** used include data analysis and generalization, structural variant synthesis and the main criterion method. The following **results** were obtained according to the objectives. The main types of component elements were generalized for different UAV classes. The analysis of potential component suppliers revealed a variety of choices, which determines the diversity and heterogeneity of possible supply structures. The formation and analysis of the main types of logistics structures made it possible to define a set of criteria for evaluation their advantages and disadvantages. A comparative analysis of various logistics structures of the enterprise is conducted, highlighting their key characteristics based on the main criteria such as minimal risk, cost efficiency, delivery speed, scalability, and flexibility in crisis conditions. A decision tree for the selection of logistics structure, considering the strategic goals of the enterprise (such as minimizing cost, delivery time, or logistics risks), was developed. A generalized multicriteria decision-making problem was formulated using the main criterion method. Three formalized models of the problem of selecting a rational logistics structure were discussed. **Conclusions**: The results of the study can be applied to improve logistics strategies in the field of UAV manufacturing, particularly in reducing delivery time (according to the established production rhythm) and minimizing risks during wartime conditions. Further research will focus on developing optimization models for combined supply and distribution structures.

Keywords: logistics structure; supplier selection; UAV production; supply routes; risks; supply chains; multicriteria selection; cost minimization; main criterion method.

Бібліографічні описи / Bibliographic descriptions

Полупан Ю. В., Малєєва О. В. Моделі вибору раціональної структури постачання комплектування для виробництва БПЛА в умовах ризику, обмежених логістичних витрат і часу. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 111–125. DOI: https://doi.org/10.30837/2522-9818.2025.3.111

Polupan, Y., Malyeyeva, O. (2025), "Models for selecting a rational supply structure of components for UAV manufacturing under risk conditions, constrained logistics costs and time limitations", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 111–125. DOI: https://doi.org/10.30837/2522-9818.2025.3.111

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

UDC 004.056.55:004.312.2 DOI: https://doi.org/10.30837/2522-9818.2025.3.126

V. RUDNYTSKYI, S. SEMENOV, N. LADA, V. BABENKO, V. LARIN, T. KOROTKYI

THE MODEL FOR CONSTRUCTING A SET OF SYMMETRIC TWO-OPERAND SET OPERATIONS THAT ALLOW PERVERSION OF OPERANDS BY COMBINING ONE-OPERAND OPERATIONS

The **subject matter** of the article is the scientific and methodological apparatus for constructing and implementing information-driven SET-operations. The **goal** of the work is to construct a model for the synthesis of symmetric two-operand SET-operations that allow permutation of operands and to determine the restrictions on their use in the development of low-resource stream ciphers. The following **tasks** are solved in the article: a model for the synthesis of symmetric two-operand SET-operations that allow permutation of operands is developed; restrictions on the use of this model are determined, which guarantee the construction of symmetric two-operand SET-operations that allow permutation of operands; it is shown that a pseudo-random change in the operating modes of the cryptographic system significantly complicates the processes of cryptanalysis of ciphergrams; modeling of SET-operations based on duplication of single-operand two-bit SET operations is carried out; the technology for constructing two-operand SET-operations based on duplication is applied to obtain a set of two-operand SET-operations. **Methods** of discrete mathematics, Set-theory and situational management have been introduced. Achieved **results**. Synthesized on the basis of this model of SET-operations allow changing the operating mode of stream cryptographic systems from symmetric to asymmetric when permuting operands in the operation. At present, the forefront of modern world experience in cryptographic science is a low-resource cryptographic system. Low-resource cryptography allows ensuring sufficiently stable cryptosystem performance with significant limitations on computational, mass-dimensional, cost and energy resources of the object of interest. One of the most relevant areas of further development of low-resource cryptography is considered to be cryptographic coding. In the process of cryptographic coding, cryptographic coding operations are implemented. The selection of operations is carried out under the control of a pseudo-random sequence, which is implemented by a cryptographic algorithm. **Conclusions**. A model for the synthesis of symmetric two-operand SET-operations that allow permutation of operands has been developed. Restrictions on the use of this model have been determined, which guarantee the construction of symmetric two-operand SET-operations that allow permutation of operands. The achieved results can be useful for the construction of low-resource stream encryption systems. Pseudo-random change of the operating modes of the cryptographic system significantly complicates the processes of cryptanalysis of ciphertexts.

Keywords: two-operand operations; cryptographic protection; low-resource cryptography; SET-encryption; SET-operations; one-operand operations.

Introduction

In the course of researching SET operation architectures, it was found that the results of their implementation depend on the results of the implementation of single-operand SET operations on which they are based [1, 2-5]. Therefore, a detailed analysis of the features of using these operations in the implementation of the simplest stream cipher algorithms, even if they are easily disclosed, is of great importance. This will significantly improve the efficiency of constructing ciphers based on multi-operand SET operations that will implement sets of various single-operand operations and their groups.

A CET operation is an operation in cryptographic encoding theory (CET). A CET operation is a numbered set of elementary functions, each of which forms a corresponding output Ci-quantum (ci-quantum from cryptographic information quantum) of information based on the processing of input Ci-quanta of information. A Ci-quantum is the minimum amount of information that a CET operation operates on [1, 2, 6]. Depending on the application of the CET operation, a ci-quantum can be a bit, byte, word, or double word.

A single-operand CET operation is a substitution table, and its use leads to the implementation of substitutions in the encryption process. In classical cryptography, there are four types of substitution ciphers [1, 4]:

- a simple substitution cipher, in which each character of the plaintext is replaced by another character from the substitution table during encryption;

- a homophonic substitution cipher, which differs from a simple transposition cipher in that each character of the plaintext can be replaced by one of several characters during encryption [5]. Its implementation is

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

associated with an increase in the number of characters used in the ciphertext compared to the plaintext;

- a polygram substitution cipher, in which groups of plaintext characters are replaced by groups of other characters from the substitution table;

- a polyalphabetic substitution cipher, in which several simple substitution ciphers are performed sequentially, the sequence of which is determined by the key [1].

Thus, the construction of a low-resource post-quantum control system in swarms of unmanned complexes and taking into account such features as: limited operating memory, reduced computing power, small physical area for implementation of the assembly, low energy capabilities, real-time response with resistance to external interference when creating a mathematical apparatus for describing linear and nonlinear transformations in permutation fields is a relevant scientific and applied problem.

Therefore, the creation of modern technology for building cryptographic systems involves the development of a theory of cryptographic encoding of information resources circulating in the network in near real time. In the course of our dissertation research, we will refer to this as CET encryption.

Analysis of the problem and existing methods

In [1], it is shown that highly efficient stream cipher systems can be built based on the use of SET operations (cryptographic encoding operations). In terms of their architecture, SET operations can be single-operand, two-operand, and multi-operand. Several approaches have been developed for constructing groups of symmetric two-operand SET operations: synthesis of groups of two-operand operations based on simulation results [3]; synthesis of groups of symmetric two-operand operations based on permutable circuits [4]; synthesis of groups of symmetric two-operand operations based on duplication of single-operand two-bit operations of the base group [5]. However, the practical implementation of these approaches to the synthesis of symmetric two-operand SET operations is associated with a number of difficulties. To synthesize groups of symmetric two-operand operations based on simulation results, it is necessary to have simulation results, but as the bit width of SET operations increases, the simulation time (required computing power) increases exponentially. The implementation of other approaches requires the availability of predefined symmetric two-operand SET operations.

Identification of previously unsolved parts of the general problem

Among stream ciphers based on SET operations, symmetric stream ciphers based on symmetric single-operand SET operations [1] are the least complex to implement. However, their disadvantage is the implementation of a single-operand SET operation of only one substitution table. The use of two-operand operations in stream ciphers provides the possibility of pseudo-random use of several substitution tables [2]. It is difficult to find or construct a new unknown symmetric two-operand operation that allows the permutation of operands for the subsequent synthesis of a group of modified SET operations with similar properties. Unfortunately, this problem has not been solved to date.

The purpose of this article is to construct a model for synthesizing symmetric two-operand SET operations that allow operand permutation and to determine the limitations on its use in the development of low-resource stream ciphers.

Main material

Let us consider the implementation of SET operation modeling based on the duplication of single-operand two-bit SET operations [1, 7-9].

When constructing two-operand operations, we will use the group of single-operand SET operations given in Table 1.

To construct a set of two-operand SET operations that allow operand permutation, we will use a two-operand SET operation [1, 9]:

$$C_{1,7,19,13}(x,y)=\begin{bmatrix}x_1\\x_2\end{bmatrix}\oplus\begin{bmatrix}y_1\\y_1\oplus y_2\end{bmatrix}. \qquad (1)$$

Using the technology of constructing two-operand SET operations based on duplication, we obtain a set of two-operand SET operations [5, 10].

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Table 1. *Discrete models of direct and inverse single-operand SET operations [1]*

$C_1(x)=\begin{bmatrix}x_1\\x_2\end{bmatrix}=C_1^{/}(x)$	$C_7(x)=\begin{bmatrix}x_1\\x_2\oplus 1\end{bmatrix}=C_7^{/}(x)$	$C_{13}(x)=\begin{bmatrix}x_1\oplus 1\\x_2\end{bmatrix}=C_{13}^{/}(x)$	$C_{19}(x)=\begin{bmatrix}x_1\oplus 1\\x_2\oplus 1\end{bmatrix}=C_{19}^{/}(x)$
$C_2(x)=\begin{bmatrix}x_1\oplus x_2\\x_2\end{bmatrix}=C_2^{/}(x)$	$C_8(x)=\begin{bmatrix}x_1\oplus x_2\\x_2\oplus 1\end{bmatrix}=C_{20}^{/}(x)$ $C_8^{/}(x)=\begin{bmatrix}x_1\oplus x_2\oplus 1\\x_2\oplus 1\end{bmatrix}=C_{20}(x)$	$C_{14}(x)=\begin{bmatrix}x_1\oplus x_2\oplus 1\\x_2\end{bmatrix}=C_{14}^{/}(x)$	$C_{20}(x)=\begin{bmatrix}x_1\oplus x_2\oplus 1\\x_2\oplus 1\end{bmatrix}=C_8^{/}(x)$ $C_{20}^{/}(x)=\begin{bmatrix}x_1\oplus x_2\\x_2\oplus 1\end{bmatrix}=C_8(x)$
$C_3(x)=\begin{bmatrix}x_1\\x_1\oplus x_2\end{bmatrix}=C_3^{/}(x)$	$C_9(x)=\begin{bmatrix}x_1\\x_1\oplus x_2\oplus 1\end{bmatrix}=C_9^{/}(x)$	$C_{15}(x)=\begin{bmatrix}x_1\oplus 1\\x_1\oplus x_2\end{bmatrix}=C_{21}^{/}(x)$ $C_{15}^{/}(x)=\begin{bmatrix}x_1\oplus 1\\x_1\oplus x_2\oplus 1\end{bmatrix}=C_{21}(x)$	$C_{21}(x)=\begin{bmatrix}x_1\oplus 1\\x_1\oplus x_2\oplus 1\end{bmatrix}=C_{15}^{/}(x)$ $C_{21}^{/}(x)=\begin{bmatrix}x_1\oplus 1\\x_1\oplus x_2\end{bmatrix}=C_{15}(x)$
$C_4(x)=\begin{bmatrix}x_2\\x_1\end{bmatrix}=C_4^{/}(x)$	$C_{10}(x)=\begin{bmatrix}x_2\\x_1\oplus 1\end{bmatrix}=C_{16}^{/}(x)$ $C_{10}^{/}(x)=\begin{bmatrix}x_2\oplus 1\\x_1\end{bmatrix}=C_{16}(x)$	$C_{16}(x)=\begin{bmatrix}x_2\oplus 1\\x_1\end{bmatrix}=C_{10}^{/}(x)$ $C_{16}^{/}(x)=\begin{bmatrix}x_2\\x_1\oplus 1\end{bmatrix}=C_{10}(x)$	$C_{22}(x)=\begin{bmatrix}x_2\oplus 1\\x_1\oplus 1\end{bmatrix}=C_{22}^{/}(x)$
$C_5(x)=\begin{bmatrix}x_2\\x_1\oplus x_2\end{bmatrix}=C_6^{/}(x)$	$C_{11}(x)=\begin{bmatrix}x_2\\x_1\oplus x_2\oplus 1\end{bmatrix}=C_{18}^{/}(x)$	$C_{17}(x)=\begin{bmatrix}x_2\oplus 1\\x_1\oplus x_2\end{bmatrix}=C_{24}^{/}(x)$	$C_{23}(x)=\begin{bmatrix}x_2\oplus 1\\x_1\oplus x_2\oplus 1\end{bmatrix}=C_{12}^{/}(x)$
$C_5^{/}(x)=\begin{bmatrix}x_1\oplus x_2\\x_1\end{bmatrix}=C_6(x)$	$C_{11}^{/}(x)=\begin{bmatrix}x_1\oplus x_2\oplus 1\\x_1\end{bmatrix}=C_{18}(x)$	$C_{17}^{/}(x)=\begin{bmatrix}x_1\oplus x_2\oplus 1\\x_1\oplus 1\end{bmatrix}=C_{24}(x)$	$C_{23}^{/}(x)=\begin{bmatrix}x_1\oplus x_2\\x_1\oplus 1\end{bmatrix}=C_{12}(x)$
$C_6(x)=\begin{bmatrix}x_1\oplus x_2\\x_1\end{bmatrix}=C_5^{/}(x)$	$C_{12}(x)=\begin{bmatrix}x_1\oplus x_2\\x_1\oplus 1\end{bmatrix}=C_{23}^{/}(x)$	$C_{18}(x)=\begin{bmatrix}x_1\oplus x_2\oplus 1\\x_1\end{bmatrix}=C_{11}^{/}(x)$	$C_{24}(x)=\begin{bmatrix}x_1\oplus x_2\oplus 1\\x_1\oplus 1\end{bmatrix}=C_{17}^{/}(x)$
$C_6^{/}(x)=\begin{bmatrix}x_2\\x_1\oplus x_2\end{bmatrix}=C_5(x)$	$C_{12}^{/}(x)=\begin{bmatrix}x_2\oplus 1\\x_1\oplus x_2\oplus 1\end{bmatrix}=C_{23}(x)$	$C_{18}^{/}(x)=\begin{bmatrix}x_2\\x_1\oplus x_2\oplus 1\end{bmatrix}=C_{11}(x)$	$C_{24}^{/}(x)=\begin{bmatrix}x_2\oplus 1\\x_1\oplus x_2\end{bmatrix}=C_{17}(x)$

To construct inverse SET operations, we use the relationship between operations:

$$C(x,y)=\begin{cases}C_i, if\ y_1=0; y_2=0\\ C_j, if y_1=0; y_2=1\\ C_n, if y_1=1; y_2=0\\ C_m, if y_1=1; y_2=1\end{cases} \rightleftarrows\rightleftarrows\rightleftarrows\rightleftarrows\Rightarrow C^{/}(x,y)=\begin{cases}C_i^{/}, if y_1=0; y_2=0\\ C_j^{/}, if y_1=0; y_2=1\\ C_n^{/}, if y_1=1; y_2=0\\ C_m^{/}, if y_1=1; y_2=1\end{cases} \quad (2)$$

The results of constructing a set of two-operand SET operations based on operation (1) are shown in Table 2.

In the process of analyzing the results of modeling SET operations based on the duplication of single-operand two-bit operations in Table 2, it was assumed that two-operand SET operations can be synthesized by combining single-operand SET operations. This assumption is based on the fact that when using the method of duplicating basic groups based on asymmetric SET operations, the resulting two-operand SET operations of the operand processing model do not coincide. Therefore, it can be argued that two-operand SET operations can be synthesized not only by duplicating basic groups, but also by combining single-operand cryptotransformation operations [1, 11-13].

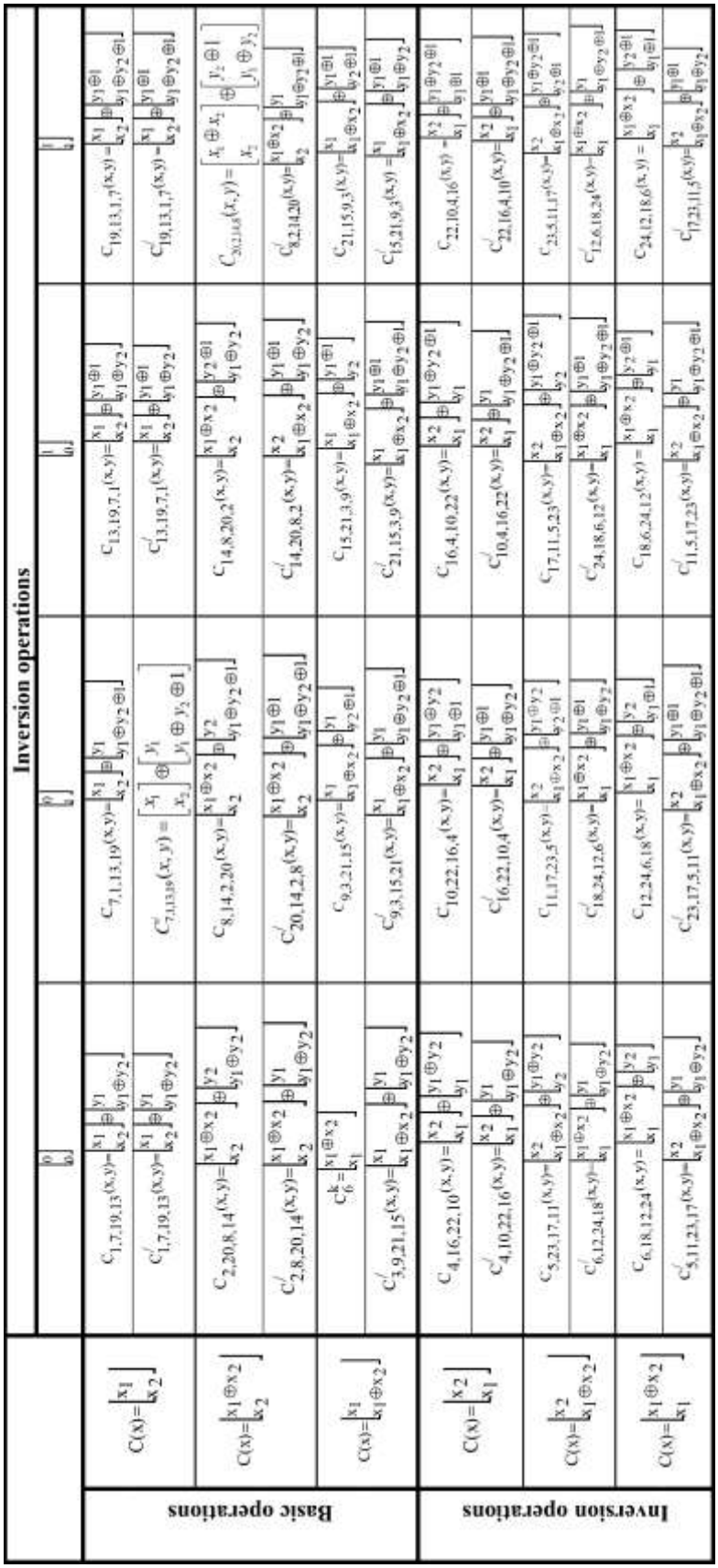
In the process of creating two-operand operations, it is possible to combine symmetric single-operand operations, symmetric and asymmetric single-operand operations, as well as asymmetric single-operand operations [1].

When forming sets of two-operand operations, the operation of processing the first operand ($_x$) is sequentially combined with the operations of processing the second operand ($_y$). Since all operations for processing the second operand are used in the synthesis of the set, the construction of a set of symmetric and asymmetric two-operand operations will depend on the first operand. Based on this, it can be assumed that when selecting a symmetric single-operand operation for processing the first operand, a set of symmetric two-operand operations can be obtained [1, 14-16].

Let's build two-operand operations by combining single-operand operations. To form a set of two-operand operations, let's select the first single-operand SET operation $C_1(x)$.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Table 2. *A synthesized set of models of asymmetric two-operand two-bit double-cycle operations based on operation duplication for the first encryption scenario [1]*



ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Let's place the first 4 two-operand SET operations based on the 4 single-operand SET operations given in the first row of Table 1.

$$C_{1,1}(x,y)=C_1\begin{pmatrix}x_1\\x_2\end{pmatrix}\oplus C_1\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}\oplus\begin{bmatrix}y_1\\y_2\end{bmatrix} \quad (3)$$

Substituting the value of the second operand into the model of operation (3), we obtain a tuple model of a two-operand SET operation [1].

$$C_{1,1}(x,y)=\begin{bmatrix}x_1\\x_2\end{bmatrix}\oplus\begin{bmatrix}y_1\\y_2\end{bmatrix}=\begin{cases}\begin{bmatrix}x_1\\x_2\end{bmatrix}, if y_1=0; y_2=0\\ \begin{bmatrix}x_1\\x_2\oplus 1\end{bmatrix}, if y_1=0; y_2=1\\ \begin{bmatrix}x_1\oplus 1\\x_2\end{bmatrix}, if y_1=1; y_2=0\\ \begin{bmatrix}x_1\oplus 1\\x_2\oplus 1\end{bmatrix}, if y_1=1; y_2=1\end{cases}=C_{1,7,13,19}(x,y).$$

the tuple model, the numbering of tuple elements coincides with the numbering of SET operations given in Table 1.

By analogy, we obtain the other three tuple models of SET operations [1].

$$C_{1,7}(x,y)=C_1\begin{pmatrix}x_1\\x_2\end{pmatrix}\oplus C_7\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}\oplus\begin{bmatrix}y_1\\y_2\oplus 1\end{bmatrix}=C_{7,1,19,13}(x,y)$$

$$C_{1,13}(x,y)=C_1\begin{pmatrix}x_1\\x_2\end{pmatrix}\oplus C_{13}\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}\oplus\begin{bmatrix}y_1\oplus 1\\y_2\end{bmatrix}=C_{13,19,1,7}(x,y)$$

$$C_{1,19}=C_1\begin{pmatrix}x_1\\x_2\end{pmatrix}\oplus C_{19}\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}\oplus\begin{bmatrix}y_1\oplus 1\\y_2\oplus 1\end{bmatrix}=C_{19,13,7,1}(x,y)$$

Let us form models of inverse SET operations based on relation (2) and find the relationship between the obtained pairs of operations. [1, 17].

As $C_1(x)\Rightarrow C_1^{/}(x)$, $C_7(x)\Rightarrow C_7^{/}(x)$; $C_{13}(x,y)\Rightarrow C_{13}(y,x)=C_{13}^{/}(x,y)$, $C_{19}(x)\Rightarrow C_{19}^{/}(x)$, then, according to expression (2), we get:

$C_{1,7,13,19}(x,y)\Rightarrow C_{1,7,13,19}^{/}(x,y)$, that means $C_{1,1}(x,y)\Rightarrow C_{1,1}^{/}(x,y)$.

$C_{7,1,19,13}(x,y)\Rightarrow C_{7,1,19,137}^{/}(x,y)$, that means $C_{1,7}(x,y)\Rightarrow C_{1,7}^{/}(x,y)$.

$C_{13,19,1,7}(x,y)\Rightarrow C_{13,19,1,7}^{/}(x,y)$, that means $C_{1,13}(x,y)\Rightarrow C_{1,13}^{/}(x,y)$.

$C_{19,13,7,1}(x,y)\Rightarrow C_{19,13,7,1}^{/}(x,y)$, that means $C_{1,19}(x,y)\Rightarrow C_{1,19}^{/}(x,y)$.

We synthesize the following four operations based on the second row of Table 1.

Let's create a SET operation $C_{1,2}(x,y)$ analogous to the operation $C_{1,1}(x,y)$.

$$C_{1,2}(x,y)=C_1\begin{pmatrix}x_1\\x_2\end{pmatrix}\oplus C_2\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}\oplus\begin{bmatrix}y_1\oplus y_2\\y_2\end{bmatrix}=C_{1,19,13,7}(x,y)$$

Then: $(C_{1,19,13,7}(x,y)\Rightarrow C_{1,19,13,7}^{/}(x,y))\Rightarrow(C_{1,2}(x,y)\Rightarrow C_{1,2}^{/}(x,y))$.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

By analogy, we will construct operations

$$C_{1,8}(x,y),\ C_{1,14}(x,y),\ C_{1,17}(x,y) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \oplus 1 \\ y_1 \oplus y_2 \end{bmatrix}$$

$$C_{1,8}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_8\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_1 \oplus y_2 \\ y_2 \oplus 1 \end{bmatrix} = C_{7,13,19,1}(x,y);$$

$$(C_{7,13,19,1}(x,y) \Rightarrow C'_{7,13,19,1}(x,y)) \Rightarrow (C_{1,8}(x,y) \Rightarrow C'_{1,8}(x,y)).$$

$$C_{1,14}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_{14}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_1 \oplus y_2 \oplus 1 \\ y_2 \end{bmatrix} = C_{13,7,1,19}(x,y);$$

$$(C_{13,7,1,19}(x,y) \Rightarrow C'_{13,7,1,19}(x,y)) \Rightarrow (C_{1,14}(x,y) \Rightarrow C'_{1,14}(x,y)).$$

$$C_{1,20}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_{20}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_1 \oplus y_2 \oplus 1 \\ y_2 \oplus 1 \end{bmatrix} = C_{20,1,7,13}(x,y);$$

$$(C_{20,1,7,13}(x,y) \Rightarrow C'_{20,1,7,13}(x,y)) \Rightarrow (C_{1,20}(x,y) \Rightarrow C'_{1,20}(x,y)).$$

Let's find patterns between operations based on the fourth row of Table 1.:

$$C_{1,4}(x,y),\ C_{1,10}(x,y),\ C_{1,16}(x,y),\ C_{1,22}(x,y)$$

$$C_{1,4}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_4\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \\ y_1 \end{bmatrix} = C_{1,13,7,19}(x,y);$$

$$(C_{1,13,7,19}(x,y) \Rightarrow C'_{1,13,7,19}(x,y)) \Rightarrow (C_{1,4}(x,y) \Rightarrow C'_{1,4}(x,y)).$$

$$(C_{20,14,8,2}(x,y) \Rightarrow C_{20,14,8,2}(y,x) = C'_{8,14,20,2}(x,y)) \Rightarrow (C_{2,19}(x,y) \Rightarrow C_{2,19}(y,x) = C'_{2,8}(x,y));$$

$$(C_{7,19,1,13}(x,y) \Rightarrow C'_{7,19,1,13}(x,y)) \Rightarrow (C_{1,10}(x,y) \Rightarrow C'_{1,10}(x,y)).$$

$$C_{1,16}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_{16}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \oplus 1 \\ y_1 \end{bmatrix} = C_{13,1,19,7}(x,y);$$

$$(C_{13,1,19,7}(x,y) \Rightarrow C'_{13,1,19,7}(x,y)) \Rightarrow (C_{1,16}(x,y) \Rightarrow C'_{1,16}(x,y)).$$

$$C_{1,22}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_{22}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \oplus 1 \\ y_1 \oplus 1 \end{bmatrix} = C_{19,7,13,1}(x,y);$$

$$(C_{19,7,13,1}(x,y) \Rightarrow C'_{19,7,13,1}(x,y) \Rightarrow (C_{1,22}(x,y) \Rightarrow C'_{1,22}(x,y)).$$

Let's form inverse operations $C'_{1,5}(x,y)$ $C'_{1,11}(x,y)$, $C'_{1,17}(x,y)$, $C'_{1,23}(x,y)$.

$$C_{1,5}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_5\begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \\ y_1 \oplus y_2 \end{bmatrix} = C_{1,19,7,13}(x,y);$$

$$(C_{1,19,7,13}(x,y) \Rightarrow C'_{1,19,7,13}(x,y)) \Rightarrow (C_{1,5}(x,y) \Rightarrow C'_{1,5}(x,y)).$$

$$C_{1,11}(x,y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_{11}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \\ y_1 \oplus y_2 \oplus 1 \end{bmatrix} = C_{7,13,1,19}(x,y);$$

$$(C_{7,13,1,19}(x,y) \Rightarrow C'_{7,13,1,19}(x,y) \Rightarrow (C_{1,11}(x,y) \Rightarrow C'_{1,11}(x,y)).$$

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

$$C_{1,17} = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_{17}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \oplus 1 \\ y_1 \oplus y_2 \end{bmatrix} = C_{13,7,19,1}(x, y);$$

$$(C_{13,7,19,1}(x, y) \Rightarrow C'_{13,7,19,1}(x, y)) \Rightarrow (C_{1,17}(x, y) \Rightarrow C'_{1,17}(x, y)).$$

$$C_{1,23}(x, y) = C_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_{23}\begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_2 \oplus 1 \\ y_1 \oplus y_2 \oplus 1 \end{bmatrix} = C_{19,1,13,7}(x, y);$$

$$(C_{19,1,13,7}(x, y) \Rightarrow C'_{19,1,13,7}(x, y)) \Rightarrow (C_{1,23}(x, y) \Rightarrow C'_{1,23}(x, y))$$

By analogy, we will create all other pairs of two-operand SET operations based on the operation $C_1(x)$. The simulation results are shown in Table 3.

The summary results of modeling operations based on the first operation of processing the first operand $C_1(x)$ (Table 3) showed that regardless of the operation of processing the second operand, if the single-operand operation of processing the first operand is symmetric, then all two-operand operations built on its basis will be symmetric [1, 18].

Let us formalize this conclusion:

$$C(x, y) = C_i(x) \oplus C_j(y) = C'(x, y) \quad (4)$$

where $C_i(x) = C'_i(x)$.

Condition (4) will be satisfied if, for a symmetric unary operation on which a set of symmetric binary operations is based, there exists a group of symmetric inversion operations [1]. Based on this restriction, the correct implementation of the synthesis model (4) can be represented as follows:

$$C_i(x, y) = C_j\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus C_1\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = C'_i(x, y) \quad (5)$$

where $C_1\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$ operation of halting the results of the operation $C_j\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ performance.

Since modifications of single-operand SET operations by gammification are placed in the rows of Table 1, only SET operations placed in rows where direct and inverse transformations coincide correspond to condition 5 [1].

For two-bit two-operand operations, only 4 sets of symmetric operations can be constructed based on single-operand operations $C_1(x)$, $C_7(x)$, $C_{13}(x)$ and $C_{19}(x)$. These operations are symmetric and represent a group of inversion operations [1]. The number of sets coincided with the results of the computational experiment.

For three-bit two-operand SET operations, 8 sets of symmetric operations can be constructed based on single-operand SET operations, and for four-bit two-operand SET operations, 16 sets of symmetric operations can be constructed.

Research results and discussion

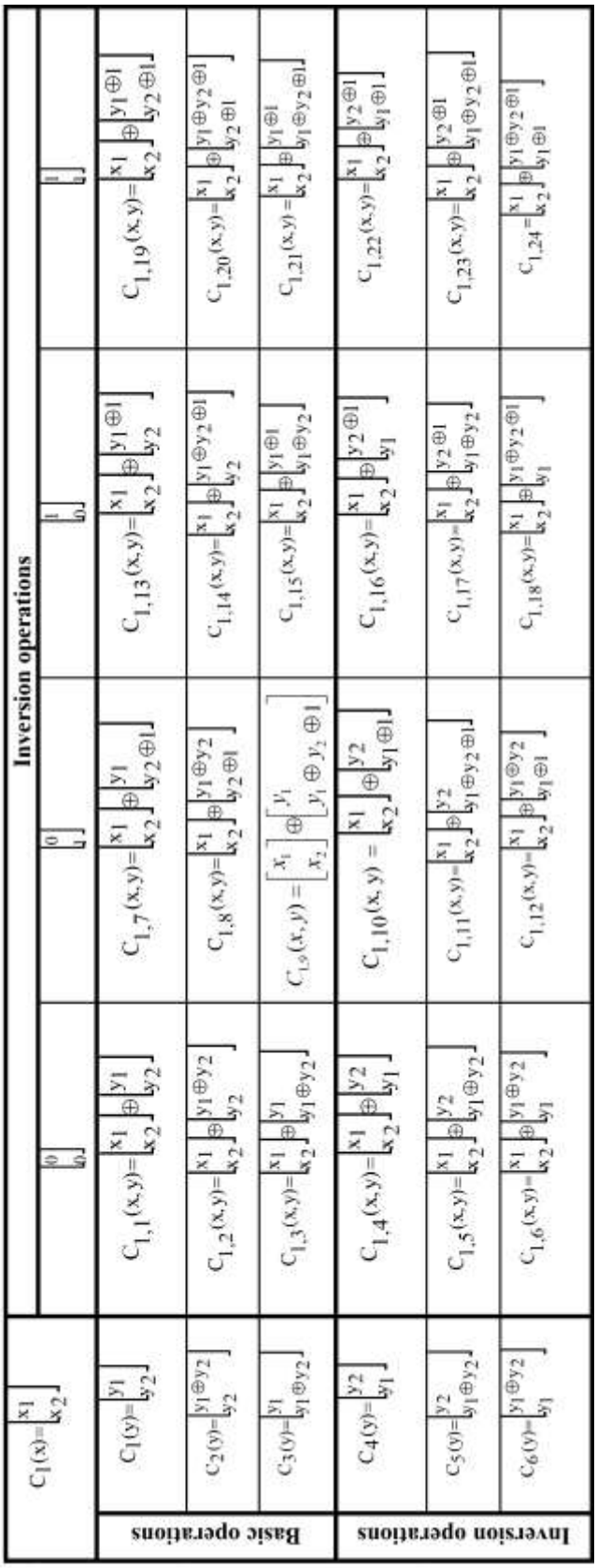
However, it should be noted that rearranging operands in synthesized operations will result in the implementation of another SET operation, which will be asymmetric.

The research revealed that for encoding and decoding operations, the first operand will be information, and the second operand will be a fading sequence. Any two-bit two-operand cryptographic transformation operation can be represented as the execution of one of four single-operand two-bit operations on the first operand, depending on the value of the two bits of the second operand.

The use of these operations in the creation of cryptographic systems will ensure the formation of low-resource stream cipher systems. The permutation of operands in SET operations will allow changing the operating mode of the cryptographic system from symmetric to asymmetric and vice versa.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Table 3. *A synthesized set of models of two-operand two-bit operations obtained by combining single-operand two-bit cryptographic transformation operations with the operation* $C_1(x)$ [1]



ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Conclusions and prospects for further development

A model for synthesizing symmetric two-operand SET operations that allow operand permutation has been developed. The restrictions on the use of this model guarantee the construction of symmetric two-operand SET operations that allow operand permutation. The use of this model provides the possibility of synthesizing SET operations for building low-resource stream cipher systems. SET operations synthesized on the basis of this model allow changing the operating mode of stream cryptographic systems from symmetric to asymmetric when swapping operands in the operation. Pseudorandom changes in the operating modes of a cryptographic system significantly complicate the processes of cryptanalysis of ciphertexts.

A promising direction for further research is the development of a method for constructing discrete-casual models of three-bit SET operations of information-controlled permutations by combining single-operand operations of two-bit SET operations.

References

1. Rudnytskyi, V. M. (2024), "Architecture of CET-operations and stream encryption technologies: monograph" / V. M. Rudnytskyi, N. V. Lada, G. A. Kuchuk, D. A. Pidlasy. Cherkasy: publisher Ponomarenko R.V., 374 p. ISBN 978-966-2554-81-6 URL: https://dndivsovt.com/index.php/monograph/issue/ view/22/22.
2. Kosenko, V., Persiyanova, E., Malyeyeva, O. (2017), “Methods of managing traffic distribution in information and communication networks of critical infrastructure systems”, *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (2), P. 48–55. DOI: https://doi.org/10.30837/2522-9818.2017.2.048
3. Zheng, Z.,Tian, K. & Liu, F. (2023), "Modern Cryptography Volume 2. A Classical Introduction to Informational and Mathematical Principle". *Springer: Singapore.* DOI: https://doi.org/10.1007/978-981-19-7644-5
4. Kumar, C., Prajapati, S. & Verma, R. (2022), "A Survey of Various Lightweight Cryptography Block ciphers for IoT devices". *IEEE International Conference on Current Development in Engineering and Technology (CCET)*, Bhopal, India, 2022, P. 1–6. DOI: https://doi.org/10.1109/CCET56606.2022.10080556
5. Rudnytskyi, V., Babenko, V., Lada, N., Tarasenko, Ya. & Rudnytska, Yu. (2022), “Constructing symmetric operations of cryptographic information encoding”. *Workshop on Cybersecurity Providing in Information and Telecommunication Systems (CPITS II 2021)*, Oct. 26, 2021. Kyiv, Ukraine: CEUR Workshop Proceedings, 2022, P. 182–194. ISSN 1613-0073.
6. Thomas, Xuan Meng, W. (2020), "Buchanan. Lightweight Cryptographic Algorithms on Resource-Constrained Devices". Preprints. DOI: https://doi.org/10.20944/PREPRINTS202009.0302.V1
7. Semenov, S., Davydov, V., Kuchuk, N. & Petrovskaya, I. (2021), "Software security threat research", *31st International Scientific Symposium Metrology and Metrology Assurance, MMA 2021.* DOI: https://doi.org/10.1109/MMA52675.2021.9610877
8. Khaled, Salah Mohamed (2020), "New Frontiers in Cryptography. Quantum, Blockchain, Lightweight, Chaotic and DNA". *Springer International Publishing*. Springer, Cham, 104. DOI: https://doi.org/10.1007/978-3-030-58996-7
9. Yalamuri, G., Honnavalli, P. & Eswaran, S. (2022), "A Review of the Present Cryptographic Arsenal to Deal with Post-Quantum Threats". *Procedia Comput. Sci*. 215, P. 834–845. DOI: https://doi.org/10.1016/j.procs.2022.12.086
10. Lada, N., Kozlovska, S., Rudnytskyi, S. (2019), "Pobudova matematychnoyi hrupy symetrychnykh operatsiy na osnovi dodavannya za modulem dva". ["The symmetric operations’ mathematical group constructing based on modulo-2 addition"]. *Modern special equipment: scientific and practical journal*. Kyiv, 2019. № 4 (59). P. 33–41. DOI: https://elar.naiau.kiev.ua/items/33ebc0b3-4177-4d6f-87a8-95719bb7fabe
11. Rudnytska, Yu., Rudnytsky, S. (2022), "Modelyuvannya symetrychnykh operatsiy kryptohrafichnoho koduvannya". ["Modeling symmetric cryptographic coding operations"]. P*roblems of informatization: Tenth International Scientific and Technical Conference: Abstracts of the Supplementary*. Cherkasy – Baku – Bielsko-Biala – Kharkiv, 24 – 25 November 2022. P. 10. available at: https://nure.ua/wp-content/uploads/2022/tom-1.pdf
12. Larin, V., Trystan, A., Hmyria, V., Strikha, S., Kostashchuk, M., Piontkivskyi, P. (2024), "Forsayt yak innovatsiynyy instrument planuvannya ta realizatsiyi naukovykh tekhnolohiy v oboronno-promyslovomu kompleksi". ["Foresight as an innovative tool for planning and implementing scientific technologies in the defence industry"]. Problems of construction, testing, application and operation of complex information systems, 26 (1), P. 93–106. DOI: https://doi.org/10.46972/2076-1546.2024.26.08

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

13. Jancarczyk, D., Rudnytskyi, V., Breus, R., Pustovit, M., Veselska, O, & Ziubina, R. (2020), "Two-Operand Operations of Strict Stable Cryptographic Coding with Different Operands' Bits". (2020) *IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS).* IEEE; 2020 Sep 17; P. 247–254. DOI: https://doi.org/10.1109/IDAACS-SWS50031.2020.9297067.
14. Zakaria, A., Azni, A., Ridzuan, F., Zakaria, N. & Maslina, H. (2023), "Daud Systematic literature review: Trend analysis on the design of lightweight block cipher". *IEEE Journal of King Saud University Computer and Information Sciences*. 35, Issue 5, May, 101550 p. DOI: https://doi.org/10.1016/j.jksuci.2023.04.003
15. Gasser, L. (2023), "Post-quantum Cryptography. In: Mulder", V., Mermoud, A., Lenders, V., Tellenbach, B. (eds) *Trends in Data Protection and Encryption Technologies. Springer*, Cham. DOI: https://doi.org/10.1007/978-3-031-33386-6_10
16. Zakaria, A., et al. (2023). Systematic literature review: Trend analysis on the design of lightweight block cipher. *Journal of King Saud University Computer and Information Sciences*, 35(5), 101550 p. DOI: https://doi.org/10.1016/j.jksuci.2023.04.003
17. Yasmin, N. & Gupta, R. (2023), "Modified lightweight cryptography scheme and its applications in IoT environment". *Int. j. inf. tecnol*. 15, P. 4403–4414. DOI: https://link.springer.com/article/10.1007/s41870-023-01486-2
18. Sabani, M., et al. (2023), "Evaluation and Comparison of Lattice-Based Cryptosystems for a Secure Quantum Computing Era". *Electronics*, 12(12), 2643 p. DOI: https://doi.org/10.3390/electronics12122643

Надійшла (Received) 25.05.2025

Відомості про авторів / About the Authors

Rudnytskyi Volodymyr – Doctor of Engineering Science, Professor Principal Researcher of State Scientific Research Institute of Armament and Military Equipment Testing and Certification, Cherkasy, Ukraine; e-mail: rvn_2008@ukr.net; ORCID ID: https://orcid.org/0000-0003-3473-7433

Semenov Serhii – Doctor of Engineering Science, Professor, Professor at University of the National Education Commission, Krakow, Poland; e-mail: serhii.semenov@uken.krakow.pl; ORCID ID: https://orcid.org/0000-0003-4472-9234

Lada Nataliia – Candidate of Technical Sciences, Leading Researcher State Scientific Research Institute of Armament and Military Equipment Testing and Certification, Cherkasy, Ukraine; e-mail: Ladanatali256@gmail.com; ORCID ID: https://orcid.org/0000-0002-7682-2970

Babenko Vira – Doctor of Engineering Science, Professor, the chef of the department of Cherkasy State Technological University, Cherkasy, Ukraine; e-mail; verababenko84@gmail.com; ORCID ID: https://orcid.org/0000-0003-2039-2841

Larin Volodymyr – Doctoral Candidate, Candidate of Technical Sciences, Associate Professor, State Scientific Research Institute of Armament and Military Equipment Testing and Certification; e-mail: l_vv83@ukr.net; ORCID ID: https://orcid.org/0000-0003-0771-2660

Korotkyi Tymofii – Post-Graduate of Cherkasy State Technological University, Cherkasy, Ukraine; e-mail: rvn_2008@ukr.net; ORCID ID: https://orcid.org/0009-0003-5159-5892

Рудницький Володимир Миколайович – доктор технічних наук, професор, головний науковий співробітник Державного науково-дослідного інституту випробувань і сертифікації озброєння та військової техніки, Черкаси, Україна.

Семенов Сергій Геннадійович – доктор технічних наук, професор, професор Університету Національної комісії з питань освіти, Краків, Польща.

Лада Наталія Володимирівна – кандидат технічних наук, провідний науковий співробітник Державного науково-дослідного інституту випробувань і сертифікації озброєння та військової техніки, Черкаси, Україна.

Бабенко Віра Григорівна – доктор технічних наук, професор, завідувач кафедри Черкаського державного технологічного університету, Черкаси, Україна.

Ларін Володимир Валерійович – кандидат технічних наук, доцент, докторант штатний Державного науково-дослідного інституту випробувань і сертифікації озброєння та військової техніки, Черкаси, Україна.

Короткий Тимофій Костянтинович – аспірант Черкаського державного технологічного університету, Черкаси, Україна.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

МОДЕЛЬ ПОБУДОВИ МНОЖИНИ СИМЕТРИЧНИХ ДВОХОПЕРАНДНИХ СЕТ-ОПЕРАЦІЙ, ЯКІ ДОПУСКАЮТЬ ПЕРЕСТАНОВКУ ОПЕРАНДІВ

Предметом дослідження в статті є науково-методологічний апарат побудови та реалізації СЕТ-операцій керованих інформацією. **Мета роботи** – побудова моделі синтезу симетричних двохоперандних СЕТ-операцій, які допускають перестановку операдів і визначення обмежень на їх використання при розробці малоресурсних потокових шифрів. У статті розв'язано такі **завдання**: розроблено модель синтезу симетричних двохоперандних СЕТ-операцій, які допускають перестановку операндів; визначено обмеження на використання даної моделі, які гарантують побудову симетричних двохоперандних СЕТ-операцій, які допускають перестановку операндів; показано, що псевдовипадкова зміна режимів роботи криптографічної системи суттєво ускладнює процеси криптоаналізу шифрограм; проведено моделювання СЕТ-операцій на снові дублювання одноопераднних двохрозрядних СЕТ-операцій; застосовано технологію побудови двохоперандних СЕТ-операцій на основі дублювання для отримання множини двохоперандних СЕТ-операцій. Упроваджено **методи** дискретної математики, теорії множин і ситуаційного управління. Досягнуті наступні **результати**: синтезовані на основі даної моделі СЕТ-операцій дозволяють змінювати режим роботи потокових криптографічних систем з симетричного на несиметричний при перестановці місцями операндів в операції. На теперешній час в авангарді сучасного світового досвіду криптографічної науки є малоресурсна криптографічна система. Малоресурсна криптографія дозволяє забезпечити достатньо стійкі показникі криптосистеми при значному обмеженні обчислювальних, масогабаритних, вартісних та енергетичних ресурсів об'єкта інтересу. Одним із найбільш атуальним напрямком подальшого розвитку малоресурсної криптографії вважається криптографічне кодування. В процесі криптографічного кодування реалізуються операції криптографічного кодування. Вибір операцій проводиться під управлінням псевдовипадкової послідовності, яка реалізується криптографічним алгоритмом. **Висновки**. Розроблена модель синтезу симетричних двохоперандних СЕТ-операцій, які допускають перестановку операндів. Визначені обмеження на використання даної моделі, які гарантують побудову симетричних двохоперандних СЕТ-операцій, які допускають перестановку операндів. Досягнуті результати можуть бути корисними для побудови малоресурсних систем потокового шифрування. Псевдовипадкова зміна режимів роботи криптографічної системи суттєво ускладнює процеси криптоаналізу шифрограм.

Ключові слова: двохоперандні операції; криптографічний захист; малоресурсна криптографія; СЕТ-шифрування; СЕТ-операції; одноперандні операції.

Бібліографічні описи / Bibliographic descriptions

Рудницький В.М., Семенов С. Г., Лада Н.В., Бабенко В.Г., Ларін В.В., Короткий Т.К. Модель побудови множини симетричних двохоперандних сет-операцій, які допускають перестановку операндів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 126–136. DOI: https://doi.org/10.30837/2522-9818.2025.3.126

Rudnytskyi, V., Semenov, S., Lada, N., Babenko, V., Larin, V., Korotkyi, T. (2025), "The model for constructing a set of symmetric two-operand set operations that allow perversion of operands by combining one-operand operations", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 126–136. DOI: https://doi.org/10.30837/2522-9818.2025.3.126

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.43:004.738.5:004.89 DOI: https://doi.org/10.30837/2522-9818.2025.3.137

О. Свєтлінський, І. Ревенчук

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ВІРТУАЛІЗАЦІЙНИХ РІШЕНЬ ДЛЯ РОЗГОРТАННЯ ЦИФРОВИХ ДВІЙНИКІВ У СИСТЕМАХ З ОБМЕЖЕНИМИ РЕСУРСАМИ

Предметом дослідження є засоби та підходи до віртуалізації в середовищах з обмеженими ресурсами, зокрема на одноплатних комп'ютерах, а також їх здатність забезпечувати якісну роботу цифрових двійників. **Мета статті** – оцінити ефективність віртуалізації для розгортання цифрових двійників на одноплатних комп'ютерах і розробити рекомендації щодо впровадження віртуалізації в середовищах з обмеженими обчислювальними ресурсами. У статті визначено такі **завдання**: аналіз сучасних апаратних рішень одноплатних комп'ютерів і наявних технологій віртуалізації; побудова експериментального середовища для практичного порівняння технологій віртуалізації на одноплатних комп'ютерах; реалізація прикладного навантаження, що імітує роботу цифрового двійника з реальними сенсорними показниками; дослідження ефективності мережної взаємодії між віртуальними середовищами та хост-системою; порівняння результатів за ключовими метриками. Упроваджено такі **методи**: технології віртуалізації *Docker*, *QEMU*, *Firecracker*; мережний моніторинг. **Досягнуті результати**: проведено порівняльний експеримент з використанням *Docker*, *QEMU* та *Firecracker* на базі *Raspberry Pi 4*; проаналізовано літературу та доступну документацію з окресленого питання; реалізовано базовий прототип цифрового двійника з під'єднанням фізичних сенсорних пристроїв, що збирають телеметричні показники в реальному часі; забезпечено поетапне автоматичне розгортання віртуальних середовищ із попередньо налаштованими компонентами для швидкого масштабування експерименту; проведено вимірювання мережних затримок, часу оброблення повідомлень та використання ресурсів платформи; визначено переваги та недоліки технологій віртуалізації в умовах обмежених обчислювальних ресурсів; оцінено практичну доцільність застосування кожного типу віртуалізації в умовах обмежених обчислювальних ресурсів. **Висновки**: *Docker* забезпечує найпростіше розгортання та найбільшу ефективність, але має меншу ізоляцію, якщо порівнювати з *QEMU* та *Firecracker*; *Firecracker* демонструє оптимальний баланс між продуктивністю, безпекою та споживанням ресурсів на одноплатних комп'ютерах; *QEMU* має найбільші накладні витрати, а тому й найменшу ефективність; використання легковагових гіпервізорів є доцільним у системах цифрових двійників на периферійних пристроях.

Ключові слова: віртуалізація; одноплатні комп'ютери; цифровий двійник; мережний моніторинг; *Raspberry Pi*.

Вступ

Світ цифрових технологій перебуває в постійній трансформації у зв'язку зі стрімким зростанням потреб користувачів, Індустрії 4.0 та поступовим переходом багатьох секторів економіки до автоматизованих рішень. Водночас класичні підходи до розроблення та підтримання програмних систем поступово втрачають свою ефективність. Сучасні проєкти – це цілісні, багаторівневі архітектури з високим рівнем абстракції для її підтримання та управління, що вимагають нових інструментів для моделювання, тестування та масштабування з огляду на наявні проблеми.

Майже завжди, коли постає питання в масштабуванні, одним із найкращих рішень буде використання віртуалізації. Вона давно вийшла за межі хмарних центрів оброблення інформації, і нині все більше проникає на рівень периферійних обчислень, зокрема в пристрої одноплатних комп'ютерів (*SBC, single board computers*), наприклад *Raspberry Pi*. Віртуалізація дає змогу знизити вартість фізичного обладнання, підвищити гнучкість і швидкість розгортання нових вузлів та забезпечити кращу масштабованість систем, що є критично важливим у швидкозмінному середовищі Інтернету речей (*IoT, internet of things*).

Особливої актуальності набуває концепція *Digital Twin* – цифрового двійника, що допомагає створити віртуальне уявлення фізичного об'єкта чи системи, уможливлюючи моделювання, аналіз та оптимізацію без утручання в реальний процес. *Digital Twin* – це відповідь одразу на кілька ключових викликів сучасної інженерії – від зниження вартості розроблення до прискорення циклу тестування та розгортання систем.

Однак застосування віртуалізації в описаній сфері на одноплатних комп'ютерах залишається маловивченим. Обмеження в обчислювальних ресурсах для деяких чи навіть більшості проєктів може стати перепоною у використанні таких комп'ютерів для

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

віртуалізації. Отже, актуальною є проблема недостатнього вивчення можливостей створення цифрових двійників на *SBC* із застосуванням сучасних технологій віртуалізації, їх продуктивності, гнучкості та потенціалу масштабованості..

Аналіз останніх досліджень і публікацій

Одними з основних функцій пристроїв *IoT* є збирання, оброблення, зберігання та передача даних з фізичного світу у віртуальний. У масивних і автономних середовищах *IoT* виникають проблеми через обмежені ресурси. Для подолання цих проблем необхідна ефективна архітектура, наприклад віртуалізація, що допомагає оптимізувати обмежені ресурси внаслідок розподілу завдань та балансування навантаження. Зі зростанням кількості обсягів інформації особливу увагу потрібно приділяти ефективному управлінню даними: що і як зберігати й передавати, що безпосередньо впливає на оптимізацію ресурсів у великомасштабних системах *IoT* [1].

У дослідженні [2], присвяченому впровадженню віртуалізації на одноплатних комп'ютерах, продемонстровано, що *Raspberry Pi 4 Model B* може успішно працювати як хост віртуалізації разом із технологію віртуалізації *VMware ESXi 7* на *ARM*, незважаючи на обмеження апаратної та програмної сумісності. Хоча платформа поступається традиційним рішенням *x86* щодо продуктивності та швидкості зберігання, вона споживає вдев'ятеро менше енергії, що робить її привабливою для енергоефективних приграничних сценаріїв. Результати підтверджують можливість повної інтеграції такої інфраструктури *ARM64* у наявні корпоративні системи. Інші вчені в роботі [3] оцінили кілька одноплатних комп'ютерів і мікроконтролерів з огляду на під'єднання недорогих датчиків і надання функціональності. Автори розширили обмежені пристрої зовнішніми модулями та адаптерами для стандартизації експериментального середовища. Результати продемонстрували, що платформи *Raspberry Pi 4* та *AN 33 BLE* мають кращу підтримку спільноти, хоча й не найширші електронні можливості. Незважаючи на високе енергоспоживання плат *SBC*, це компенсується більшою інтеграцією функціональності, що робить їх придатними для гнучких рішень *IoT*. Хоча *Raspberry Pi 4B* не є найпотужнішим одноплатним комп'ютером на базі *ARM*, доступним на ринку, він виявився надійною та стабільною платформою, особливо у використанні з гіпервізором *KVM* (*kernel-based virtual machine*). Це було досліджено в статті [4], де один із ключових висновків полягає в тому, що обидві платформи здатні запускати віртуальні машини й мати високу працездатність завдяки застосуванню *KVM*.

Автори праці [5] переконують, що сучасні одноплатні комп'ютери досить продуктивні для запуску типових робочих навантажень. Завдяки своїй низькій вартості та енергоефективності кластери одноплатних комп'ютерів стають реальним інструментом для периферійних обчислень, зокрема рішення на базі *Raspberry Pi* дають змогу створювати недорогі, компактні та масштабовані системи, які можна розгортати поза традиційними центрами оброблення інформації. Автори дослідження [6] використовували *Docker Swarm* на одноплатних комп'ютерах у кластерах і дійшли висновків, що *SBC* забезпечують ефективну платформу для первинного оброблення інформації з метою мінімізації обсягу даних, що будуть надіслані до хмари чи інших серверів.

Щодо потенційної ефективності *MQTT* на *SBC* було проведено дослідження [7], у якому за допомогою розробленої системи моніторингу для сервера *MQTT* на базі брокера *Mosquitto* було проаналізовано продуктивність і характеристики комунікації *MQTT*. Система містила три компоненти: генератор навантаження з 11 *Raspberry Pi*, програма для моніторингу на сервері та вебпанель для візуалізації інформації в реальному часі. Експерименти підтвердили стабільну роботу системи, що свідчить про доцільність використання *Raspberry Pi* в середовищах із високим навантаженням по *MQTT*. Група вчених у межах дослідження [8] довела, що прототип на базі *Raspberry Pi 3* може реалістично виконувати складні робочі навантаження мережі з оброблення інформації з кількох датчиків, інтегруючись із *NMAP*, *MQTT*, *InfluxDB* та *MariaDB*. Система продемонструвала низьку затримку на оброблення показників і загальний відгук. Це є позитивним результатом для використання одноплатних комп'ютерів у контексті цифрового двійника.

У наступному дослідженні [9] увагу зосереджено на застосуванні цифрових двійників в управлінні сільськогосподарськими водними ресурсами, що є яскравим прикладом необхідності розвитку напряму граничних обчислень. Використання *SBC* є корисним не тільки в розподілених інфраструктурах для складних енергоефективних рішень у сільськогосподарському секторі, а й у ширшому контексті граничних

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

обчислень. Окрім цього, кіберфізичні системи (*CPS, cyber-physical systems*) також отримують переваги від застосування цифрового двійника. Відповідно до дослідження [10] цифрові двійники відіграють ключову роль у розвитку Індустрії 4.0, поєднуючи цифровий та фізичний світи, забезпечуючи динамічне моделювання та інтеграцію інформації реального світу. Це робить технологію надзвичайно актуальним інструментом для сучасних систем *CPS* у промисловості, медицині та міському плануванні].

Мета роботи й завдання

З поданого вище аналізу можна визначити, що використання віртуалізації на одноплатних комп'ютерах досліджено, але більшість авторів зосереджують увагу на окремих аспектах, таких як запуск однієї віртуальної машини, загальне енергоспоживання або базова продуктивність. Як правило, у поданих роботах не розглядається повне "прикладне навантаження", що відповідало б реальному використанню цифрового двійника з потоками інформації, взаємодією датчиків, моделюванням, запуском декількох віртуальних машин на одній платформі.

Крім того, ефект поступового зниження ефективності роботи *SBC* залежно від запуску додаткових віртуальних машин залишається значною мірою невивченим. Це є критичним фактором у разі, коли один фізичний пристрій має емулювати одночасно декілька цифрових двійників або працювати в багатокомпонентній системі.

У зв'язку з цим виникає потреба в комплексному дослідженні, у межах якого необхідно не лише вивчити технічну можливість запуску віртуальних машин на одноплатних комп'ютерах, а й оцінити їх ефективність в умовах навантажень, наближених до реальних.

Мета дослідження – аналіз ефективності створення цифрових двійників на одноплатних комп'ютерах із використанням віртуалізації, а також формулювання рекомендацій щодо доцільного впровадження цієї технології в умовах обмежених ресурсів.

Отже, для досягнення мети необхідно виконати такі завдання:

– проаналізувати сучасні апаратні рішення одноплатних комп'ютерів щодо підтримки технологій віртуалізації;

– розглянути та порівняти наявні технології віртуалізації, звертаючи увагу на їх вимоги, сумісність і продуктивність;

– дослідити наявні середовища виконання, придатні для віртуалізації на *SBC*;

– визначити критерії оцінювання ефективності віртуалізації потреб цифрових двійників;

– сформулювати й реалізувати прикладне навантаження, що імітуватиме роботу цифрового двійника в типовому сценарії;

– створити експериментальну систему, яка дасть змогу вимірювати продуктивність зі змінною кількістю віртуальних машин;

– розробити рекомендації щодо оптимального впровадження віртуалізації на *SBC* у контексті створення цифрового двійника.

Матеріали й методи

Аналіз апаратного забезпечення

На ринку існує чимало одноплатних комп'ютерів із різними характеристиками та призначенням. Саме тому особливу увагу необхідно приділити вибору платформи, щоб визначити найбільш ефективну для завдання. Вона має відповідати низці вимог: достатній обсяг оперативної пам'яті, підтримка апаратної віртуалізації, наприклад *ARM Virtualization Extensions*, розширені можливості введення-виведення та низьке енергоспоживання. Розглянемо основні сучасні рішення.

Raspberry Pi 4 Model B [11] є однією з найпопулярніших платформ одноплатних комп'ютерів. Вона має чотириядерний процесор, варіації до 8 ГБ оперативної пам'яті, підтримку *USB 3.0*, *Gigabit Ethernet*, *microSD* або *USB SSD* для зберігання інформації. Завдяки високій популярності та доступності, наявності безлічі готових образів ОС та підтримці 64-бітної архітектури ця модель стала фактичним стандартом для експериментів.

Raspberry Pi 5 [12] – наступна ітерація в лінійці *Raspberry Pi*, має варіації до 16 ГБ оперативної пам'яті, підтримку *PCIe 2.0*. Ця модель забезпечує потенційну продуктивність удвічі та більше разів вищу, ніж *Pi 4* в реальних завданнях, зокрема у віртуалізації. З першого погляду це робить її набагато привабливішою для сценаріїв дослідження *Digital Twin*, особливо коли на одному пристрої планується запускати кілька віртуальних машин. Однак важливо зважати на деякі технічні особливості, наприклад необхідність активного охолодження. Процесор *Pi 5* має більш високу теплову проєктну потужність (*TDP, thermal design power*), і без вентилятора або великого радіатора він починає тротліти (*throttle*) навіть в умовах середнього

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

навантаження. Отже, для стабільної роботи в завданнях віртуалізації охолодження є не рекомендацією, а вимогою. Крім цього, навіть якщо *Raspberry Pi 5* вже декілька років на ринку, *Raspberry Pi 4* має більш стабільну екосистему, оскільки вона підтримується з 2019 року. Звичайно, зі збільшеною потужністю висуваються й підвищені вимоги до живлення. *Pi 4* є більш енергоефективним, оскільки використовує блок живлення 5В/3А порівняно з 5В/5А у *Pi 5*.

Radxa Rock 5B [13] – потужний одноплатний комп'ютер на базі восьмиядерного процесора, що підтримує до 32 ГБ оперативної пам'яті, має вбудований *eMMC*, слот для *NVMe SSD, HDMI 2.1*. Отже, *Rock 5B* є високопродуктивною альтернативою лінійці *Raspberry Pi* для завдань ШІ, віртуалізації, оброблення медіа, або навіть ігор, але ціна значно вища за інші варіанти.

Libre Computer Board AML-S905X-CC [14] – бюджетна альтернатива з меншими можливостями. Має чотириядерний процесор, 2 ГБ оперативної пам'яті, є дешевшим варіантом, якщо порівнювати з *Pi 4* та *Pi 5*, але віртуалізація тут можлива лише в базовому вигляді.

Orange Pi 5 [15] – більш потужний аналог *Raspberry Pi 5*, але від іншого виробника. Наявна підтримка до 16 ГБ оперативної пам'яті, *eMMC, M.2 NVMe SSD* з восьмиядерним процесором. Цю модель можна ефективно використовувати в периферійних проєктах, бо також підтримує віртуалізацію на апаратному рівні. *Orange Pi 5* має значно вищу ціну, ніж *Raspberry Pi*, і за обчислювальними можливостями наближена до *Radxa Rock 5B*.

Порівняння потенційних платформ для проведення експерименту подано в табл. 1.

Таблиця 1. *Характеристики одноплатних комп'ютерів*

Платформа	Процесор	Оперативна пам'ять	Підтримка апаратної віртуалізації	Живлення	Охолодження	Орієнтовна ціна
Raspberry Pi 4 Model B	4 ядра	до 8 ГБ, LPDDR4	так	5V/3A	пасивне, опціонально активне	$35–75
Raspberry Pi 5	4 ядра	до 16 ГБ, LPDDR4X	так	5V/5A	обов'язково активне	$50–120
Radxa Rock 5B	8 ядер	до 32 ГБ, LPDDR4X	так	5V/5A	обов'язково активне	$130–270
Orange Pi 5	8 ядер	до 16 ГБ, LPDDR4X	так	5V/4A	рекомендовано активне	$85–130
Libre AML -S905X-CC	4 ядра	до 2 ГБ, DDR3	ні	5V/2A	пасивне	$20–30

Унаслідок аналізу платформ обрано модель *Raspberry Pi 4 Model B* з 4 ГБ оперативної пам'яті. Цей вибір зумовлений низкою факторів. Модель з 4 ГБ оперативної пам'яті забезпечує достатньо ресурсів для запуску кількох віртуалізованих середовищ, зберігаючи водночас помірну вартість. Окрім цього, ця модель доступна на ринку з 2019 року, що дало змогу розробникам досягти стабільного середовища підтримки як у спільноті користувачів, так і серед дистрибутивів операційних систем, драйверів та інструментів віртуалізації. На відміну від *Raspberry Pi 5*, четверта модель може працювати тривалий час навіть за умови пасивного охолодження. І хоча *Pi 4* поступається *Pi 5* за обчислювальною потужністю, у завданнях, де основне навантаження – мережний обмін, зберігання інформації, взаємодія з датчиками, продуктивність *Pi 4* залишається більш ніж достатньою, а завдяки своїй низькій вартості та широкій доступності *Pi 4* допомагає легко масштабувати експерименти або розгортати аналогічні системи в інших середовищах.

Аналіз середовищ виконання для віртуалізованих систем

У контексті створення цифрових двійників на базі *Raspberry Pi* та вивчення віртуалізації ключовим моментом є вибір відповідного віртуалізованого середовища, що самостійно запускатиметься всередині віртуальних машин чи контейнерів. Основні вимоги до такого середовища:

– мінімальне споживання ресурсів;

– можливість автоматизації та налаштування без графічного інтерфейсу;

– наявність мережних інструментів і базових сервісів, зокрема *MQTT, Node-RED* тощо;

– сумісність із архітектурою *ARM*.

Отже, маємо критерії щодо вибору найкращого середовища для віртуалізованих систем.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

Зауважимо, що використовувати *Windows* для домашніх комп'ютерів на *SBC* не є вигідною альтернативою з погляду ефективності. І хоча існує *Windows 10/11 IoT Core* для *ARM*, це все ж таки дуже обмежена версія зі своєю специфікою та не найпрозорішою підтримкою технологій віртуалізації. Емуляція ж середовища, що запустило б дистрибутиви, робить практичне тестування продуктивності платформи неможливим.

MacOS – це закрита платформа, що за межами офіційного обладнання від *Apple* не підтримує жодних версій *ARM* та технічно неможлива для легального використання чи досліджень на *Raspberry Pi*. Отже, єдиною альтернативою, що можна раціонально розглядати, є застосування дистрибутивів *Linux*.

Розглянемо деякі доступні варіанти віртуалізованого середовища на *ARM*.

Debian – класичний вибір дистрибутиву *Linux*. Він стабільний, з багатою підтримкою архітектури *ARM*, має чи не найбільший репозиторій пакетів, але займає багато місця на диску між інших дистрибутивів. Загальне використання – системи, де важливі стабільність і стандартизація, що можуть собі дозволити такий обсяг дистрибутиву.

Ubuntu Server – похідний дистрибутив від *Debian*, що також має активну спільноту, надмірну документацію та офіційну підтримку *ARM*, проте, на відміну від *Debian*, містить більшу кількість задач, що працюють на фоні, та потребує більш потужну платформу, що не позитивно впливатиме на ефективність роботи віртуалізованих середовищ.

Arch Linux ARM – значно легкий і гнучкий, на відміну від попередніх дистрибутивів, але вимагає чимало часу на налаштування та має недолік з нестабільністю бази пакетів, а це може дуже негативно позначитися на автоматизованому розгортанні віртуальних машин.

Alpine Linux – дистрибутив, що відрізняється від інших вкрай малим розміром образу, дуже низьким споживанням оперативної пам'яті та стабільною підтримкою архітектури *ARM64*. Він не має недоліків з автоматизованим розгортанням, як *Arch Linux*, і має суттєвий репозиторій пакетів, що робить налагодження практичного навантаження дуже простим. Цей дистрибутив є найбільш підхожий варіант для завдань дослідження.

Завдяки окресленим характеристикам *Alpine Linux* обраний як основне середовище виконання у віртуалізованих системах, що дає змогу зосередитися на архітектурі цифрового двійника, не витрачаючи ресурсів на непотрібну інфраструктуру.

Аналіз технологій віртуалізації

Вибір технології віртуалізації – один із найважливіших аспектів у дослідженні продуктивності віртуалізації на *SBC*. Він визначає, наскільки ефективно можна реалізувати ізольовані середовища для моделювання, аналізу й тестування систем. Однак на платформах *SBC* цей вибір ускладнений обмеженнями апаратної архітектури *ARM64*, нестачею ресурсів, а також частковою або експериментальною підтримкою деяких інструментів.

Docker [16] на платформах *ARM*, зокрема *Raspberry Pi*, посідає особливе місце завдяки своїй легкості, швидкому запуску контейнерів та відносній простоті використання. Ця технологія не створює повноцінних віртуальних машин, а замість цього застосовує так звані контейнери (*Docker containers*), у цьому разі процеси користувача поділяються на рівні ядра. Це означає високу продуктивність та низьке споживання пам'яті, але знижений рівень ізоляції, що спричиняє потенційні проблеми з безпекою.

QEMU дає змогу створювати й запускати повноцінні віртуальні машини, що є ізольованими операційними системами, які мають власне ядро та диск. Емулятор має відкриту кодову базу, що дає змогу додавати, редагувати та вилучати функціонал за побажанням користувача [17]. У разі, коли хост-система (*host system*) підтримує апаратну віртуалізацію, *KVM* допомагає значно прискорити *QEMU* внаслідок використання гіпервізора (*hypervisor*) на рівні ядра. У поєднанні з *KVM* віртуальні машини потенційно можуть майже не відрізнятися від реальних машин із виділеним апаратним забезпеченням.

Firecracker [18] – гіпервізор, що створює так звані мікровіртуальні машини (*micro VM*). Він розроблений з огляду на безсерверні обчислення (*serverless computing*) та забезпечує надзвичайно швидкий запуск із мінімальним використанням ресурсів. Хоча підтримка *ARM64* в *Firecracker* поки не так розвинена, як *x86_64*, вона вже дає змогу створювати ізольовані середовища, що можуть запускатися дуже швидко. Це робить її перспективною альтернативою для систем, які потребують масштабування в реальному часі, наприклад для оброблення подій датчиків на великій кількості паралельних цифрових двійників.

Крім описаних вище технологій, існують інші, що з певних причин менш ефективні для використання

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

на *Raspberry Pi*. Наприклад, *KubeVirt* запускає віртуальні машини в середовищі *Kubernetes*, але потребує складної інфраструктури. *KubeVirt* орієнтується саме на хмарні обчислення. *Xen* – відносно легкий гіпервізор – активно застосовувався в системах, що вбудовувалися в минулому, але тепер менш поширений на платформах з архітектурою *ARM64*.

Тому вибір технології віртуалізації для *Digital Twin* на *SBC* залежить від компромісу між продуктивністю, ізоляцією, швидкістю запуску, сумісністю та масштабованістю. *Docker* надає легке контейнерне середовище, тоді як *QEMU* з *KVM* забезпечує повну віртуалізацію. *Firecracker*, також з використанням *KVM*, незважаючи на свою експериментальність на *ARM64*, відкриває перспективу мікровіртуальної архітектури для цифрових двійників. Разом ці інструменти дають змогу обрати рішення відповідно до конкретного завдання, що стане предметом подальших досліджень.

Аналіз ефективності віртуалізації

Аналіз рівня ефективності різних технологій віртуалізації на *Raspberry Pi*, що є обчислювально обмеженою платформою, вимагає використання чітко визначених і відтворюваних метрик для отримання точних експериментальних показників. У дослідженні обрано низку ключових критеріїв, що дають змогу кількісно оцінити як продуктивність самої віртуалізованої системи, так і ступінь навантаження на апаратне забезпечення.

Час запуску (*startup time*) – це період від запуску ініціалізації віртуального середовища до повної робочої готовності, що передбачає запуск операційної системи, служб усередині віртуального середовища, виклик користувацьких скриптів. Цей критерій є критично важливим, коли віртуальні середовища мають запускатися динамічно, наприклад, у процесі масштабування або у відповідь на зовнішні події. У разі цифрового двійника це дасть змогу миттєво додавати нові сенсори чи їх групи, не потерпаючи від тривалого запуску середовища.

Використання оперативної пам’яті – ступінь застосування оперативної пам’яті хост-системи віртуальними середовищами. Знаючи максимальний обсяг пам’яті, критерій допомагає оцінити, скільки середовищ може бути запущено одночасно на обмеженій платформі без перевантаження пам’яті, що негативно вплине на швидкодію всієї системи. Хоча більшість гіпервізорів дають змогу та рекомендують виставляти незмінні обмеження на пам’ять, критерій може бути корисним для знаходження та аналізу витоків пам’яті у процесі застосування віртуалізованих середовищ.

Час проходження *ping*-пакета (*RTT, round trip time*) – це середній час, необхідний для надсилання *ICMP*-пакета та отримання відповіді. Критерій допомагає оцінити базову якість мережної взаємодії віртуального середовища, зокрема затримки, які могли бути спричинені нагромадженням віртуальними мережними інтерфейсами або додатковою маршрутизацією. Це важливо для цифрових двійників, яким необхідно отримувати та передавати інформацію в режимі реального часу.

RTT для повідомлень *MQTT* – час від надсилання повідомлення *MQTT* до моменту, коли копія повертається відправнику назад. На відміну він часу проходження *ping*-пакета, цей критерій відтворює не тільки затримку мережі, але й здатність віртуального середовища швидко обробляти події в протоколі, що часто використовується для Інтернету речей. Це особливо важливо в системах цифрових двійників у процесі передачі телеметрії.

Затримка доступу до диска (*disk delay*) – це час виконання типових операцій читання та запису у віртуальне файлове сховище незалежно від архітектури віртуального середовища. Продуктивність диска суттєво впливає на системи, що потребують роботи з базами даних. Продуктивність диска суттєво впливає на ефективність “цифрових двійників”, через їх активну роботу з базами даних.

Використання центрального процесора – середній рівень його застосування під час виконання типових завдань чи в режимі очікування. Він показує, наскільки ефективно гіпервізор розподіляє ресурси й чи не витрачає він занадто багато обчислювальної потужності на існування. Простий критерій, який може швидко, але не дуже точно, визначити ступінь результативності певної технології віртуалізації.

Температура центрального процесора – це температура чипа під час роботи системи. Як і з використанням центрального процесора – це простий, але показовий критерій, що може допомогти звернути увагу на потенційні проблеми ефективності роботи цього елемента. Питання охолодження не є суттєвим, зокрема для *Raspberry Pi 4*, але загалом *SBC* можуть значно перегріватися, що допомагає відслідковувати цей критерій.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Кожен із наведених вище критеріїв обраний з огляду на особливості реального застосування цифрових двійників, тобто систем, які мають бути автономними, реагувати в реальному часі, забезпечувати ізоляцію, але водночас легко масштабуватися та, якщо можливо, не сильно завантажувати загальну систему. Комплексне використання цих метрик дає змогу об'єктивно порівнювати різні підходи до віртуалізації та формувати рекомендації щодо їх практичного впровадження.

Проведення експерименту

Для порівняльного аналізу ефективності віртуалізації з використанням різних технологій віртуалізації в контексті цифрових двійників на базі одноплатного комп'ютера необхідно було реалізувати експериментальне середовище, що передбачає як підготовку апаратного забезпечення, так і програмну автоматизацію проведення дослідження з додатковим первинним аналізом результатів.

Як було зазначено вище, обрана платформа – це *Raspberry Pi 4 Model B* з 4 ГБ оперативної пам'яті. На *SD*-карту обсягом 32 ГБ був записаний 64-бітний дистрибутив *Debian* для архітектури *ARM*, попередньо налаштований для використання *KVM*. Без увімкнення функціоналу *KVM* всі вимірювання будуть заздалегідь далекі від оптимальних значень.

До платформи було під'єднано зовнішні сенсори, а саме п'ять пристроїв: *GY-302*, *GY-87*, *DHT22*, *MQ-7* та мікросхема АЦП *MCP3008*. Загалом комп'ютер мав доступ до шести сенсорів різної направленості: сенсор освітленості *BH1750FVI*, гіроскоп *MPU6050*, компас *QMC5883L*, сенсор атмосферного тиску *BMP180*, сенсор вологості та температури *DHT22*, датчик газу *MQ-7*. Для постійного зчитування значень сенсорів та відправлення цих показників за допомогою *MQTT* було розроблено скрипт мовою програмування *Python*, що збирає телеметричні показники з різною для кожного сенсора періодичністю.

Зв'язок із датчиками найчастіше виконувався через шину *I2C*, деякі використовували спеціальні протоколи через *GPIO* та частина сенсорів була під'єднана до шини *SPI*.

Для кожної з досліджуваних технологій віртуалізації, а саме *QEMU*, *Firecracker*, *Docker*, створено окремі середовища, максимально схожі одне на одне. Були спроби скопіювати середовище *QEMU* для роботи на *Firecracker*, але це виявилося нетривіальним і виходить за межі дослідження. У кожному з випадків образ містив дистрибутив *Alpine Linux* з мінімальним набором системних утиліт, що потім за потреби розширювалися. Для всіх середовищ був обмежений розмір диска до 1 ГБ, а обсяг оперативної пам'яті становив максимум до 256 МБ. Що менше місця займає образ без втрат у швидкості, то більше паралельно працюючих віртуальних машин може бути.

Як практичне навантаження було реалізовано просту цифрову подвійну комунікацію: значення, надходячи з фізичних пристроїв, опиняються в межах хост-истеми та надсилаються віртуальною локальною мережею через протокол *MQTT* усім запущеним віртуальним машинам чи контейнерам. У віртуальному середовищі за допомогою технології *Node-RED* такі повідомлення обробляються та зберігаються в базі даних тимчасових рядів *InfluxDB*. Ця група технологій імітує оброблення телеметричної інформації пристроїв *IoT* у віртуальному контексті, щоб отримати експериментальні показники, максимально наближені до реальних.

Для мережної взаємодії між віртуальними середовищами та хост-системою використано два підходи: інтерфейси *TAP* для *QEMU* та *Firecracker*, і мережний міст у разі *Docker*. Кожній віртуальній машині призначався окремий інтерфейс, що забезпечує пряме під'єднання до мережного простору хост-системи, даючи змогу обмінюватися повідомленнями *MQTT*, *ICMP*-пакетами та збирати метрики затримки.

Для того, щоб запобігти людській помилці та неточності у виконанні випробування, розроблено групу скриптів *Bash*, які могли бути запущені експериментатором одноразово, і система самостійно виконувала повний цикл:

– початок роботи скрипта *Python* для отримання показників із сенсорів;

– запуск ініціалізації набору віртуальних середовищ;

– старт сервісів *Node-RED* та *InfluxDB* у віртуалізованих середовищах;

– збір системних метрик за описаними вище критеріями;

– зберігання отриманої інформації у файлах формату *.csv*;

– завершення роботи скриптів і віртуалізованих середовищ.

Кожен експеримент проводився кілька разів з метою досягнення статистично стабільних результатів.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Результати дослідження

Для проведення експерименту застосовано *Raspberry Pi 4 Model B*, переваги якого вже описані в розділі про аналіз апаратного забезпечення. За допомогою з'єднувальних дротів було під'єднано всі потрібні пристрої, що використовують шини *I2C, SPI* та *GPIO*. Як носій пам'яті обрано карту *MediaRange MicroSDHC Class 10* на 32 ГБ. Вона не змінювалася між експериментами над різними технологіями віртуалізації, тому будь-які недоліки карти пам'яті позначилися на всіх експериментальних показниках, що не заважає їх порівнювати в межах дослідження. Сама програмна база операційної системи *Pi 4* також залишалась незмінною, тому отримана інформація максимально відповідає вимогам порівняльного аналізу між обраними технологіями. Результат установки зображено на рис. 1. Ліворуч розташований *Raspberry Pi 4*, що видно із символу малини на друкованій платі, а праворуч, на макетній платі, подано всі периферійні пристрої.

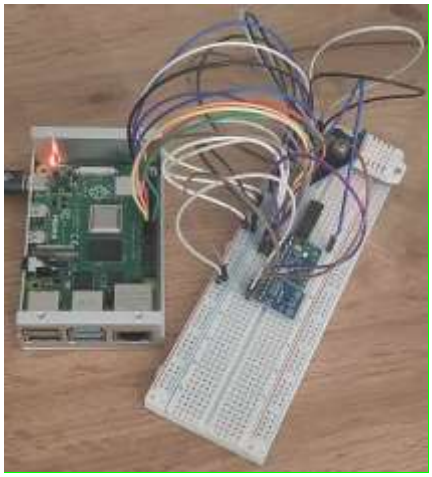


Рис. 1. Експериментальна апаратна установка з *Raspberry Pi 4* та пристроями *GY-302, GY-87, DHT22, MQ-7, MCP3008*

Загалом проведено чотири запуски експерименту для кожної з технології віртуалізації. Показники були усереднені з метою запобігання отримання некоректних значень або впливу на експеримент зовнішніх факторів. Підготовлено вісім віртуальних середовищ, але під час проведення експериментів платформа показувала сильну нестабільність, екстремальне погіршення продуктивності та надвисокі показники температури процесора в умовах восьми одночасно запущених віртуальних машин *QEMU*. З огляду на це було прийнято рішення обійтися сімома середовищами та порівнювати інформацію саме з них, щоб залишати *Pi 4* працездатним.

У межах усіх експериментів для кожного почергового запуску віртуального середовища виміряно час від запуску ініціалізації середовища до виконання скрипта користувача із середини запущеного середовища. Час вимірювався способом збереження часової мітки у двох випадках, а повідомлення до хост-системи було реалізовано за допомогою *ping*-пакета, надісланого віртуалізованим середовищем (див. табл. 2).

Таблиця 2. *Час почергового запуску віртуальних середовищ*

Кількість запущених віртуальних середовищ	Час запуску Docker (с)	Час запуску Firecracker (с)	Час запуску QEMU (с)
1	0.761	3.176	11.944
2	0.84	3.246	11.786
3	0.94	3.34	12.313
4	1.088	3.455	16.731
5	1.507	3.652	20.784
6	1.642	3.733	27.01
7	1.779	3.948	44.393

З отриманих експериментальних показників можемо спостерігати вагому різницю між кожною технологією. Поки *Docker* і *Firecracker* запускаються впродовж декількох секунд чи навіть швидше, *QEMU* відносно тривалий час проходить ініціалізацію. Частково це можна пояснити рівнем емуляції та безпеки, що забезпечує *QEMU*, втрачаючи здатність швидко розгортати нові віртуальні машини.

Окрім цього, помітна загальна тенденція в надлишковому часі запуску в кожній технології приблизно після четвертого запущеного середовища. Імовірно, це пов'язано з тим, що платформа – це чотириядерний процесор.

Щоб зменшити короткострокові коливання, які були сильно помітні під час первинного огляду, покращити їх візуальну інтерпретацію та сформулювати висновки про загальні тенденції, для всіх наступних експериментальних показників застосовано їх згладжування за допомогою процесу експоненціального рухомого середнього (*exponential moving average*). Водночас було збережено часову динаміку:

$$EMA_t = \alpha \cdot x_t + (1-\alpha) \cdot EMA_{t-1}, \quad (1)$$

де EMA_t – експоненціальна рухома середня в момент часу t;

α – коефіцієнт згладжування;

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

x_t – поточний елемент показників;

EMA_{t-1} – попередня експоненціальна рухома середня.

Експериментально визначено оптимальне значення коефіцієнта згладжування. У всіх лінійних графіках воно дорівнює приблизно 0.18.

Обрана модель платформи мала 4 ГБ оперативної пам'яті, що спочатку здавалося дуже недостатньо для такого типу експерименту, але *Firecracker* та *QEMU* досить ефективно виконували завдання цифрового двійника з виділеними 256 МБ оперативної пам'яті на кожну з машин (див. рис. 2).

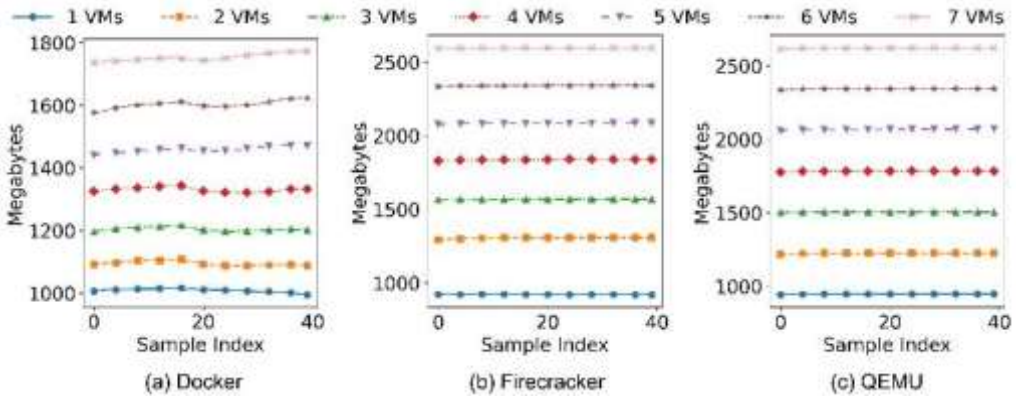


Рис. 2. Порівняння використання оперативної пам'яті під час почергового запуску віртуальних середовищ

Отже, спостерігаємо паритет між двома технологіями, а *Docker* зі свого боку динамічно підвищував використання пам'яті від навантаження, що виявилося суттєво менше за інші варіанти. Приблизно 100 МБ за контейнер, тому теоретично можна було б запустити до 30 контейнерів або 12 віртуальних машин, зважаючи тільки на оперативну пам'ять.

Наступним критерієм, за яким проведено експеримент, є час проходження *ping*-пакета (див. рис. 3).

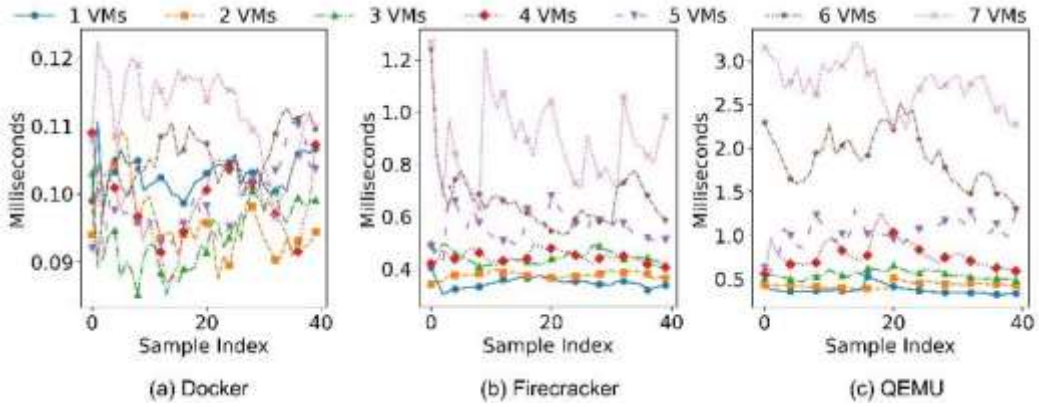


Рис. 3. Порівняння середнього часу проходження *ping*-пакета внутрішньою мережею між хост-системою та віртуалізованими середовищами

Вимірювання проведено в масштабі мікросекунд, але перетворені в мілісекунди для кращої візуалізації. *Docker* продемонстрував майже незалежність часу проходження пакета від кількості запущених контейнерів. Завжди значення коливаються приблизно зі швидкістю 0.1 мілісекунда, що є найкращим показником серед альтернатив. *Firecracker* та *QEMU* мали гірші результати за цим критерієм за умови 0.6 мілісекунди та 1.5 мілісекунди в середньому відповідно до альтернатив. Стабільне та наднизькі значення проходження *ping*-пакета *Docker* вказує на мінімальні мережні витрати, типові для контейнеризації, яка не потребує повної мережної віртуалізації. У *Firecracker* і *QEMU* додаткові віртуальні машини спричиняють затримки, що зростають. Це вказує на вищу вартість мережної ізоляції.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Більш затратною за ресурсами операцією є проходження *MQTT*-повідомлення, що також дає змогу оцінити стан мережі в процесі віртуалізації, але в іншому контексті. Було застосовано схожі інструменти вимірювання часу між цим і попереднім критерієм через їх функціональну схожість. Досягнуті результати дещо віддзеркалюють показники від *ping*-пакетів, але додають новий погляд на ефективність віртуалізації (див. рис. 4).

З отриманих показників видно, що *Docker* демонструє кращу продуктивність і стабільність, що повторює результати *ping*-пакетів. *Firecracker* повільніше обробляє повідомлення, і кожне почергове середовище сильніше погіршує швидкість оброблення, але все ж таки залишається досить стабільним, у межах 100–200 мілісекунд. Це відповідає пріоритету технології – забезпечити ізоляцію, подібну до віртуальної машини, з продуктивністю, подібною до контейнера. *QEMU* сильно сповільнюється за умови шести-семи запущених машин, але в разі чотирьох машин тримається майже стабільно, як і інші альтернативи. Знову ж таки, це може бути пов'язано з чотириядерним процесором платформи. Результати *QEMU* вказують на значні витрати на віртуалізацію мережі, пов'язану з емуляцією мережного інтерфейсу.

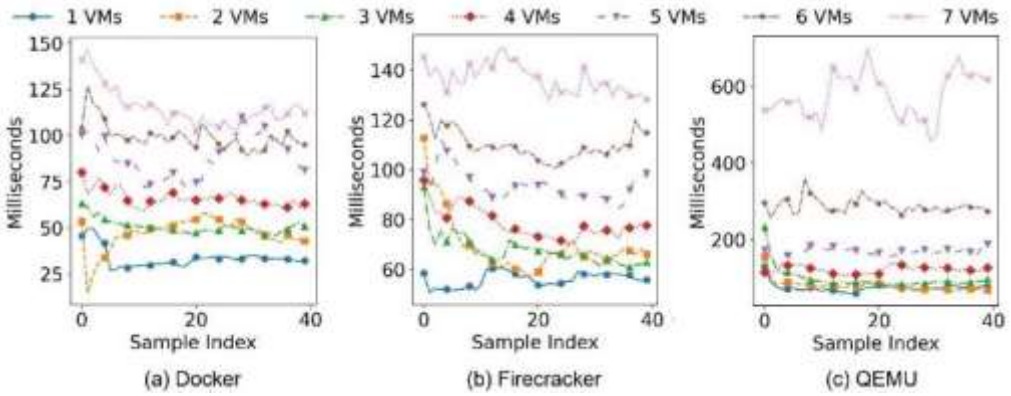


Рис. 4. Порівняння середнього часу проходження *MQTT*-повідомлення від хост-системи до кожного з віртуалізованих середовищ і у зворотному напрямку

Затримка доступу до диска відіграє важливу роль у порівнянні технологій віртуалізації. У контексті цифрового двійника необхідно завжди зберігати отримані показники, вчасно оновлювати налаштування. За умови значних навантажень це може суттєво вплинути на швидкодію окремого цифрового двійника, тому отримані показники відіграють важливу роль у загальному порівнянні технологій (див. рис. 5).

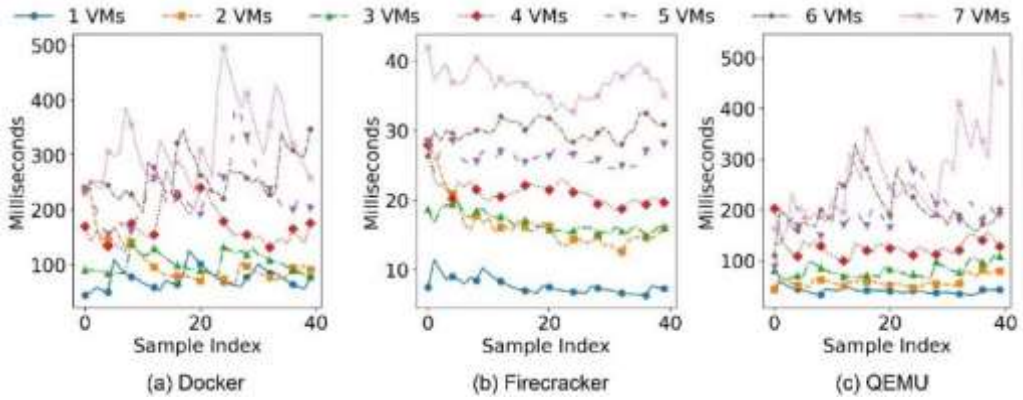


Рис. 5. Порівняння середньої затримки доступу до диска між усіма запущеними віртуальними середовищами під час проведення експерименту

Firecracker демонструє найкращу продуктивність диска, що пов'язано з використанням мінімалістичного дизайну, що означає менше емуляції. *Firecracker* використовує віртуальні пристрої безпосередньо,

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

а не повну апаратну емуляцію, як *QEMU*. *Docker*, хоч і швидко обробляє повідомлення, але працює на файловій системі хост-системи й часто використовує накладну файлову систему (*overlay filesystem*), що створює додаткові витрати під час кожного запису / читання. Крім цього, наявні недоліки в разі значної кількості контейнерів, бо вони конкурують за одні й ті самі ресурси. Через те, що *Docker* не має ізоляції введення-виведення на рівні ядра, як це можливо у *Firecracker*, спостерігаємо сильну різницю в отриманих показниках. *QEMU* загалом схожий на *Firecracker* за показниками, але через надмірну емуляцію в межах віртуалізації на одноплатному комп'ютері все ж таки поступається останньому. *QEMU* так чи інакше створює високі накладні витрати на операції введення-виведення, а можливість кешувати дані, що відсутня у *Docker*, робить його кращим за цим критерієм, ніж контейнери.

Загальним критерієм, що тільки побічно інформує про ефективність віртуалізації, є навантаження центрального процесора. За декілька хвилин експерименту для кожної кількості віртуальних середовищ відсоток навантаження зміг стабілізуватися (див. рис. 6).

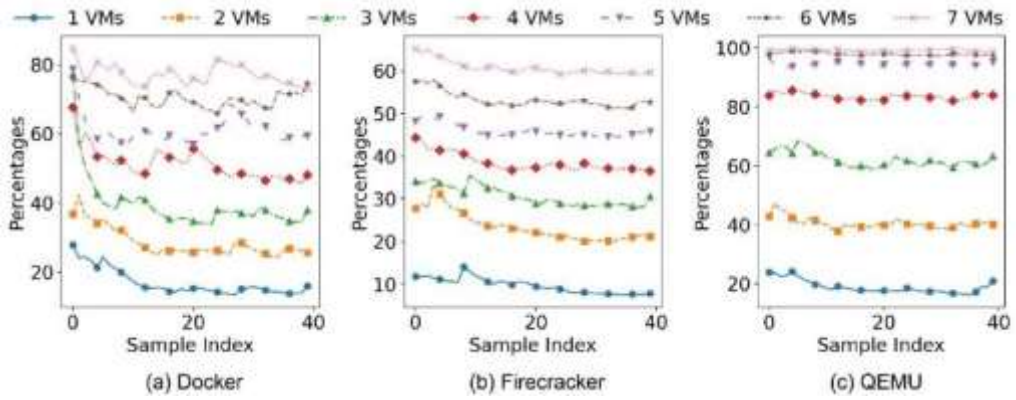


Рис. 6. Порівняння завантаженості центрального процесора під час проведення експерименту з віртуалізації цифрових двійників

Відразу спостерігаємо недостатню потужність у *Pi 4* для підтримки семи віртуальних машин. Уже на п'яти машинах платформа була навантажена майже на максимум. Саме ці показники переконали відмовитися від восьми віртуальних середовищ. Максимальна ізоляція та емуляція девайсів і ресурсів спричиняє надмірні витрати часу центрального процесора. *Docker* демонструє вищі пікові навантаження, якщо порівнювати з *Firecracker*. Це можна пояснити прямим доступом контейнерів до ресурсів хост-системи та відсутністю ізоляції. Хоча *Firecracker* показує кращі результати щодо завантаження процесора, але беручи до уваги інші критерії, не можна сказати, що *Firecracker* виконує задачі швидше за *Docker*.

Схожа ситуація і з температурою центрального процесора, що прямо залежить від його навантаження. Кожний наступний цифровий двійник сильніше навантажує систему, що призводить до більшого нагрівання чипа (див. рис. 7).

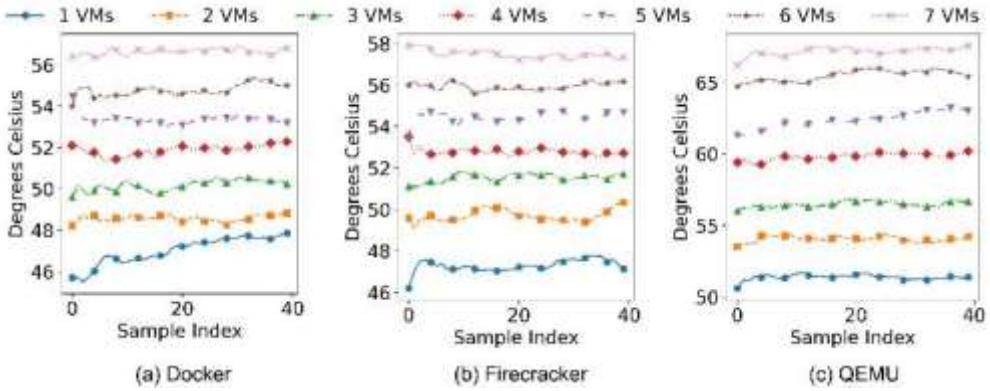


Рис. 7. Порівняння температури центрального процесора платформи під час навантаження віртуалізованих цифрових двійників

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Docker та *Firecracker* майже однаково нагрівають процесор, тоді як *QEMU*, дійшовши до максимального навантаження, розігріває платформу до 70ºС. Це не максимально допустима температура, що становить 85 ºС, але, зрештою, значно вища за альтернативи.

Висновки

Результати дослідження продемонстрували суттєві розбіжності в ефективності використання обчислювальних ресурсів, у стабільності та затримках передачі інформації в умовах упровадження обраних технологій віртуалізації *Docker*, *Firecracker* та *QEMU* і застосування цифрового двійника на одноплатному комп'ютері *Raspberry Pi 4 Model B*.

У межах поставленого завдання, а саме віртуалізації легкого й базового цифрового двійника, *Docker* виявився найкращим варіантом для сценаріїв із високими вимогами до продуктивності та мінімальних затримок. Його легка архітектура дала змогу досягти найменших затримок у мережі та обміну повідомленнями *MQTT*, а також помірного навантаження на процесор. Однак це супроводжувалося вищим навантаженням на центральний процесор, що свідчить про інтенсивну взаємодію з хост-системою.

Визначено, що *Firecracker* має більше накладних витрат, ніж *Docker*, але також залишається дуже вдалим рішенням, що поєднує високу ефективність доступу до диска, швидкий час запуску. Незважаючи на дещо вищу мережну затримку, ця технологія забезпечує ізоляцію та стабільність, що є вкрай важливим для безпеки хост-системи та інших віртуалізованих середовищ.

Найменш ефективною технологією в межах проведеного експерименту є *QEMU*. Попри всі переваги повної віртуалізації, обрана платформа не може впоратися з таким навантаженням на процесор, тому масштабування до рівня *Firecracker* чи *Docker* не є можливим. Використання *QEMU* призвело до високих температур чипа та значних затримок як у процесі мережної комунікації, так і під час роботи з диском. У застосуванні лише декількох віртуальних машин *QEMU* є гарною альтернативою, особливо коли ізоляція та емуляція є важливими компонентами.

Отже, вибір технології віртуалізації на одноплатному комп'ютері залежно від потреб насамперед буде на користь *Docker* або *Firecracker*. Перший кращий щодо налагоджування та функціоналу за більш прості проєкти чи системи, а *Firecracker* ефективний, коли ізоляція та підтримка саме окремого диска мають значення в прийнятті рішень. *QEMU* можна застосовувати в окремих випадках, які малоймовірно будуть дотичними до контексту цифрового двійника.

Перспективи подальшого дослідження

У статті проаналізовано ефективність віртуалізації на одноплатних комп'ютерах у контексті цифрових двійників, однак є низка напрямів, що заслуговує на подальше опрацювання та розширення.

По-перше, хоча було обрано *Raspberry Pi 4* як платформу для проведення експериментів, існують і інші альтернативи, що можуть бути більш придатними платформами за інших критеріїв, наприклад *Raspberry Pi 5*. Нова апаратна платформа пропонує значно вищу потужність обчислень завдяки кращому процесору. Проте в цьому разі вона потребує більш складного охолодження та є менш стабільною, що пояснюється її новизною. Порівняльний експеримент на *Pi 5* виявив би не тільки різницю в продуктивності, а й у стабільності, масштабованості та температурних обмеженнях, що суттєво розширило б межі дослідження в контексті цифрових двійників на *SBC*.

По-друге, з обраних критеріїв порівняння видно, що було зосереджено увагу на стані всієї платформи та майже всі показники отримані ззовні віртуалізованого середовища, а не зсередини. Один із напрямів подальшого розвитку міг бути присвячений глибшому вивченню саме внутрішнього стану віртуалізованих середовищ. Перехід від моделі "чорної скриньки" до "білої скриньки", де збираються внутрішні метрики віртуалізованої ОС, дало б змогу краще зрозуміти вплив шару віртуалізації на продуктивність машини.

По-третє, було б цілком доцільно винести частину зібраних телеметричних показників із сенсорів за межі експериментальної платформи, наприклад на окремий мікроконтролер, як-от *ESP32*, який передаватиме всі значення по *MQTT* у віртуальні середовища. Отже, експериментальна платформа цілком зосереджена на віртуалізації цифрових двійників, і такий розподіл завдань є іншою загальноприйнятою альтернативою в розподілених системах *IoT*, де інформація надходить у хмару чи пригранічні сервери з окремих пристроїв.

По-четверте, практичне навантаження може бути ускладнене та розширене. Замість запропонованої

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

пари технологій *Node-RED* та *InfluxDB*, можна моделювати більш складні цифрові двійники, що передбачатимуть аналіз аномалій, візуалізацію, навчання моделей машинного навчання, передачу команд по *MQTT* безпосередньо до сенсорів тощо. Якщо ускладнювати практичне навантаження, можна все ближче й ближче відтворювати реальні приклади з *IoT*, але водночас зберігати передбачуваність і контрольованість експериментальної системи.

Усі окреслені напрями не тільки розширюють прикладну цінність експерименту, але й відкривають перспективи подальших наукових досліджень у сфері периферійних обчислень, ефективної віртуалізації та цифрових двійників у системах *IoT*.

Список літератури

1. Saniya Z., Roohie N. M. Resource management in pervasive Internet of Things: A survey. *Journal of King Saud University Computer and Information Sciences*. 2021. Vol. 33. № 8. P. 921–935. DOI: https://doi.org/10.1016/j.jksuci.2018.08.014
2. Lambropoulos G., Mitropoulos S., Douligeris C., Maglaras L. Implementing Virtualization on Single-Board Computers: A Case Study on Edge Computing. *Computers*. 2024. Vol. 13. №. 2. 54 p. DOI: https://doi.org/10.3390/computers13020054
3. Álvarez J.L., Mozo J.D., Durán E. Analysis of Single Board Architectures Integrating Sensors Technologies. *Sensors*. 2021. Vol. 21. 6303 p. DOI: https://doi.org/10.3390/s21186303
4. Gamess E., Parajuli M., Shah S. Performance evaluation of the KVM hypervisor running on ARM-based single-board computers. *International Journal of Computer Networks & Communications (IJCNC)*. 2023. Vol. 15. №. 2. P. 147–164. DOI: 10.5121/ijcnc.2023.15208
5. Johnston J. S., Basford J. P., Perkins S. C., Herry H., Tso F. P., Pezaros D., Mullins D. R., Yoneki E., Cox J. S., Singer J. Commodity single board computer clusters and their applications. *Future Generation Computer Systems*. 2018. Vol. 89. P. 201–212. DOI: https://doi.org/10.1016/j.future.2018.06.048
6. Penchalaiah N., Al-Humaimeedy A.S., Mashael M., Babu C.J., Khan O.I., Aldhyani T.H.H. Clustered Single-Board Devices with Docker Container Big Stream Processing Architecture. *Computers, Materials and Continua*. 2022. Vol. 73. №. 3. P. 5349--5365. DOI: https://doi.org/10.32604/cmc.2022.029639
7. Hwang K., Jung I.H., Lee J.M. Monitoring of MQTT-based Messaging Server. *Webology*. 2022. Vol. 19. №. 1. P. 4724–4735. DOI: 10.14704/WEB/V19I1/WEB19316
8. Alam T., Rokonuzzaman M., Sarker S., Azim Z., Das T., Hossain M.I. Internet of Things-based Home Automation with Network Mapper and MQTT Protocol. *Computers and Electrical Engineering*. 2024. Vol. 120. 109807 p. DOI: https://doi.org/10.1016/j.compeleceng.2024.109807
9. Ahsen R., Di Bitonto P., Novielli P., Magarelli M., Romano D., Diacono D., Monaco A., Amoroso N., Bellotti R., Tangaro S. Harnessing. Digital Twins for Sustainable Agricultural Water Management: A Systematic Review. *Applied Sciences*. 2025. Vol. 15. 4228 p. DOI: https://doi.org/10.3390/app15084228
10. Vidyalakshmi G., Gopikrishnan S., Wadii B., Anis K., Gautam S. Digital Twins and Cyber-Physical Systems: A New Frontier in Computer Modeling. *CMES – Computer Modeling in Engineering and Sciences*. 2025. Vol. 143. №. 1. P. 51–113. DOI: https://doi.org/10.32604/cmes.2025.057788
11. Raspberry Pi 4 Official Website. URL: https://www.raspberrypi.com/products/raspberry-pi-4-model-b (дата звернення: 13.09.2024)
12. Raspberry Pi 5. The everything computer. Optimised. URL: https://www.raspberrypi.com/products/raspberry-pi-5/ (дата звернення: 14.09.2024)
13. Radxa ROCK 5B. URL: https://radxa.com/products/rock5/5b (дата звернення: 01.10.2024)
14. Le Potato AML-S905X-CC. URL: https://libre.computer/products/aml-s905x-cc (дата звернення: 01.10.2024)
15. Orange Pi 5. URL: http://www.orangepi.org/html/hardWare/computerAndMicrocontrollers/details/Orange-Pi-5.html (дата звернення: 01.10.2024)
16. Docker Homepage. URL: https://www.docker.com/ (дата звернення: 06.01.2025)
17. Oracle Virtual Box vs VMware Workstation vs QEMU. URL: https://www.starwindsoftware.com/blog/type-2-hypervisors-comparison-virtualbox-vmware-qemu/ (дата звернення: 25.11.2024)
18. Getting started with Firecracker on Raspberry Pi. URL: https://dev.to/l1x/getting-started-with-firecracker-on-raspberry-pi-1pbc (дата звернення: 10.02.2025)

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

References

1. Saniya, Z., Roohie, N. (2021), "Resource management in pervasive Internet of Things: A survey", *Journal of King Saud University* Computer and Information Sciences, No. 33(8), P. 921–935. DOI: https://doi.org/10.1016/j.jksuci.2018.08.014
2. Lambropoulos, G., Mitropoulos, S., Douligeris, C., Maglaras, L. (2024), "Implementing Virtualization on Single-Board Computers: A Case Study on Edge Computing", *Computers*, No. 13(2), 54 p. DOI: https://doi.org/10.3390/computers13020054
3. Álvarez, J., Mozo, J., Durán, E. (2021), "Analysis of Single Board Architectures Integrating Sensors Technologies", *Sensors*, No. 21, 6303 p. DOI: https://doi.org/10.3390/s21186303
4. Gamess, E., Parajuli, M., Shah, S. (2023), "Performance evaluation of the KVM hypervisor running on ARM-based single-board computers", *International Journal of Computer Networks & Communications (IJCNC)*, No. 15(2), P. 147–164. DOI: 10.5121/ijcnc.2023.15208
5. Johnston, J., Basford, J., Perkins, S., Herry, H., Tso, F., Pezaros, D., Mullins, D., Yoneki, E., Cox, J., Singer, J. (2018), "Commodity single board computer clusters and their applications", *Future Generation Computer Systems*, No. 89, P. 201–212. DOI: https://doi.org/10.1016/j.future.2018.06.048
6. Penchalaiah, N., Al-Humaimeedy, A., Mashael, M., Babu, C., Khan, O., Aldhyani, T. (2022), "Clustered Single-Board Devices with Docker Container Big Stream Processing Architecture", *Computers*, Materials and Continua, No. 73(3), P. 5349–5365. DOI: https://doi.org/10.32604/cmc.2022.029639
7. Hwang, K., Jung, I., Lee, J. (2022), "Monitoring of MQTT-based Messaging Server", *Webology*, No. 19(1), P. 4724–4735. DOI: 10.14704/WEB/V19I1/WEB19316
8. Alam, T., Rokonuzzaman, M., Sarker, S., Azim, Z., Das, T., Hossain, M. (2024), "Internet of Things-based Home Automation with Network Mapper and MQTT Protocol", *Computers and Electrical Engineering*, No. 120. 109807 p. DOI: https://doi.org/10.1016/j.compeleceng.2024.109807
9. Ahsen, R., Di Bitonto, P., Novielli, P., Magarelli, M., Romano, D., Diacono, D., Monaco, A., Amoroso, N., Bellotti, R., Tangaro, S. (2025), "Digital Twins for Sustainable Agricultural Water Management: A Systematic Review", *Applied Sciences*, No. 15, 4228 p. DOI: https://doi.org/10.3390/app15084228
10. Vidyalakshmi, G., Gopikrishnan, S., Wadii, B., Anis, K., Gautam, S. (2025), "Digital Twins and Cyber-Physical Systems: A New Frontier in Computer Modeling", *CMES - Computer Modeling in Engineering and Sciences*, No. 143(1), P. 51–113. DOI: https://doi.org/10.32604/cmes.2025.057788
11. Raspberry Pi 4 Official Website, available at: https://www.raspberrypi.com/products/raspberry-pi-4-model-b (last accessed: 13.09.2024)
12. Raspberry Pi 5. The everything computer. Optimised, available at: https://www.raspberrypi.com/products/raspberry-pi-5/ (last accessed: 14.09.2024)
13. Radxa ROCK 5B, available at: https://radxa.com/products/rock5/5b (last accessed: 01.10.2024)
14. Le Potato AML-S905X-CC, available at: https://libre.computer/products/aml-s905x-cc (last accessed: 01.10.2024)
15. Orange Pi 5, available at: http://www.orangepi.org/html/hardWare/computerAndMicrocontrollers/details/Orange-Pi-5.html (last accessed: 01.10.2024)
16. Docker Homepage, available at: https://www.docker.com/ (last accessed: 06.01.2025)
17. Oracle Virtual Box vs VMware Workstation vs QEMU, available at: https://www.starwindsoftware.com/blog/type-2-hypervisors-comparison-virtualbox-vmware-qemu/ (last accessed: 25.11.2024)
18. Getting started with Firecracker on Raspberry Pi, available at: https://dev.to/l1x/getting-started-with-firecracker-on-raspberry-pi-1pbc (last accessed: 10.02.2025)

Надійшла (Received) 22.05.2025

Відомості про авторів / About the Authors

Свєтлінський Олег Андрійович – аспірант, Харківський національний університет радіоелектроніки, кафедра програмної інженерії, Харків, Україна; e-mail: oleh.svietlinskyi@nure.ua; ORCID ID: https://orcid.org/0009-0003-4221-9297

Ревенчук Ілона Анатоліївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, Харків, Україна; e-mail: ilona.revenchuk@nure.ua; ORCID ID: https://orcid.org/0000-0002-5188-9538

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Svietlinskyi Oleh – Postgraduate , Kharkiv National University of Radio Electronics, Department of Software Engineering, Kharkiv, Ukraine.

Revenchuk Ilona – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Software Engineering, Kharkiv, Ukraine.

RESEARCH ON THE EFFICIENCY OF VIRTUALIZATION SOLUTIONS FOR DEPLOYING DIGITAL TWINS IN SYSTEMS WITH LIMITED RESOURCES

The **subject matter** of the article is the means and approaches to virtualization in environments with limited resources, in particular on single-board computers, as well as their ability to ensure the effective operation of digital twins. The **goal** of the work is to evaluate the effectiveness of virtualization for deploying digital twins on single-board computers and develop recommendations for virtualization usage in environments with limited computing resources. The following **tasks** were solved in the article: analysis of modern hardware solutions for single-board computers and existing virtualization technologies; construction of an experimental environment for practical comparison of virtualization technologies on single-board computers; implementation of an application load that simulates the operation of a digital twin with real sensor data; study of the efficiency of network interaction between virtual environments and the host system; comparison of results by key metrics. The following **methods** used are: virtualization technologies Docker, QEMU, Firecracker; network monitoring; thorough analysis of literature and available documentation on the topic; comparative analysis of the achieved results. The following **results** were obtained: a comparative experiment using Docker, QEMU and Firecracker based on Raspberry Pi 4 was carried out; thorough analysis of literature and available documentation on the topic was conducted; a basic prototype of a digital twin with the connection of physical sensor devices that collect telemetry data in real time was implemented; phased automatic deployment of virtual environments with pre-configured components was provided for rapid scaling of the experiment; network delays, message processing time and platform resource usage were measured; the advantages and limitations of virtualization technologies in conditions of limited computing resources were determined; the practical feasibility of using each type of virtualization in conditions of limited computing resources was assessed. **Conclusions**: Docker provides the easiest deployment and highest efficiency, but has less isolation compared to QEMU and Firecracker; Firecracker demonstrates the optimal balance between performance, security, and resource consumption on single-board computers; QEMU has the highest overhead, and therefore the lowest efficiency; the use of lightweight hypervisors is advisable in digital twin systems on peripheral devices.

Keywords: virtualization; single-board computers; digital twin; network monitoring; Raspberry Pi.

Бібліографічні описи / Bibliographic descriptions

Світлінський О.А., Ревенчук І.А. Дослідження ефективності віртуалізаційних рішень для розгортання цифрових двійників у системах з обмеженими ресурсами. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 137–151. DOI: https://doi.org/10.30837/2522-9818.2025.3.137

Svietlinskyi, O, Revenchuk, I. (2025), "Research on the efficiency of virtualization solutions for deploying digital twins in systems with limited resources", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 137–151. DOI: https://doi.org/10.30837/2522-9818.2025.3.137

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

УДК 004.85:004.94 DOI: https://doi.org/10.30837/2522-9818.2025.3.152

О. Соловей

ЗВАЖЕНА МЕТРИКА ЧУТЛИВОСТІ ДЛЯ ПРОГНОЗУВАННЯ ЗАТРИМКИ В *KAFKA*-КЛАСТЕРІ

Предметом дослідження є комбінований підхід до аналізу чутливості для комплексного оцінювання впливу конфігураційних параметрів *Kafka*-кластера на кінцеву затримку в системах потокового оброблення інформації. **Мета роботи** – розроблення нового підходу до аналізу чутливості, що поєднує класичні методи оцінювання впливу параметрів (методи Морріса та Соболя) з метриками, що відтворюють структурні зміни розподілу вихідних даних (евклідова відстань, хеллінгерова відстань, *J*-дивергенція). Такий підхід дає змогу дослідити вплив параметра не лише з погляду амплітуди його ефекту, а й щодо змін у формі та структурі ймовірнісного розподілу результатів. Для досягнення мети розв'язуються такі **завдання**: формальне визначення нового підходу до аналізу чутливості; розроблення мережі Баєса для моделювання наскрізної затримки *Kafka*-кластера; аналіз чутливості за запропонованим підходом; експериментальне дослідження нового підходу з використанням розрахованих відповідно до нього ваг впливу для ініціалізації матриці ваг нейронної мережі, що прогнозує наскрізну затримку в *Kafka*-кластері залежно від обраної конфігурації. Для реалізації поставлених завдань у дослідженні впроваджено такі **методи**: теорія експериментів, евклідова геометрія, статистична теорія розподілів, інформаційна теорія, теорія машинного навчання, баєсівська статистика й теорія графів. **Досягнуті результати.** Для оцінювання ефективності підходу проведено порівняльне навчання нейронної мережі з різними стратегіями ініціалізації ваг. Аналіз функції витрат, побудованої за критерієм мінімізації середньоквадратичної похибки, продемонстрував, що найменших значень вона досягає саме для моделі, яка була ініціалізована вагами, отриманими за запропонованим підходом до оцінювання впливу параметрів на вихідну змінну моделі. **Висновки.** У дослідженні запропоновано новий підхід до аналізу чутливості. Новизна підходу полягає в інтеграції переваг як причинно-орієнтованих, так і дисперсійно-оцінних методів у межах єдиної зваженої метрики чутливості. Практична цінність підходу полягає в тому, що його застосування під час аналізу чутливості або ініціалізації матриці ваг нейронної мережі дає змогу підвищити точність оцінювання впливу параметрів, покращити збіжність моделі та скоротити час її навчання.

Ключові слова: метод Морріса; індекс Соболя; евклідова відстань; хеллінгерова відстань; *J*-дивергенція; аналіз чутливості.

Вступ

Інформаційні технології для управління великими даними є ключовим інструментом управління проєктами міського будівництва, оскільки забезпечують оброблення значних обсягів різнорідної інформації з різних джерел. До таких джерел належать системи управління бізнес-процесами (ERP), взаємодією із зацікавленими сторонами (CRM), охороною праці та ризиками на будівництві, експлуатацією об'єктів, а також системи для проєктування, створення моделей просторових об'єктів, інженерного аналізу й застосунки доповненої та віртуальної реальності (AR/VR) (рис. 1). Ефективне функціювання цих технологій вимагає створення гнучкої, масштабованої та надійної інформаційної архітектури, здатної обробляти структуровані, напівструктуровані та неструктуровані потоки даних, які частково надходять у режимі реального часу й частково в режимі пакетного оброблення [1].

Отже, архітектура інформаційної технології для управління інформацією проєктів міського будівництва передбачає такі шари: прийом потокових даних; прийом пакетних даних; оброблення та збереження потокових даних; аналітика потокових даних (рис. 2).

Шар прийому потокових даних (рис. 2) включає кластер *Apache Kafka*, що забезпечує отримання інформації в реальному часі з низькою затримкою. Висока пропускна здатність у процесі передачі повідомлень від *Kafka*-продюсера до *Kafka*-споживача є однією з ключових вимог до інформаційних технологій, призначених для управління різнорідними даними проєктів міського будівництва. Конфігурація *Kafka*-кластера забезпечує високу продуктивність за умови відповідності її параметрів потребам системи. Втім, велика гнучкість у налаштуванні конфігурації *Kafka* спричиняє суттєву невизначеність у продуктивності кластера *Apache Kafka*, зокрема щодо кінцевої затримки, яка визначається як час від моменту надсилання інформації продюсером до їх отримання споживачем.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Рис. 1. Джерела даних "Інформаційні технології для управління великими даними проєктів міського будівництва"

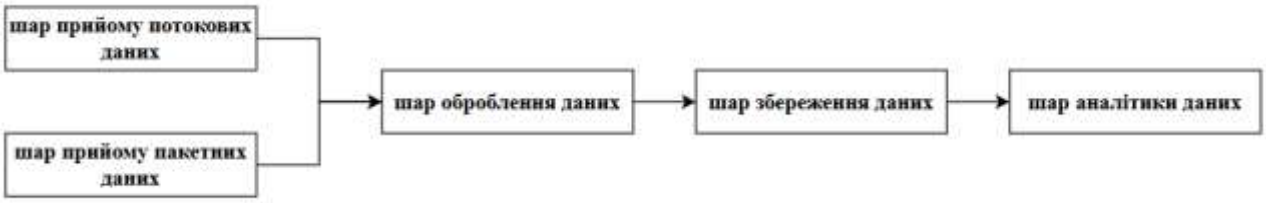


Рис. 2. Архітектура інформаційної технології для управління даними проєктів міського будівництва

Наявні дослідження здебільшого спрямовані на покращення окремих показників продуктивності, таких як затримка або відмовостійкість, залишаючи поза увагою приховані взаємозв'язки між параметрами конфігурації та чутливість системи до локальних змін значень. Унаслідок цього можуть формуватися субоптимальні або нестійкі конфігурації, що погано масштабуються в умовах реального використання.

Попри важливість цього аспекту, вибір оптимальної *Kafka* конфігурації залишається складним завданням. Труднощі полягають в необхідності динамічного налаштування параметрів з метою запобігання перевантаженням, мінімізації витрат, а також забезпечення високої доступності й низької затримки.

Аналіз останніх досліджень і публікацій

Коли модель глибокого навчання використовується для визначення конфігурації кластера *Apache Kafka* на основі прогнозування затримки в кластері, вибір початкової матриці ваг має значний вплив на збіжність навчання та кінцеву продуктивність моделі, оскільки суттєво впливає на активацію нейронів [2]. За замовчуванням застосовується метод Ксав'єра (*Xavier Uniform*) [3], що ґрунтується на рівномірному розподілі ваг залежно від кількості нейронів на вході та виході:

$$w \sim U\left(-\sqrt{\frac{6}{n_{in}+n_{out}}}, \sqrt{\frac{6}{n_{in}+n_{out}}}\right), \qquad (1)$$

де U – рівномірний розподіл; *nin, nout* – кількість нейронів, що додані в шари та пов'язані матрицею ваг.

Однак нещодавні дослідження [4] вказують на необхідність зважати на причинно-наслідкові зв'язки в прийнятті рішень, що актуалізує пошук нових підходів до ініціалізації ваг. Серед альтернативних методів: ініціалізація на основі підкласового аналізу дискримінантів (SDA) [5], *ZerO*-ініціалізація з використанням матриць Адамара [6], методи масштабування ваг [7] та адаптивне масштабування під час навчання [8]. Проте зазначені методи

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

здебільшого не беруть до уваги причинно-наслідкові впливи, що обмежує їх ефективність у роботі з неповними або структурно складними даними.

Аналіз чутливості є важливим підходом для визначення впливу вхідних параметрів на вихідні характеристики системи або моделі. У праці [9] запропоновано чотириетапну структуру, що поєднує метамоделювання з методами Морріса та Соболя. Це дає змогу ефективно виявляти ключові фактори, що впливають на енергоспоживання, знижуючи водночас обчислювальні витрати. У роботі [10] порівняно різні методи чутливості, зокрема DLOR і FAST, для аналізу ознак, релевантних до продуктивності моделей CNN. Найбільш ефективними виявились методи Морріса та Соболя.

Подальші дослідження вдосконалюють методику обчислення індексів Соболя. Автори праці [10] запропонували спрощений підхід із використанням індексів Шеплі, де глобальні ефекти параметрів оцінюються першими, а решта – виводяться лінійними перетвореннями. Дослідження [11] демонструє застосування глобального аналізу чутливості для визначення пріоритетних параметрів у життєвому циклі виробництва. У статті [12] запропоновано підхід для аналізу дискретної баєсівської мережі за допомогою евклідової відстані, хеллінгерової відстані та *J*-дивергенції. Ці метрики дають змогу оцінити схожість і розбіжності між імовірнісними розподілами, що формуються внаслідок варіацій параметрів.

Метод Морріса дає змогу оцінити локальний вплив кожного параметра через варіацію вихідної функції, тоді як індекси Соболя дають глобальну уяву про внесок параметра до загальної дисперсії. Їх комбіноване застосування розкриває повну картину впливів, однак не зважає на причинно-наслідкові залежності між параметрами.

Метрики відстані між імовірнісними розподілами (евклідова, хеллінгерова відстані, *J*-дивергенція) можуть бути інтегровані в причинно-орієнтовані моделі, зокрема баєсівські мережі, що забезпечує взяття до уваги структурних впливів на зміну розподілів. Проте вони не фіксують змінність самої функції результату, як це роблять класичні методи аналізу чутливості.

Отже, наразі відсутній уніфікований підхід, який би поєднував такі фактори:

1) взяття до уваги структурної подібності між результатами на основі метричних відстаней;

2) оцінювання прямого впливу параметрів (за допомогою індексів Морріса μ_i);

3) взяття до уваги взаємодій між параметрами (за допомогою індексів Соболя S_i).

Це створює передумови для розроблення нового підходу до аналізу чутливості, який поєднує переваги як причинно-орієнтованих, так і дисперсійно-оцінних методів у єдиній зваженій метриці.

Мета роботи й завдання

Мета дослідження – запропонувати новий підхід для аналізу чутливості, який поєднує класичні методи оцінювання впливу параметрів (методи Морріса та Соболя) з метриками, що відтворюють структурні зміни розподілу вихідних даних (евклідова відстань, хеллінгерова відстань, *J*-дивергенція). Такий підхід дасть змогу дослідити вплив параметра не лише з погляду амплітуди його ефекту, але й щодо змін у формі та структурі ймовірнісного розподілу результатів.

Для досягнення окресленої мети необхідно розв’язати такі завдання:

1) формально визначити новий підхід до аналізу чутливості, який поєднає методи Морріса μ_i та Соболя S_i з одним із методів (евклідова відстань, хеллінгерова відстань або *J*-дивергенція);

2) розробити мережу Баєса для моделювання наскрізної затримки *Kafka*-кластера;

3) проаналізувати чутливість за методами евклідової відстані, хеллінгерової відстані, *J*-дивергенції на розробленій мережі Баєса;

4) проаналізувати чутливість за методами Морріса μ_i та Соболя S_i на моделі *XGBRegressor*;

5) проаналізувати чутливість за запропонованим підходом;

6) провести експериментальне дослідження запропонованого підходу, використовуючи ваги впливу для ініціалізації матриці ваг нейронної мережі, що прогнозує наскрізну затримку в *Kafka*-кластері залежно від обраної конфігурації.

Матеріали й методи

Нехай $L(X)$ – це затримка в кластері *Kafka* з конфігурацією X, змодельована функцією $f(X)$, де $X = \{x_1, x_2, \ldots, x_n\}$ – набір незалежних параметрів

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

кластера *Kafka*. Загальна дисперсія затримки $L(X)$, розкладена за вхідними параметрами, описується як

$$Var\left(L(X)\right)=\sum_{i=1}^{N}V_i+\sum_{1\le i<j\le N}V_{ij}+\ldots+V_{1,2,\ldots n}, \quad (2)$$

$$V_i=Var_{X_i}E_{X_{\sim i}}\left(f(X)\,|\,X_i\right), \quad (3)$$

$$V_{ij}=Var_{X_{ij}}E_{X_{\sim ij}}\left(f(X)\,|\,X_i,X_j\right)-V_i-V_j, \quad (4)$$

де $V_{1,\,2,\ldots,\,N}$ – дисперсія, що виникає внаслідок взаємодії всіх ознак у наборі X.

Індекс Соболя першого порядку оцінює, яка частка дисперсії затримки $L(X)$, зумовлена впливом лише одного вхідного параметра x_i:

$$S_i=\frac{V_i}{Var(Y)}. \quad (5)$$

Метод Морріса для обчислення μ_i:

$$\mu_i=\frac{1}{R}\sum_{r=1}^{R}\left|d_i^r\right|, \quad (6)$$

де d_i^r – це елементарний ефект вхідного параметра на r-й ітерації.

$$d_i^r=\frac{f\left(x^r+\Delta e_i\right)-f\left(x^r\right)}{\Delta}, \quad (7)$$

де x^r – випадково обраний вектор на r-й ітерації;

Δe_i – одиничний вектор, що змінює тільки параметр *xi*;.

Δ – крок у сітці пошуку.

Нехай загальний розподіл затримки незалежно від окремих значень параметрів визначено:

$$P(L)=\sum_X P(L|X)P(X). \quad (8)$$

Умовний розподіл за фіксованого параметра $x_i=a$:

$$P(L\,|\,x_i=a)=\sum\nolimits_{X_{\sim i}}P(L|x_i=a,X_{\sim i})\cdot P(X_{\sim i}), \quad (9)$$

де $X_{\sim i}$ – всі параметри, крім x_i.

Евклідова відстань (D_E) двох розподілів імовірностей визначається за сумою квадратів різниць розподілів:

$$D_E=\sqrt{\sum_{l\in L}\left(P(L=l)-P(L=l\,|\,x_i=a)\right)^2}. \quad (10)$$

Відстань Хеллінгера (D_H) вимірює відстань між двома ймовірнісними розподілами й основана на квадратичній різниці їх імовірностей:

$$D_H=\frac{1}{\sqrt{2}}\sqrt{\sum_{l\in L}\left(\sqrt{P(L=l)}-\sqrt{P(L=l\,|\,x_i=a)}\right)^2}. \quad (11)$$

Відстань *J*-дивергенції (D_J) оцінює, наскільки один розподіл відрізняється від іншого за допомогою порівняння їх "втрат" інформації щодо середнього розподілу:

$$D_J=\sum_{l\in L}P(L=l)\log\frac{P(L=l)}{P(L=l\,|\,x_i=a)}+ +P(L=l\,|\,x_i=a)\log\frac{P(L=l\,|\,x_i=a)}{P(L=l)}. \quad (12)$$

Підхід до аналізу чутливості, який поєднує методи Морріса μ_i (5) та Соболя S_i (6) з одним із методів (10)–(12), формально опишемо як зважену комбінацію:

$$W_{D_E}=\alpha\cdot S_i+\beta\cdot\mu_i+\gamma\cdot D_E; \quad (13)$$

$$W_{D_H}=\alpha\cdot S_i+\beta\cdot\mu_i+\gamma\cdot D_H; \quad (14)$$

$$W_{D_j}=\alpha\cdot S_i+\beta\cdot\mu_i+\gamma\cdot D_J. \quad (15)$$

Вибір значень для параметрів α, β, γ залежить від завдання, а саме: варто збільшити параметр α, якщо необхідно визначити параметри з найвищим глобальним впливом; потрібно збільшити параметр β, якщо необхідно взяти до уваги середній вплив параметра незалежно від його локальних або глобальних властивостей; потрібно збільшити параметр γ, якщо необхідно взяти до уваги структурну подібність між результатами на основі метричних відстаней.

Для порівняльного аналізу результатів оцінювань за методами (13)–(15) скористаємося класифікацією:

$$O(W_i)=\begin{cases}"H", & \text{якщо } W_i\ge Q_{67}^i\\ "M", & \text{якщо } Q_{33}^i\le W_i<Q_{67}^i\\ "L", & \text{якщо } W_i<Q_{33}^i\end{cases}, \quad (16)$$

де $"H"$, $"M"$, $"L"$ – класи оцінки W_i – "сильний", "середній" та "низький";

$W_i=\left(W_{ij}-\min(W_i)\right)/\left(\max(W_i)-\min(W_i)\right)$ – нормалізована оцінка;

i – індекс методу з множини $O\left[W_{De},W_{D_H},W_{D_j}\right]$;

j – індекс параметра конфігурації з множини X;

Q_{67}^i, Q_{33}^i – 67-й, 33-й процентилі, обчислені на W_i.

Для ініціалізації нейронів першого прихованого шару нейронної мережі з метою прогнозування затримок *Kafka*-кластера з вагами впливу $O(W_i)$,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

обчисленими відповідно до (16), запропонуємо правило "більшості голосів":

$$W(x) = argmax_{o\epsilon O} \mid i : O(W_i) = o \mid . \quad (17)$$

Ініціалізація нейронів першого прихованого шару нейронної мережі вагами W виконується за виразом

$$h^1 = f\left(W\vec{X} + b^1\right), \quad (18)$$

де h^1 – вихід першого прихованого шару нейронної мережі;

b^1 – вектор зсуву для першого прихованого шару;

$f(\bullet)$ – функція активації, що застосовується до лінійної комбінації вхідних даних.

Архітектура *Kafka*-кластера містить сервери, застосунки, що надсилають потокову інформацію на сервер (продюсерів) та клієнтів (*consumers*), які під'єднуються для отримання даних та сервісу *ZooKeeper*, який використовується *Kafka* для управління та координації серверів, доданих у *Kafka*-кластер [14, 15]. Наскрізну затримку (*end-to-end latency*) потокового оброблення подій у *Kafka*-кластері можна виміряти часом *TL*, який дорівнює сумі таких часових інтервалів: час, необхідний продюсеру для збирання подій у пакет перед їх відправленням на сервер (далі – *Tproducer*); час від моменту отримання події в буфері сервера до її збереження на диску сервера-лідера (далі – *Tleadercommit*); час від моменту збереження повідомлення на диску сервера-лідера до завершення його копіювання на інші сервери в кластері *Kafka* (далі – *Treplica*); час від моменту, коли продюсер *Kafka* отримав підтвердження від сервера-лідера (параметр "*acks_config*" = 1 або "*acks_config*" = "*all*") або коли певна кількість інформації була збережена на диску сервера-лідера ("*acks_config*" = 0), до моменту, коли клієнт *Kafka* отримує подію (далі – *Tfetch*). Отже, формальне визначення наскрізної затримки в кластері *Kafka* має такий вигляд:

$$TL = T_{producer} + T_{leadercommit} + T_{replica} + T_{fetch}. \quad (19)$$

На рис. 3 подано запропоновану структуру дискретної мережі Баєса для моделювання наскрізної затримки в кластері *Kafka*, побудовану на основі аналізу архітектури *Kafka*-кластера. Ця структура містить шар спостережувальних змінних $X = \{x_1, \ldots, x_{11}\}$, що відповідають параметрам конфігурації *Kafka*-кластера, а саме: $x_1(acks\_config)$ – визначає рівень підтвердження, який вимагає сервер-лідер від продюсерів; $x_2(buffer.memory)$ – визначає обсяг пам'яті в байтах, яку продюсер може використовувати для формування пакета даних перед надсиланням на сервер; $x_3(max.inflight.requests.per.connection)$ – встановлює максимальну кількість непідтверджених запитів, які можна надіслати на сервер за допомогою одного з'єднання; $x_4(socket.receive.buffer.bytes)$ – задає розмір мережевого буфера сокета для прийняття інформації сервером; $x_5(max.poll.records)$ – визначає максимальну кількість записів, що *Kafka*-клієнт може обробити за один виклик методу $poll()$; $x_6(log.flush.interval.ms)$ – встановлює час, протягом якого повідомлення може залишатися в пам'яті сервера перед записом на диск; $x_7(replica.fetch.min.bytes)$ – вказує на мінімальний обсяг інформації, який має бути скопійований на сервері перед надсиланням запиту на отримання нових даних від сервера-лідера; $x_8(linger.ms)$ – визначає, як довго продюсер накопичуватиме інформацію для формування пакета перед відправленням на сервер-лідер; $x_9(batch.size)$ – максимальний розмір пакета даних у байтах, який продюсер надсилає на сервер-лідер за один раз; $x_{10}(compression.type)$ – тип стиснення інформації, який використовує продюсер перед надсиланням; $x_{11}(fetch.min.bytes)$ – мінімальна кількість байтів, що має бути записана на диск сервера-лідера перед тим, як стати доступною для клієнтів.

Обрані параметри конфігурації $X = \{x_1, \ldots, x_{11}\}$ безпосередньо впливають на показники продуктивності кластера *Kafka* $Y = \{y_1, \ldots, y_6\}$. Визначимо їх як приховані змінні мережі Баєса та включимо у перший прихований шар, а саме: $y_1(RequestQueueTimeMs)$ – протягом якого дані, надіслані продюсером, перебувають у черзі запитів перед прийняттям для збереження в журналі подій сервера; $y_2(ResponseQueueTimeMs)$ – час у мілісекундах, протягом якого дані, збережені на сервері, перебувають у черзі відповідей; $y_3(LogFlushRateAndTimeMs)$ – частота й час, з якими дані, збережені в журналі логів сервера, переміщуються з пам'яті сервера на диск; $y_4(BytesInPerSec)$ – загальна кількість байтів, які Kafka сервер отримує щосекунди від усіх продюсерів. Високі значення можуть свідчити про необхідність додати обчислювальні ресурси або скоригувати конфігурацію продюсерів; $y_5(BytesOutPerSec)$ –

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

вимірює швидкість, з якою дані надсилаються з сервера до клієнтів; $y_6(batch\_size\_avg)$ – середній розмір пакета повідомлень, який продюсер надсилає до брокера *Kafka*.

Другий прихований шар містить складники виразу (19), оскільки їх значення формуються зі значень показників продуктивності $Y=\{y_1,\ldots,y_6\}$, а саме: y_1^1 – загальний час продюсера (*Tproducer*) в мілісекундах; y_2^1 – загальний час збереження на диску сервера-лідера (*Tleadercommit*) в мілісекундах; y_3^1 – загальний час копіювання даних на інші сервери (*Treplica*) в мілісекундах; y_4^1 – загальний час для отримання даних клієнтами (*Tfetch*) в мілісекундах. Вихідний шар містить єдину змінну – наскрізну затримку в кластері *Kafka* (*TL*), для якої запропонована мережа Баєса здійснюватиме прогнозування. Для побудови дискретної баєсівської мережі значення кожного вузла має бути дискретизоване. Для цього подаємо кожну змінну за нотацією

$$\langle N,\ S,\ G(S)\rangle\ , \qquad (20)$$

де N – назва змінної; S – множина станів $S_{i=\overline{1,N}}=\{s_i\}$, поданих у вигляді лінгвістичних термів; $G(S)$ – множина значень для кожного стану $s_i\in S$.

Для визначення таблиць умовних імовірностей для вузлів мережі Баєса скористаємося алгоритмом максимальної правдоподібності (MLE) [16].

Для обчислення оцінок впливу за методами Соболя та Морріса необхідна модель, здатна точно прогнозувати значення вихідної функції як для випадкових варіацій вхідних змінних, згенерованих за методом Морріса, так і для дисперсій вхідних змінних, згенерованих за методом Соболя. Для цього дослідження оберемо модель *XGBRegressor*, алгоритм якої полягає в побудові ансамблю дерев рішень, де кожне нове дерево додається до ансамблю з метою мінімізації помилок, припущених попередніми моделями дерев, збільшуючи в такий спосіб точність моделі [17]. Модель *XGBRegressor* побудуємо для параметрів конфігурації, що належать до множини X спостережувальних змінних мережі Баєса та мають неперервні значення.

Результати досліджень та їх обговорення

На рис. 4–6 подано обчислені за методами (10)–(12) оцінки впливу на запропонованій мережі Баєса (рис. 3). Порівняння результатів отриманих значень сил впливу за методами *DE, DH, DJ* вказує на певні розбіжності.

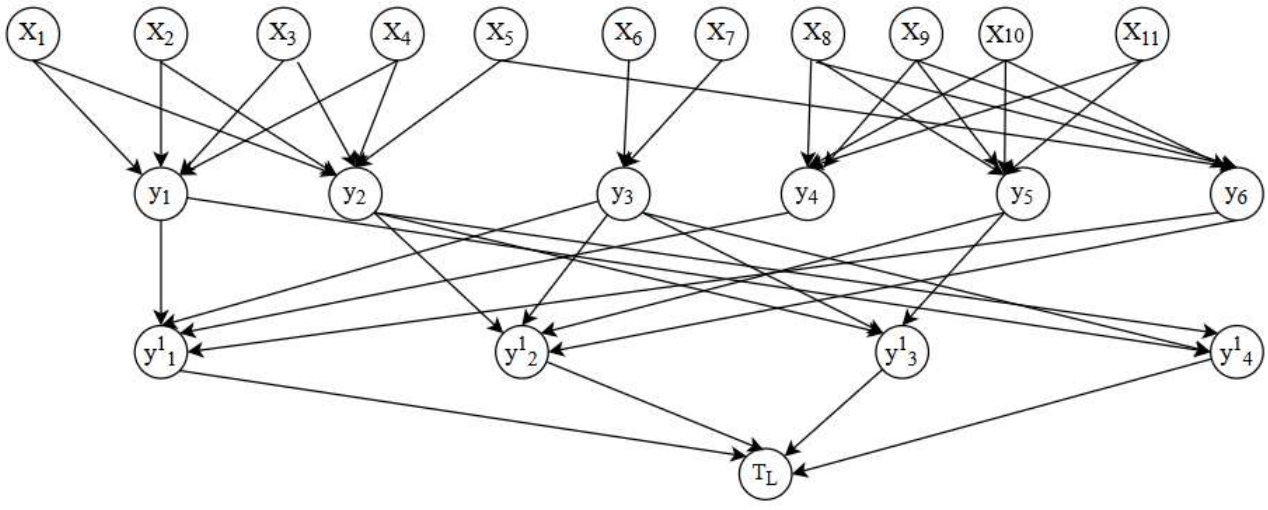


Рис. 3. Розроблена структура дискретної мережі Баєса для моделювання наскрізної затримки в кластері *Kafka*

1. Шар із спостережувальними змінними $X=\{x_1,\ \ldots,x_{11}\}$.

1.1. За методом D_E сильний вплив мають: $x_6(log.flush.interval.ms)$ на $y_3(LogFlushRateAndTimeMs)$; $x_8(linger.ms)$ на $(y_3(LogFlushRateAndTimeMs);$ $y_4(BytesInPerSec))$; $x_9(batch.size)$ на $y_4(BytesInPerSec)$; $x_{10}(compression.type)$ на $y_4(BytesInPerSec)$. Середній вплив оцінено для $x_8(linger.ms)$ на $y_5(BytesOutPerSec)$; $x_9(batch.size)$ на $y_5(BytesOutPerSec)$; $x_{10}(compression.type)$ на $y_5(BytesOutPerSec)$.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

1.2. За методом D_H сильний вплив мають: $x_5(max.poll.records)$ на $y_6(batch\_size\_avg)$; $x_8(linger.ms)$ на $y_6(batch\_size\_avg)$; $x_9(batch.size)$ на $y_6(batch\_size\_avg)$; $x_{10}(compression.type)$ на $y_6(batch\_size\_avg)$. Середній вплив оцінено $x_1(acks\_config)$ на $y_1(RequestQueueTimeMs)$; $x_2(buffer.memory)$ на $y_1(RequestQueueTimeMs)$; $x_9(batch.size)$ на $(y_4(BytesInPerSec); y_5(BytesOutPerSec))$; $x_8(linger.ms)$ на $(y_4(BytesInPerSec); y_5(BytesOutPerSec))$; $x_{10}(compression.type)$ на $(y_4(BytesInPerSec); y_5(BytesOutPerSec))$; $x_3(max.inflight.requests.per.connection)$ на $y_2(ResponseQueueTimeMs)$; $x_4(socket.receive.buffer.bytes)$ на $y_1(RequestQueueTimeMs)$.

1.3. За методом D_J сильний вплив мають: $x_5(max.poll.records)$ на $y_6(batch\_size\_avg)$; $x_8(linger.ms)$ на $y_6(batch\_size\_avg)$; $x_9(batch.size)$ на $y_6(batch\_size\_avg)$; $x_{10}(compression.type)$ на $y_6(batch\_size\_avg)$. Середній вплив оцінено $x_1(acks\_config)$ на $y_1(RequestQueueTimeMs)$; $x_2(buffer.memory)$ на $y_1(RequestQueueTimeMs)$; $x_9(batch.size)$ на $(y_4(BytesInPerSec); y_5(BytesOutPerSec))$; $x_{11}(fetch.min.bytes)$ на $(y_4(BytesInPerSec); y_5(BytesOutPerSec))$; $x_8(linger.ms)$ на $(y_4(BytesInPerSec); y_5(BytesOutPerSec))$; $x_3(max.inflight.requests.per.connection)$ на $y_1(RequestQueueTimeMs)$; $x_4(socket.receive.buffer.bytes)$ на $y_1(RequestQueueTimeMs)$.

2. Перший шар з прихованими змінними $Y = \{y_1, ..., y_6\}$.

2.1. За методом *DE* сильний вплив мають: $y_1(RequestQueueTimeMs)$ на $y_4^1(Tfetch)$; $y_2(ResponseQueueTimeMs)$ на $(y_3^1(Treplica); y_4^1(Tfetch))$; $y_3(LogFlushRateAndTimeMs)$ на $(y_3^1(Treplica); y_4^1(Tfetch))$; $y_5(BytesOutPerSec)$ на $y_3^1(Treplica)$. Середній вплив оцінено $y_6(batch\_size\_avg)$ на $(y_1^1(Tproducer); y_2^1(Tleadercommit))$; $y_4(BytesInPerSec)$ на $y_1^1(Tproducer)$; $y_5(BytesOutPerSec)$ на $y_2^1(Tleadercommit)$; $y_1(RequestQueueTimeMs)$ на $y_1^1(Tproducer)$; *y2(ResponseQueueTimeMs)* на $y_2^1(Tleadercommit)$; *y3(LogFlushRateAndTimeMs)* на $(y_1^1(Tproducer); y_2^1(Tleadercommit))$.

2.2. За методом *DH* сильний вплив мають: *y1 (RequestQueueTimeMs)* на y_4^1 *(Tfetch)*; *y2 (ResponseQueueTimeMs)* на $(y_2^1$ *(Tleadercommit)*; y_4^1 *(Tfetch))*; *y3 (LogFlushRateAndTimeMs)* на $(y_2^1$ *(Tleadercommit)*; y_4^1 *(Tfetch))*, *y5 (BytesOutPerSec)* на y_2^1*(Tleadercommit)*; *y6 (batch_size_avg)* на y_2^1 *(Tleadercommit)*. Середній вплив оцінено для *y2 (ResponseQueueTimeMs)* на y_3^1 *(Treplica)*; *y5 (BytesOutPerSec)* на y_3^1 *(Treplica)*.

2.3. За методом *DJ* сильний вплив мають: *y1 (RequestQueueTimeMs)* на y_4^1 (Tfetch); *y3 (LogFlushRateAndTimeMs)* на $(y_2^1$ *(Tleadercommit)*; y_4^1 *(Tfetch))*; *y5 (BytesOutPerSec)* на y_2^1 *(Tleadercommit)*; *y6 (batch_size_avg)* на y_2^1 *(Tleadercommit)*.

3. Другий шар з прихованими змінними $\{y_1^1, y_2^1, y_3^1, y_4^1\}$.

Оцінка впливу є узгодженою між трьома методами. Усі змінні другого прихованого шару мають сильний вплив на вихідну змінну *TL*.

У табл. 1 наведено оцінки, отримані за методами Морріса та Соболя, обчислені на моделі *XGBRegressor*, і на їх основі визначено зважені оцінки *WDE*, *WDH*, *WDJ* за формулами (13)–(15), де значення *α, β, γ* становлять 0.2, 0.2 та 0.6 відповідно. Значення параметрів підібрано так, щоб брати до уваги структурні подібності між результатами, отриманими на основі метричних відстаней. У табл. 2 визначено класи зважених оцінок, обчислені за класифікацією (16). Порівняльний аналіз вказує на повну узгодженість оцінок за методами *WDH*, *WDJ*, які визначили сильний вплив параметрів *batch.size*, *fetch.min.bytes*, *linger.ms*, *max.poll.records* та сердній вплив параметрів *acks_config*, *buffer.memory*,

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

max.inflight.requests.per.connection, *socket.receive.buffer.bytes.*

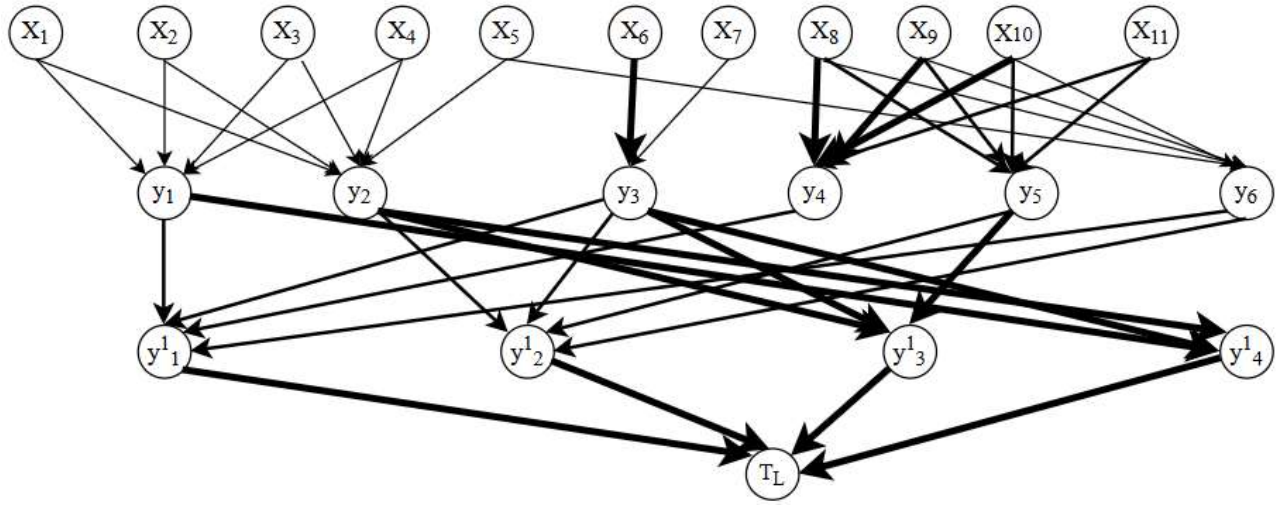


Рис. 4. Оцінювання впливу в мережі Баєса за методом евклідової відстані (*DE*)

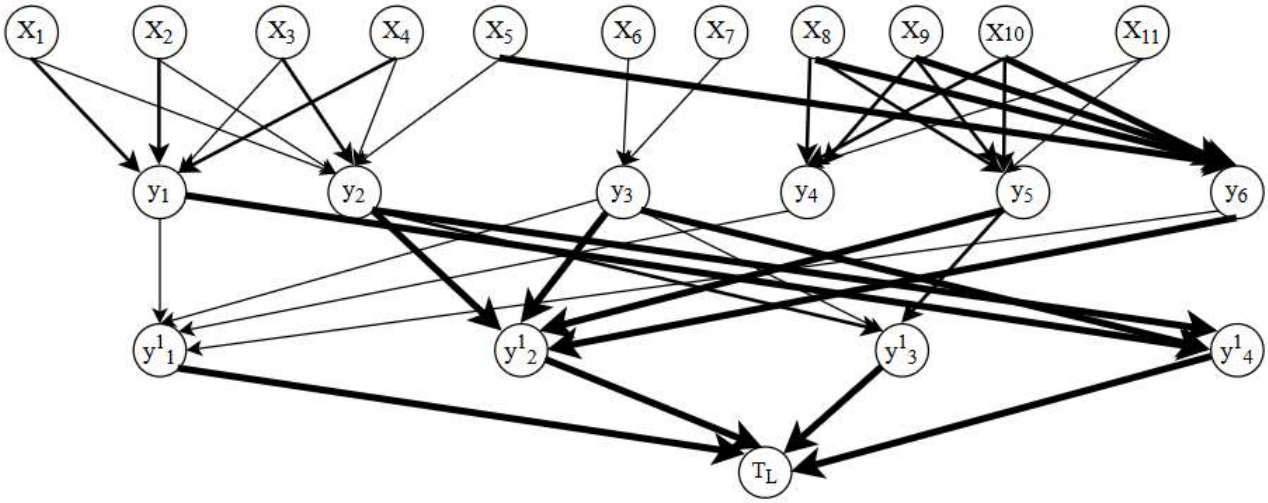


Рис. 5. Оцінювання впливу в мережі Баєса за методом хеллінгерової відстані (*DH*)

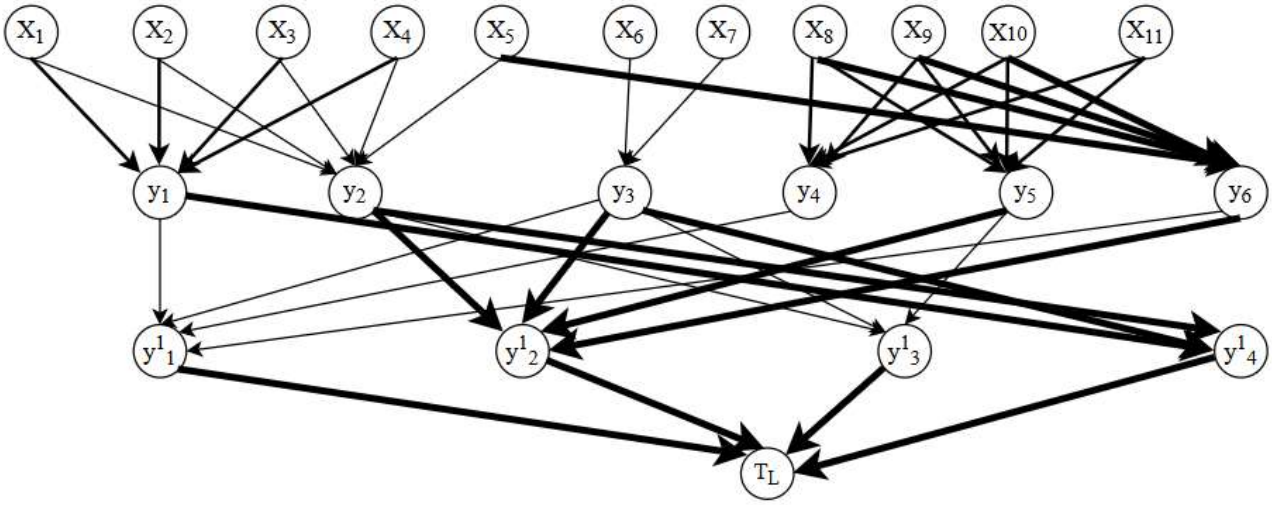


Рис. 6. Оцінювання впливу в мережі Баєса за методом *J*-дивергенції (*DJ*)

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Класи зважених оцінок за методом *WDE* відрізняються від результатів за методами *WDH, WDJ*, а саме: параметр конфігурації *log.flush.interval.ms* має сильний вплив за методом *WDE* і низький за методами *WDH, WDJ*; параметр *max.poll.records* має низький вплив за методом *WDE* і високий за методами *WDH, WDJ*; параметр *replica.fetch.min.bytes* має сердній вплив за методом *WDE*, але низький за методами *WDH, WDJ*.

Таблиця 1. *Оцінки впливу параметрів конфігурації Kafka-кластера, обчислені на мережі Баєса та моделі XGBRegressor*

Назва вузла мережі Баєса	*DE*	*DH*	*DJ*	Індекс Морріса, μi	Індекс Соболя, *Si*	Оцінка за методом, *WDE*	Оцінка за методом, *WDH*	Оцінка за методом, *WDJ*
acks_config	0.003	0.042	0.287	0	0	0.002	0.025	0.172
batch.size	0.036	0.127	0.739	1	0.953	0.412	0.467	0.834
batch_size_avg	0.014	0.060	0.369	n/a	n/a	n/a	n/a	n/a
buffer.memory	0.003	0.042	0.287	0	0	0.002	0.025	0.172
BytesInPerSec	0.000	0.016	0.123	n/a	n/a	n/a	n/a	n/a
BytesOutPerSec	0.038	0.073	0.247	n/a	n/a	n/a	n/a	n/a
compression type	0.036	0.127	0.739	n/a	n/a	n/a	n/a	n/a
fetch.min.bytes	0.022	0.063	0.369	0	0	0.013	0.038	0.222
linger.ms	0.036	0.127	0.739	0.2	0.046	0.070	0.124	0.492
log.flush.interval.ms	0.025	0.009	0.000	0.00	0	0.015	0.005	0.000
LogFlushRateAndTimeMs	0.083	0.169	0.739	n/a	n/a	n/a	n/a	n/a
max.inflight.requests.per.connection	0.003	0.042	0.287	0	0	0.002	0.025	0.172
max.poll.records	0.002	0.060	0.410	0	0	0.001	0.036	0.246
replica.fetch.min.bytes	0.007	0.000	0.000	0	0	0.004	0.000	0.000
RequestQueueTimeMs	0.037	0.089	0.492	0	0	0.022	0.054	0.295
ResponseQueueTimeMs	0.073	0.146	0.616	0	0	0.044	0.087	0.370
socket.receive.buffer.bytes	0.003	0.042	0.287	0	0	0.002	0.025	0.172
Tfetch	1.000	1.000	1.000	n/a	n/a	n/a	n/a	n/a
Tleadercommit	1.000	1.000	1.000	n/a	n/a	n/a	n/a	n/a
Tproducer	0.995	0.997	1.000	n/a	n/a	n/a	n/a	n/a
Treplica	1.000	1.000	1.000	n/a	n/a	n/a	n/a	n/a

Познака "n/a" означає, що параметри прихованих шарів баєсівської мережі не додані до моделі *XGBRegressor*, оскільки значення цих параметрів не задаються під час конфігурації *Kafka*-кластера, а формуються як результат впливу визначеної конфігурації.

На рис. 7 зображено вплив кожного параметра конфігурації з множини *X* на затримку *Kafka*-кластера, обчислений за методами на основі метричних відстаней *DE, DH, DJ*, прямого впливу параметрів (за допомогою індексів Морріса μi) та з огляду на взаємодію між параметрами (за допомогою індексів Соболя Si). Різна висота стовпців ілюструє неузгодженості в оцінках, отриманих відповідно до зазначених методів. Лінії є нормалізованими оцінками впливу параметрів, побудованими на основі зважених метрик *WDE, WDH, WDJ*. Близькість ліній для методів *WDH, WDJ* вказує на узгодженість оцінок.

За правилом класифікації "більшість голосів", відповідно до виразу (18), отримано узгоджені класи оцінки впливу *W(X)*, які після надання числових значень були використані для ініціалізації матриці ваг нейронної мережі прямого поширення, що прогнозує наскрізну затримку в *Kafka*-кластері залежно від обраної конфігурації. Для визначення ефективності запропонованого підходу оцінювання впливу *W(X)* на навчання нейронної мережі також застосовано під час

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

ініціалізації матриці ваг за оцінками, отриманими на основі запропонованого підходу (13)–(15), і в процесі ініціалізації матриці ваг нейронної мережі прямого поширення за методом *Xavier*, який упроваджується за замовчуванням у нейронних мережах прямого поширення.

Таблиця 2. *Класи оцінок впливу параметрів, визначені за запропонованою класифікацією та правилом "більшості голосів"*

Назва вузла мережі Баєса	***WDJ***	***WDH***	***WDE***	**Клас оцінки, *WDJ***	**Клас оцінки, *WDJ***	**Клас оцінки, *WDJ***	**Клас оцінки за правилом "більшості голосів"**
acks_config	0.172	0.025	0.002	M	M	M	M
batch.size	0.834	0.467	0.412	H	H	H	H
batch_size_avg	0.222	0.036	0.008	M	M	M	M
buffer.memory	0.172	0.025	0.002	M	M	M	M
fetch.min.bytes	0.222	0.038	0.013	H	H	H	H
linger.ms	0.492	0.124	0.070	H	H	H	H
log.flush.interval.ms	0.000	0.005	0.015	L	L	H	L
max.inflight.requests.per.connection	0.172	0.025	0.002	M	M	M	M
max.poll.records	0.246	0.036	0.001	H	H	L	H
replica.fetch.min.bytes	0.000	0.000	0.004	L	L	M	L
socket.receive.buffer.bytes	0.172	0.025	0.002	M	M	M	M

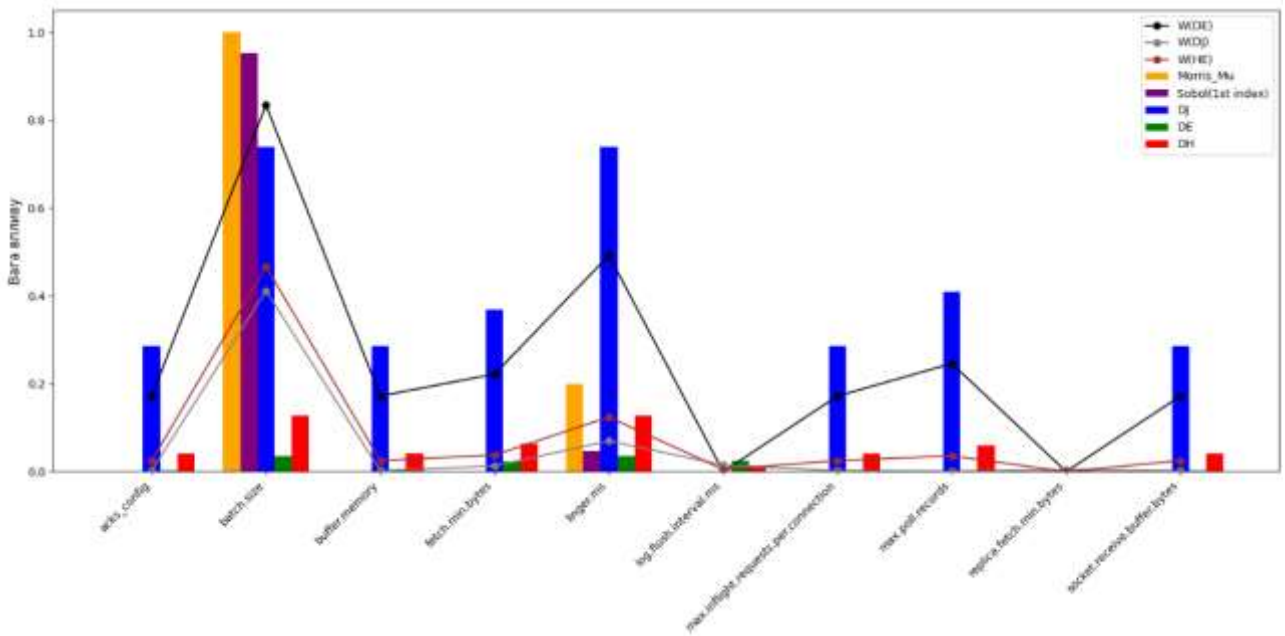


Рис. 7. Порівняння оцінок впливу параметрів конфігурації на затримку *Kafka*-кластера за різними методами

На рис. 8 згладжена функція витрат, отримана на основі критерію мінімізації середньоквадратичної похибки, демонструє найменші значення протягом перших 40 ітерацій навчання нейронної мережі, яка була ініціалізована вагами узгоджених класів оцінки впливу *W(X)*. Це свідчить про ефективність запропонованого підходу, оскільки що швидше мережа досягає мінімуму функції витрат, то ефективнішою є ініціалізація ваг.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

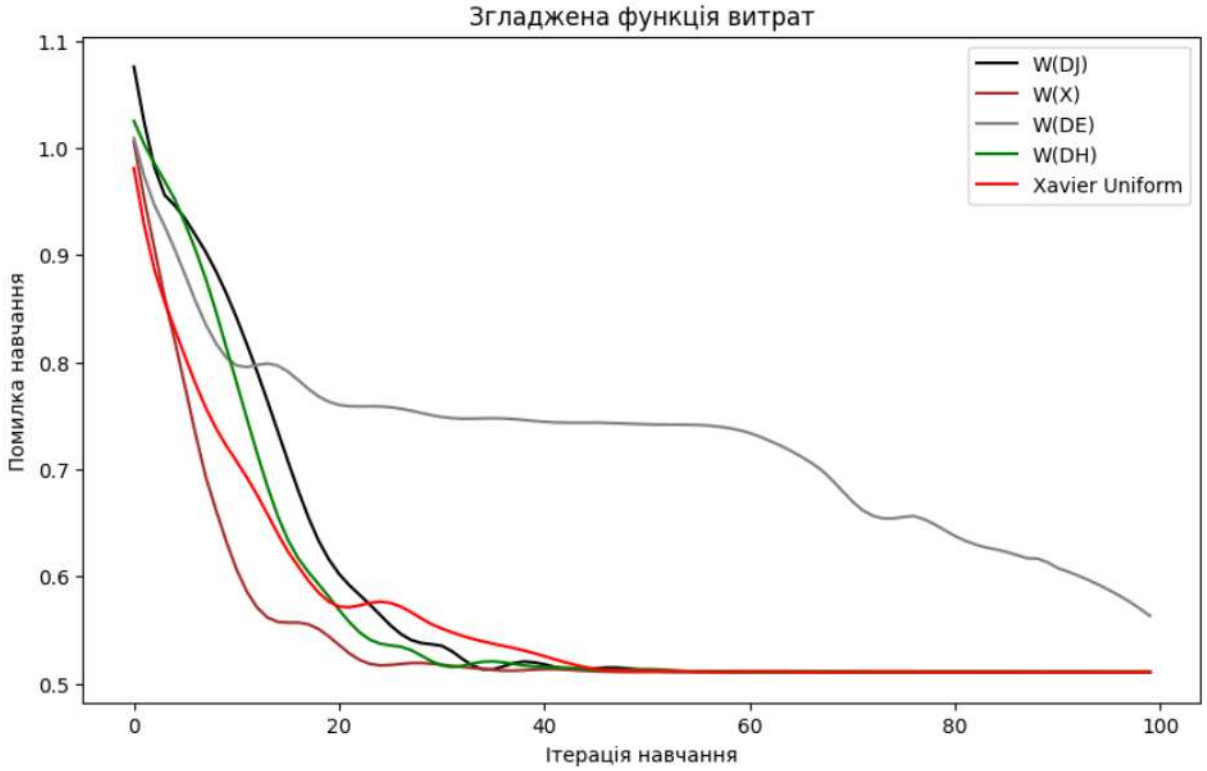


Рис. 8. Згладжена функція витрат, яка отримана на основі критерію мінімізації середньоквадратичної похибки в процесі навчання нейронної мережі

Висновки й перспективи подальшого дослідження

У статті запропоновано новий підхід до аналізу чутливості, який поєднує оцінку прямого впливу параметрів (за допомогою індексу Морріса μi), взяття до уваги взаємодій між параметрами (за допомогою індексу Соболя Si) і причинно-наслідкові зв'язки між параметрами, що оцінюються за допомогою методів, застосовуваних у межах причинно-орієнтованих моделей, зокрема в баєсівських мережах (евклідова відстань, хеллінгерова відстань, *J*-дивергенція).

Запропонований підхід формалізовано у вигляді зваженої комбінації вищезазначених методів. Для його апробації розроблено класифікацію оцінок чутливості, а для узагальнення результатів – правило "більшості голосів".

Порівняльний аналіз оцінок впливу параметрів конфігурації *Kafka*-кластера на наскрізну затримку, проведений за допомогою метрик евклідової відстані, хеллінгерової відстані та *J*-дивергенції на моделі баєсівської мережі, показав узгодженість у визначенні сильного впливу змінних другого прихованого шару. Водночас виявлено розбіжності щодо оцінок спостережуваних змінних та змінних першого прихованого шару, де вплив здебільшого оцінювався як середній. Це свідчить про різницю в чутливості зазначених методів і потребує їх уточнення або адаптації.

Запропонований підхід на основі зважених метрик продемонстрував абсолютну узгодженість оцінок за методами *WDH* (зважена хеллінгерова відстань) і *WDJ* (зважена *J*-дивергенція), що підтверджує перевагу комбінованого підходу над окремими метриками *DE, DH, DJ*.

Для оцінювання ефективності підходу проведено порівняльне навчання нейронної мережі з різними стратегіями ініціалізації ваг: на основі оцінок *W(DE), W(DH), W(DJ)*; за стандартним методом ініціалізації *Xavier*; за узгодженими класами оцінки впливу *W(X)*. Аналіз згладженої функції витрат, побудованої за критерієм мінімізації середньоквадратичної похибки, продемонстрував, що найменших значень функція досягає впродовж перших 40 ітерацій саме в моделі, ініціалізованій вагами з узгоджених оцінок *W(X)*. Це свідчить про покращену збіжність мережі та більш якісну ініціалізацію ваг.

Новизна підходу полягає в інтеграції переваг причинно-орієнтованих і дисперсійно-оцінних методів у межах єдиної зваженої метрики чутливості.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Практична цінність підходу полягає в тому, що його застосування під час аналізу чутливості або ініціалізації матриці ваг нейронної мережі дає змогу підвищити точність оцінок впливу параметрів, покращити збіжність моделі та скоротити час її навчання.

Подальші дослідження будуть спрямовані на впровадження метрик структурної подібності для аналізу чутливості та формалізацію критерію узгодженості оцінок, отриманих різними методами.

Список літератури

1. Solovei O., Honcharenko T., Fesan A. "Technologies to manager big data of urban building projects", *Management of Development of Complex Systems*, No. 60, P.121–128, 2024. DOI: 10.32347/2412-9933.2024.60.121-128
2. Narkhede, M. V., Bartakke, P. P., Sutaone, M. S. "A review on weight initialization strategies for neural networks", *Artificial intelligence review*, No. 55(1), P. 291-322. 2022. DOI: 10.1007/s10462-021-10033-z
3. Wong, K., Dornberger, R., Hanne, T. "An analysis of weight initialization methods in connection with different activation functions for feedforward neural networks", *Evolutionary Intelligence*, No. 17(3), P.2081-2089, 2024. DOI: 10.1007/s12065-022-00795-y
4. Brand J.E., Zhou X., Xie Y. "Recent developments in causal inference and machine learning", *Annual Review of Sociology*, No. 49 (1), P. 81-110. 2023. DOI: 10.1146/annurev-soc-030420-015345
5. Chumachenko K., Iosifidis A., and Gabbouj M. "Feedforward neural networks initialization based on discriminant learning", *Neural Networks*, No. 146, P. 220-229. 2022. DOI: 10.1016/j.neunet.2021.11.020
6. Zhao J., Schäfer F., and Anandkumar A. "Zero initialization: Initializing neural networks with only zeros and ones", *Published in Transactions on Machine Learning Research*, № 11. 2021. URL: arXiv preprint arXiv:2110.12661
7. Pan Y., Wang C., Wu Z., Wang Q., Zhang M., and Xu Z. "IDInit: A Universal and Stable Initialization Method for Neural Network Training", 2025. URL: arXiv preprint arXiv:2503.04626
8. Zhu C., Ni R., Xu Z., Kong K., Huang W. R., and Goldstein T. "Gradinit: Learning to initialize neural networks for stable and efficient training", *Advances in Neural Information Processing Systems*, No. 34, P.16410-16422. 2021. DOI: https://doi.org/10.3390/app15042008
9. Nouri A., van Treeck C., & Frisch J. "Sensitivity Assessment of Building Energy Performance Simulations Using MARS Meta-Modeling in Combination with Sobol'Method", *Energies*, No. 17(3), 695 p. 2024. DOI: https://doi.org/10.3390/en17030695
10. Sadeghi Z., Matwin S. "A Review of Global Sensitivity Analysis Methods and a comparative case study on Digit Classification". 2024. URL: arXiv preprint arXiv:2406.16975
11. Mazo G., Tournier L. "An inference method for global sensitivity analysis", *Technometrics*, No. 67(2), P. 270-282. 2025.
12. Kozniewski M., Kolendo Ł., Chmur S., Ksepko M. "Impact of Parameters and Tree Stand Features on Accuracy of Watershed-Based Individual Tree Crown Detection Method Using ALS Data in Coniferous Forests from North-Eastern Poland", *Remote Sensing*, No. 17(4), 575 p. 2025. DOI: https://doi.org/10.3390/rs17040575
13. Kaddoura M., Majeau-Bettez G., Amor B., & Margni M. "Global sensitivity analysis reduces data collection efforts in LCA: A comparison between two additive manufacturing technologies", *Science of the Total Environment*, No. 975, 179269 p. 2025. DOI: https://doi.org/10.1016/j.scitotenv.2025.179269
14. Raptis T. P., Passarella A. "A survey on networked data streaming with apache kafka", *IEEE Access*, No. 11, P. 85333-85350. 2023. DOI: 10.1109/ACCESS.2023.3303810
15. Kafka Producer Configuration Reference for Confluent Platform. URL: https://docs.confluent.io/platform/current/installation/configuration/producer-configs.html.
16. Wang J., Chen Z., Song Y., Liu Y., He J., Ma S. "Data-Driven Dynamic Bayesian Network Model for Safety Resilience Evaluation of Prefabricated Building Construction", *Buildings*, No. 14, 570 p. 2024. DOI: 10.3390/buildings14030570
17. Echabarri S., Do P, Vu H., Bornand B. "Machine learning and Bayesian optimization for performance prediction of proton-exchange membrane fuel cells", *Energy and AI*, No. 17, 100380 p. DOI: https://doi.org/10.1016/j.egyai.2024.100380

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

References

1. Solovei, O., Honcharenko, T., Fesan, A. (2024), "Technologies to manager big data of urban building projects", *Management of Development of Complex Systems*, No. 60, P.121–128, DOI: 10.32347/2412-9933.2024.60.121-128
2. Narkhede, M. V., Bartakke, P. P., Sutaone, M. S. (2022), "A review on weight initialization strategies for neural networks", *Artificial intelligence review*, No. 55(1), P. 291-322. DOI: 10.1007/s10462-021-10033-z
3. Wong, K., Dornberger, R., Hanne, T. (2024), "An analysis of weight initialization methods in connection with different activation functions for feedforward neural networks", *Evolutionary Intelligence*, No. 17(3), P.2081-2089, DOI: 10.1007/s12065-022-00795-y
4. Brand, J.E., Zhou, X., Xie, Y. (2023), "Recent developments in causal inference and machine learning", *Annual Review of Sociology*, No. 49 (1), P. 81-110. DOI: 10.1146/annurev-soc-030420-015345
5. Chumachenko, K., Iosifidis, A., Gabbouj, M. (2022), "Feedforward neural networks initialization based on discriminant learning", *Neural Networks*, No. 146, P. 220-229. DOI: 10.1016/j.neunet.2021.11.020
6. Zhao, J., Schäfer, F., Anandkumar, A. (2021), "Zero initialization: Initializing neural networks with only zeros and ones", *Published in Transactions on Machine Learning Research*, № 11. available at: arXiv preprint arXiv:2110.12661
7. Pan, Y., Wang, C., Wu, Z., Wang, Q., Zhang, M., Xu, Z. (2025), "IDInit: A Universal and Stable Initialization Method for Neural Network Training", available at: arXiv preprint arXiv:2503.04626
8. Zhu, C., Ni, R., Xu, Z., Kong, K., Huang, W. R., Goldstein, T. (2021), "Gradinit: Learning to initialize neural networks for stable and efficient training", *Advances in Neural Information Processing Systems*, No. 34, P.16410-16422. DOI: https://doi.org/10.3390/app15042008
9. Nouri, A., van Treeck, C., Frisch, J. (2024), "Sensitivity Assessment of Building Energy Performance Simulations Using MARS Meta-Modeling in Combination with Sobol'Method", *Energies*, No. 17(3), 695 p. DOI: https://doi.org/10.3390/en17030695
10. Sadeghi, Z., Matwin, S. (2024), "A Review of Global Sensitivity Analysis Methods and a comparative case study on Digit Classification". available at: arXiv preprint arXiv:2406.16975
11. Mazo, G., Tournier, L. (2025), "An inference method for global sensitivity analysis", *Technometrics*, No. 67(2), P. 270-282.
12. Kozniewski, M., Kolendo, Ł., Chmur, S., Ksepko, M. (2025), "Impact of Parameters and Tree Stand Features on Accuracy of Watershed-Based Individual Tree Crown Detection Method Using ALS Data in Coniferous Forests from North-Eastern Poland", *Remote Sensing*, No. 17(4), 575 p. DOI: https://doi.org/10.3390/rs17040575
13. Kaddoura, M., Majeau-Bettez, G., Amor, B., Margni, M. (2025), "Global sensitivity analysis reduces data collection efforts in LCA: A comparison between two additive manufacturing technologies", *Science of the Total Environment*, No. 975, 179269 p. DOI: https://doi.org/10.1016/j.scitotenv.2025.179269
14. Raptis, T. P., Passarella, A. (2023), "A survey on networked data streaming with apache kafka", *IEEE Access*, No. 11, P. 85333-85350. DOI: 10.1109/ACCESS.2023.3303810
15. "Kafka Producer Configuration Reference for Confluent Platform". available at: https://docs.confluent.io/platform/current/installation/configuration/producer-configs.html.
16. Wang, J., Chen, Z., Song, Y., Liu, Y., He, J., Ma S. (2024), "Data-Driven Dynamic Bayesian Network Model for Safety Resilience Evaluation of Prefabricated Building Construction", *Buildings*, No. 14, 570 p. DOI: 10.3390/buildings14030570
17. Echabarri, S., Do, P, Vu, H., Bornand, B. (2024), "Machine learning and Bayesian optimization for performance prediction of proton-exchange membrane fuel cells", *Energy and AI*, No. 17, 100380 p. DOI: https://doi.org/10.1016/j.egyai.2024.100380

Надійшла (Received) 28.05.2025

Відомості про авторів / About the Authors

Соловей Ольга Леонідівна – кандидат технічних наук, Київський національний університет будівництва і архітектури, доцент кафедри інформаційних технологій проєктування та прикладної математики, Київ, Україна; e-mail: solovey.ol@knuba.edu.ua; ORCID ID: https://orcid.org/0000-0001-8774-7243

Solovei Olga – PhD (Technical Sciences), Kyiv National University of Construction and Architecture, Associate Professor at the Department of Information Technology Design and Applied Mathematic, Kyiv, Ukraine.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

A WEIGHTED SENSITIVITY METRIC FOR PREDICTING LATENCY IN A KAFKA CLUSTER

The subject of this article is sensitivity analysis methods used to assess how variations in input parameters affect the output results of a model or system. The aim of the study is to develop a new approach to sensitivity analysis that combines classical parameter impact assessment methods (Morris and Sobol methods) with metrics that capture structural changes in the distribution of output data (Euclidean distance, Hellinger distance, Jensen divergence). This approach allows for evaluating the influence of a parameter not only in terms of the amplitude of its effect, but also in terms of changes in the shape and structure of the probability distribution of the results. To achieve this objective, the article addresses the following tasks: formal definition of a new sensitivity analysis approach; development of a Bayesian network for modeling end-to-end latency in a Kafka cluster; performing sensitivity analysis using the proposed approach; and conducting an experimental study using the calculated parameter influence weights to initialize the weight matrix of a neural network that predicts Kafka cluster latency based on selected configuration parameters. To accomplish these tasks, the study applied methods from the theory of experiments, Euclidean geometry, statistical distribution theory, information theory, machine learning, Bayesian statistics, and graph theory. Results: To evaluate the effectiveness of the proposed approach, comparative training of a neural network was conducted using various weight initialization strategies. Analysis of the loss function, constructed using the mean squared error minimization criterion, showed that the lowest values were achieved by the model initialized with weights obtained using the proposed parameter influence estimation approach. Conclusions: The study proposes a novel approach to sensitivity analysis. The innovation lies in integrating the strengths of both causal-oriented and variance-based methods within a unified weighted sensitivity metric. The practical value of this approach is that its application in sensitivity analysis or neural network weight initialization improves the accuracy of parameter impact assessment, enhances model convergence, and reduces training time.

Keywords: Morris method; Sobol index; Euclidean distance; Hellinger distance; Jensen divergence; sensitivity analysis.

Бібліографічні описи / Bibliographic descriptions

Соловей О.Л. Зважена метрика чутливості для прогнозування затримки в KAFKA-кластері. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 152–165. DOI: https://doi.org/10.30837/2522-9818.2025.3.152

Solovei, O. (2025), " A weighted sensitivity metric for predicting latency in a KAFKA cluster", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 152–165. DOI: https://doi.org/10.30837/2522-9818.2025.3.152

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

УДК 004.056:338 DOI: https://doi.org/10.30837/2522-9818.2025.3.166

А. Тімошин, Л. Калєніченко, Ю. Гнусов, І. Хавіна, М. Цуранов, І. Довгань

ІНТЕГРОВАНА МОДЕЛЬ УПРАВЛІННЯ РИЗИКАМИ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ НА ОСНОВІ AHP ТА БАЙЄСОВИХ МЕРЕЖ

Предмет дослідження — управління ризиками інформаційної безпеки в умовах сучасного цифрового середовища, де необхідна інтеграція стратегічних та тактичних підходів для забезпечення адаптивного захисту. **Мета роботи** — розробка гібридної моделі управління кіберризиками шляхом поєднання методологічного аналізу, експертних оцінок, ймовірнісного моделювання та технічного моніторингу. **Завдання дослідження** полягають у: (1) аналізі взаємодоповнюваності методології CRAMM і систем SIEM; (2) побудові процедури кількісної пріоритезації загроз і вразливостей на основі аналітичного методу ієрархій (AHP); (3) інтеграції отриманих оцінок у байєсові мережі (BN) для ймовірнісного прогнозування ризиків; (4) реалізації запропонованого підходу за допомогою сучасних інструментів автоматизації. Методи, використані в роботі, включають: методологію CRAMM для ідентифікації активів, загроз і вразливостей; AHP Томаса Сааті для кількісної оцінки пріоритетів на основі експертних суджень із вимірюванням узгодженості за допомогою коефіцієнта конкордації Кендалла; математичне моделювання причинно-наслідкових зв'язків за допомогою байєсових мереж (BN); а також використання систем класу SIEM для оперативного моніторингу подій безпеки. Практична реалізація підходу здійснювалася за допомогою Python, зокрема бібліотек Numpy, SciPy, pgmpy, та веб-інтерфейсу Streamlit. **Результати.** Розроблено інтегрований підхід, що об'єднує CRAMM, AHP, BN та SIEM у єдину адаптивну систему управління ризиками. Показано, що AHP дозволяє перетворити суб'єктивні експертні оцінки в об'єктивні вагові коефіцієнти, що підвищує надійність аналізу. На основі цих даних побудовано байєсову мережу для оцінки ризику фінансових збитків, яка враховує наявність загрози, вразливості та можливий інцидент. Модель реалізовано програмно, продемонстровано процес факторизації спільного розподілу та маргіналізації прихованих змінних для отримання апостеріорних ймовірностей. Веб-інтерфейс на базі Streamlit забезпечує зручність використання інструменту непрофесійними користувачами. **Висновки.** Запропонований гібридний підхід дозволяє ефективно поєднати стратегічне планування (CRAMM), експертні оцінки (AHP), ймовірнісне моделювання (BN) та оперативний моніторинг (SIEM), формуючи проактивну, науково обґрунтовану систему управління ризиками. Така інтеграція забезпечує високий рівень адаптивності та точності в умовах динамічного загрозного ландшафту, що робить модель практично застосовною для організацій різного рівня.

Ключові слова: Методологія CRAMM, Системи SIEM, Аналітичний метод ієрархій (AHP), Байєсові мережі (BN), Експертні оцінки, Аналіз загроз та вразливостей.

Вступ

У сучасному цифровому середовищі кібербезпека набуває стратегічного значення, виходячи за межі технічних аспектів і стаючи невід'ємною складовою національної безпеки, економічної стійкості та захисту прав громадян. Хоча традиційно управління кіберризиками сприймається як прерогатива держави, реалії глобалізованого кіберпростору, постійне ускладнення атак та швидкий технологічний прогрес ставлять під сумнів ефективність виключно регульованого державного підходу. Державні структури часто відстають від темпів розвитку загроз, обмежені в ресурсах та експертній компетенції, тоді як самі загрози носять транснаціональний характер і вимагають оперативної, гнучкої та міжсекторної взаємодії. У цих умовах приватний сектор, який є лідером інновацій у сфері IT-безпеки, виступає ключовим партнером у забезпеченні комплексного захисту.

Ефективність сучасних систем управління ризиками інформаційної безпеки значною мірою базується на міжнародних стандартах, зокрема ISO/IEC 27005, який визнано одним із найповніших підходів до структурованого аналізу ризиків [6]. Однак, як показують дослідження, успішне впровадження таких стандартів вимагає інтеграції з гнучкими методологічними та технічними інструментами [19]. У контексті кібербезпеки велике значення має також аналітика великих даних, яка дозволяє поєднувати стратегічне планування з оперативним реагуванням [23].

Мета статті — розробити та обґрунтувати інтегрований гібридний підхід до управління ризиками інформаційної безпеки, який поєднує переваги методологічного аналізу, експертних оцінок,

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

ймовірнісного моделювання та реального часу моніторингу для формування адаптивної, проактивної системи захисту.

Завдання дослідження полягає у:

- аналізі ролі та взаємодоповнюваності методології CRAMM і систем SIEM у стратегічному та тактичному управлінні ризиками;
- розробці процедури кількісної пріоритезації загроз та вразливостей на основі аналітичного методу ієрархій (AHP) із вимірюванням узгодженості експертних думок;
- побудові ймовірнісної моделі управління ризиками за допомогою байєсових мереж (BN) з використанням даних, отриманих через AHP;
- демонстрації практичної реалізації запропонованого підходу з використанням сучасних інструментів автоматизації (Python, pgmpy, Streamlit).

1. Аналіз останніх досліджень і публікацій

На даний час розроблені потужні методології оцінки та управління ризиками інформаційної безпеки. Одна з таких методологій – це CRAMM, яка розроблена у Великій Британії. Вона фокусується на систематичному аналізі активів організації, ідентифікації загроз та вразливостей, а також оцінці потенційного впливу ризиків, що описується в детальних посібниках та документації. Ці матеріали надають покрокові інструкції щодо ідентифікації та оцінки активів, виявленні загроз та вразливостей, оцінки ризиків, вибору та впровадженні заходів безпеки на основі каталогів, моніторинг та перегляд. Програмний комплекс CRAMM версії 5.0 використовує кількісні та якісні методи оцінки ризиків. Ретельний аналіз методів CRAMM наведено в роботі [1].

В деяких роботах пропонуються методи і засоби оцінювання ризиків та управління ризиками інформаційної безпеки для конкретних сфер діяльності. Наприклад в роботі [2] проаналізовано шляхи зменшення рівня ризиків у сфері електронного бізнесу, де особливо вразливими є інформаційні та фінансові транзакції. Математична модель ризиків в цій роботі спирається на схеми незалежних випробувань Бернуллі та схему Пуассона. Тобто, здійснюється ймовірнісне прогнозування результатів. Ще одним прикладом окремого випадку щодо безпеки даних є робота [3], яка досліджує сучасні підходи та інструменти для забезпечення безпеки API у веб-застосунках, реалізованих за допомогою JavaScript. Дослідженню кібербезпеки інформаційних ресурсів підприємства в IT-галузі присвячена робота [4], в якій для мінімізації ризиків інформаційної безпеки використовується такий показник, як рівень витрат (в матеріальному або вартісному вираженні) на відновлення працездатності системи у разі її відмови за одним або декількома напрямками. Така модель зводиться до вирішення багатопараметричної задачі лінійного програмування. Взагалі, формула "*Ризик = Імовірність загрози × Потенційний збиток*" є однією з найпоширеніших у сфері управління ризиками. Вона використовується у різних стандартах і підходах до оцінки інформаційної безпеки, фінансового аналізу та управління ризиками. Наприклад, в роботі [5], розвинуто дискретну ймовірнісну модель багатокритеріального аналізу ушкодженості об'єкта захисту за припущення про незалежність атак та засобів захисту. Для випадкової величини кількості ушкоджень за фіксований проміжок часу отримано представлення у вигляді суми біноміально розподілених випадкових величин, які залежать від параметрів атак та захисту. Описано випадкові величини економічних втрат, часу відновлення та затрат на відновлення, для яких знайдено в аналітичному вигляді математичні сподівання та дисперсії.

У статті [6] розглядається структура оцінки ризиків інформаційної безпеки (ISRA), яка є основою для порівняння різних методів оцінки ризиків. Автори пропонують комплексну структуру, яка дозволяє порівнювати різні методи, додаючи нові завдання з кожного розглянутого методу. Якщо завдання було присутнє в дослідженому методі ISRA, але не в CURF, воно додавалося до моделі, що давало змогу виміряти повноту досліджуваних методів. За результатами дослідження, "ISO/IEC 27005 Information Security Risk Management" є найбільш повним підходом на даний момент.

Одним з показників, який відображає поточний стан кібербезпеки держави є Національний індекс кібербезпеки (National Cyber Security Index, NCSI). Це глобальний оперативний індекс, що оцінює готовність країн запобігати кіберзагрозам та керувати кіберінцидентами. NCSI також є базою даних із загальнодоступними доказовими матеріалами та інструментом для нарощування національного потенціалу кібербезпеки. В роботі [7] пропонується комплекс формальних математичних моделей, які

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

забезпечують опис завдання визначення національного рівня цифрового розвитку країн з урахуванням національного рівня кібербезпеки та кіберзахисту з позицій системного аналізу. Потрібно зауважити, що NSCI структуровано 46 індикаторами, з яких найбільш суттєвими є розробка політики кібербезпеки, аналіз кіберзагроз, професійний розвиток, захист цифрових послуг, електронна ідентифікація та довірчі послуги, захист персональних даних, боротьба з кіберзлочинністю, військові кібероперації. Зокрема, Україна посідає 15-те місце в цьому рейтингу з індексом 80.83. Це свідчить про відносно високий рівень кібербезпеки в країні.

З аналізу наявних підходів випливає, що більшість із них або надто загальні, або надто спеціалізовані, або засновані на статичних моделях, що не враховують динаміку кіберзагроз. Крім того, спостерігається розрив між стратегічними методологіями (наприклад, CRAMM) та тактичними системами (наприклад, SIEM), що призводить до фрагментарності в управлінні ризиками.

На основі виявлених прогалин, актуальною є задача розробки інтегрованого гібридного підходу до управління ризиками інформаційної безпеки, який:

- поєднує переваги методологічного аналізу (CRAMM) та оперативного моніторингу (SIEM);
- перетворює суб'єктивні експертні оцінки в об'єктивні кількісні показники за допомогою аналітичного методу ієрархій (AHP) із перевіркою узгодженості (коефіцієнт Кендалла);
- будує ймовірнісну модель причинно-наслідкових зв'язків між загрозами, вразливостями та інцидентами за допомогою байєсових мереж (BN);
- забезпечує автоматизовану реалізацію моделі з використанням сучасних інструментів (Python, pgmpy, Streamlit) для підвищення доступності та точності аналізу.

2. Матеріали та методи

2.1. CRAMM і SIEM як інструменти управління ризиками в інформаційній безпеці

Ефективне управління ризиками в інформаційній безпеці передбачає поєднання методологічних підходів до аналізу та оцінки загроз із практичними інструментами моніторингу й реагування. У цій площині важливу роль відіграють такі рішення, як програмний комплекс CRAMM та системи управління інформаційною безпекою класу SIEM.

CRAMM (CCTA Risk Analysis and Management Method) – це одна з перших стандартизованих методик аналізу та управління ризиками у сфері інформаційних технологій [1]. Вона була розроблена у Великій Британії Центральним агентством телекомунікацій (Central Computer and Telecommunications Agency, CCTA) ще у 1980-х роках та продовжує застосовуватись у сучасних версіях, зокрема у CRAMM 5.0. Її трьоетапний процес (ідентифікація активів, аналіз загроз, розрахунок ризиків) детально описаний у офіційному користувацькому посібнику [13]. Хоча CRAMM забезпечує комплексний підхід, його ефективність залежить від якості експертних суджень, що потребує доповнення формальними методами їхньої валідації [16].

Основна ідея CRAMM полягає у трьохетапному підході до управління ризиками:

1. Ідентифікація активів – визначення інформаційних ресурсів, технічних засобів та процесів, які підлягають захисту.
2. Аналіз загроз і вразливостей – оцінка можливих сценаріїв атаки чи випадкових інцидентів, що можуть вплинути на активи.
3. Розрахунок ризиків і вибір контрзаходів – визначення рівня ризику шляхом співставлення вартості активу, ймовірності реалізації загрози та потенційних наслідків.

CRAMM реалізує підхід, що поєднує формалізовані питання-анкети для експертів та базу знань із типовими загрозами та контрзаходами. Важливо, що методика орієнтована не лише на оцінку технічних аспектів безпеки, але й на управлінські та організаційні заходи. Таким чином, CRAMM забезпечує комплексний підхід до управління ризиками, хоча її головним недоліком залишається висока залежність від суб'єктивності експертних оцінок.

Системи управління інформаційною безпекою класу SIEM (Security Information and Event Management) з'явилися на початку 2000-х років як відповідь на потребу в централізованому зборі та аналізі подій безпеки. На відміну від CRAMM, що має аналітично-методологічний характер, SIEM орієнтовані на практичне виявлення та попередження інцидентів у режимі реального часу.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Системи SIEM є ключовим інструментом сучасного моніторингу безпеки, забезпечуючи збір, кореляцію подій і автоматизоване реагування [14]. Згідно з останніми дослідженнями, їх еволюція включає інтеграцію з SOAR та використання машинного навчання для виявлення аномалій [23]. Особливо ефективні SIEM-системи при протидії складним атакам, таким як DDoS, де необхідна швидка кореляція подій [24].

Функціонал сучасних SIEM-систем зазвичай включає:

- збір та нормалізація даних з різних джерел: журнали подій ОС, мережеві пристрої, антивіруси, системи контролю доступу тощо.
- кореляція подій з метою виявлення складних атак, які неможливо розпізнати за допомогою окремих повідомлень.
- виявлення інцидентів і формування сповіщень для аналітиків безпеки.
- аналіз та звітність, що дає змогу формувати картину загроз і відповідати вимогам регуляторів.
- інтеграція з системами реагування, включно з автоматизацією дій (SOAR — Security Orchestration, Automation and Response).

Розвиток SIEM-систем останніх поколінь включає елементи машинного навчання та поведінкової аналітики, що дає можливість виявляти аномальні дії користувачів чи внутрішні загрози.

Попри відмінності у підходах, CRAMM і SIEM можна розглядати як взаємодоповнюючі інструменти в рамках загальної системи управління ризиками:

- CRAMM відповідає за методологічний рівень: виявлення цінних активів, аналіз можливих загроз і формування політики безпеки.
- SIEM реалізує практичний рівень: моніторинг середовища, збір доказів інцидентів і їх оперативне виявлення.
- об'єднавши результати CRAMM та SIEM, організація отримує як стратегічну, так і тактичну складову управління ризиками: від оцінки потенційних загроз – до їхнього реального підтвердження у вигляді подій.

Таким чином, CRAMM можна розглядати як інструмент планування та оцінки ризиків, тоді як SIEM – як інструмент оперативного контролю та моніторингу. У комплексі вони створюють основу для більш складних моделей аналізу, зокрема із залученням багатокритеріальних методів (AHP) та ймовірнісних підходів (Bayesian Networks), що розглядатимуться у наступних розділах статті.

2.2. Метод попарних порівнянь

Аналітичний метод ієрархій (AHP), запропонований Т. Сааті [16], є одним із найпоширеніших багатокритеріальних методів прийняття рішень у сфері інформаційної безпеки. Важливим елементом є оцінка узгодженості експертних думок, для чого використовується коефіцієнт конкордації Кендалла [8]. Як показано в [10], цей підхід дозволяє кількісно оцінювати ризики, зменшуючи суб'єктивність. Для покращення узгодженості матриць попарних порівнянь можуть застосовуватися алгоритми оптимізації [11]. Останні дослідження демонструють, що комбінація AHP із байєсовими мережами дозволяє проводити кількісну оцінку ризиків навіть за умов неповних даних [25].

2.2.1. Класична модель метода Сааті. Припустимо, що перелік загроз T сформовано ($|T| = n$), тоді їх ранжування для інформаційної системи може бути виконано різними способами. Якщо експерти досягли спільної згоди щодо класифікації попарним порівнянням загроз за рівнем ризику, то складається квадратна матриця попарних порівнянь A розміром $(n \times n)$. Елементи матриці $A = (a_{ij})$ – це експертні оцінки важливості одного елемента щодо іншого. Чим більше значення $a_{ij} > 1$ (як правило, це цілі числа), тим елемент (загроза) T_i важливіший за T_j. При цьому, $a_{ji} = {}^{1}/_{a_{ij}}, a_{ii} = 1$. Часто використовується 9-бальна шкала Сааті: 1 – однакове значення; 3 – слабка перевага; 5 – сильна перевага; 7 – дуже сильна перевага; 9 – абсолютна перевага. Проміжні значення 2, 4, 6, 8 використовуються для уточнень. Побудована таким чином матриця попарних порівнянь називається класичною матрицею Сааті. Для отримання ваги відносної важливості кожної із загроз (вектора ваг), достатньо нормалізувати класичну матрицю (кожен елемент поділити на суму відповідного стовпця) і обчислити середнє значення по рядках.

2.2.2. Реалізація метода Сааті на основі матриці ранжувань. Інший підхід щодо реалізації метода попарних порівнянь починається з побудови *матриці ранжування* загроз $R = (r_{ij})$ (іншими словами, таблиці експертних оцінок). Рядкам матриці ранжування відповідають об'єкти (загрози), а

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

стовпцям – експерти (надалі, E). Припустимо, що кількість загроз $|T| = n$, а кількість експертів $|E| = m$. Тоді маємо матрицю $R(n \times m)$. Чисельність групи експертів, як правило, становить 7-8 осіб. Кожний експерт заповнює свій стовпець, оцінюючи загрози шляхом надання їм рангів. Ранжування працює так, що 1 – це ранг для найбільш небезпечної загрози; n – ранг для найменш небезпечної загрози; іншим загрозам надаються якісь проміжні (цілі) значення на розсуд експерта. Більшість методик AHP припускають, що вищий пріоритет загрози має нижчий ранг. Це відповідає стандартам оцінки ризиків.

Для визначення ступеня згоди між експертами щодо ранжування загроз, використовують коефіцієнт конкордації Кендалла [8].

Наступним кроком обчислюють матриці попарних порівнянь A^r $(r = \overline{1,m})$ по кожному експерту. Порівнюються між собою ранги матриці ранжування, окремо по кожному стовпцю, і формується квадратна матриця розміром $(n \times n)$. Тобто, кожному стовпцю відповідає матриця. Якщо загроза T_i більш небезпечна ніж T_j, то приймають $a_{ij} = 1,\ a_{ji} = 0,\ a_{ii} = 0.5,\ i,j = \overline{1,n}$. Щоб дістати групову оцінку ступеня впливу кожної з загроз на результат, будується матриця математичних сподівань X_r оцінок кожної з пар чинників. Якщо матриці попарних порівнянь A^r пораховані, то матриця математичних сподівань є сума цих матриць помножена на $1/m$. Завдання оцінювання коефіцієнтів відносної важливості зводиться до визначення максимального власного значення матриці X_r та відповідного йому власного вектора k з використанням ступеневого (ітераційного) алгоритму [9, 10].

Подальша процедура виявлення найбільш критичної загрози полягає в оцінки вразливостей. Спочатку формується перелік вразливостей і далі експерти, кожний окремо, оцінюють вразливості по тому ж самому принципу, як це було у випадку загроз. Далі також визначають ступень згоди між експертами, і, остаточно, обчислюють вектор коефіцієнтів відносної важливості вразливостей.

На підставі векторів коефіцієнтів відносної важливості загроз та коефіцієнтів відносної важливості вразливостей будують порівняльну матрицю (таблицю) загроз та вразливостей. Рядки таблиці відповідають вразливостям, стовпці – загрозам. Якщо елемент вектора загроз T є більшим за відповідний елемент вектора вразливостей W, то в таблиці на перетині стовпця загроз з рядком вразливостей ставиться знак "+", у протилежному випадку – "0". Та загроза, яка має найбільше число поєднань з вразливостями (стовпець з найбільшою кількістю знаку "+") є самою пріоритетною [10].

Отже, загальна схема оцінки ризиків методом попарних порівнянь представлена на рис. 1.

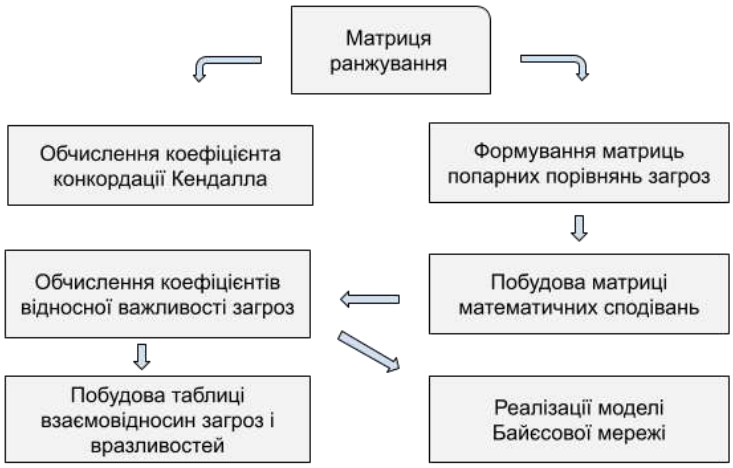


Рис. 1. Загальна схема оцінки ризиків методом попарних порівнянь

2.3. Математична модель BN

Байєсові мережі (BN) є потужним інструментом моделювання причинно-наслідкових зв'язків у системах управління ризиками [17]. Вони дозволяють враховувати невизначеність і динаміку загрозного середовища, що робить їх особливо придатними для аналізу кіберризиків [18]. Як показано в [20], гібридні

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

моделі, що поєднують AHP і BN, забезпечують високу точність прогнозування. Посилання на референсну модель управління ризиками [21] підтверджує, що BN можуть бути частиною повноцінної системи ISRM. Нещодавні дослідження також підтверджують їхню ефективність у випадках неповних даних [25].

Нехай маємо множину випадкових змінних (вузлів) $X = X_1, \dots, X_n$. Байєсова мережа (BN) – це орієнтовний ациклічний граф $G = (X, E)$, де E – множина дуг.

Кожен вузол X_i має умовний розподіл

$$P(X_i \mid Pa(X_i)), \quad (8)$$

де $Pa(X_i)$ – множина "батьків" вузла X_i у G (вузли з яких у графі виходять стрілки в X_i).

Повна факторизація для спільного розподілу наступна:

$$P(X_i, \dots, X_n) = \prod_{i=1}^{n} P(X_i)|Pa(X_i)) \quad (9)$$

На практиці явно побудувати цей розподіл досить складно, так як кількість комбінацій зростає експоненційно з числом вузлів у мережі. Це робить прямий обчислювальний підхід практично неможливим навіть для відносно невеликих моделей. Щоб уникнути цієї проблеми, у байєсових мережах застосовують спеціальні алгоритми *інференсу*, які дозволяють отримати потрібні апостеріорні ймовірності без явного формування всього розподілу. Серед таких алгоритмів можна виділити:

- точні методи: Variable Elimination, Junction Tree;
- наближені методи: Sampling (Gibbs Sampling, Likelihood Weighting), Belief Propagation.

Байєсові мережі забезпечують ефективне обчислення умовних ймовірностей, використовуючи локальні залежності між змінними та *факторизацію* глобального розподілу, або іншими словами – "структурну формулу" моделі.

У байєсовій мережі ми розділимо вузли графу так: $X = Q \cup H \cup E$, де Q – запитувані (цільові) вузли, H – приховані вузли, E – спостережені вузли (evidence). Приховані вузли (hidden variables) – це ті змінні X_i, для яких ми не маємо фактичних даних, але вони впливають на залежності між запитуваними і спостереженими змінними.

І тепер, основна задача – знайти маргінальний (умовний) розподіл цільових вузлів при наявності спостережень e:

$$P(Q \mid e) = \frac{\sum_H P(Q,H,e)}{\sum_{Q,H} P(Q,H,e)} \quad (10)$$

Тобто, інференс зводиться до *маргіналізації* – сумування по всіх можливих комбінаціях прихованих змінних H. Зробимо уточнення чисельнику формули (10):

$$\sum_H P(Q,H,e) = \sum_H \prod_{i=1}^{n} P\left(X_i \middle| Pa(X_i)\right) \quad (11)$$

Знаменник формули (10) є просто нормалізаційною константою.

Нагадаємо, що марковським покриттям (MB – Markov blanket) вузла X_i вважається мінімальна множина вузлів, яка робить X_i умовно незалежним від усіх інших у мережі. Воно складається з "батьків" вузла X_i (тобто, це $Pa(X_i)$), "дітей" вузла X_i (це $Ch(X_i)$) і "батьків дітей" вузла X_i. Таким чином,

$$MB(X_i) = Pa(X_i) \cup Ch(X_i) \cup \bigcup_{Y \in Ch(X_i)} Pa(Y) \quad (12)$$

Тим самим ми підкреслюємо те, що нам важливі значення вузлів із $MB(X_i)$, і інші вузли мережі G не несуть додаткової інформації про вузол X_i. Це є ключовим моментом в інференсі, тобто обчисленні апостеріорних ймовірностей P(гіпотеза | спостереження), коли знання про вузли оновлюються при отриманні нових даних.

Серед ключових моментів побудови байєсової мережі є матриця умовних ймовірностей (CPT – Conditional Probability Table), рядки якої відповідають можливим комбінаціям значень "батьків" $Pa(X_i)$, а стовпці – можливим значенням самого вузла X_i. Отже,

$$\Theta_i = \begin{bmatrix} P(X_i = x_{i1}|Pa(X_i) = p_{i1}) & \cdots & P(X_i = x_{im}|Pa(X_i) = p_{i1}) \\ \vdots & \ddots & \vdots \\ P(X_i = x_{i1}|Pa(X_i) = p_{ik}) & \cdots & P(X_i = x_{im}|Pa(X_i) = p_{ik}) \end{bmatrix}. \quad (13)$$

При цьому,

$$\sum_{j=1}^{m} P\left(X_i = x_{ij} \middle| Pa(X_i) = p_{ir}\right) = 1, \forall r \in \{1, \dots, k\}. \quad (14)$$

CPT задає локальні правила ймовірності для кожного вузла, в той час коли маргінальні розподіли дозволяють об'єднати всі локальні ймовірності у глобальний розподіл мережі, щоб відповісти на питання "яка ймовірність певної події/стану вузла Q, якщо ми знаємо деякі спостереження e".

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

3. Результати дослідження та експерименти

3.1. Практична реалізація методу попарних порівнянь

Обчислення коефіцієнтів відносної важливості на основі матриці ранжування можна здійснити з використанням бібліотек Python, і, веб-інтерфейсу Streamlit для зручності роботи експертів [11, 15].

Будемо вважати, що ми отримали оцінки загроз від експертів і маємо матрицю ранжувань $R = (r_{ij})$. Приклад матриці ранжувань в табличному виді представлено в табл. 1.

Таблиця 1. *Формування матриці експертних оцінок*

	E1	E2	E3	E4	E5	E6
T1	1	3	2	3	3	3
T2	2	2	4	2	4	2
T3	3	1	5	1	2	1
T4	4	6	6	5	6	4
T5	5	5	1	4	1	5

де, T – загрози, E – експерти. Нагадаємо, що 1 – це ранг для найбільш *небезпечної* загрози. Треба зауважити, що при *класичному* ранжуванні кожній загрозі надається унікальний ранг. У випадку *рангування з повторами* експерт може вказати однаковий ранг для загроз, якщо на його думку ці загрози мають однакову важливість. У цьому випадку, при розрахунку коефіцієнта Кендалла або створенні матриць попарних порівнянь це допустимо. При однакових рангах, в матриці попарних порівнянь, пара оцінюється як 0.5 (часткова перевага).

Програма на Python обчислення коефіцієнтів відносної важливості (SAAT-ваг) буде складатися з декілька модулів (функцій). Перша функція, kendall_concordance(expert_ratings), рахує коефіцієнт конкордації Кендалла (Kendall's W) для визначення ступеня згоди між експертами. Параметром функції є матриця ранжувань. Формула для обчислення Kendall's W:

$$K_{con} = {S}/{S_{max}}, \quad (1)$$

де S – дисперсія сум ранжувань (по кількості експертів), $S_{max} = \frac{n^3-n}{12}$ – максимально можлива сума квадратів відхилень (n – кількість загроз). Обчислення середніх рангів для кожної загрози (по рядкам транспонованої матриці R) реалізовано в бібліотеки Numpy (Python) так → mean_ranks = np.mean(matrix.T, axis=0). Сума квадратів відхилень середніх рангів від середнього значення рангу → S = np.sum((mean_ranks – np.mean(mean_ranks))**2).

Для перевірки значущості коефіцієнта використовується критерій χ-квадрат, який обчислюється так:

$$\chi^2 = m \cdot (n-1) \cdot K_{con}. \quad (2)$$

Значення χ^2 порівнюється з критичним значенням χ^2_{crit} при рівні значущості α і ступенях свободи $n-1$. Якщо $\chi^2 > \chi^2_{crit}$, то узгодженість оцінок експертів вважається статистично значущою. Як правило, рівень значущості α вибирають 0.05, тобто 95% довірчий рівень. Для визначення χ^2_{crit} скористаємось кодом Python, який формує таблицю критичних значень для зазначених рівнів значущості і ступенів свободи. Для цього використовується модуль chi2 бібліотеки scipy.stats. Остаточно, функція kendall_concordance залишає повідомлення зі значеннями χ^2, χ^2_{crit} та висновком про узгодженість оцінок експертів.

Наступна функція, create_pairwise_comparison_matrix(expert_ratings), формує матриці попарних порівнянь загроз. Елементи матриці визначаються визначаються за формулою (1).

$$A_r = (a^r_{ij}) = \begin{cases} 1, & \text{якщо } T^r_i < T^r_j \\ 0.5, & \text{якщо } T^r_i \approx T^r_j \\ 0, & \text{якщо } T^r_i > T^r_j \end{cases} \quad (3)$$

де r – номер експерта, T^r_i – ранг Ti -ої загрози Er -го експерта.

Функція create_pairwise_comparison_matrix реалізована за рахунок звичайних операторів циклу та умовного оператора.

Матриця математичних сподівань X обчислюється функцією Python build_expectation_matrix(pairwise_matrices). Її параметром є список квадратних матриць попарних порівнянь, pairwise_matrices. Фактично,

$$X_T = \frac{1}{m}\sum A_r, \quad (4)$$

і тому ця функція містить звичайний цикл для додавання всіх матриць A_r.

Формування вектора відносної важливості $\vec{k} = [k_1, k_2, \dots, k_n]^T$ зводиться до визначення максимального власного значення матриці XT та відповідного йому власного вектора $\vec{k}$ з використанням ступеневого (ітераційного) алгоритму.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Схема ітераційного алгоритму наступна:

1. Початкова умова t = 0, $k^{[0]} = (1, 1, \dots, 1)^T$.

2. Рекурентні співвідношення задані формулами (2):

$$\lambda^{[t]} = (1, 1, \dots, 1) \cdot Y^{[t]}, \quad k^{[t]} = \frac{1}{\lambda^{[t]}} \cdot Y^{[t]}, \qquad (5)$$

де $\lambda^{[t]} = X \cdot k^{[t-1]}$, $t = \overline{1, m}$.

3. Ознакою закінчення алгоритму є умова:

$$max\left|k^{[t]} - k^{[t-1]}\right| < E, \qquad (6)$$

де E – задана точність (наприклад, 0.001).

Функція iterative_algorithm(X, epsilon=1e-3) обчислює вектор $\vec{k}$ відповідно до алгоритму.

Варіант веб-сторінки в Streamlit з розрахунками представлено на рисунку 2.

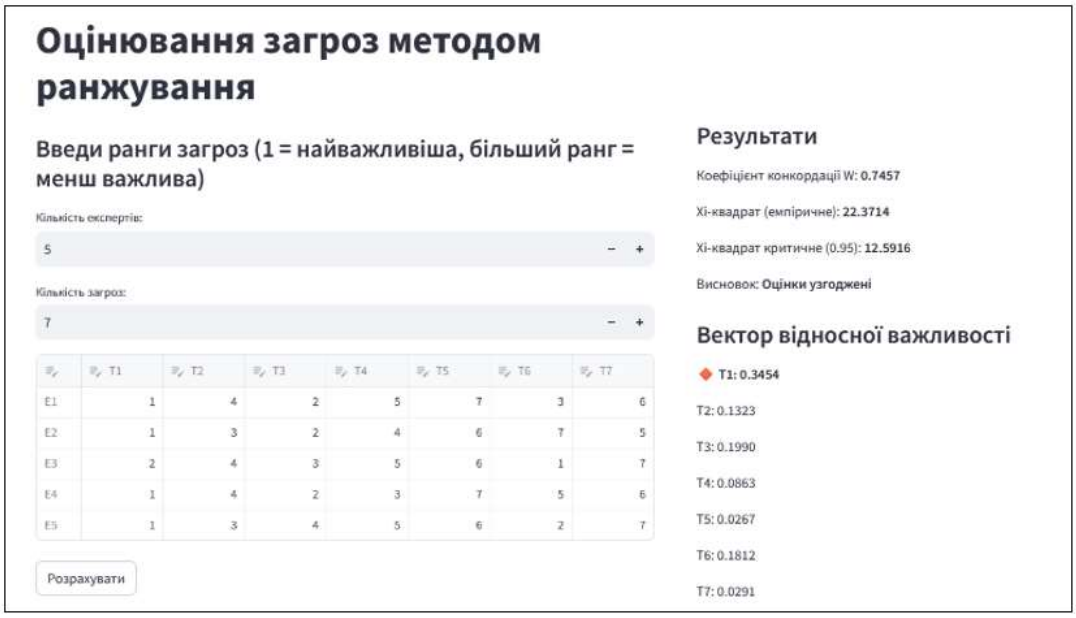


Рис. 2. Веб-інтерфейс застосунку в Streamlit для оцінювання загроз

Застосунок було протестовано на багатьох прикладах, зокрема на прикладі з 7 загрозами і 5 експертами. В результаті отримано коефіцієнт узгодженості W=0.7457, який був статистично значущим при

$$\chi^2 = 22.3714 > \chi^2_{crit} = 12.5916. \qquad (7)$$

Найбільший коефіцієнт важливості мала загроза T1, що відображено у веб-інтерфейсі.

3.2. Приклад побудови баєсової мережі

Розглянемо 4 бінарні змінні (вузли):

T – загроза атаки ({yes, no}).

V – вразливість в системі ({yes, no}).

I – інцидент (проникнення / витік даних) ({yes, no}).

L – значні фінансові втрати ({yes, no}).

Нехай маємо наступний граф G (мережа ризику кіберінциденту):

$$T \longrightarrow I \longrightarrow L, \qquad V \longrightarrow I \qquad (15)$$

Поставимо задачу оцінки ризику збитків L. Тоді граф G поділиться наступним чином:

$$Q = \{L\}, E = \{T, V\}, H = \{I\} \qquad (16)$$

Інформація про загрози T та вразливості V може бути отримана методом AHP, який розглянуто у розділі 2.

Спільна ймовірність для всіх вузлів факторизується як:

$$P(T, V, I, L) = P(T)P(V)P(I \mid T, V)P(L \mid I) \qquad (17)$$

Будуємо CPT для вузла I . Інцидент залежить від T і V, тому маємо наступну таблицю:

$P(I|T,V) =$

	$V=0$	$V=1$
$T=0$	$P(I=1\|0,0)$	$P(I=1\|0,1)$
$T=1$	$P(I=1\|1,0)$	$P(I=1\|1,1)$

При цьому, $P(I = 0 \mid T, V) = 1 - P(I = 1 \mid T, V)$).

Збитки залежать тільки від I:

$P(L|I) =$

I	$P(L=1\|I)$
0	α
1	β

Основна задача полягає в обчисленні ймовірності:

$$P(L \mid T, V) = \sum_I P(L, I \mid T, V). \qquad (18)$$

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Але, згідно попереднім міркуванням, виконаємо факторизацію:

$$P(L \mid T,V) = \sum_I P(L|I)P(I \mid T,V). \quad (19)$$

Якщо перейти до числових значень, то нехай $\alpha = 0.1, \beta = 0.8$.

Припустимо, що

$P(I = 1 \mid T = 1, V = 1) = 0.9$
$P(I = 1 \mid T = 1, V = 0) = 0.6$
$P(I = 1 \mid T = 0, V = 1) = 0.5$
$P(I = 1 \mid T = 0, V = 0) = 0.1$

Імовірності для збитків:

$P(L = 1 \mid I = 1) = 0.8$
$P(L = 1 \mid I = 0) = 0.1$

Тоді,

$$P(L = 1 \mid T = 1, V = 1) = (0.8 \cdot 0.9) + (0.1 \cdot 0.1) = 0.72 + 0.01 = 0.73.$$
$$P(L = 1 \mid T = 1, V = 0) = (0.8 \cdot 0.6) + (0.1 \cdot 0.4) = 0.48 + 0.04 = 0.52.$$
$$P(L = 1 \mid T = 0, V = 1) = (0.8 \cdot 0.5) + (0.1 \cdot 0.5) = 0.4 + 0.05 = 0.45.$$
$$P(L = 1 \mid T = 0, V = 0) = (0.8 \cdot 0.1) + (0.1 \cdot 0.9) = 0.08 + 0.09 = 0.17.$$

Тепер можемо записати вже CPT для L після маргіналізації прихованого вузла *I*:

$P(L = 1|T, V)$ =

	$V = 0$	$V = 1$
$T = 0$	0.17	0.45
$T = 1$	0.52	0.73

Ми отримали готову таблиця умовних імовірностей, яку можна використовувати для оцінки ризику збитків L.

Наведений приклад побудови BN «T → I → L ← V» відповідає сучасним практикам моделювання кіберризиків [17]. Подібні структури використовуються в гібридних фреймворках для оцінки фінансових наслідків інцидентів [20]. Використання даних AHP для заповнення CPT підтверджується дослідженнями, що демонструють можливість інтеграції експертних оцінок у ймовірнісні моделі [25].

3.3 Реалізація побудови баєсової мережі на Python

Для реалізації BN використана бібліотека pgmpy — одна з найпоширеніших у науковому співтоваристві для роботи з ймовірнісними графічними моделями [15]. Її функціонал дозволяє будувати мережі, визначати CPT та виконувати інференс за допомогою алгоритмів, таких як Variable Elimination [17]. Подібні підходи успішно застосовуються в задачах оцінки ризиків [20].

Загальна структура коду наступна:

1. Створення моделі: model = BayesianNetwork([('T', 'I'), ('V', 'I'), ('I', 'L')]);
2. Вказуємо значення CPT для вузлів I та L, використовуючи клас TabularCPD;
3. Додаємо CPT до моделі: model.add_cpds(cpd_T, cpd_V, cpd_I, cpd_L);
4. Перевіряємо модель: ok = model.check_model();
5. Готуємо ймовірнісний висновок: infer = VariableElimination(model);
6. Виводимо результати:

```
q1 = infer.query(variables=['L'], evidence={'T': 1, 'V': 1})
print("\nP(L | T=1, V=1):\n", q1)
q2 = infer.query(variables=['L'], evidence={'T': 0, 'V': 1})
print("\nP(L | T=0, V=1):\n", q2)
```

Висновки й перспективи подальшого дослідження

У статті розглянуто комплексний підхід до управління ризиками інформаційної безпеки, заснований на інтеграції методологічного планування, експертних оцінок, кількісного моделювання та технічного моніторингу. Показано, що ефективна система кібербезпеки повинна поєднувати стратегічні та тактичні інструменти, забезпечуючи як довгострокове планування, так і оперативне реагування на загрози.

Методологія CRAMM виявилася ефективним інструментом для структурованого аналізу активів, загроз і вразливостей, а також для формування політики безпеки. Однак її суб'єктивність потребує доповнення об'єктивними кількісними методами. Системи класу SIEM, навпаки, забезпечують реальний час моніторингу, кореляцію подій і швидке

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

виявлення інцидентів, але не вирішують завдання пріоритезації ризиків на стратегічному рівні. Саме тому їх слід розглядати не як конкурентні, а як взаємодоповнюючі компоненти єдиної системи управління ризиками.

Для усунення суб'єктивності експертних оцінок запропоновано використання аналітичного методу ієрархій (AHP) за Т. Сааті, реалізованого на основі матриці ранжувань. Цей підхід дозволяє отримати вектор ваг (SAAT-ваг), що кількісно відображає пріоритетність загроз і вразливостей. Особливу увагу приділено вимірюванню узгодженості експертних думок за допомогою коефіцієнта конкордації Кендалла, що підвищує надійність результатів. Реалізація методу на Python із використанням бібліотек Numpy, SciPy та веб-інтерфейсу Streamlit робить його доступним і практично застосовним для організацій будь-якого рівня.

Найбільш інноваційним елементом дослідження є інтеграція отриманих кількісних оцінок у байєсові мережі (BN). Байєсові мережі дозволяють моделювати причинно-наслідкові зв'язки між загрозами, вразливостями, інцидентами та фінансовими збитками, забезпечуючи гнучкий імовірнісний інференс. На прикладі побудови мережі «T → I → L ← V» показано, як на основі даних AHP можуть бути сформовані таблиці умовних ймовірностей (CPT) і отримані точні оцінки ризику через маргіналізацію прихованих змінних. Використання бібліотеки pgmpy для Python дозволяє автоматизувати процес моделювання та ймовірнісного висновування.

Запропонований гібридний підхід відповідає сучасним тенденціям інтеграції стратегічних та тактичних інструментів управління ризиками [19]. Поєднання CRAMM, AHP, BN та SIEM дозволяє створити адаптивну систему, яка враховує як організаційні, так і технічні аспекти безпеки [23]. Перспективним напрямом є подальше розширення моделі шляхом інтеграції з SOAR-системами та використанням методів машинного навчання для динамічного оновлення параметрів BN [20].

Список літератури

1. Сидоркін П. Г., Горліченко С. О., Некоз В. С., Шилан М. В. Методи управління ризиками інформаційної безпеки CRAMM та COBIT 5 FOR RISK. *Сучасні інформаційні технології у сфері безпеки та оборони.* 2023. № 2 (47). С. 41 – 47. DOI: https://doi.org/10.33099/2311-7249/2023-47-2-41-47
2. Берко А. Ю., Висоцька В. А., Рішняк І. В. Методи та засоби оцінювання ризиків безпеки інформації в системах електронної комерції. Інформаційні системи та мережі: [збірник наукових праць]: Видавництво Львівської політехніки. 2008. № 610 (1). С. 20 – 33. URL: https://science.lpnu.ua/sites/default/files/journal-paper/2019/apr/16336/vis610inform-syst-20-33.pdf
3. Залива В. В. Методики захисту API за допомогою JavaScript: математичні моделі для підвищення безпеки. *Телекомунікаційні та інформаційні технології.* 2024. № 3 (84). С. 4 – 11. DOI: 10.31673/2412-4338.2024.030411
4. Карпович І.М., Гладка О.М., Наконечна Ю.А. Аналіз ризиків безпеки інформаційної системи IT-підприємства. *Вчені записки ТНУ імені В.І. Вернадського.* 2020. Том 31 (70) № 5 С. 9-74. DOI https://doi.org/10.32838/2663-5941/2020.5/12
5. A multicriterial analysis of the efficiency of conservative information security systems / Dudykevych V., Prokopyshyn I., Chekurin V., Opirskyy I., Lakh Yu., Kret T., Ivanchenko Ye., Ivanchenko I., *Eastern-European Journal of Enterprise Technologies.* 2019.Vol. 3, Issue 9 (99). P. 6–13. DOI: https://doi.org/10.15587/1729-4061.2019.166349
6. Gaute Wangen, Christoffer Hallstensen, Einar Snekkenes. A framework for estimating information security risk assessment method completeness. *Int. J. Inf. Secur.* 2018. Vol. 17. P. 681 – 699. DOI: https://doi.org/10.1007/s10207-017-0382-0
7. Барченко Н. Л., Любчак В. О., Лаврик Т. В. Модель індикаторів оцінки національного рівня цифровізації та кібербезпеки держав світу. *КІБЕРБЕЗПЕКА: освіта, наука, техніка.* 2022. № 2(18). С. 73 – 85.
8. Amanda Gearharta, D. Terrance Booth, Kevin Sedivec and Christopher Schauer. Use of Kendall's coefficient of concordance to assess agreement among observers of very high resolution imagery. Geocarto International 2013. Vol. 28, No. 6, P. 517–526. DOI: 10.1080/10106049.2012.725775
9. Бурячок В. Л., Толубко В. Б., Хорошко В. О., Толюпа С. В. Інформаційна та кібербезпека: соціотехнічний аспект: за заг. ред. д-ра техн. наук, професора В. Б. Толубка. Київ: ДУТ. 2015. 288 с.
10. Дзюба Л. Ф., Чмир О. Ю. Оцінювання ризиків інформаційної безпеки з використанням методів математичної статистики. Львівський державний університет безпеки життєдіяльності. Вісник ЛДУБЖД. 2022. №26. С. 47–54. URL: http://www.irbis-nbuv.gov.ua/cgi-

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILA=&2_S21STR=Vldubzh_2022_26_8

11. Олецький О. В. Підвищення узгодженості матриць попарних порівнянь у методі аналізу ієрархій на основі розв'язків систем лінійних алгебраїчних рівнянь. Наукові записки НаУКМА. Комп'ютерні науки. 2022. Том 5. 2022. С. 85-91. DOI: 10.18523/2617-3808.2022.5.85-91
12. Wilson, Simon and De Persis, Cristina and Bosque, José Luis and Huertas, Irene and Sillero Denamiel, Maria Remedios, Quantitative System Risk Assessment from Incomplete Data with Belief Networks and Pairwise Comparison Elicitation. URL: https://ssrn.com/abstract=4577878 (дата звернення 01.07.2025)
13. CRAMM Version 5.1 User Guide. URL: https://pdfcoffee.com/cramm-version-51-user-guide-pdf-free.html (дата звернення 01.07.2025
14. Смірнова Т. В., Константинова Л. В., Конопліцька-Слободенюк О. К., Козлов Я. О., Кравчук О. В., Козірова Н. Л., Смірнов О. А. Дослідження сучасного стану SIEM-систем. "Кібербезпека: освіта, наука, техніка" No 1(25), P. 6-18. 2024 DOI: https://doi.org/10.28925/2663-4023.2024.25.618
15. Ankur Ankan, Abinash Panda. pgmpy: Probabilistic Graphical Models using Python. 2015. URL: https://www.researchgate.net/publication/328778465_pgmpy_Probabilistic_Graphical_Models_using_Python (дата звернення 01.07.2025)
16. Saaty T. L. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. *McGraw-Hill*. 1980. URL: https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1895817
17. Fenton N., Neil M. Risk Assessment and Decision Analysis with Bayesian Networks. *CRC Press*. 2012. 4 p. DOI: https://doi.org/10.1080/15598608.2014.847770
18. Khakzad N., Khan F., Amyotte P. Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliability Engineering & System Safety*, №111, 2013. P. 81–92. DOI: https://doi.org/10.1016/j.ress.2012.10.015
19. Borg A., Feldt R., Hansson K. Cyber security risk assessments: Systematic development of a risk assessment process using the ISO. *IEC 27005 standard. Computers & Security*, № 47, P. 128–143. 2014. DOI: https://doi.org/10.1016/j.cose.2014.07.003
20. Sharma S., Singh S., Sharma A. A hybrid framework for cyber risk assessment using fuzzy AHP and Bayesian networks. *Journal of Information Security and Applications*, № 52, 102492 p. 2020. DOI: https://doi.org/10.1016/j.jisa.2020.102492
21. Cherdantseva Y., Hilton J. A reference model of information security risk management. *Proceedings of the 2013 9th International Conference on Availability, Reliability and Security (ARES)*, P. 546–555. 2013. DOI: https://doi.org/10.1109/ARES.2013.72
22. Onwuegbuzie A. J., Collins K. M. T., Jiao, Q. G. The role of theory in advanced mixed research designs. *International Journal of Multiple Research Approaches*, № 4(1), P. 8–22. 2010. DOI: https://doi.org/10.5172/mra.4.1.8
23. Akhgar B., Chalkidis G., Hessami A. G. Cybersecurity Big Data Analytics: Governance and strategic decision making in cyberspace. *Springer*. 2020. DOI: https://doi.org/10.1007/978-3-030-43541-7
24. Zargar S. T., Joshi J., Tipper D. A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE Communications Surveys & Tutorials*, 15(4), P. 2046–2069. 2013. DOI: https://doi.org/10.1109/SURV.2013.031413.00127
25. Wilson S. P., De Persis C. Quantitative system risk assessment from incomplete data with belief networks and pairwise comparison elicitation. *Risk Analysis*, 42(8), P. 1683–1702. 2022. DOI: https://doi.org/10.1111/risa.13878

References

1. Sydorkin, P. H., Horlichenko, S. O., Nekoz, V. S., Shylan, M. V. (2023), "Methods of information security risk management CRAMM and COBIT 5 FOR RISK". ["Metody upravlinnia ryzykamy informatsiinoi bezpeky CRAMM ta COBIT 5 FOR RISK"]. *Modern Information Technologies in the Sphere of Security and Defense*, № 2(47), P. 41–47. DOI: https://doi.org/10.33099/2311-7249/2023-47-2-41-47
2. Berko, A. Yu., Vysotska, V. A., Rishniak, I. V. (2008), "Methods and means of assessing information security risks in e-commerce systems". [Metody ta zasoby otsiniuvannia ryzykiv bezpeky informatsii v systemakh elektronnoi komertsii]. *Information Systems and Networks: [collection of scientific papers]: Lviv Polytechnic Publishing House*, 610(1), P. 20–33. available at: https://science.lpnu.ua/sites/default/files/journal-paper/2019/apr/16336/vis610inform-syst-20-33.pdf

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

3. Zalyva, V. V. (2024), "API protection methods using JavaScript: mathematical models for enhancing security". ["Metodyky zakhystu API za dopomohoiu JavaScript: matematychni modeli dlia pidvyshchennia bezpeky"], *Telecommunication and Information Technologies*, № 3(84), P. 4–11. DOI: 10.31673/2412-4338.2024.030411
4. Karpovych, I. M., Hladka, O. M., Nakonechna, Yu. A. (2020), "Analysis of information system security risks of an IT enterprise". ["Analiz ryzykiv bezpeky informatsiinoi systemy IT-pidpryiemstva"]. *Scientific Notes of V.I. Vernadsky TNU*, № 31(70)(5), P. 9–74. DOI https://doi.org/10.32838/2663-5941/2020.5/12
5. Dudykevych, V., Prokopyshyn, I., Chekurin, V., Opirskyy, I., Lakh, Yu., Kret, T., Ivanchenko, Ye., Ivanchenko, I. (2019), "A multicriterial analysis of the efficiency of conservative information security systems". *Eastern-European Journal of Enterprise Technologies*, № 3(9)(99), P. 6–13. DOI: https://doi.org/10.15587/1729-4061.2019.166349
6. Wangen, G., Hallstensen, C., Snekkenes, E. (2018), "A framework for estimating information security risk assessment method completeness". *International Journal of Information Security*, № 17, P. 681–699. DOI: https://doi.org/10.1007/s10207-017-0382-0
7. Barchenko, N. L., Liubchak, V. O., Lavryk, T. V. (2022), "Model of indicators for assessing the national level of digitalization and cybersecurity of world states". ["Model indykatoriv otsinky natsionalnoho rivnia tsyfrovizatsii ta kiberbezpeky derzhav svitu"] *Cybersecurity: education, science, technology*, 2(18), P. 73–85.
8. Gearharta, A., Booth, D. T., Sedivec, K., Schauer, C. (2013), "Use of Kendall's coefficient of concordance to assess agreement among observers of very high resolution imagery". *Geocarto International*, № 28(6), P. 517–526. DOI: 10.1080/10106049.2012.725775
9. Buriachok, V. L., Tolubko, V. B., Khoroshko, V. O., Toliupa, S. V. (2015), "Information and cybersecurity: sociotechnical aspect". [Informatsiina ta kiberbezpeka: sotsiotekhnichnyi aspekt]. *Kyiv: DUT*. 288 p.
10. Dziuba, L. F., Chmyr, O. Yu. (2022), "Assessment of information security risks using methods of mathematical statistics". ["Otsiniuvannia ryzykiv informatsiinoi bezpeky z vykorystanniam metodiv matematychnoi statystyky"]. *Bulletin of Lviv State University of Life Safety*, № 26, P. 47–54. available at: http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILA=&2_S21STR=Vldubzh_2022_26_8
11. Olietskyi, O. V. (2022), "Improving the Consistency of Pairwise Comparison Matrices in the Analytic Hierarchy Process Based on Solutions of Systems of Linear Algebraic Equations". *Scientific Notes of NaUKMA. Computer Sciences*. Vol. 5. P. 85–91. DOI: 10.18523/2617-3808.2022.5.85-91
12. Wilson, Simon, De Persis, Cristina, Bosque, José Luis, Huertas, Irene, Sillero, Denamiel, Maria, Remedios "Quantitative System Risk Assessment from Incomplete Data with Belief Networks and Pairwise Comparison Elicitation". available at: https://ssrn.com/abstract=4577878 (last accessed 01.07.2025)
13. "CRAMM Version 5.1 User Guide". available at: https://pdfcoffee.com/cramm-version-51-user-guide-pdf-free.html (last accessed 01.07.2025
14. Smirnova, T. V., Konstantinova, L. V., Konoplitska-Slobodeniuk, O. K., Kozlov, Y. O., Kravchuk, O. V., Kozirova, N. L., Smirnov, O. A. (2024), "Research on the Current State of SIEM Systems". *Cybersecurity: Education, Science, Technology*, *1*(25). P. 6-18. DOI: https://doi.org/10.28925/2663-4023.2024.25.618
15. "Ankur Ankan, Abinash Panda. pgmpy: Probabilistic Graphical Models using Python". 2015. available at: https://www.researchgate.net/publication/328778465_pgmpy_Probabilistic_Graphical_Models_using_Python (last accessed 01.07.2025)
16. Saaty, T. L. (1980), "The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation". *McGraw-Hill*. available at: https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1895817
17. Fenton, N., Neil, M. (2012), "Risk Assessment and Decision Analysis with Bayesian Networks". *CRC Press*. 4 p. DOI: https://doi.org/10.1080/15598608.2014.847770
18. Khakzad, N., Khan, F., Amyotte, P. (2013), "Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches". *Reliability Engineering & System Safety*, № 111, P. 81–92. DOI: https://doi.org/10.1016/j.ress.2012.10.015
19. Borg, A., Feldt, R., & Hansson, K. (2014), "Cyber security risk assessments: Systematic development of a risk assessment process using the ISO". *IEC 27005 standard. Computers & Security*, № 47, P. 128–143. DOI: https://doi.org/10.1016/j.cose.2014.07.003

20. Sharma, S., Singh, S., & Sharma, A. (2020), "A hybrid framework for cyber risk assessment using fuzzy AHP and Bayesian networks". *Journal of Information Security and Applications*, № 52, 102492 p. DOI: https://doi.org/10.1016/j.jisa.2020.102492
21. Cherdantseva, Y., Hilton, J. (2013), "A reference model of information security risk management". *Proceedings of the 2013 9th International Conference on Availability, Reliability and Security (ARES)*, P. 546–555. DOI: https://doi.org/10.1109/ARES.2013.72
22. Onwuegbuzie, A. J., Collins, K. M. T., Jiao, Q. G. (2010), "The role of theory in advanced mixed research designs". *International Journal of Multiple Research Approaches*, № 4(1), P. 8–22. DOI: https://doi.org/10.5172/mra.4.1.8
23. Akhgar, B., Chalkidis, G., Hessami, A. G. (2020), "Cybersecurity Big Data Analytics: Governance and strategic decision making in cyberspace". *Springer*. DOI: https://doi.org/10.1007/978-3-030-43541-7
24. Zargar, S. T., Joshi, J., & Tipper, D. (2013), A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE Communications Surveys & Tutorials*, № 15(4), P. 2046–2069. DOI: https://doi.org/10.1109/SURV.2013.031413.00127
25. Wilson, S. P., De Persis, C. (2022), "Quantitative system risk assessment from incomplete data with belief networks and pairwise comparison elicitation". *Risk Analysis*, № 42(8), P. 1683–1702. DOI: https://doi.org/10.1111/risa.13878

Надійшла (Received) 05.04.2025

Відомості про авторів / About the Authors

Тімошин Анатолій Сергійович, кандидат фізико-математичних наук, доцент, Харківський національний університет внутрішніх справ, доцент кафедри інформаційних систем та технологій, e-mail: timxxvii@gmail.com, ORCID ID: https://orcid.org/0009-0005-6916-8252

Калєніченко Лідія Іванівна, доктор юридичних наук, професор, Харківський національний університет внутрішніх справ, завідувач кафедри інформаційних систем та технологій ННІ №4, e-mail: Kalenichenkolida@gmail.com, ORCID ID: https://orcid.org/0000-0003-4068-4729,

Гнусов Юрій Валерійович, кандидат технічних наук, доцент, Харківський національний університет внутрішніх справ, завідувач кафедри кібербезпеки та DATA-технологій, e-mail: duke6969@i.ua, ORCID ID: https://orcid.org/0000-0002-9017-9635

Хавіна Інна Петрівна, кандидат технічних наук, доцент, Харківський національний університет внутрішніх справ, доцент кафедри кібербезпеки та DATA-технологій, e-mail: inna.khavina25@gmail.com, ORCID ID: https://orcid.org/0000-0002-1856-1186

Цуранов Михайло Віталійович, Харківський національний університет внутрішніх справ, старший викладач кафедри кібербезпеки та DATA-технологій, e-mail: ukrbear2006@gmail.com, ORCID ID: https://orcid.org/0000-0002-2115-7029

Довгань Ірина Артурівна, Харківський національний університет внутрішніх справ, викладач кафедри інформаційних систем та технологій, e-mail: dovganirisha@gmail.com, ORCID ID: https://orcid.org/ 0009-0001-0440-9810

Timoshyn Anatolii, Associate Professor, Kharkiv National University of Internal Affairs, Associate Professor of the Department of Information Systems and Technologies.

Kalienichenko Lidia, Doctor of Law, Professor, Kharkiv National University of Internal Affairs, Department of Information Systems and Technologies (Head of Department).

Gnusov Yurii, Candidate of technical science, Associate Professor, Kharkiv National University of Internal Affairs, Head of the Department of Cybersecurity and DATA Technologies

Khavina Inna, Candidate of technical science, Associate Professor, Kharkiv National University of Internal Affairs, Associate Professor of the Department of Cybersecurity and DATA Technologies

Tsuranov Mykhailo, Kharkiv National University of Internal Affairs, Senior Lecturer, Department of Cybersecurity and DATA Technologies

Dovhan Iryna, Kharkiv National University of Internal Affairs, Teacher of the Department of Information Systems and Technologies.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

INTEGRATED INFORMATION SECURITY RISK MANAGEMENT MODEL BASED ON AHP AND BAYESIAN NETWORKS

The subject of the study is information security risk management in a modern digital environment, where the integration of strategic and tactical approaches is necessary to ensure adaptive protection. The purpose of the work is to develop a hybrid model of cyber risk management by combining methodological analysis, expert assessments, probabilistic modeling and technical monitoring. The objectives of the study are: (1) analysis of the complementarity of the CRAMM methodology and SIEM systems; (2) construction of a procedure for quantitative prioritization of threats and vulnerabilities based on the analytical hierarchy process (AHP); (3) integration of the obtained estimates into Bayesian networks (BN) for probabilistic risk forecasting; (4) implementation of the proposed approach using modern automation tools. The methods used in the work include: CRAMM methodology for identifying assets, threats and vulnerabilities; Thomas Saati's AHP for quantitative assessment of priorities based on expert judgments with measurement of consistency using the Kendall concordance coefficient; mathematical modeling of causal relationships using Bayesian networks (BN); and the use of SIEM-class systems for operational monitoring of security events. The practical implementation of the approach was carried out using Python, in particular the Numpy, SciPy, pgmpy libraries, and the Streamlit web interface. Results. An integrated approach was developed that combines CRAMM, AHP, BN, and SIEM into a single adaptive risk management system. It is shown that AHP allows you to transform subjective expert assessments into objective weighting factors, which increases the reliability of the analysis. Based on these data, a Bayesian network was built to assess the risk of financial losses, which takes into account the presence of a threat, vulnerability, and a possible incident. The model is implemented programmatically, demonstrating the process of factoring the joint distribution and marginalizing latent variables to obtain posterior probabilities. The web interface based on Streamlit ensures the ease of use of the tool by non-professional users. Conclusions. The proposed hybrid approach allows for the effective combination of strategic planning (CRAMM), expert assessments (AHP), probabilistic modeling (BN) and operational monitoring (SIEM), forming a proactive, scientifically sound risk management system. Such integration provides a high level of adaptability and accuracy in a dynamic threat landscape, which makes the model practically applicable for organizations of various levels.

Keywords: CRAMM methodology, SIEM systems, Analytic Hierarchy Process (AHP), Bayesian Networks (BN), Expert evaluation, Threat and vulnerability analysis.

Бібліографічні описи / Bibliographic descriptions

Тімошин А.С., Калєніченко Л.І., Гнусов Ю.В., Хавіна І.П., Цуранов М.В., Довгань І.А. Інтегрована модель управління ризиками інформаційної безпеки на основі ahp та байєсових мереж. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 166–179. DOI: https://doi.org/10.30837/2522-9818.2025.3.166

Timoshyn, A., Kalienichenko, L., Gnusov, Y., Khavina, I., Tsuranov, M., Dovhan, I. (2025), "Integrated information security risk management model based on ahp and bayesian networks", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 166–179. DOI: https://doi.org/10.30837/2522-9818.2025.3.166

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

UDC 004.77/.72 DOI: https://doi.org/10.30837/2522-9818.2025.3.180

I. CHEPURNA, D. FROLOV

A METHOD FOR INCREASING THE PRODUCTIVITY OF A DISTRIBUTED FIREWALL BASED ON PROXMOX IN CORPORATE COMPUTER NETWORKS

The subject of the study in the article is a method for increasing the performance of a distributed firewall based on LXC containers of the Proxmox VE environment for corporate computer networks. The goal of the work is to develop approaches to ensure a high level of efficiency of a distributed firewall for monitoring and managing traffic in corporate networks and virtualized networks, enabling the minimization of delays during traffic filtering and ensuring reliable operation of the corporate network under conditions of limited hardware resources. To solve the problem, the following research methods were applied: theoretical analysis of literature sources; analysis of the features of the application of containerization technology for implementing dynamic network traffic control, study of methods to improve computational resource utilization efficiency in environments with limited hardware resources, analysis of the advantages of distributed firewall regarding minimizing data transmission delays, increasing system throughput, and reducing unauthorized access risks; experimental validation of the functionality and efficiency of the distributed firewall. The results obtained indicate that the proposed method allows minimizing delays during traffic filtering and provides automatic scaling of the firewall`s functionality while maintaining the integrity of the network security system. The proposed approach provides a high level of CCM protection by segmenting the network with the assignment of a separate LXC container to serve each local subnet, which allows for targeted traffic filtering and flexible access policy management. Conclusions: the paper proposes a configuration of a distributed firewall in the Proxmox environment, including setting up a basic set of filtering rules to ensure the effective operation of a corporate computer network. The scientific novelty of the method lies in improvement of security mechanisms in scalable environments with limited hardware resources, enabling a high level of protection against external and internal threats, while maintaining fault tolerance and reliability of the network infrastructure. Experimental validation of the method's functionality and efficiency confirmed the feasibility of its implementation to ensure stable and controlled access to the corporate computer network's resources.

Keywords: method; distributed firewall; container; Proxmox; virtualization; delay; traffic filtering.

Introduction

Under the conditions of rapid development of information technologies and growth of network infrastructure, the issue of ensuring security in corporate computer networks becomes particularly relevant [1]. Modern network infrastructure is becoming increasingly scalable, which increases the risk of unauthorized access to nodes and services that require effective and continuous protection from internal and external threats.

A distributed firewall provides network traffic control, detection, and blocking of potential threats, making it an effective tool for protecting distributed infrastructure. However, the complexity and number of filtering rules directly affect network performance, security system response time, and service stability.

The use of virtualization technologies enables the creation of a flexible and scalable network infrastructure. In particular, the use of the Proxmox platform with containerization support helps to implement a managed distributed protection system, which allows this approach to be applied in conditions of limited hardware resources and high performance and security requirements.

At the same time, the growth in network load, incoming traffic volumes, threat complexity, and the number of filtering rules significantly reduces the effectiveness of traditional approaches to IT infrastructure protection [2]. This can lead to delays in information transfer and a decrease in overall system performance. Therefore, the search for practical solutions that ensure an adequate level of protection system performance in real time is a pressing task.

The use of effective optimization methods and the implementation of a distributed firewall help to achieve a balance between the level of protection, traffic processing speed, and rational use of server environment resources. Thus, improving firewall mechanisms is a key step toward increasing the reliability and stability of server infrastructure in the context of growing demands for corporate computer network protection.

Problem statement

The main directions of development the modern information society are formed in the context of dynamic IT development, which is a key factor in digital transformation and the basis for the formation of effective

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

cybersecurity systems. At the same time, corporate networks are facing the problem of creating a scalable infrastructure with a high level of information security in the case of limited computing resources [3]. IT ecosystems in the public and commercial sectors increasingly require reliable protection against internal and external threats, as well as support for high performance, given the complexity of network architecture and rising hardware costs.

The deployment and use of firewalls in modern IT environments is accompanied by a number of problems that make it difficult to ensure an adequate level of cybersecurity. In distributed environments, security policy management must be centralized, flexible, and adaptive, but the increasing complexity of corporate networks and the use of network virtualization technologies complicate traditional methods of manual security rule configuration, which can lead to errors, vulnerabilities, and lengthy configuration times [4, 5].

The growing number of IoT devices, particularly in healthcare networks, increases the risk of cyberattacks as these networks become more open and vulnerable [6]. This increases the load on traditional firewalls, which may not be able to effectively control traffic at all levels of the architecture. In addition, the use of packet filtering firewalls leads to exponential growth in the size of policy rules, which increases the likelihood of anomalies in their execution and makes it difficult to maintain a continuous level of security [7].

The economic aspects of cybersecurity also play a key role in the implementation of firewalls, as the costs of their proper configuration, management, and updating are constantly increasing. For example, financial losses from cyberattacks in the UK increased by 31% in one year, highlighting the critical need for effective protection mechanisms [8]. At the same time, environmental aspects are becoming increasingly relevant, in particular the increased energy consumption caused by the high computational complexity of manual firewall policy administration, which is particularly relevant for distributed systems [9].

Therefore, the implementation of firewalls in corporate networks requires a comprehensive approach that takes into account a number of important factors: determining the nature of potential threats, requirements for network infrastructure architecture, integration of centralized security policy management mechanisms, and ensuring the necessary level of automation for deploying and configuring firewall settings. Achieving an adequate level of corporate network protection using firewalls is possible with the implementation of automated and centralized solutions that guarantee high performance and ensure network resilience to external and internal threats.

Analysis of recent studies and publications

Current research in the field of computer networks, particularly virtualized ones, is largely focused on analyzing aspects of firewall deployment and their role in ensuring the security of such environments. Particular attention is paid to the effectiveness of network traffic filtering, security policy optimization, and research into the impact of firewalls on the performance of high-speed networks. The solutions discussed in the scientific literature draw attention to the need to increase the performance of firewalls, which can be achieved by automating the configuration of security rules, minimizing delays in the processing of network traffic, and implementing methods for dynamically scaling the network infrastructure. The importance of adapting protection mechanisms to traffic changes and increasing the number of devices in virtualized environments is emphasized separately.

According to the authors of [10], traditional firewalls play a key role in ensuring the security of modern networks, but their configuration is still mostly done manually. This creates a risk of errors, which can lead to security breaches and significant time spent on reconfiguration. In virtualization-based networks, this problem becomes even more relevant due to the increasing complexity and dynamism of the infrastructure, which makes manual administration of security rules difficult.

In addition, traditional firewalls are typically implemented as separate dedicated devices that sequentially apply security policies to each incoming packet. The authors of [11] argue that processing filtering rules can create a significantly greater load on the system than normal routing. This problem becomes particularly relevant in the context of increasing network traffic speeds, which requires constant optimization of packet filtering mechanisms. Effective configuration of security policies is critical to ensuring the performance and security of high-speed networks, especially in the context of dynamically changing threats.

In study [12], the authors also emphasize that firewalls can become a bottleneck for network performance and a primary target for attacks, particularly

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

denial-of-service attacks. The firewall filtering mechanism is based on checking each incoming packet against a defined set of rules to decide whether to block or forward it. The performance of such a check depends on the average number of rules that need to be processed before a decision is made. When the firewall is under heavy load, such as during peak traffic periods or DoS attacks, processing delays can increase, leading to packet loss and network performance degradation.

The introduction of virtualization and containerization technologies [13] is an effective approach to solving the problem of network infrastructure scalability. Recent research in the field of virtualization and containerization is aimed at improving the efficiency and reliability of distributed firewalls in containerized environments, such as Proxmox, in order to ensure secure access to corporate networks [14]. *Proxmox*, in particular, offers built-in tools for managing network resources, including server clustering and live migration of virtual machines [15]. This allows network infrastructure to be scaled without the need for additional investment in new equipment, which is important for small and medium-sized businesses. However, issues related to traffic processing performance and optimal resource utilization remain relevant and require new approaches.

The paper [16] analyzes the performance of architectures using one and several virtual servers on the *Proxmox VE* platform. The results of the study show that the architecture with multiple virtual servers demonstrates higher availability, specifically 80.25% with 100 concurrent users, compared to the single-server architecture, where availability is 78.4% with 80 users. The results confirm the feasibility of using scalable containerization-based solutions for the effective deployment of distributed firewalls.

Thus, an analysis of recent studies and publications [10–16] shows that scientific developments in the field of virtualized computer networks are aimed at improving security mechanisms, particularly in the context of distributed firewalls. The main problems associated with the deployment and use of firewalls are aimed at improving traffic filtering efficiency, automating security rule settings, and adapting to changing network conditions. Traditional firewalls, while playing a key role in network protection, have limitations in terms of traffic processing speed and configuration flexibility, which can lead to delays and increased system load. Virtualization and containerization technologies, such as the *Proxmox VE* platform, offer effective solutions for dynamically scaling network infrastructure and optimizing resource utilization. They contribute to increased fault tolerance, network management flexibility, and security in corporate environments. However, the issue of performance in network traffic processing remains relevant and requires further research and improvement of automated optimization methods.

Main research material

Algorithmic provision

The proposed method for improving the performance of a distributed firewall, aimed at ensuring effective monitoring and management of network traffic in corporate and virtualized networks, is based on the integration of modern virtualization technologies with mechanisms for dynamic distribution of network resources. The scientific novelty of the approach lies in the use of virtualization to build a multi-level network infrastructure with flexible segmentation and controlled access to resources, which provides an increased level of security. The implementation of continuous traffic monitoring, support for adaptive data flow management, and the ability to apply differential access policies contribute to the growth of overall network performance and ensure an adequate level of security in accordance with the requirements of modern corporate environments.

The method involves a series of sequential steps, each of which is aimed at implementing a structured and scalable approach to building a distributed traffic filtering system. Let's take a look at these steps.

1. Analyze the existing infrastructure for logical network segmentation based on the functional purpose of the nodes and the criticality of the data. This allows you to optimally determine the required initial number and placement of distributed firewall containers.
2. Deploy containers and configure routing to ensure proper interaction between segments and with the external network.
3. Configure traffic filtering policies to form a set of filtering rules, including restricting access to critical resources, allowing specific protocols and ports, and detecting anomalies.
4. Testing the distributed firewall with the initial configuration to analyze the status of the distributed firewall, delays, and evaluate resource usage for further optimization or configuration adjustments to ensure the required level of throughput and system stability.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

The proposed method is appropriate for use in the following conditions:

- the need for centralized security coordination and unification of access policies at all levels of the corporate network;
- increased requirements for the protection of critical corporate network resources to ensure their integrity and confidentiality;
- limited hardware resources, requiring optimization of the use of available infrastructure in view of reliability and efficiency requirements.

Thus, the implementation of the proposed method makes it possible not only to optimize the use of network resources and ensure a high level of protection, but also to create a flexible environment capable of adapting to changes in the topology of the corporate network and increased requirements for its stability.

Assessment of the developed method's effectiveness

The overall effectiveness of a distributed firewall in a corporate network is determined by its ability to minimize the number of unwanted connections, i.e., those considered potentially dangerous, as well as by reducing delays in network traffic processing. At the same time, with the increasing load on the network infrastructure, it is critical to be able to scale the firewall with minimal deployment and configuration costs.

Thus, the optimization task of organizing a distributed firewall can be formalized as the task of selecting the optimal set of configuration parameters that minimize the total traffic processing time while achieving a high level of security and scalability of the system. Formally, this task can be expressed as follows:

$$\min_{P} C(R), \quad (1)$$

where $R = \{r_1, r_2, ..., r_n\}$ – a set of filtering rules applied in a distributed firewall;

$C(R)$ – target function that reproduces the generalized costs of traffic processing, taking into account security and scalability requirements.

Accordingly, the effectiveness $E(R)$ of a distributed firewall can be presented as a function inversely proportional to the value of the target function $C(R)$:

$$E(R) = \frac{1}{\omega_1 \cdot C(R) + \omega_2 \cdot U_{CPU} + \omega_3 \cdot U_{RAM} + \omega_4 \cdot S + \omega_5 \cdot M}, \quad (2)$$

where ω_1, ω_2, ω_3, ω_4, ω_5 – weight coefficients that determine the priorities of criteria formed from the requirements and needs in the application of a distributed firewall;

U_{CPU} – level of containers' processor resources usage in the environment *Proxmox*;

U_{RAM} – RAM usage level;

S – assessment of security level, which can be defined as the number of blocked (potentially harmful) network packets;

M – scaling efficiency coefficient, which indicates the system's ability to adapt to increased load (for example, the number of successfully scaled containers without loss of performance).

The components of the optimization problem of organizing a distributed firewall (2) are presented below.

1. The total traffic processing time by the firewall $C(R)$ is a target function that reproduces the generalized traffic processing costs. Minimizing packet processing time can be achieved by automating the deployment of rules for new containers and distributing traffic across segments, which allows you to clearly build traffic filtering rules and reduce the time it takes to redirect packets to their destination. In this case, traffic processing time will be presented as

$$C(R) = T_f + T_r, \quad (3)$$

where T_f – average processing time for one packet in the firewall container, ms;

T_r – average time to redirect a packet to its destination after filtering, ms.

2. The use of processor resources U_{CPU} determines the likelihood of firewall container overload, which leads to increased packet processing time, process scheduling queues, and traffic loss during peak loads, as well as increased power consumption in conditions of limited resources. U_{CPU} is the inverse of the number of firewall containers and can be defined as follows:

$$U_{CPU} = \frac{1}{n} \sum_{i=1}^{n} \frac{P_i}{P_{\max}}, \quad (4)$$

where n – number of firewall containers;

P_i – actual processor load in the *i*-th container;

$P_{\max}$ – maximum allowable processor load specified by hardware resources.

3. The use of RAM U_{RAM} determines the effective management of the firewall's traffic filtering rules. The level of RAM usage also affects container stability and

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

scalability. In case of insufficient memory, this can lead to containers crashing, the inability to add new filtering rules, or the refusal to deploy new firewall containers. The level of RAM usage can be calculated as

$$U_{RAM} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{M_{\max}}, \quad (5)$$

where n – number of firewall containers;

M_i – amount of RAM used in the *i*-th container;

$M_{\max}$ – maximum available memory size in the system.

4. The security S level assessment will depend on the analysis of attacks, their complexity, and the effectiveness of traffic filtering rules. The security S level assessment will be calculated as

$$S = 1 - \sum_{i=1}^{k}\left(\frac{p_i}{c_i}\right), \quad (6)$$

where k – number of protection levels provided by a distributed firewall;

p_i – probability of a successful attack at the *i*-th level;

c_i – computational complexity of conducting an attack at the *i*-th level.

5. The efficiency of firewall container scaling M can be estimated over a period of time for deploying a firewall container and adding a set of filtering rules to it, and can be expressed as

$$M = \frac{1}{T_d + T_c}, \quad (7)$$

where T_d – average time to deploy a new container;

T_c – average time to add a set of filtering rules to it.

The proposed approach provides a comprehensive assessment of the effectiveness of a distributed firewall implementation, taking into account security and scalability requirements. This approach is key to determining the optimal architecture and operating parameters of a distributed firewall under variable load and limited resources.

Setting up the experiment

At the initial stage, the hardware requirements for the test environment were analyzed. It showed that for productive operation of a distributed firewall, a processor with at least four cores and a maximum frequency of 2.9 GHz can be used. The amount of RAM depends on the number of connected devices and data flow. From a practical point of view, 16 GB of RAM allows you to effectively serve three containers used to deploy a distributed firewall and serve up to three segments of a corporate network that are built as separate local subnets. The availability of such hardware resources ensures the deployment of a distributed firewall in corporate networks of small and medium-sized businesses with the ability to scale segments, as well as scale distributed firewall containers. The communication channel bandwidth must be at least 1 Gbit/s. To service a distributed firewall, you need to have two network interfaces that will be used to control the firewall and traffic to corporate network segments, which is sufficient to serve users of an experimental corporate network. The hard drive where the *Proxmox* virtualization platform with a distributed firewall is deployed must have at least 500 GB of free space.

Within the experiment, all functions of the distributed firewall have been implemented, and the possibility of scaling firewall containers as the load on the corporate network increases has been provided. Figure 1 shows a generalized diagram of a distributed firewall based on LXC containers of the *Proxmox VE* virtualized environment.

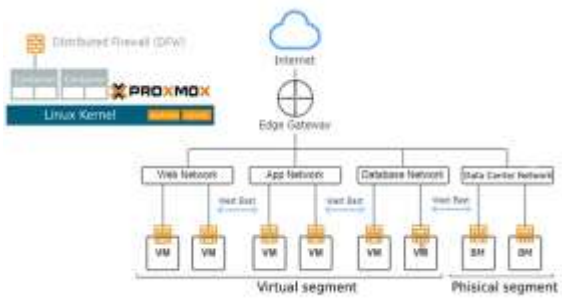


Fig. 1. Generalized diagram of a distributed firewall for a corporate network

At the initial stage, an isolated virtualized environment was created using LXC container technology within the *Proxmox VE* platform. Each container acts as a separate distributed firewall node with predefined network traffic filtering rules. Ubuntu 20.04 is installed as the base operating system in each container, providing a stable foundation for further configuration of access control policies.

The distributed firewall architecture involves the use of two types of network interfaces. The external interface with the IP address 192.168.31.50/24 provides a

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

connection to the global Internet and is the access interface to the *Proxmox VE* platform. Internal interfaces (vmbr1, vmbr2, and vmbr3) were configured for distributed firewall containers, each of which has its own IP address: 192.168.31.101/24, 192.168.31. 102/24, and 192.168.31.103/24.

To ensure isolation and increase the security level of network traffic in the corporate network, access to a separate local network served by the corresponding firewall node was configured:

- subnet 172.16.16.0/24 is served by container 101;
- subnet 172.16.17.0/24 is served by container 102;
- subnet 172.16.18.0/24 is served by container 103.

Thus, each container processes network traffic coming to the corresponding subnet, implementing the distribution and segmentation of access control policies (Fig. 2).

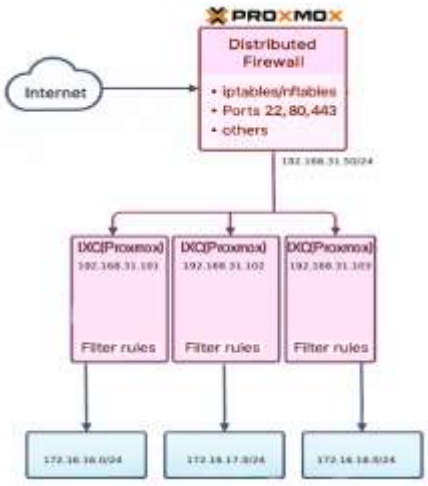


Fig. 2. Distributed firewall diagram for an experimental corporate network

In the second stage, network traffic filtering mechanisms are configured in each container using the iptables utility in accordance with defined security policies. The basic configuration of rules provides for opening ports 22/TCP (SSH), 80/TCP (HTTP), and 443/TCP (HTTPS), as well as allowing ICMP packet processing, which provides remote administrative access and supports the functioning of web services.

Filtering rules are configured using the following utility *iptables* commands:

iptables -A FORWARD -p icmp -j ACCEPT
iptables -A FORWARD -p tcp --dport 22 -j ACCEPT
iptables -A FORWARD -p tcp --dport 80 -j ACCEPT
iptables -A FORWARD -p tcp --dport 443 -j ACCEPT

Containerization enables rapid system scaling in case of increased request volume or more complex filtering policies, particularly when expanding the range of open ports, implementing VPN connections, or processing special network protocols. This approach allows creating additional containers without changing the underlying system architecture. In addition, it is possible to integrate automation and orchestration tools for the distributed firewall container cluster using the built-in *Proxmox VE* API.

After completing the deployment of the distributed firewall, a comprehensive check of its functionality was performed. At the initial stage of testing, ICMP diagnostics were performed from the test container to the services of the local network served by the distributed firewall, in particular to the deployed *nginx* web server. For additional diagnostics of network service availability, the *mtr* utility was used, which provides route verification and analysis of loss and delay statistics in real time (Fig. 3). Successful completion of these tests confirmed the correct configuration of the distributed firewall, the correct implementation of filtering rules, the compliance of network routes, and the availability of Internet access from each individual container. This, in turn, confirmed the reliability of the access control system in accordance with the specified security architecture.

Fig. 3. Firewall test results using route checking with the utility *mtr*

Within the scope of this work, the *M/M/n* queuing model and Jackson's open network model [17] were used to model the performance of a distributed firewall. The use of these models is due to the need for quantitative analysis of the performance of the distributed firewall infrastructure in the event of increased load and complexity of traffic filtering policies.

The *M/M/n* model allows us to correctly describe the behavior of a queuing system in which incoming traffic filtering requests are sent to a limited number of firewall containers that function as independent processors. Taking into account parameters such as input flow intensity, average request processing time, and the number of parallel service channels allows us to evaluate system properties such as average delay, load level, and

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

the probability of queues. This helps to determine the minimum number of containers required to achieve a given level of performance under varying loads.

At the same time, Jackson's open network model is used to model the more complex interaction between the components of a distributed firewall. It provides the probability of describing the system as a set of nodes between which requests can circulate, replicating the structure of a real firewall with multiple modules or levels of protection, such as traffic inspection, authorization verification, threat analysis, etc. Taking into account internal traffic between containers allows us to evaluate the impact of complex multi-level security policies on overall system performance.

The combined use of the *M/M/n* and Jackson models helped to compare centralized and distributed firewall architectures and to establish optimal configurations for the container infrastructure. The simulation results showed that scaling the number of containers reduces the average request processing delay. This directly improves the efficiency of the traffic filtering system.

Thus, the dependence of firewall efficiency on the number of filtering rules, as well as on the level of scalability in conditions of increasing load and security requirements, is confirmed (Figs. 4, 5).

The experiment was conducted at the laboratory of computing systems and network technologies of the Department of Electronic Computers at Kharkiv National University of Radio Electronics.

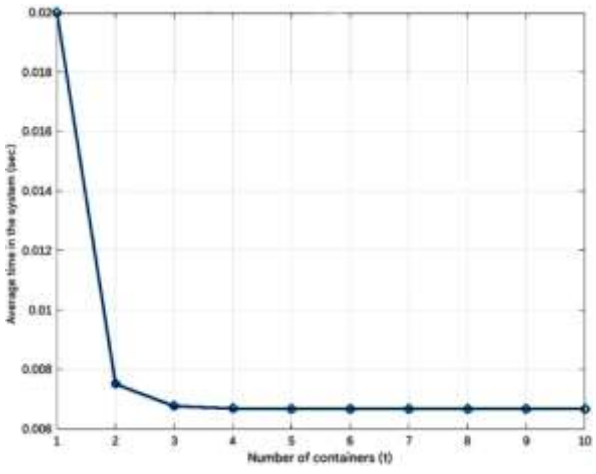


Fig. 4. Graph showing the dependence of distributed firewall efficiency on the number of containers for the open Jackson model

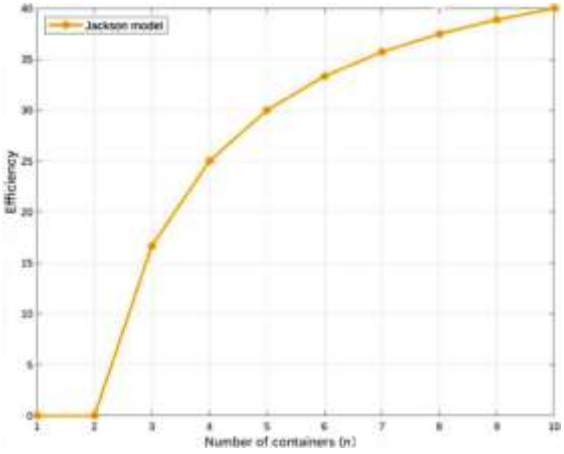


Fig. 5. Graph showing the dependence of distributed firewall efficiency on the number of containers for the *M/M/n* model

Conclusions

The article proposes and analyzes in detail a method for improving the performance of a distributed firewall in order to ensure a high level of security in virtualized environments or computer networks with scalable infrastructure. The proposed approach minimizes delays during traffic filtering and enables automatic scaling of firewall functionality while maintaining a comprehensive network security system.

The scientific novelty of the method lies in the improvement of security mechanisms in scalable environments in the case of limited hardware resources, which contributes to achieving a high level of protection against external and internal threats, while maintaining the fault tolerance and reliability of the protective infrastructure.

The results of the effectiveness analysis showed that when scaling the infrastructure, it is critically important to consider the deployment time of the protection system, as this stage can be potentially vulnerable. However, reducing delays in the scaling process allows for continuous protection, and the even distribution of the load on active infrastructure elements helps maintain its stability and efficiency.

Practical implementation of the method involves taking into account the characteristics of the transmitted information, the level of protection required, available resources, as well as filtering parameters and the number of security rules applied. The proposed approach is suitable for small and medium-sized businesses that seek to ensure reliable network protection in environments with remote access and limited resource capacity, while

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

maintaining the ability to scale and use automated configuration mechanisms.

Further research should focus on optimizing the proposed method by integrating intelligent monitoring and event processing systems, such as Prometheus and Grafana. The use of these tools will enable effective metric collection, information visualization, and rapid anomaly detection, which will help increase the response speed of the distributed firewall to potential threats.

References

1. Vazhynskyi, V. B., Tkachov, V. M. (2023), "Problematyka bezpeky ta kryterii khnadiinosti multykhmarnykh seredovyshch. Natsionalnyi universytet “Poltavska politekhnika imeni Yuriia Kondratiuka". *National University “Yuri Kondratyuk Poltava Polytechnic”*, 75 p. available at: http://repositsc.nuczu.edu.ua/bitstream/123456789/19451/1/issue-galley-73%20%281%29.pdf
2. Swati, Roy, S., Singh, J., Mathew, J. (2025), "Securing IIoT systems against DDoS attacks with adaptive moving target defense strategies". *Scientific Reports,* 15(1), 9558 p. DOI: https://doi.org/10.1038/s41598-025-93138-7
3. Bytsiv, M. M. (2021), "Znachennia informatsiinykh tekhnolohii yak chynnyka innovatsii u diialnosti maloho ta serednoho biznesu", *Biznes, innovatsii, menedzhment: problemy ta perspektyvy: zbirnyk tez dopovidei II Mizhnarodnoi nauk.-prakt. konferentsii,* Natsionalnyi tekhnichnyi universytet Ukrainy «Kyivskyi politekhnichnyi instytut imeni Ihoria Sikorskoho», P. 206-207. available at: https://confmanagement-proc.kpi.ua/article/view/231790
4. Bringhenti, D., Marchetto, G., Sisto, R., Valenza, F., Yusupov, J. "Automated Firewall Configuration in Virtual Networks," *in IEEE Transactions on Dependable and Secure Computing,* Vol. 20, No. 2, P. 1559-1576, DOI: 10.1109/TDSC.2022.3160293
5. Zhurylo, O. Liashenko, O. (2024), "Arkhitektura ta systemy bezpeky IoT na osnovi tumannykh obchyslen", *Innovative Technologies and Scientific Solutions for Industries*, (1(27), P. 54–66. DOI: 10.30837/ITSSI.2024.27.054
6. Anwar, R. W., Abdullah, T., Pastore, F. (2021), "Firewall Best Practices for Securing Smart Healthcare Environment: A Review". *Applied Sciences*, 11(19), 9183 p. DOI: https://doi.org/10.3390/app11199183
7. Togay, C., Kasif, A., Catal, C., Tekinerdogan, B. (2022), "A Firewall Policy Anomaly Detection Framework for Reliable Network Security", *in IEEE Transactions on Reliability* , Vol. 71, P. 339-347, DOI: 10.1109/TR.2021.3089511
8. Kaur, H., Atif, M., Chauhan, R. (2020), "An Internet of Healthcare Things (IoHT)-Based Healthcare Monitoring System". *In Advances in Intelligent Computing and Communication; Springer*: Berlin, Germany, P. 475–482.
9. Bringhenti, D., Valenza, F. (2024), "GreenShield: Optimizing Firewall Configuration for Sustainable Networks," *in IEEE Transactions on Network and Service Management,* Vol. 21, No. 6, P. 6909-6923, DOI: 10.1109/TNSM.2024.3452150
10. Bringhenti, D., Marchetto, G., Sisto, R., Valenza, F. Yusupov, J. (2023), "Automated Firewall Configuration in Virtual Networks," *in IEEE Transactions on Dependable and Secure Computing* , Vol. 20, № 2, P. 1559-1576, DOI: 10.1109/TDSC.2022.3160293
11. Tiwari, Aman, Sivani, Papini, Hemamalini, V. (2022), "An enhanced optimization of parallel firewalls filtering rules for scalable high-speed networks." *Materials Today: Proceedings* № 62, P. 4800-4805. DOI:10.1016/j.matpr.2022.03.346
12. Sinha, M., Bera, P., Satpathy, M. (2021), "An Anomaly Free Distributed Firewall System for SDN", *2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)* , Dublin, Ireland, P. 1-8, DOI: 10.1109/CyberSA52016.2021.9478256
13. Novorodovskyi, V. (2020), "Informatsiina bezpeka Ukrainy v umovakh rosiiskoi ahresii. Society". *Document. Communication.* № 9. P. 150–179. DOI: https://doi.org/10.31470/2518-7600-2020-9-150-1179
14. Tkachov, V. M., Chepurna, I. S., Fesenko, T. H. (2024), "Metod multyrivnevoho vpn-tuneliuvannia dlia zabezpechennia viddalenoho dostupu do vuzliv ekstranet-merezhi", *Visnyk Khersonskoho natsionalnoho tekhnichnoho universytetu*. №. 3 (90). P. 299-308. DOI: https://doi.org/10.35546/kntu2078-4481.2024.3.37
15. "Proxmox VE. Proxmox VE." (2025), available at: https://pve.proxmox.com/mediawiki/index.php?title=Main_Page&oldid=12223 (last accessed: 26.04.2025)
16. Ariyanto, Y. (2023), "Single server-side and multiple virtual server-side architectures: Performance analysis on Proxmox VE for e-learning systems". *ITEGAM-JETIA*, №9(44), P. 25-34. DOI: https://doi.org/10.5935/jetia.v9i44.903
17. Ghandour, O., El Kafhali, S., Hanini, M. (2024), "Adaptive workload management in cloud computing for service level agreements compliance and resource optimization". *Computers and Electrical Engineering,* № 120, 109712 p. DOI:10.1016/j.compeleceng.2024.109712

Надійшла (Received) 25.05.2025

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

Відомості про авторів / About the Authors

Chepurna Iryna – Kharkiv National University of Radio Electronics, Assistant Lecturer at the Department of Electronic Computers, Kharkiv, Ukraine; e-mail: iryna.chepurna@nure.ua; ORCID ID: https://orcid.org/0009-0008-2442-6221
Frolov Dmytro – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Informatics, Leading Engineer at the IOC, Kharkiv, Ukraine; e-mail: dmytro.frolov@nure.ua; ORCID ID: https://orcid.org/0009-0009-3291-3561

Чепурна Ірина Сергіївна – Харківський національний університет радіоелектроніки, асистентка кафедри електронних обчислювальних машин, Харків, Україна.
Фролов Дмитро Євгенович – Харківський національний університет радіоелектроніки, аспірант кафедри інформатики, провідний інженер ІОЦ, Харків, Україна.

МЕТОД ПІДВИЩЕННЯ ПРОДУКТИВНОСТІ РОЗПОДІЛЕНОГО БРАНДМАУЕРА НА БАЗІ *PROXMOX* У КОРПОРАТИВНИХ КОМП'ЮТЕРНИХ МЕРЕЖАХ

Предметом дослідження в статті є метод підвищення продуктивності розподіленого брандмауера на базі LXC-контейнерів у середовищі *Proxmox VE* для корпоративних комп'ютерних мереж. **Мета роботи** – розроблення підходів до забезпечення високого рівня ефективності розподіленого брандмауера для моніторингу та управління трафіком у корпоративних і віртуалізованих мережах, що дає змогу мінімізувати затримки під час фільтрації трафіка та забезпечити надійне функціювання корпоративної мережі в умовах обмежених апаратних ресурсів. Для розв'язання завдань упроваджено такі **методи дослідження:** теоретичний аналіз літературних джерел; аналіз особливостей застосування технології контейнеризації для реалізації динамічного контролю мережного трафіка; вивчення методів підвищення ефективності застосування обчислювальних ресурсів у середовищах з обмеженими апаратними ресурсами; аналіз переваг розподіленої архітектури брандмауера щодо мінімізації затримок під час передачі інформації, підвищення пропускної здатності системи та зниження ризиків несанкційного доступу; експериментальна перевірка працездатності та ефективності розподіленого брандмауера. **Досягнуті результати.** Запропонований метод дає змогу мінімізувати затримки під час фільтрації трафіка та забезпечити автоматичне масштабування функційності брандмауера в умовах збереження цілісної системи безпеки мережі. Розроблений підхід забезпечує високий рівень захисту ККМ способом сегментації мережі з призначенням окремого контейнера LXC для обслуговування кожної локальної мережі, що допомагає здійснювати цілеспрямовану фільтрацію трафіка та гнучке управління політиками доступу. **Висновки.** У роботі запропоновано конфігурацію розподіленого брандмауера в середовищі *Proxmox* разом із налаштуванням базового набору правил фільтрації для забезпечення ефективного функціювання корпоративної комп'ютерної мережі. Наукова новизна методу полягає в удосконаленні механізмів забезпечення безпеки в масштабованих середовищах за умов обмеженості апаратних ресурсів, що дає змогу досягти захисту високого рівня від зовнішніх і внутрішніх загроз, зберігаючи водночас відмовостійкість і надійність мережної інфраструктури. Експериментальна перевірка працездатності та ефективності методу підтвердила доцільність його впровадження для забезпечення стабільного й контрольованого доступу до мережних ресурсів ККМ.

Ключові слова: метод; розподілений брандмауер; контейнер; Proxmox; віртуалізація; затримка; фільтрація трафіка..

Бібліографічні описи / Bibliographic descriptions

Чепурна І. С., Фролов Д. Є. Метод підвищення продуктивності розподіленого брандмауера на базі *Proxmox* у корпоративних комп'ютерних мережах. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 180–188. DOI: https://doi.org/10.30837/2522-9818.2025.3.180

Chepurna, I., Frolov, D. (2025), "A method for increasing the productivity of a distributed firewall based on Proxmox in corporate computer networks", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 180–188. DOI: https://doi.org/10.30837/2522-9818.2025.3.180

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

UDC 681.326
DOI: https://doi.org/10.30837/2522-9818.2025.3.189

O. SHKIL, O. FILIPPENKO, D. RAKHLIS, I. FILIPPENKO, V. KORNIIENKO

OPTIMIZATION OF SOFTWARE CODE FOR HIGH-LEVEL SYNTHESIS DURING HARDWARE IMPLEMENTATION OF THE COMPUTATIONALLY-LOADED ALGORITHMS

The **subject matter** of the work is the impact of code optimization methods of highly intensive algorithms, used in digital signal processing, on hardware costs and performance when implemented on different platforms. The goal of the work is to conduct a comparative analysis of the impact of the effects of three C-code optimization approaches: loop unrolling, switching to fixed-point arithmetic, and their combinations, on performance and hardware costs when implementing matrix multiplication, fast Fourier transform, and wavelet transform algorithms using high-level synthesis (HLS) tools on system-on-chip (SoC) platforms, personal computers (PCs), and single-board computers. The following **tasks** were solved in the article: implementation of highly intensive algorithms based on selected hardware platforms and using HLS; comparison of execution time of algorithms with and without different optimization methods; comparison of hardware costs for algorithms' implementations with and without different variants of optimization; formulate conclusions about the impact of different C-code optimization methods on performance and hardware costs on different target platforms. The following **methods were** used: C/C++ code optimization methods, diagnostic experiments using high-level synthesis tools to implement digital signal processing algorithms on the selected hardware platform, and statistical data collection using Python. The following **results** were obtained: for algorithms based on arithmetic operations, code optimization provided up to 30% reduction in execution time on ARM platforms. For algorithms based on the Fourier transform, complex optimization reduced execution time by up to 90% on processor devices. For programmable logic (FPGA), none of the optimization methods provided a significant execution acceleration. However, the transition to fixed arithmetic reduced hardware costs by 40–80% regardless of the algorithm type. **Conclusions**. The choice of a C code optimization strategy significantly impacts the efficiency of algorithm implementation on processor architectures. In contrast, optimizing the data types used plays a key role for FPGAs. In contrast, for FPGAs, optimizing the data types used plays a key role.

Keywords: embedded systems; high-level synthesis; C code optimization; System-on-Chip.

Introduction

In modern embedded systems, in particular, in high-performance solutions based on system-on-chip (SoC) technology, optimization of algorithms implemented in a high-level programming language is becoming increasingly important in order to efficiently use available hardware resources. In the process of developing such systems, an important task is to achieve a balance between computing speed, hardware resource consumption, and data processing accuracy.

In digital signal processing systems (DSP), algorithms that have a high computational load attract special attention. Such algorithms include matrix multiplication, fast Fourier transform (FFT), and wavelet transform. These algorithms have numerous applications ranging from computer vision and image processing to telecommunications and real-time signal analysis.

These algorithms are considered computationally intensive due to their complexity and scale of data processing. For example, matrix multiplication in the classical version is characterized by computational complexity $O(N^2)$, which leads to an exponential increase in the number of operations as the size of the input matrices increases. Even the use of optimized variants, such as the Coppersmith-Winograd method with a complexity of about $O(N^{2.5})$, does not eliminate the significant burden on computing resources. The fast Fourier transform, which is widely used in spectral analysis, has a complexity of $O(N \cdot \log(N))Nlog\left(N\right)$ but performs a large number of operations with complex numbers, which creates an additional load on arithmetic units. Similarly, the wavelet transform involves multilevel signal decomposition, which is implemented by successive iterations and requires efficient memory and data flow management, especially in real time.

Implementation of such algorithms in high-level synthesis environments, where the functionality is described in C/C++, opens up wide opportunities for applying various optimization methods. Among the most common methods are looping, switching to fixed-point arithmetic, and a combination of these. Such optimization techniques can reduce the logic depth, improve parallelization, reduce power consumption, and improve

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

performance, especially when implemented on field-programmable gate arrays (FPGAs), embedded ARM processors, or single-board computers.

At the same time, the effectiveness of the same optimization methods depends significantly on the computing platform. For example, optimization techniques that demonstrate a significant reduction in execution time on ARM processors may not have the same effect on FPGAs, where the nature of computing operations is different. This justifies the need for a detailed comparative analysis of the impact of software optimization methods on performance and hardware costs for each type of implementation platform.

It is also important to consider that the effectiveness of different optimization methods is closely related to the nature of the algorithm itself. In the case of predominantly arithmetic computations, optimization methods that reduce the number of unnecessary operations and optimize memory access are critical. Whereas for algorithms based on transformations, special attention should be paid to those optimization methods that affect the efficient organization of multi-level data processing and minimize delays in processing large amounts of information.

Embedded systems that are designed to perform specialized application tasks are usually characterized by limited computing power, memory capacity, and energy efficiency. Given these limitations, it is important to carefully plan and optimally utilize available hardware resources, such as embedded memory and specialized computing modules, when developing such systems.

There are two main ways to improve the performance of an embedded system, which is determined by reducing the execution time of the target algorithm: hardware and software.

The hardware way involves upgrading the computing platform, but this often contradicts the economic and energy constraints inherent in many classes of embedded systems. Instead, software optimization involves improving the runtime environment and the target program code. If the choice of the execution environment is determined by the initial technical specification, the focus is on optimizing the implementation of the algorithm itself.

Universal approaches to software optimization include the use of more efficient algorithms with minimization of redundant calculations and memory accesses, optimization of data structures, involvement of low-level program code with specialized libraries, and reduction of calculation accuracy while maintaining the required functionality. It is worth noting that the effectiveness of the proposed approaches depends on the specific problem statement and the chosen algorithm [1].

In the case of using Xilinx ZYNQ-7000 SoCs as the hardware basis, the architecture of which integrates the Processing System (PS) and Programmable Logic (PL), it is advisable to consider optimizing the software implementation in both PS and PL. This approach allows for flexible distribution of the computational load between general-purpose processors and hardware-accelerated modules.

Thus, the object of research is methods for optimizing the high-level description of computationally intensive algorithms. The subject of the study is the impact of the selected optimization methods on hardware costs and performance when implementing these algorithms on different platforms.

The purpose of the article is to compare the impact of three approaches to C-code optimization, namely loop unrolling, transition to fixed-point arithmetic, and their combination, on the efficiency of implementing matrix multiplication, fast Fourier transform, and wavelet transform algorithms using high-level synthesis tools on SoC, PC, and single-board computers.

Literature review

Paper [2] explores the potential of using Large Language Models (LLM) for automated adaptation of program code to HLS-compliant. One of the significant contributions of the researchers is the implementation of a framework for correcting program compilation errors, which is guided by the LLM and combines automatic code generation with various hardware-oriented optimization methods. According to the conclusions, an experimental evaluation on 24 applications demonstrates that the proposed framework significantly exceeds the indicators of successful adaptation of program code compared to traditional scripts and direct application of LLM.

Study [3] analyzes the potential of code optimization specific to high-level synthesis to achieve higher performance and energy efficiency compared to traditional implementations in the field of High Performance Computing (HPC). The authors propose a library of optimized primitives for high-level synthesis with an analysis of individual optimization techniques aimed at reducing memory access time, scalability of the architecture, and optimization of the execution pipeline. The analysis procedure includes the study of individual

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

optimization techniques and their impact on high-performance hardware computers.

In [4], a new FPGA implementation of the Strassen algorithm for matrix multiplication is considered, which demonstrates significant performance advantages over traditional approaches. The authors of the paper test the algorithm implementation on Alveo U50 and U280 hardware platforms for matrix sizes up to 256 elements. They also study the effect of the matrix data type on the speed of the proposed and implemented architecture.

The authors of the study [5] consider the growing role of software components in complex heterogeneous embedded systems, where development time, flexibility, and reuse are important factors. The authors note that due to the regularity of processing multimedia and DSP applications, statically scheduled devices, in particular VLIW (Very Long Instruction Word) processors, are promising options for such systems compared to dynamically scheduled processors, such as modern superscalar General Purpose Processors (GPPs).

Paper [6] presents a promising study of the use of high-level synthesis tools for the development of hardware accelerators focused on digital signal processing tasks with an emphasis on fast Fourier transform. The authors emphasize the importance of behavioral models for achieving efficient and high-performance hardware architectures. They also analyze the proposed architecture in terms of performance and hardware costs.

The authors of [7] consider the shortcomings of traditional multiplication algorithms for very large numbers and their impact on the performance of circuits, in particular in digital filters. The authors propose the use of the Schönhage-Strassen Algorithm (SSA) to optimize the operation of filters with a finite impulse response (FIR), which are the main components of many DSP processors.

The study [8] proposed an efficient hardware architecture for the two-dimensional discrete Fourier transform (SDFT), which is focused on minimizing the use of hardware resources and optimizing performance for real-time tasks. The proposed architecture significantly reduces the need for hardware resources compared to the Park method.

The authors of [9] consider the process of code optimization at the compilation stage. The authors emphasize that code optimization tries to improve the target code without changing its output or causing side effects. The paper describes classical optimization techniques, such as eliminating common nested expressions, removing dead code, and collapsing constants, which are widely used in compilers. The article also analyzes the challenges faced by compiler developers and the latest code optimization methods for such systems.

The study [10] considers the improvement of the key encapsulation mechanism based on a ring with a restricted polynomial of degree N (NTRU-KEM). The authors point out that, despite its high reliability, NTRU-KEM has increased storage and computing requirements compared to classical cryptography, which leads to significant memory consumption and performance degradation. This paper proposes a hardware-software co-design approach that allows for the customization of computations to meet variable requirements for running time and number of iterations. The main contribution of the work is the development of a new hardware acceleration technique focused on optimizing the use of the SoC bus. Experimental results confirm the effectiveness of the proposed approach.

Paper [11] presents the POLSCA compilation system, which improves the process of automatic optimization of nested affine cycles for high-level synthesis. The system performs a preliminary decomposition of the project to balance code complexity and parallelism, and automatically modifies memory interfaces to better support HLS tools. This avoids the need to manually add directives and simplifies integration. Experiments on the Polybench/C benchmarks have shown that the approach provides an average speed up of 1.5 times.

Study [12] proposes a Design Space Exploration (DSE) method for high-level synthesis that takes into account information from the scheduling stage. The authors use this information to improve the efficiency of a genetic algorithm implemented on the basis of their own HLS tool. The proposed approach allows finding more Pareto-optimal solutions compared to methods that do not take into account scheduling data. The method outperforms the traditional Genetic Algorithm (GA) approach, reducing the execution time by a factor of four and achieving 95.7% of optimal solutions while exploring only 0.18% of the Pareto space.

The authors of [13] also propose an approach to studying DSE in high-level synthesis based on Contrastive Learning (CL) to determine the dominance relationship between projects. Unlike traditional methods that evaluate absolute values of productivity and cost, the proposed method classifies projects by their relative efficiency. The CL-method is integrated with three

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

modern DSE approaches, which significantly reduces the number of syntheses without losing the quality of the results. Comparative experiments have shown the advantage of classification methods over regression methods in predicting Pareto dominance.

The study [14] presents a method for automatic optimization of HLS designs based on domain-specific knowledge and does not require quality assessment models or metaheuristics. The proposed approach automatically selects efficient directive configurations for source code, which is especially useful for users without hardware experience. The method has been tested on more than 100 examples from benchmarks and GitHub running on Xilinx ZCU104 FPGA. On average, we achieved a speedup of ×7.2 compared to manually optimized projects and ×1.35 compared to overprovisioning methods. Comparison with modern DSE methods showed similar quality of results, but at a speed that exceeds traditional approaches by 100 to 1000 times.

Paper [15] presents Stream-HLS, a new methodology and framework for automated creation of high-performance hardware architectures based on C/C++ or PyTorch code. The solution is built on the Multi-Level Intermediate Representation (MLIR) infrastructure and addresses key HLS issues, including support for multi-core applications, global loop scheduling, and graph pipelining. Stream-HLS automatically generates an optimized dataflow architecture and host code for FPGAs using an analytical performance model. Experiments on standard and real-world problems (Transformers, Convolutional Neural Networks (CNN), Multilayer Perceptrons (MLP)) demonstrate an acceleration of up to 79.4× compared to state-of-the-art solutions and up to 10.6× compared to manual optimization.

The authors of [16] propose an approach to the design of Application-Specific Instruction-set Processors (ASIP), fully implemented at the level of the ANSI C programming language standard using HLS tools. The authors combine the description of the Central Processing Unit (CPU) and the hardware accelerator to synthesize the entire system together, which reduces the area and power consumption due to optimal resource allocation. HLS also provides the ability to automatically generate different ASIP variants with different balance between performance, area, and power consumption. The results show that the proposed approach outperforms traditional methods, providing lower overhead, higher performance, and a significant reduction in power consumption compared to a basic RISC-V processor and state-of-the-art analogs.

The presented review covers a wide range of modern research in the field of high-level synthesis, optimization of hardware architectures, and intelligent design automation methods that are relevant for high-performance computing systems, embedded solutions, and digital signal processing.

Therefore, the issue of optimizing the code of highly loaded algorithms used in digital signal processing in order to reduce hardware costs and increase performance when implemented on different target platforms remains relevant.

Problem statement

It is necessary to analyze how different ways of optimizing the C-code of the algorithm description in the high-level synthesis of embedded systems on a chip affect the performance and hardware costs of the resulting devices. The impact of the proposed optimization methods will also be analyzed for classic PCs and single-board computers.

Different types of computationally intensive algorithms will be used as examples:

– matrix multiplication (classical algorithm, computational complexity of which is O(n(3)), Winograd algorithm, computational complexity of which is O(n2,5))

– series decomposition using harmonic and wave functions (fast Fourier transform (FFT transform, computational complexity of which is O(N log N)) and wavelet transform, computational complexity of which is O(N log N)).

These algorithms are analyzed for objects of different sizes.

Taking into account the tasks, we introduce the following notations:

1. $A = \{A_1, A_2, ..., A_m\}$ $A = \{A_1, A_2, \ldots, A_m\}$ - a set of algorithms, where eachA_i is a computationally intensive algorithm (e.g., matrix multiplication, FFT, wavelet transform).

2. $D = \{D_1, D_2, ..., D_n\}$ $D = \{D_1, D_2, \ldots, D_n\}$- set of objects, where eachD_j is the size of the data volume for a particular algorithm. For matrix multiplication algorithms, these are matrix sizes, for FFT the length of the transformation, for wavelet transform the size of the input buffer and the level of decomposition.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

3. $O = \{O_1, O_2, ..., O_k\}$ $O = \{O_1, O_2, \ldots, O_k\}$ - a number of ways to optimize C code (looping, fixed point, combined optimization).

4. $P = \{P_1, P_2, ..., P_l\}$ $P = \{P_1, P_2, \ldots, P_l\}$ - a variety of platforms on which the algorithms will be tested, including classic PCs, single-board computers, and SoCs.

Thus, the implementation of the algorithmA_i on the objectD_j with the optimizationO_k on the platformP_l can be described as:

$$R_{ijkl} = Run(A_i, D_j, O_k, P_l). \quad (1)$$

The objective of the study **is to** investigate the impact of different ways to optimize O_k on performance and hardware costs when implementing $A_1, A_2, \ldots, A_m$ algorithms on platforms $P_1, P_2, ..., P_l$ $P_1, P_2, \ldots, P_l$.

The classical ways of optimizing C/C++ code were chosen, namely, loop unrolling, transition to fixed-point arithmetic, and a hybrid version that combines both approaches. The implementation of fixed-point arithmetic was performed using the fixed_16_16 type. This data type is used in many implementations of digital signal processing libraries and highly optimized linear algebra libraries.

Let's introduce a normalized indicator of reducing the time T of executing computational algorithms for different ways of optimizing program code (T_{opt}/ T_{base}). In this case, the average normalized index of the execution time of the algorithm K_i for objects of different sizes is calculated as follows:

$$K_i = \sum_{j=1}^{m} \frac{T_{опт}}{T_{базове}} \Big/ m, \quad (2)$$

where m is the number of objects under consideration;T_{opt} is the index of reducing the execution time of computing algorithms for different methods of optimizing program code in microseconds (μs);T_{base} is the execution time of the algorithm without optimization in microseconds (μs).

The averages of the algorithm execution time for different optimization methods are calculated as follows:

$$M_j = \sum_{i=1}^{n} K_{ij} \Big/ n, \quad (3)$$

where n is the number of optimization considered algorithms.

The mathematical component of the algorithms under consideration

Consider the matrix multiplication algorithm with complexity $O(N^3)$, where N is the size of the input data in the units required to represent it. It is based on the classical formula for multiplying two matrices $C = A \cdot B$, whose elements are calculated by the formula:

$$C_{ij} = \sum_{k=1}^{m} A_{ij} \cdot B_{kj}, \quad (4)$$

where C is the result matrix of size ; $n \times p$ A is the result matrix of size ; $n \times m$ B is the result matrix of size $m \times p$

Let's consider the Coppersmith-Winograd matrix multiplication algorithm, which is an improved modification of the classical matrix multiplication algorithm aimed at reducing the number of multiplication operations. Its key idea is to pre-calculate auxiliary values, which allows you to optimize and speed up the multiplication, especially when working with large square matrices.

For matrices of size $n \times m$ and $m \times p$, the following iterations of the algorithm are performed:

1) for each row of matrix A, the auxiliary matrix rowFactor is calculated

$$rowFactor[i] = \sum_{k=1}^{m/2} A[i][2k-2] \cdot A[i][2k-1]; \quad (5)$$

2) for each column of matrix B, the auxiliary matrix colFactor is calculated:

$$colFactor[j] = \sum_{k=1}^{m/2} B[2k-2][j] \cdot B[2k-1][j]; \quad (6)$$

3) the main multiplication is performed

$$C[i][j] = -rowFactor[i] - colFactor[j] = \sum_{k=1}^{m/2} (A[i][2k-2] + \cdots + B[2k-1][j]) \cdot (A[i][2k-1] + \cdots + B[2k-2][j]); \quad (7)$$

4) if the number of columns m is odd, an additional adjustment is performed

$$C[i][j] += A[i][m-1] + B[m-1][j]. \quad (8)$$

Let's take a look at the Fourier transform. The Fourier transform is a mathematical operation that projects a function (signal) from the time domain to the frequency domain by decomposing it into a basis of complex exponents. It is based on the principle that any energy-limited function (signal) can be represented as a linear combination of harmonic functions of sine and cosine waves of corresponding frequencies and amplitudes.

In general, the complex Fourier transform is defined as

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi kn}{N}}, \quad npu\ k = 0,1,2,\ldots,N-1, \quad (9)$$

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

where $X(k)$ is the result of the transformation in the frequency range; $x(n)$ is the input complex sequence; N is the length of the transformation; $e^{-j\frac{2\pi kn}{N}}$ is the complex exponential basis (twiddle factor).

The length of the Fourier transform determines the possible frequency resolution. That is, with a transform length of 512 and a sampling frequency of 48 kHz, according to the Nyquist-Shannon theorem, each frequency interval in the Fourier transform spectrum corresponds to a frequency bin with a step of 46.875 Hz.

Consider the wavelet transform. The Discrete Wavelet Transform (DWT) is a formalized implementation of the wavelet transform based on a discrete set of scaling and shifting coefficients that follow certain mathematical rules. This transformation allows the signal to be decomposed into a basis formed by mutually orthogonal wavelet functions. This approach is fundamentally different from the Continuous Wavelet Transform (CWT), as well as from its variations adapted to discrete time series, in particular the so-called Discrete-Time Continuous Wavelet Transform (DT-CWT).

The design of a wavelet function is based on the so-called scaling function, which defines the scaling properties of a wavelet. The requirement of orthogonality of the scaling function to its discrete translations imposes a number of mathematical conditions on it, among which the dilation equation plays a key role. This equation provides a recursive definition of the scaling function through its shifts and the corresponding filter coefficients, and is the basis for constructing orthogonal wavelet bases, such as Dobeshi wavelets.

The discrete wavelet transform is defined as

$$\Psi_x^{\Psi}[n,a^j] = \sum_{m=0}^{N-1} x[m] \cdot \Psi_j^*[m-n], \qquad (10)$$

where $\Psi_x^{\Psi}[n,a^j]$ is the wavelet coefficient for the signal x[n] at scale a(j) and offset n, i.e., the result of projecting the signal onto a wavelet; $x[m]$ is the input signal; $\Psi_j^*[m-n]$ is a complex-conjugate wavelet function shifted by n and scaled by j; $\sum_{m=0}^{N-1}$ is a discrete convolution of the signal with a wavelet function.

In turn, the wavelet function $\psi_j[n]$ is defined as.

Description of the experimental part

To study the impact of different methods of optimizing computationally intensive algorithms, we selected objects of different dimensions in order to compare the execution time and hardware costs for each of the proposed implementation platforms. When measuring the execution time of algorithms on each platform, the high-resolution timer available on the platform was taken into account.

To obtain statistics on the algorithm execution time, a template function was developed that uses the timer and executes the algorithm for a specified number of iterations, and then calculates the statistics on the execution time of the selected algorithm.

The analysis of hardware costs for the PL part was performed by analyzing the post-synthesis report in Vivado after exporting the IP core from Vitis HLS to Vivado.

The set of matrices was analyzed on all platforms for square matrices of (16x16), (24x24) and (32x32) elements. The implementation of the Fourier transform was compared at the transform lengths of 512, 4096, and 8192, which is determined by the number of N-analyzing components of the FFT series, according to (9). The wavelet transform was analyzed using the db4 wave function, decomposition level 5, and signal length 256 on all platforms.

The methods of code optimization used in the study are denoted as follows:

- C - basic implementation of the algorithm for the float type without optimization;
- C1 - loop unrolling (loop unrolling, float type);
- C2 - a basic implementation of using fixed-point arithmetic instead of the float type;
- C3 - complex optimization with loop unrolling and the use of fixed-point arithmetic instead of the float type.

It should be taken into account that when using the basic implementation of the algorithm, the result of measuring its execution time is equal to Tbasic, so for the method of optimizing the code C, the average time Ki is not calculated.

The following algorithms will be used as algorithms for further analysis:

- Matrix n(3) isa classical matrix multiplication algorithm;
- Winograd n(2,5) - an optimized matrix multiplication algorithm that reduces the number of multiplication operations due to preliminary calculations;
- FFT is an algorithm based on the Fast Fourier Transform;

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

– Wavelet - an algorithm based on the wavelet transform using the db4 wave function and decomposition 5.

Table 1 shows the results of measuring the execution time of the considered algorithms in microseconds (µs) and the corresponding averaged K1/K2/K3 for different methods of optimizing C1/C2/C3 when implemented on a PC platform with an Apple M1 processor, with a matrix size of 32x32 elements and a Fourier transform length of 512.

Table 1. *Results of measuring the execution time of algorithms on a PC*

Algorithms	Optimization methods and their average algorithm execution time						
	C	C1	K(1)	C2	K(2)	C3	K(3)
Matrix n3	4 µs	4 µs	1	4 µs	1	3 µs	0.75
Grapes n2,5	3 µs	2 µs	0.67	3 µs	1	2 µs	0.67
FFT	16 µs	14 µs	0,86	3 µs	0,19	3 µs	0,19
Wavelet	7 µs	5 µs	0,71	6 µs	0,86	5 µs	0,71

The results shown in Table 1 demonstrate a reduction in execution time for some algorithms when applying different optimization methods: in particular, for the FFT transform, a gradual decrease in time is observed (0.19 ≤ 0.19 < 0.86, i.e., K((3) <) (K) ((2) () <) (K) (1)), and for the classical matrix multiplication algorithm (Matrix n3), there is an improvement when moving to fixed-comma (C2) and hybrid optimization (C3). Instead, for the Wavelet transform and the Winograd n2,5matrix multiplication algorithm, the optimizations did not significantly reduce the execution time on the used hardware platform.

Table 2 shows the results of measuring the execution time of the considered algorithms in microseconds (µs) and the corresponding averaged K1/K2/K(3) for different C1/C2/C3 optimization methods when implemented on the Raspberry Pi platform with an ARM Cortex A53 core for matrix sizes of 32x32 elements, FFT of 512 elements.

Table 2. *Results of measuring the execution time of algorithms on the Raspberry Pi platform*

Algorithms	Optimization methods and their average algorithm execution time						
	C	C1	K(1)	C2	K(2)	C3	K(3)
Matrix n3	406 µs	395 µs	0,97	310 µs	0,76	290 µs	0,71
Grapes n2,5	303 µs	303 µs	1	263 µs	0,87	243 µs	0,81
FFT	719 µs	718 µs	1	51 µs	0,07	48 µs	0,07
Wavelet	61 µs	61 µs	1	73 µs	1,2	73 µs	1,2

The results shown in Table 2 demonstrate a significant improvement in the performance of some algorithms on the Raspberry Pi platform due to the use of various optimization methods.

The most pronounced effect is observed for the FFT transform: the algorithm execution time decreases from 719 µs to 48 µs (respectively, K1= 1, K(3) = 0.07).

For the classical Matrix n3matrix multiplication, an improvement was also recorded - 406 µs versus 290 µs (respectively, K1= 0.97, K3= 0.71), as well as for the Winograd n2.5algorithm - 303 µs versus 243 µs (K1= 1, K3= 0.81).

Instead, for the Wavelet transform, the execution time increased from 61 µs to 73 µs (K1= 1, K3= 1.2), which indicates the inefficiency of optimization in this case.

Table 3 shows the results of measuring the execution time of the considered algorithms in microseconds (µs) and the corresponding averaged execution times K1/K2/K(3) for different methods of optimizing C1/C2/C3 when implemented on the Zynq-7000 platform, namely on its PS part, with a matrix size of 128×128 elements and a Fourier transform length of 8192.

Table 3. *Results of measuring the execution time of algorithms on the PS part of ZYNQ*

Algorithms	Optimization methods and their average algorithm execution time						
	C	C1	K(1)	C2	K(2)	C3	K(3)
Matrix n3	35380 µs	34845 µs	0,98	28786 µs	0,81	25496 µs	0,72
Grapes n2,5	36666 µs	35818 µs	0,97	31129 µs	0,85	23466 µs	0,64
FFT	12075.6 µs	12076.1 µs	1	1550 µs	0,13	1550 µs	0,13
Wavelet	754.1 µs	755.1 µs	1	1867.5 µs	2,5	1867.4 µs	2,5

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

The results shown in Table 3 show the effectiveness of various ways to optimize the code of computational algorithms on the PS part of ZYNQ. In particular, for the FFT transformation, we observe a significant reduction in execution time from 12076.1 μs to 1550 μs (respectively, K1= 1, K3= 0.13), and for the classical matrix multiplication algorithm Matrix n(3),we see a moderate improvement (K1= 0.98, K(3) = 0.72) when switching to fixed-point calculations and hybrid optimization. The Winograd n2.5algorithm also demonstrates a slight decrease in time from 35818 μs to 23466 μs (respectively K1= 0.97, K3= 0.64), while for Wavelet transforms, optimizations proved to be ineffective - the execution time even increased (K1= 1, K3= 2.5).

Table 4 shows the results of measuring the execution time of the considered algorithms in microseconds (μs) and the corresponding averaged K1/K2/K(3) for different methods of optimizing C1/C2/C3 when implemented on the Zynq-7000 platform, namely on the PL part, with a matrix size of 24x24 elements and an FFT size of 512 elements.

Table 4. *Results of measuring the execution time of algorithms on the PL part of ZYNQ*

Algorithms	Optimization methods and their average algorithm execution time						
	C	C1	K(1)	C2	K(2)	C3	K(3)
Matrix n3	41.192 μs	40.128 μs	0,97	39.14 μs	0,95	39.54 μs	0,96
Grapes n2,5	41.118 μs	41.118 μs	1	39.54 μs	0,96	39.54 μs	0,96
FFT	1500.97 μs	1500.7 μs	1	1325.8 μs	0,88	1288.8 μs	0,86
Wavelet	324.066 μs	339.465 μs	1,05	225.97 μs	1	319.15 μs	0,98

The results presented in Table 4 show that for the FFT transformation, there is a moderate decrease in the algorithm execution time from 1500.9 μs to 1288.8 μs (respectively, K1= 1, K(3) = 0.86), which indicates a limited effect of different code optimization methods. For the classical Matrix n3matrix multiplication and the Winograd n2.5algorithm, the changes are insignificant - the time is reduced by only a few microseconds, and the improvement factors remain close to one (for example, for Matrix n^3 K1= 0.97 and K(3) = 0.96). The most noticeable improvement is observed for the wavelet transform - the algorithm execution time decreases from 324.07 μs to 225.97 μs (C2), which corresponds to a decrease in the coefficient from K1= 1.05 to K(2) = 1. However, in general, all optimization methods have a less pronounced effect in the PL part compared to the PS part.

Table 5 summarizes the time characteristics of different C-code optimization methods M1/M2/M(3) for different algorithms on the corresponding hardware implementation platforms, namely, on a PC with an Apple M1 processor, on a Raspberry Pi single-board computer with an ARM Cortex A53 core, and on the PS and PL parts of the Zynq-7000 SoC.

Table 5 *Summary of time characteristics of different code optimization methods*

Hardware platform	Matrix n3			FFT			Wavelet		
	M(1)	M(2)	M(3)	M(1)	M(2)	M(3)	M(1)	M(2)	M(3)
PC	0,88	0,81	0,63	0,93	0,3	0,3	0.64	0,85	0,69
Raspberry Pi	1	0,715	0,68	0,99	0,078	0,078	1	1,15	1,15
PS part of Zynq	0,99	0,87	0,68	1	0,083	0,083	1	2,55	2,55
PL part of Zynq	0,98	0,96	0,95	1	0,86	0,85	1,05	1	0,98

To visualize the results shown in Table 5, let's draw graphs.

Figure 1 shows the dependence of the normalized execution time of the algorithms on different ways of optimizing the C++ code for matrix multiplication and Fourier transform implemented on the PS part of the ZYNQ SoC, Figure 2 - on a PC with an Apple M1 processor, and Figure 3 - on the PL part of the ZYNQ SoC.

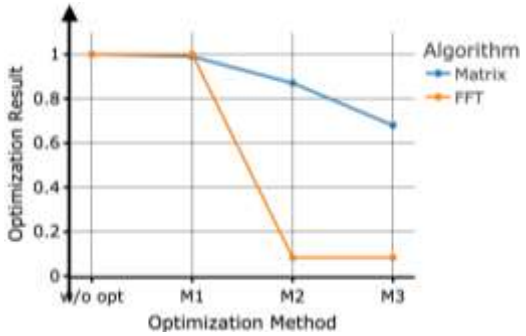


Fig. 1. Normalized execution time of algorithms for different ways of optimizing C/C++ code for the PS part of Zynq

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Figure 1 shows that the optimization with the transition to fixed-point calculations on the Zynq PS platform significantly reduces the execution time of the FFT transformation (orange graph), while matrix multiplication remains almost unchanged (blue graph).

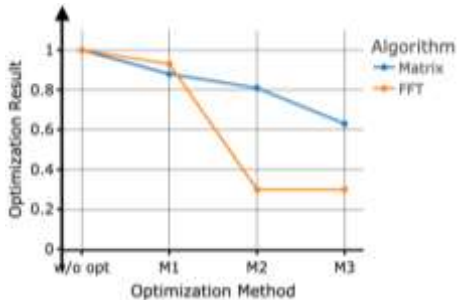


Fig. 2. Normalized execution time of algorithms for different ways of optimizing C/C++ code for PC

Figure 2 demonstrates similar behavior when implemented on a PC with an ARM M1 processor, where the transition to fixed point significantly reduces the execution time of the algorithms (orange graph).

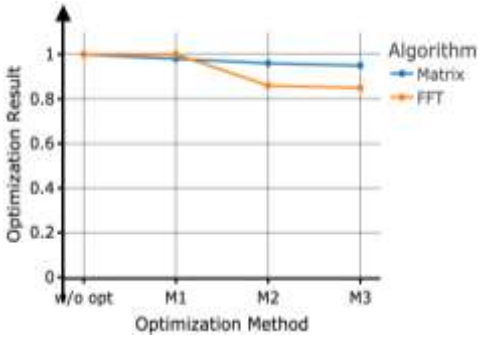


Fig. 3. Normalized algorithm execution time for different ways of optimizing C/C++ code for the PL part of ZYNQ

Figure 3 shows that when implementing algorithms on the PL part of Zynq, code optimization almost did not affect the execution time of the algorithms. A slight reduction in execution time is present for the FFT transform (orange graph).

Let's analyze the hardware costs of implementing various algorithms on the PL part of the Zynq-7000 SoC using different ways of optimizing the code for two input data options:

- option 1: matrix size 24x24 and Fourier transform length 512;
- option 2: matrix size 32x32 and Fourier transform length 4096.

The data in the cells is given as a percentage (%) of the total number of corresponding primitives available on the ZYNQ 7000, namely

- number of FPGA basic logic units (Look-Up-Table, LUT): 53200;
- number of synchronous memory elements (Flip-Flop, FF): 106400;
- number of digital signal processing units (Digital Signal Processing DSP): 220;
- number of block memory (Block Random Access Memory, BRAM): 140.

Table 6 shows the absolute (Abs column) and normalized (Norm column) hardware costs in % when implementing algorithms on the PL part of the Zynq-7000 SoC using different ways of optimizing the C1/C2/C3 code. The size of the matrix is 24x24 elements, the length of the Fourier transform is 512.

Table 7 shows the absolute (Abs column) and normalized (Norm column) hardware costs in % when implementing the algorithms on the PL part of the Zynq-7000 SoC using different ways to optimize the C1/C2/C3 code. The size of the matrix is 32x32, the length of the Fourier transform is 4096.

Figures 4 - 5 show the normalized results of hardware costs for the implementation of the Fourier transform of 512 samples and 4096 samples, respectively, on the PL part of the ZYNQ for all types of available ZYNQ resources, namely LUTs, FFs, DSP blocks and built-in BRAM.

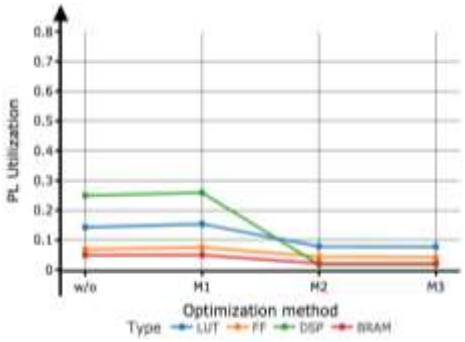


Figure 4. Normalized results of hardware costs for the implementation of the Fourier transform with a length of 512 samples

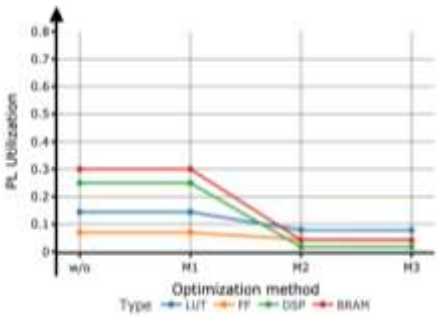


Fig. 5. Normalized results of hardware costs for implementing the Fourier transform with a length of 4096 samples

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Table 6. *Hardware costs for different optimization methods on the ZYNQ PL part (option 1)*

Methods of of optimization	Resources PL ZYNQ	Algorithms							
		Matrix n3		Grapes n2,5		FFT		Wavelet	
		Abs	*Norm*	*Abs*	*Norm*	*Abs*	*Norm*	*Abs*	*Norm*
C	LUT	7933	0.149	14477	0.0272	7604	0.143	18927	0.356
	FF	10063	0.095	7133	0.161	7362	0.069	20712	0.195
	DSP	120	0.545	108	0.491	55	0.25	89	0.405
	BRAM	52	0.371	54	0.386	7	0.05	31	0.221
C1	LUT	7984	0.15	14478	0.0272	8194	0.154	26442	0.497
	FF	10149	0.095	17133	0.161	7996	0.075	29158	0.274
	DSP	120	0.545	108	0.491	57	0.259	120	0.545
	BRAM	52	0.371	54	0.386	7	0.05	45	0.321
C2	LUT	2438	0.046	3629	0.068	4178	0.079	15307	0.288
	FF	2977	0.028	3353	0.032	4684	0.044	9368	0.088
	DSP	72	0.327	108	0.491	4	0.018	30	0.136
	BRAM	52	0.371	54	0.386	3	0.021	19	0.136
C3	LUT	2438	0.046	3630	0.068	4122	0.077	15785	0.297
	FF	2977	0.028	3353	0.032	4504	0.042	9753	0.092
	DSP	72	0.327	108	0.491	4	0.018	30	0.136
	BRAM	52	0.371	54	0.386	3	0.021	19	0.136

Table 7. *Hardware costs for different optimization methods on the ZYNQ PL part (option 2)*

Methods of of optimization	Resources PL ZYNQ	Algorithms					
		Matrix n3		Grapes n2,5		FFT	
		Abs	*Normal*	*Abs*	*Norm*	*Abs*	*Norm*
C	LUT	10057	0.189	18758	0.353	7666	0.144
	FF	12630	0.119	22068	0.207	7454	0.07
	DSP	160	0.727	144	0.655	55	0.25
	BRAM	68	0.486	70	0.5	42	0.3
C1	LUT	10062	0.189	18762	0.353	7666	0.144
	FF	12713	0.119	22068	0.207	7454	0.07
	DSP	160	0.727	144	0.655	55	0.25
	BRAM	68	0.486	70	0.5	42	0.3
C2	LUT	2734	0.051	4366	0.082	4190	0.079
	FF	3277	0.031	3659	0.034	4690	0.044
	DSP	96	0.436	144	0.655	4	0.018
	BRAM	68	0.486	70	0.5	6	0.043
C3	LUT	2734	0.051	4366	0.082	4153	0.078
	FF	3277	0.031	3659	0.034	4510	0.042
	DSP	96	0.436	144	0.655	4	0.018
	BRAM	68	0.486	70	0.5	6	0.021

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Figures 4 and 5 show that for the implementation of the Fourier transform, the transition to fixed-point calculations significantly reduces hardware costs.

Figures 6 and 7 show the normalized results of hardware costs for the implementation of the Winograd algorithm for 24x24 and 32x32 matrices, respectively, on the PL part of ZYNQ for all types of available ZYNQ resources, namely LUTs, FFs, DSP blocks, and on-chip BRAM.

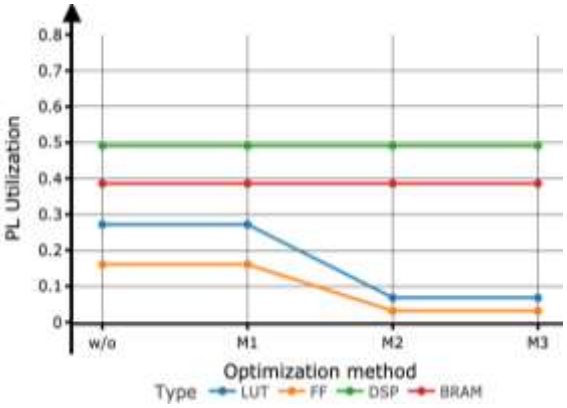


Fig. 6. Normalized results of hardware costs for implementing the Winograd algorithm for 24x24 matrices

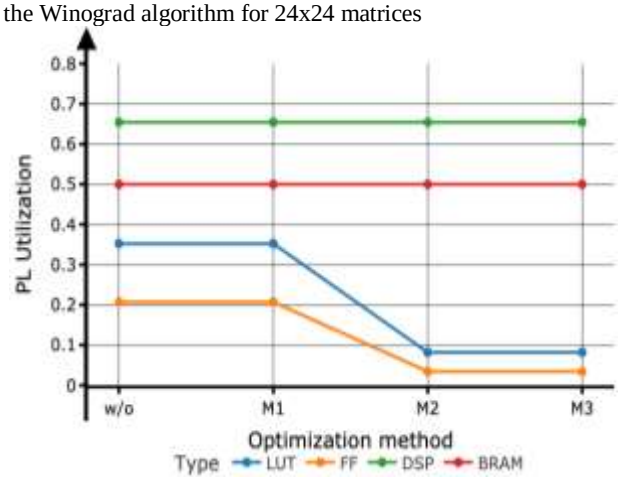


Fig. 7. Normalized results of hardware costs for implementing the Winograd algorithm for 32x32 matrices

When analyzing Figures 6 and 7, it should be noted that in the case of implementing matrix multiplication by the Winograd algorithm, a significant reduction occurs only in terms of the costs of such PL primitives as LUT and FF, while optimization does not affect DSP blocks and the use of internal BRAM memory.

Figure 8 shows the normalized results of hardware costs when implementing the Wavelet transform using the wave function db4, decomposition level 5 and input sequence length 256 on the ZYNQ PLU for all types of available ZYNQ resources, namely LUTs, FFs, DSP blocks and internal BRAM memory.

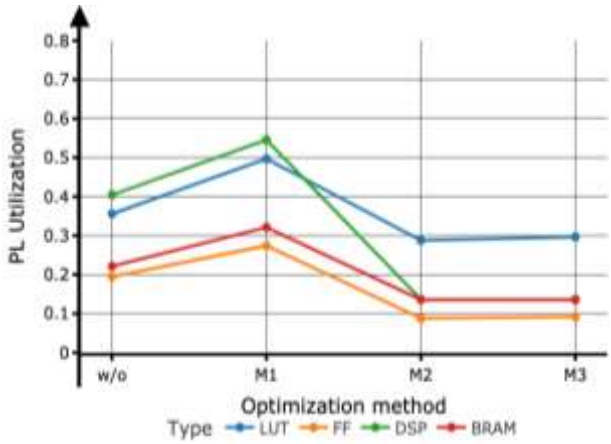


Fig. 8. Normalized results of hardware costs for Wavelet transform implementation

Figure 8 shows that optimization with loop deployment alone leads to an increase in hardware costs for implementing the Wavelet transform. However, the use of a fixed-comma switch also leads to a reduction in hardware costs.

Research results and discussion

After analyzing the research results from the point of view of the hardware platform, we can draw the following conclusions.

1. When using an Apple PC with an ARM M1 processor, all optimization methods for all algorithms work with a reduction in execution time from 30% to 70%.

2. When using a single-board computer Raspberry Pi and the PS part of the Zynq-7000 SoC, the time reduction is approximately the same (this is due to the Cortex processor cores underlying both platforms). For the Fourier transform, it is 90%, for matrices - 30%, for the wavelet transform, there is no time reduction.

3. When using the PL part of the Zynq-7000 SoC, the time reduction is observed only for the Fourier transform and not more than 15%.

From the point of view of the most highly loaded algorithms, the following can be noted.

1. For the group of matrix multiplication algorithms that use only arithmetic operations of addition, subtraction, and multiplication, all optimization methods work approximately the same and provide a time reduction of 5% to 35%.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

2. The group of algorithms related to series decomposition is divided into 2 parts: the Fourier transform and the wavelet transform, which are differently affected by the code optimization methods under study. Optimization of FFT algorithms for processor devices provides 70% to 90% time reduction. For the wavelet transform, the stated optimization methods do not work, except for the PC implementation (30% on average).

3. The hardware costs for all algorithms implemented on the PL part of the SoC increase linearly with the size of the implementation objects.

We summarize the general results of the analyzed code optimization methods as follows.

1. For the wavelet transform, the stated optimization methods do not work, except for the variant with the PC implementation (on average, 30% reduction in algorithm execution time).

2. The optimization method C1 (loop deployment + float type) does not have a significant effect for all algorithms of embedded systems (except for PC). It also does not affect the hardware costs of the PL part of the Zynq SoC.

3. Optimization methods C2 (basic implementation + fixed-point arithmetic instead of float type) and C3 (loop deployment + fixed-point arithmetic instead of float type) have a high effect (up to 90% reduction in execution time) for Fourier transform algorithms (FFT) and up to 35% for matrix multiplication (for embedded systems). For wavelet transform, the stated optimization methods do not work.

4. For programmable logic, the time reduction for C2 and C3 is observed only for FFT and gives no more than 15% reduction in execution time. At the same time, there is a significant reduction in hardware costs up to 80% for matrix algorithms and up to 40% for FFT algorithms.

Conclusions and prospects for further development

As part of the study, we analyzed the impact of typical C/C++ code optimization techniques to determine the impact of each on algorithm execution time and hardware costs on several target platforms and made the following general conclusions

- for algorithms based on arithmetic operations, code optimization gives the effect of reducing the execution time by up to 30% on ARM processor devices;
- for algorithms based on the Fourier transform, comprehensive code optimization gives the effect of reducing the execution time by up to 90% on processor devices;
- for programmable logic devices, code optimization (for selected methods) does not significantly reduce the time;
- the use of fixed-point arithmetic for programmable logic devices gives a significant reduction in hardware costs (from 40% to 80%) for all types of algorithms.

Thus, we have identified code optimization options that are effective in terms of reducing hardware costs when replacing the float data type with an equivalent fixed-point type (fixed_16_16) used for calculations during high-level synthesis.

It was found that in terms of performance, when using the PL part of the Zynq SoC, different ways of optimizing the code have almost no effect on the result. This may be due to the fact that the developed implementation of the algorithm should be optimized for the specific properties of the SoC hardware.

That is why one of the directions for further research may be to analyze the properties of standard HLS-IP cores of the PL-device for use in high-level synthesis.

References

1. Tratt, L. (2025), "Four kinds of optimisation", available at: https://tratt.net/laurie/blog/2023/ four_kinds_of_optimisation.html (last accessed: 10.02.2025).
2. Xu, K., Zhang, G. L., Yin, X., Zhuo, C., Schlichtmann U., Li B. (2024), "Automated C/C++ program repair for high-level synthesis via large language models", *ACM/IEEE international symposium on machine learning for CAD (MLCAD '24)*, 09-11 September 2024, Salt Lake City, USA, P. 1–9. DOI: https://doi.org/10.1109/mlcad62225.2024.10740262
3. Licht, J. de Fine, Besta, M., Meierhans, S., Hoefler, T. (2021), “Transformations of high-level synthesis codes for high-performance computing”, *IEEE Transactions on Parallel and Distributed Systems*, 2021,Vol. 32 (No. 5), P. 1014–1029. DOI: https://doi.org/10.1109/tpds.2020.3039409

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) *ISSN 2524-2296 (online)*

4. Ahmad A., Du L., Zhang W. (2024), "Fast and practical Strassen's matrix multiplication using FPGAs", *34th International Conference on Field-Programmable Logic and Applications (FPL'24)*, 2-6 September 2024, Torino, Italy, P. 311–317. DOI: https://doi.org/10.1109/fpl64840.2024.00050
5. Khan G. N., Iniewski K. (2017), Embedded systems code optimization and power consumption: chapter in *Embedded and Networking System*, 2017, P. 97–116. DOI: https://doi.org/10.1201/b15497-8
6. Almorin, H., Gal, B. L., Crenne, J., Jego, C., Kissel, V. (2022), "High-throughput FFT architectures using HLS tools", *29th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 24–26 October 2022, Glasgow, United Kingdom, P. 1–4. DOI: https://doi.org/10.1109/ icecs202256217.2022.9970886
7. Gayathri, S. (2022), "Improved FIR filter using Schonhage-Strassen algorithm based multipliers", *International Journal of Science and Research (IJSR)*, 2022, Vol. 11 (No. 11), P. 1103–1106. DOI: https://doi.org/10.21275/sr221120132156
8. Juang, W.-H., Wu, M.-C., Sheu, Y.-H., Shieh, J.-Y., Hsieh, T.-H. (2023), "A cost-efficient hardware accelerator design for 2D sliding discrete fourier transform", *International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, 17-19 July 2023, PingTung, Taiwan, P. 595–596. DOI: https://doi.org/10.1109/icce-taiwan58799.2023.10227037
9. Kumar, A. (2015), "New trends and challenges in source code optimization", *International Journal of Computer Applications*, 2015, Vol. 131(No. 16), P. 27–32. DOI: https://doi.org/10.5120/ijca2015907609
10. Lee, Y., Youn, J., Nam, K., Oh, H., Paek, Y. (2023), "Optimizing hardware resource utilization for accelerating the NTRU-KEM algorithm", *Computers*, 2023, Vol. 12 (No. 259), P. 1–14. DOI: https://doi.org/10.3390/computers12120259
11. Zhao, R., Cheng, J., Luk, W., Constantinides, G. A. (2022), "POLSCA: polyhedral high-level synthesis with compiler transformations", *32nd International Conference on Field-Programmable Logic and Applications (FPL)*, 29 August – 2 September 2022, Belfast, United Kingdom, P. 235–242. DOI: https://doi.org/10.1109/fpl57034.2022.00044
12. Qian, X., Shi, J., Shi, L., Zhang, H., Bian, L., Qian, W. (2022), "Scheduling information-guided efficient high-level synthesis design space exploration", *IEEE 40th International Conference on Computer Design (ICCD)*, 23-26 October 2022, Olympic Valley, USA, P. 203–206. DOI: https://doi.org/10.1109/iccd56317.2022.00038
13. Hong, H., Xiao, C., Wang, S. (2024), "Rethinking high-level synthesis design space exploration from a contrastive perspective", *42nd International Conference on Computer Design (ICCD)*, 18-20 November 2024, Milan, Italy, P. 179–182. DOI: https://doi.org/10.1109/iccd63220.2024.00035
14. Ferikoglou, A., Kakolyris, A., Kypriotis, V., Masouros, D., Soudris, D., Xydis, S. (2023), "Data-driven HLS optimization for reconfigurable accelerators", *61st ACM/IEEE Design Automation Conference*, 23-27 June 2023, San Francisco, USA, P. 1–16. DOI: https://doi.org/10.1145/3649329.3658471
15. Basalama, S., Cong, J. (25), "Stream-HLS: towards automatic dataflow acceleration", *ACM/SIGDA International Symposium on Field Programmable Gate Arrays (FPGA '25)*, 27 February 2025 – 1 March 2025, Monterey, USA, P. 103–114. DOI: https://doi.org/10.1145/3706628.3708878
16. Si, Q., Schaefer, C. B. (2023), "ADVICE: automatic design and optimization of behavioral application specific processors", *Great Lakes Symposium on VLSI (GLSVLSI'23)*, 5-7 June 2023, Knoxville, USA, P. 327–332. DOI: https://doi.org/10.1145/3583781.3590214

Надійшла (Received) 22.05.2025

Відомості про авторів / About the Authors

Shkil Oleksandr − PhD, associated professor, associated professor of design automation department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; e-mail: oleksandr.shkil@nure.ua; ORCID ID: https://orcid.org/0000-0003-1071-3445
Filippenko Oleh − PhD, associated professor, associate professor of infocommunication engineering department named by V.V. Popovsky, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; e-mail: oleh.filippenko@nure.ua; ORCID ID: https://orcid.org/0000-0003-4616-250X
Rakhlis Dariia − PhD, associated professor, associated professor of design automation department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; e-mail: dariia.rakhlis@nure.ua; ORCID ID: https://orcid.org/0000-0002-6652-1840
Filippenko Inna − PhD, associated professor, associated professor of design automation department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; e-mail: inna.filippenko@nure.ua; ORCID ID: https://orcid.org/0000-0002-3584-2107
Korniienko Valentyn – PhD student of design automation department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; e-mail: valentyn.korniienko1@nure.ua; ORCID ID: https://orcid.org/0000-0001-7070-5127

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

Шкіль Олександр Сергійович – кандидат технічних наук, доцент, доцент кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Харків, Україна.
Філіпенко Олег Ігорович – кандидат технічних наук, доцент кафедри інфокомунікаційної інженерії ім. В.В. Поповського, Харківський національний університет радіоелектроніки, Харків, Україна.
Рахліс Дарія Юхимівна – кандидат технічних наук, доцент, доцент кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Харків, Україна.
Філіпенко Інна Вікторівна – кандидат технічних наук, доцент кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Харків, Україна.
Корнієнко Валентин Русланович – аспірант кафедри автоматизації проектування обчислювальної техніки, Харківський національний університет радіоелектроніки, Харків, Україна.

ОПТИМІЗАЦІЯ ПРОГРАМНОГО КОДУ ДЛЯ ВИСОКОРІВНЕВОГО СИНТЕЗУ ПРИ АПАРАТНІЙ РЕАЛІЗАЦІЇ ОБЧИСЛЮВАЛЬНО-НАВАНТАЖЕНИХ АЛГОРИТМІВ

Предметом дослідження є вплив методів оптимізації коду високонавантажених алгоритмів, що застосовуються у цифровій обробці сигналів, на апаратні витрати та швидкодію при реалізації на різних платформах. **Мета.** Порівняльний аналіз впливу трьох підходів до оптимізації С-коду, а саме розгортання циклів, перехід до арифметики з фіксованою комою та їх комбінації, на ефективність реалізації алгоритмів множення матриць, швидкого перетворення Фур'є та вейвлет-перетворення за допомогою засобів високорівневого синтезу (HLS) на платформі SoC, персональних комп'ютерах (ПК) та одноплатних комп'ютерів. У статті вирішуються такі **завдання**: реалізація високонавантажених алгоритмів на базі обраних апаратних платформ та з використанням HLS; порівняння часу виконання алгоритмів із застосуваннями трьох підходів до оптимізації та без; порівняння апаратних витрат для реалізацій алгоритмів з різними варіантами оптимізації коду та без; сформулювати висновки про вплив різних способів оптимізації С-коду на швидкодію та апаратні витрати на різних цільових платформах. Використовуються такі **методи**: методи оптимізації C/C++ коду, діагностичний експеримент за допомогою інструментарію високорівневого синтезу для реалізації алгоритмів цифрової обробки сигналів на обраній апаратній платформі та збору статистичних даних з використанням Python. **Результати.** Для алгоритмів на основі арифметичних операцій оптимізація коду забезпечила до 30% зменшення часу виконання на ARM-платформах. Для алгоритмів на основі перетворення Фур'є комплексна оптимізація дозволила скоротити час виконання до 90% на процесорних пристроях. Для програмованої логіки (FPGA) жоден з методів оптимізації не забезпечив значного прискорення виконання, однак перехід до фіксованої арифметики зумовив зменшення апаратних витрат на 40–80% незалежно від типу алгоритму. **Висновки.** Вибір стратегії оптимізації С-коду має суттєвий вплив на ефективність реалізації алгоритмів на процесорних архітектурах, тоді як для FPGA ключову роль відіграє оптимізація використаних типів даних. Отримані результати можуть слугувати практичними рекомендаціями для проєктування вбудованих систем із застосуванням HLS з метою прискорення алгоритмів.

Ключові слова: вбудовані системи; високорівневий синтез; оптимізація коду С; система на кристалі.

Бібліографічні описи / Bibliographic descriptions

Шкіль О.С., Філіпенко О.І., Рахліс Д.Ю., Філіпенко І.В., Корнієнко В.Р. Оптимізація програмного коду для високорівневого синтезу при апаратній реалізації обчислювально-навантажених алгоритмів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 189–202. DOI: https://doi.org/10.30837/2522-9818.2025.3.189

Shkil, O., Filippenko, O., Rakhlis, D. Filippenko, I., Korniienko, V. (2025), "Optimization of software code for high-level synthesis during hardware implementation of the computationally-loaded algorithms", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 189–202. DOI: https://doi.org/10.30837/2522-9818.2025.3.189

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 681.3:621.8:004.94
DOI: https://doi.org/10.30837/2522-9818.2025.3.203

А. Пастушенко, Л. Коваль

КЕРУВАННЯ РОБОТОМ-МАНІПУЛЯТОРОМ ЗА ДОПОМОГОЮ СИГНАЛІВ ПОВЕРХНЕВОЇ ЕЛЕКТРОМІОГРАФІЇ

Предметом дослідження є методи керування роботизованим маніпулятором на основі сигналів поверхневої електроміографії (ЕМГ) із застосуванням індивідуалізованої нормалізації та класифікації жестів згортковою нейронною мережею (CNN). **Мета роботи** – створення системи управління, що поєднує попередню нормалізацію ЕМГ-сигналів та глибоке навчання для підвищення точності розпізнавання жестів і стабільності в реальному часі. Для досягнення поставленої мети виконано такі **завдання**: здійснено збір ЕМГ-сигналів за допомогою міобраслета *Myo Armband*, проведено їх попереднє оброблення з використанням мінімакс-нормалізації та нормалізації з нульовим середнім, реалізовано перетворення сигналів у зображення за допомогою ковзного вікна, після чого побудовано та навчено згорткову нейронну мережу на основі архітектури *ResNet* з оптимізацією алгоритмом *Adam*, а також проаналізовано та порівняно точність CNN та класичних моделей машинного навчання (SVM і *Random Forest*). **Досягнуті результати** показали точність класифікації 97,27% у тестовому середовищі та 91,71% – у реальному часі. Модель CNN перевищила традиційні методи на 18–19%, а нормалізація нульового середнього підвищила точність на 2,34% порівняно з мінімакс-нормалізацією. Система зберігала стабільність навіть за умови варіацій положення браслета завдяки індивідуальній нормалізації. **Висновки.** Запропонована система продемонструвала високу точність, надійність і адаптивність у реальному часі. Наукова новизна полягає в поєднанні індивідуалізованої нормалізації ЕМГ-сигналів із *ResNet*, що забезпечує стабільність і перевищує точність традиційних алгоритмів. У майбутньому заплановано розширення набору жестів, дослідження більш складних умов та оптимізація нейронних мереж для вбудованих систем.

Ключові слова: електроміографія; керування жестами; роботизований маніпулятор; згорткова нейронна мережа; *ResNet*; нормалізація даних; біонічні протези; управління в реальному часі.

Вступ

Сучасний розвиток робототехніки та біомедичних технологій активно спрямований на вдосконалення інтерфейсів "людина – машина", які забезпечують інтуїтивне та ефективне управління роботизованими пристроями. Одним із перспективних напрямів є використання сигналів поверхневої електроміографії (ЕМГ) для керування роботами-маніпуляторами, що дає змогу відтворювати природні рухи людини за допомогою електричних сигналів м’язової активності. Проте точне й стабільне розпізнавання жестів на основі ЕМГ залишається викликом через індивідуальні особливості користувачів, варіативність сигналів і вплив зовнішніх факторів.

Одним із шляхів підвищення точності класифікації ЕМГ-сигналів є застосування глибоких нейронних мереж, зокрема згорткових нейронних мереж (CNN), які здатні ефективно виділяти складні просторово-часові властивості сигналів. Водночас індивідуалізована попередня нормалізація сигналів відіграє важливу роль у зниженні впливу шумів і варіацій, що підвищує стабільність та якість роботи системи. Поєднання цих двох підходів відкриває нові можливості для створення більш точних і адаптивних систем керування роботизованими маніпуляторами.

У цьому дослідженні поставлено за мету розробити та дослідити метод управління роботом-маніпулятором, оснований на поєднанні індивідуалізованої попередньої нормалізації ЕМГ-сигналів і глибокої згорткової нейронної мережі. Реалізація такого підходу має забезпечити підвищену точність класифікації жестів і стабільність роботи системи в реальному часі, що є важливим кроком для практичного впровадження біонічних пристроїв і протезування.

Мета й завдання

Постановка мети полягає в розробленні системи управління роботизованим маніпулятором на основі сигналів поверхневої електроміографії, яка забезпечує високу точність і стабільність розпізнавання жестів у реальному часі. Для цього необхідно дослідити вплив різних методів індивідуалізованої попередньої нормалізації ЕМГ-сигналів на якість класифікації, а також розробити та навчити ефективну згорткову нейронну мережу, здатну адаптуватися до особливостей

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

кожного користувача. Крім того, потрібно проаналізувати й порівняти досягнуті результати з традиційними методами машинного навчання, щоб обґрунтувати переваги запропонованого підходу.

Аналіз сучасних публікацій

У сучасних дослідженнях активно вивчаються методи керування роботизованими маніпуляторами за допомогою сигналів поверхневої електроміографії (ЕМГ) у поєднанні з глибоким навчанням, зокрема згортковими нейронними мережами (CNN). Ці підходи дають змогу створювати інтуїтивні та точні інтерфейси для керування протезами й реабілітаційними пристроями.

Guo et al. [1] запропонували легку згорткову нейронну мережу (Lw-CNN) для розпізнавання ЕМГ-сигналів і реального часу керування роботом-маніпулятором. Модель досягла середньої точності на 90% в офлайн-режимі та 84% у режимі онлайн, що демонструє її ефективність у розпізнаванні намірів руху верхньої кінцівки.

Wan et al. [2] розробили систему керування біонічним маніпулятором на основі CNN, яка забезпечує точність класифікації понад 91% для здорових осіб і близько 89% для ампутованих. Система працює в реальному часі із затримкою менше ніж 100 мс, що робить її придатною для практичного застосування.

У дослідженні *Bao et al.* [3] створено гібридну модель CNN-LSTM для оцінювання кінематики зап'ястя на основі ЕМГ-сигналів. Поєднання CNN для просторового аналізу та LSTM для оброблення часових залежностей дало змогу покращити точність оцінювання складних рухів зап'ястя.

Проєкт, поданий на *Hackster.io* [4], демонструє реальне впровадження CNN на мікроконтролері *Sony Spresense* для керування біонічною рукою в реальному часі. Система використовує *Myo Armband* для збору ЕМГ-сигналів та забезпечує точне розпізнавання жестів із низькою затримкою.

У роботі *Meng et al.* [5] запропоновано метод компенсації руху на основі sEMG для реабілітаційного роботизованого пристрою верхньої кінцівки. Система забезпечує активну участь пацієнта в тренуванні, адаптуючи допоміжну силу відповідно до інтенсивності ЕМГ-сигналів.

Atzori et al. [6] розробили відкритий набір даних з ЕМ-сигналами для керування роботизованими протезами рук. Ця інформація містить різноманітні жести та учасників, що дає змогу досліджувати універсальні методи розпізнавання намірів руху. Застосування такого набору даних сприяє стандартизації тестування алгоритмів та порівнянню ефективності різних моделей машинного навчання в завданнях керування біонічними пристроями.

Banzi & Shiloh [7] описали платформу *Arduino* як універсальне середовище для прототипування електронних систем. У контексті керування роботизованими маніпуляторами *Arduino* використовується для інтеграції сенсорів, таких як ЕМГ-браслети, з виконавчими механізмами, забезпечуючи швидку реалізацію та тестування прототипів систем управління.

Patel & Patel [9] досліджують методи нормалізації даних у задачах дата-майнінгу та показують їх значний вплив на продуктивність алгоритмів. У контексті роботи це підтверджує необхідність застосування мінімаксу та нульового середнього для оброблення ЕМГ-сигналів, що підвищує стабільність і точність класифікації жестів.

Bhandari в праці [10] детально пояснює концепцію стандартного нормального розподілу та його застосування в статистичному аналізі. Це знання корисне в обробленні ЕМГ-сигналів, оцінюванні відхилень та інтерпретації результатів нормалізації, що впливає на ефективність моделей глибокого навчання.

Сучасні публікації [11–18] свідчать, що методи керування роботизованими маніпуляторами на основі сигналів поверхневої електроміографії (ЕМГ) у поєднанні з глибокими згортковими нейронними мережами (CNN) є перспективним напрямом досліджень. Наявні роботи демонструють можливість створення високоточних систем розпізнавання жестів і намірів руху в режимі реального часу, що підтверджується результатами з точністю класифікації від 84 до понад 91%. Поєднання CNN з іншими архітектурами, такими як LSTM, дає змогу ефективно зважати як на просторові, так і на часові властивості ЕМГ-сигналів, що покращує якість оцінки складних рухів. Практична реалізація таких систем, зокрема на вбудованих платформах, показує їх потенціал для застосування в реабілітації та протезуванні, забезпечуючи низьку затримку та стабільність роботи. Загалом, сучасні дослідження підтверджують доцільність використання глибокого навчання для підвищення точності та швидкодії систем керування роботизованими маніпуляторами на основі ЕМГ.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Опис системи

Система збирання та оброблення сигналів поверхневої електроміографії основана на міопов'язці *Myo Armband*, яка знімає восьмиканальні ЕМГ-сигнали з частотою 200 Гц. Отримані результати попередньо обробляються для стабілізації сигналів за умови зміни положення пристрою. Оброблені сигнали перетворюються в послідовні матриці у вигляді зображень, що подаються на вхід згорткової нейронної мережі, яка забезпечує класифікацію жестів. Результати застосовуються для керування роботизованим маніпулятором за допомогою мікроконтролера *Arduino* в реальному часі.

Обладнання

Для збирання електроміографічних (ЕМГ) сигналів використовувалася наручна пов'язка *Myo Armband* [6], оснащена вісьмома сухими поверхневими ЕМГ-датчиками та інерційним вимірювальним блоком із дев'ятьма ступенями свободи. Частота дискретизації сигналів становила 200 Гц. Незважаючи на те, що медичні ЕМГ-датчики зазвичай працюють на частотах не менше ніж 1 кГц, простота використання *Myo* (зокрема можливість швидкого надягання пов'язки без додаткових налаштувань) зробила цей пристрій оптимальним для нашого експерименту. Роботизований маніпулятор, застосований у дослідженні, виготовлений за допомогою 3D-друку. Конструкція передбачала два пальці, кожен з яких рухався за допомогою мотора й пластикової нитки, яка виконувала роль приводу. Для керування моторами використовувався мікроконтролер *Arduino* [7] разом із драйверами двигунів. На цьому мікроконтролері було завантажено програму, що приймала індекси жестів і відповідно до них задавала кутові положення для кожного двигуна. З'єднання між *Myo, Arduino* та комп'ютером здійснювалося через послідовний інтерфейс: комп'ютер отримував сигнали від *Myo*, обробляв їх, розпізнавав жест за допомогою нейронної мережі, після чого передавав результат (ідентифікатор жесту) до *Arduino* для відповідного керування рухом маніпулятора.

База даних

У дослідженні базою даних використовувалася *NinaPro* [8] – відкрита платформа, розроблена для сприяння вивченню сучасних міоелектричних протезів верхніх кінцівок. Ця база широко застосовується в наукових дослідженнях, пов'язаних із розпізнаванням жестів на основі ЕМГ-сигналів. Типовий підхід до збору інформації в *NinaPro* передбачає, що учасник утримує кожний жест протягом короткого проміжку часу, повторюючи його кілька разів. Для запобігання м'язовій втомі жести чергуються із фазами відпочинку. Однак, якщо жест утримується впродовж більш тривалого часу (наприклад, для моделювання реального повсякденного використання), у сигналі можуть з'являтися шуми, пов'язані із втомою м'язів, що впливає на стабільність та якість зчитуваної інформації.

Попереднє оброблення даних

У зібраних нами наборах даних використовувався один міобраслет, який у процесі кожного зчитування передавав вісім сигналів від електродів. Для оброблення інформації обрано вісім послідовних зчитувань і сформовано з них матрицю розміром 8×8, де кожен рядок відповідав одному зчитуванню з восьми каналів.

Далі ці значення масштабувалися в діапазон від 0 до 255, що давало змогу подавати отриману матрицю як зображення у відтінках сірого. Отже, кожна матриця могла бути візуалізована як "кадр" або графік ЕМГ-сигналу. Зображення формувалося поступово: перший графік містив рядки 1–8-й, потім донизу додавали новий 9-й рядок, найстаріший рядок вилучався, і в такий спосіб формувався наступний графік з 2–9-го рядків. Такий підхід створював ковзне подання сигналів у вигляді послідовних "кадрів" ЕМГ, придатних для подання в нейронну мережу, яка працює із зображеннями.

Для нормалізації оброблення графіка ЕМГ використовувалися метод мінімакс-нормалізації та нормалізації нульового середнього.

Мінімально-максимальна нормалізація – це техніка лінійної нормалізації, що широко використовується в попередньому обробленні зображень. Цей метод не змінить розподіл набору даних [9]:

$$z=\left(x-\min\left(x\right)\right)/\left(\max\left(x\right)-\min\left(x\right)\right). \quad (1)$$

Мінімакс-нормалізація може відтворювати дані ЕМГ у вигляді 0 і 1.

Нормалізація нульового середнього є одним із найпоширеніших методів нормалізації, і дані, оброблені цим методом, підходять до стандартного нормального розподілу [10]:

$$z=\left(x-\mu\right)/\sigma, \quad (2)$$

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

де μ і σ – середнє й стандартне відхилення повністю введених даних ЕМГ відповідно. Дані ЕМГ після нормалізації нульового середнього відповідають стандартному нормальному розподілу із середнім значенням 0 і стандартним відхиленням 1.

Під час тестування помітили, що амплітуда ЕМГ-сигналів змінюється щоразу, коли міопов'язку знімають і знову надягають. Через це під час кожного нового налаштування пристрою (наприклад, після ініціалізації *Myo*), ми просили користувача виконати певний жест з міткою 12 протягом 5 с. На підставі отриманих показників визначено максимальне й мінімальне значення сигналу, після чого обчислювалися середнє значення та дисперсія. Ці параметри використовувалися для подальшої нормалізації ЕМГ-сигналів, щоб забезпечити стабільність вхідної інформації для моделі незалежно від зміни положення або контакту пов'язки.

Класифікація

Основним класифікатором у дослідженні була згорткова нейронна мережа із залишковими зв'язками (*ResNet*-архітектура) [11]. Її структура містила чотири згорткових шари, один залишковий блок, а також чотири повнозв'язані шари, доповнені механізмом випадкового відключення (*dropout*), і фінальний шар *softmax*, що забезпечував як класифікацію жестів, так і обчислення ймовірностей для кожного з них. Як функція втрат використовувалась нормалізована експоненціальна функція (*softmax loss*), що дає змогу ефективно працювати з багатокласовими задачами. Навчання мережі здійснювалося за допомогою оптимізатора *Adam* [12], що використовує зворотне поширення помилки для оновлення ваг кожного шару. У повнозв'язаних шарах застосовувались ReLU-активації, які краще розв'язують проблему зникнення градієнта порівняно з традиційними сигмоїдальними чи *tanh*-функціями. Щоб уникнути перенавчання, у цих шарах також застосовувалось *dropout*-відключення з імовірністю збереження нейронів 0,5. У залишковому блоці використовувалась нормалізація екземпляра (*instance normalization*) – аналог нормалізації до нульового середнього, але з тією різницею, що середнє значення (μ) та стандартне відхилення (σ) обчислюються лише на основі поточного вхідного графіка, властивого для завдань, пов'язаних зі стилізацією зображень. Уся модель була реалізована з використанням фреймворку *TensorFlow* [13], а навчену модель збережено у вигляді бінарного файлу. Це дало змогу безпосередньо інтегрувати класифікатор у систему керування роботизованим маніпулятором із застосуванням розпізнаних жестів як команд.

Збирання навчальних даних

У збиранні даних для навчання класифікатора жестів брали участь п'ятеро здорових добровольців віком від 22 до 30 років. Усі молоді люди не мали відомих порушень нервово-м'язової системи й дали згоду на участь у дослідженні.

Перед початком експерименту кожному учаснику на передпліччя одягали міопов'язку, що забезпечувала зчитування електроміографічних (ЕМГ) сигналів із м'язів руки. Основним завданням учасників було виконання двох заздалегідь визначених жестів руки. Кожен жест повторювався по 10 разів, крім того, між повтореннями передбачалися короткі перерви для запобігання м'язової втоми, що могло б вплинути на якість сигналу.

Для кожного виконаного жесту зчитувалися ЕМГ-сигнали, які безперервно записувалися у вигляді часових серій. Щоб забезпечити точну прив'язку сигналів до конкретних жестів, час виконання жесту маркувався відповідною міткою. Отримані сигнали розбивалися на короткі часові вікна фіксованої довжини, кожне з яких розглядалося як окремий зразок для подальшого аналізу.

Для покращення якості показників проводилася нормалізація сигналів. Оскільки амплітуда ЕМГ-сигналів може змінюватися залежно від положення пов'язки та ступеня контакту зі шкірою, після кожного нового надягання пристрою учасник виконував спеціальний тестовий жест упродовж 5 с. За цією інформацією визначалися параметри нормалізації, які використовувалися для приведення вхідних сигналів до стабільного діапазону. Такий підхід давав змогу зменшити вплив зміщень і забезпечити порівняність сигналів між різними сесіями експерименту.

Отримані нормалізовані часові вікна разом із відповідними мітками жестів формували базу даних, яка застосовувалася для навчання згорткової нейронної мережі. Завдяки такій методології було забезпечено збирання репрезентативних даних із чіткою класифікацією жестів для подальшої ефективної роботи системи розпізнавання.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Аналіз результатів навчання

Нейронна мережа була навчена за допомогою 5000 ітерацій, графік результату й втрат зображений на рис. 1. Із графіків видно, що продуктивність мережі й помилки почали погіршуватися після 2000 ітерацій, з чого можна зробити висновок, що мережа навчена після 2000 ітерацій.

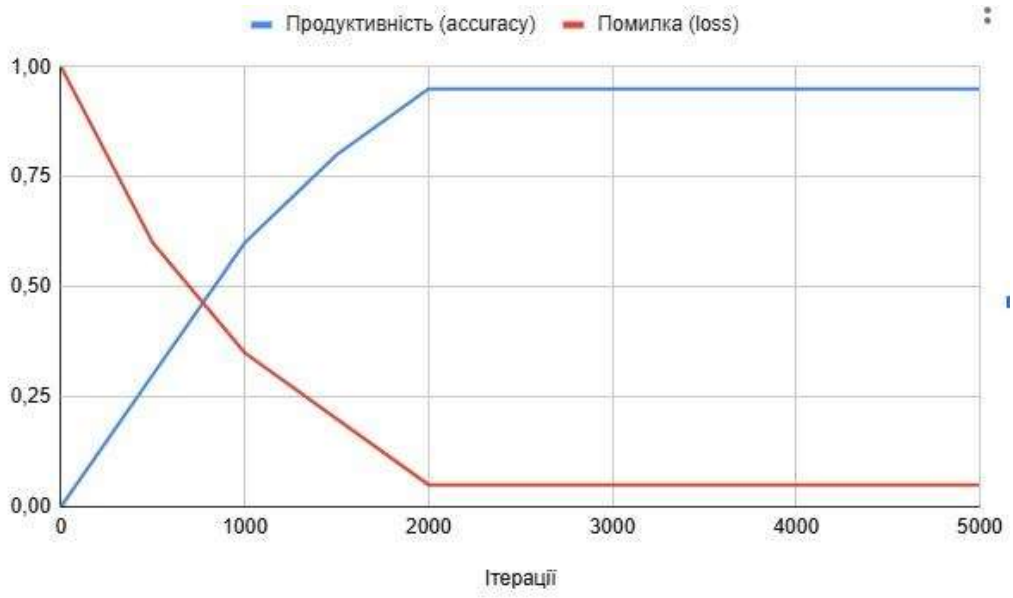


Рис. 1. Значення точності і втрат ітерацій

Оцінювання методів нормалізації

На рис. 2 продемонстровано порівняння загальної точності методів попереднього оброблення даних. Можна зробити висновок, що нормалізація нульового середнього на 2,3362% вища, ніж мінімально-максимальна нормалізація.

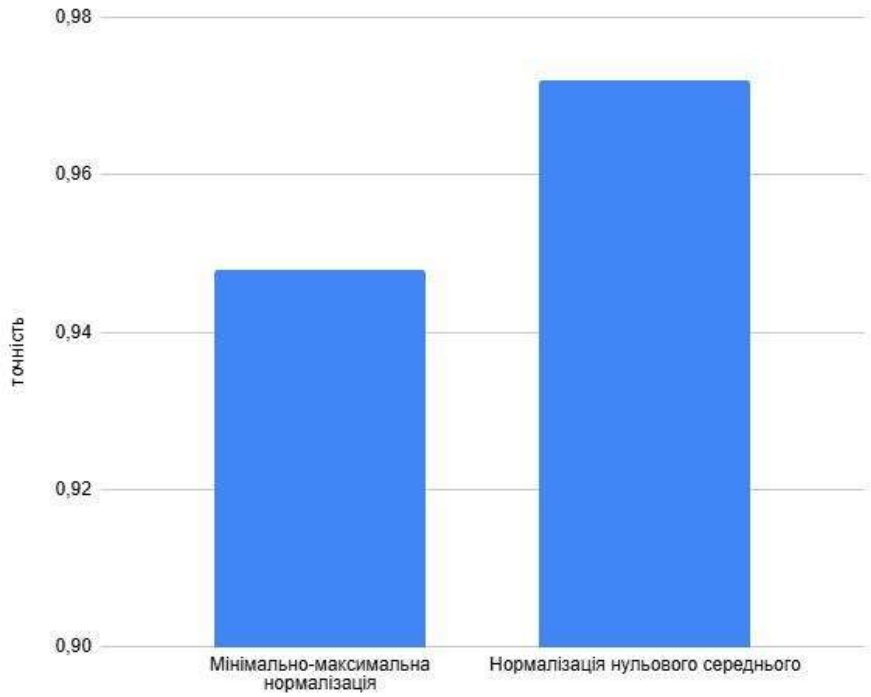


Рис. 2. Результати оцінювання точності методів нормалізації

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

Оцінювання ефективності класифікаторів

З метою порівняння різних методів машинного навчання для оцінювання ефективності було протестовано кілька алгоритмів і методів аналізу функцій.

У дослідженні використано кілька методів машинного навчання для аналізу та класифікації даних. Зокрема застосовувалися RF (випадкові ліси), TD (часова різниця), mDWT (багатовимірне дискретне перетворення вейвлетів), SVM (*Support Vector Machine*, метод опорних векторів), а також CNN (згорткова нейронна мережа).

На рис. 3 подано результати тестування різних класифікаторів. Можна зробити висновок, що метод CNN, використаний у дослідженні, демонструє на 19,39% вищу ефективність порівняно з SVM і mDWT і на 18,37% вищу, ніж RF і TD. Унаслідок тестування класифікатора CNN досягнуто точність 97,27%.

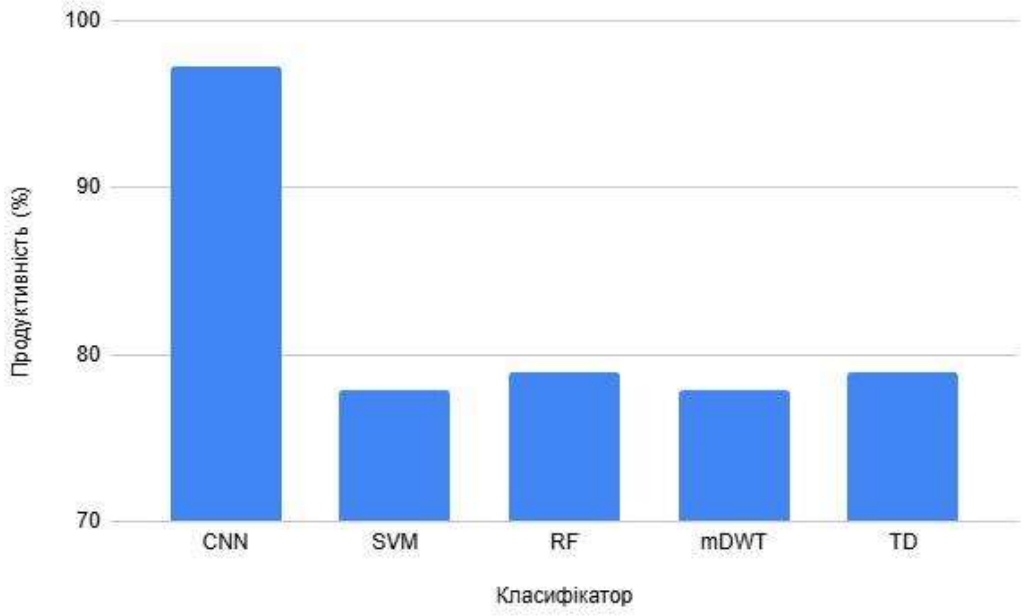


Рис. 3. Результати точності класифікаторів

Аналіз тестування в режимі реального часу

Для оцінювання роботи системи розпізнавання жестів у реальному часі було проведено серію тестів із керування роботизованим маніпулятором. Учасники експерименту одягали міопов'язку *Myo*, яка збирала ЕМГ-сигнали, що в реальному часі оброблялися згортковою нейронною мережею (CNN) для класифікації жестів. Розпізнані жести миттєво передавалися до мікроконтролера *Arduino*, що забезпечував керування рухами двопальцевого маніпулятора. Під час тестування учасники виконували набір заздалегідь визначених жестів, які система мала розпізнати та правильно інтерпретувати для відповідного руху робота.

Результати

У процесі тестування в режимі реального часу було досягнуто середню точність розпізнавання жестів на рівні 91,7%.

Порівняння

Розроблена система досягла середньої точності розпізнавання жестів у режимі реального часу, яка становила 91,7%. Це є конкурентним результатом порівняно із сучасними дослідженнями. У праці [1] продемонстровано 90% точності офлайн і 84% онлайн. Автори роботи [3] досягли точності понад 91% для здорових користувачів. У дослідженні [14] отримали схожу точність (91,25%) з використанням легкої моделі CNN+LSTM, оптимізованої для вбудованих систем. У праці [15] автори продемонстрували 94% точності, використовуючи карти активації м'язів (TMA), що може бути корисним для покращення розпізнавання складних жестів. У дослідженні [16] досягнуто 98,4% точності за умови застосування *Myo Armband* та CNN для розпізнавання шести жестів, що свідчить про високу ефективність запропонованого підходу. У праці [17] продемонстровано 93% точності з використанням EMGNet і трансферного навчання, що дає змогу покращити цю властивість за умови обмежених даних. Автори роботи [18] досягли

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33) ISSN 2524-2296 (online)

90,3% точності за допомогою моделі CNN з увагою до каналів, що може бути корисним для покращення точності в реальному часі.

Недоліки та обмеження

Основним обмеженням системи є чутливість до тривалих змін стану м'язів, таких як втома, що може впливати на стабільність сигналів і знижувати точність розпізнавання. Крім того, хоча індивідуальна нормалізація після кожного надягання пов'язки знижує вплив неточного розташування датчиків, повністю усунути цей фактор наразі не вдалося. Обмеження також пов'язані з кількістю підтримуваних жестів – у цьому дослідженні протестовано лише два основні жести, що обмежує застосовність системи в більш складних сценаріях.

Практичне значення для подальших досліджень

Досягнуті результати демонструють життєздатність використання міопов'язки в поєднанні зі згортковими нейронними мережами для реального часу керування роботизованими пристроями. Це відкриває перспективи для подальшого розширення набору розпізнаваних жестів, покращення алгоритмів адаптивної нормалізації та впровадження систем у практичні застосування, такі як протези, реабілітаційні пристрої або інтерфейси "людина – машина". Особливо перспективним є розвиток методів, які підвищать стійкість системи до фізіологічних змін і тривалої експлуатації в реальних умовах.

Висновки

У цьому дослідженні розроблено та протестовано систему розпізнавання жестів на основі сигналів поверхневої електроміографії (ЕМГ), зібраних за допомогою міопов'язки *Myo Armband*. Застосування восьмиканального пристрою з частотою дискретизації 200 Гц дало змогу отримати достатньо якісні сигнали для аналізу, а попереднє оброблення даних, зокрема перетворення сигналів у сірі зображення, зробило можливим ефективне використання згорткової нейронної мережі для класифікації жестів.

Оброблення сигналів передбачало нормалізацію за допомогою методів мінімально-максимальної нормалізації та нормалізації нульового середнього. Аналіз продемонстрував, що нормалізація нульового середнього забезпечує вищу точність класифікації на 2,34%, що свідчить про її перевагу в цьому контексті. Важливим кроком стала увага до змін амплітуди сигналів після кожного нового надягання браслета, що дало змогу забезпечити стабільність даних за допомогою індивідуального калібрування.

Класифікація жестів здійснювалася за допомогою згорткової нейронної мережі *ResNet*-типу, яка досягла точності 97,27% на тестовій вибірці. Порівняльний аналіз з іншими методами (RF, SVM, mDWT, TD) підтвердив перевагу CNN, яка продемонструвала на 18–19% вищу точність розпізнавання. Навчання моделі з використанням оптимізатора *Adam* показало стабільну збіжність, а результати втрат свідчили про досягнення оптимального навчання після приблизно 2000 ітерацій.

Під час тестування в режимі реального часу було досягнуто середньої точності 91,7%, що підтверджує можливість практичного застосування системи для керування роботизованими пристроями. Це значення є конкурентним порівняно із сучасними аналогами, більшість з яких демонструють точність у межах 84–94%. Результати дослідження підтверджують ефективність запропонованого підходу навіть за умови обмеженої кількості жестів.

Однак система має певні обмеження. Вона чутлива до м'язової втоми та змін розташування датчиків, що знижує стабільність розпізнавання в разі тривалої експлуатації. Крім того, поточна реалізація підтримує лише два жести, що обмежує її функціональність у більш складних сценаріях використання.

З практичного погляду результати відкривають подальші перспективи для досліджень. Можливе розширення кількості розпізнаваних жестів, удосконалення методів нормалізації, підвищення стійкості до фізіологічних змін користувача, а також інтеграція з реальними пристроями, такими як біонічні протези або інтерфейси "людина – машина". Отримані напрацювання можуть стати основою для майбутніх досліджень у сфері реабілітації, допоміжних технологій та безконтактного управління технікою.

Перспективи подальших досліджень

Результати дослідження свідчать про значний потенціал системи розпізнавання жестів на основі сигналів ЕМГ, однак також окреслюють низку напрямів, які можуть бути суттєво розвинуті в подальших роботах.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)

По-перше, перспективним є розширення кількості розпізнаваних жестів. Поточна реалізація підтримує лише два жести, що обмежує можливості практичного застосування. Додавання більшої кількості жестів потребуватиме збільшення обсягу навчальної інформації, покращення архітектури нейронної мережі та оптимізації підходів до калібрування, щоб забезпечити збереження високої точності розпізнавання.

По-друге, важливо дослідити методи підвищення стійкості системи до змін зовнішніх умов. Зокрема варто розробити алгоритми, які будуть менш чутливими до неправильного надягання браслета, м'язової втоми, потовиділення чи інших фізіологічних змін користувача. Одним із можливих рішень є використання адаптивного навчання або персоналізованих моделей, які беруть до уваги індивідуальні особливості користувача в реальному часі.

Третім напрямом є оптимізація оброблення сигналів. Поглиблений аналіз часових і частотних властивостей ЕМГ-сигналів може дати змогу сформувати більш інформативні вхідні ознаки для моделі. Також доцільно дослідити альтернативні способи перетворення сигналів у зображення або тривимірне подання, що може покращити ефективність нейронної мережі.

Крім того, перспективним є поєднання ЕМГ-даних з іншими типами сенсорної інформації, наприклад, із показниками акселерометра чи гіроскопа. Мультимодальні підходи можуть значно підвищити точність класифікації та дати змогу розпізнавати складні або динамічні жести.

Також варто приділити увагу практичній інтеграції системи в реальні пристрої: біонічні протези, інвалідні візки, системи віртуальної або доповненої реальності. Це передбачає не лише технічну адаптацію, а й дослідження ергономіки, зручності використання та тривалості взаємодії користувача із системою.

Список літератури

1. Guo B., Ma Y., Yang J., Wang Z., Zhang X. Lw-CNN-Based Myoelectric Signal Recognition and Real-Time Control of Robotic Arm for Upper-Limb Rehabilitation. *Computational Intelligence and Neuroscience,* 2020, 8846021 p. DOI: https://doi.org/10.1155/2020/8846021
2. Wan Y., Han Z., Zhong J., Chen G. (2018). Pattern recognition and bionic manipulator driving by surface electromyography signals using convolutional neural network. *Advances in Mechanical Engineering,* 10 (10), P. 1–11. 2018. DOI: https://doi.org/10.1177/1729881418802138
3. Bao T., Zaidi S. A. R., Xie S., Yang P., Zhang Z. A CNN-LSTM Hybrid Framework for Wrist Kinematics Estimation Using Surface Electromyography. *IEEE Transactions on Instrumentation and Measurement*. 82 p. 2019. DOI: https://doi.org/10.1109/TIM.2020.3036654
4. Real-time Bionic Arm Control Via CNN-based EMG Recognition. URL: https://www.hackster.io/emgarm/real-time-bionic-arm-control-via-cnn-based-emg-recognition-b013d3
5. Meng Q., Yue Y., Li S., Yu H. Electromyogram-based motion compensation control for the upper limb rehabilitation robot in active training. *Mechanical Sciences,* 13(2), 2022. P. 675–685. DOI: https://doi.org/10.5194/ms-13-675-2022
6. Atzori M., Gijsberts A., Castellini C., Caputo B., Mittaz Hager A.G., Elsig S., Giatsidis G., Bassetto F., Müller H. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Scientific Data*, № 1, 140053 p. 2014. DOI: https://doi.org/10.1038/sdata.2014.53
7. Banzi M., Shiloh M. Getting Started with Arduino: The Open Source Electronics Prototyping Platform. Maker Media. 2014. URL: https://www.sarcnet.org/files/Getting%20Started%20With%20Arduino.pdf
8. Patel S., Patel V. A study on normalization techniques in data mining. *International Journal of Computer Applications*, 179(14), 2018. P. 1-4. DOI: https://doi.org/10.5120/ijca2018917759
9. Bhandari P. The Standard Normal Distribution Calculator, Examples & Uses. Scribbr. 2020. URL: https://www.scribbr.com/statistics/standard-normal-distribution/
10. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016. P. 770-778. DOI: https://doi.org/10.1109/CVPR.2016.90
11. Kingma D. P., Ba J. Adam: A method for stochastic optimization. *In Proceedings of the International Conference on Learning Representations (ICLR).* 2014. DOI: https://doi.org/10.48550/arXiv.1412.6980
12. Abadi M., Barham P., Chen J., Chen Z., Davis A., Dean J., Devin M., Ghemawat S., Irving G., Isard M., Kudlur M., Levenberg J., Monga R., Moore S., Murray D., Steiner B., Zheng X. Tensor Flow: A system for large-scale machine learning. *In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI),* 2016. P. 265–283. URL: https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

13. Dianchun Bai, Tie Liu, Xinghua Han, Hongyu Yi. Application Research on Optimization Algorithm of sEMG Gesture Recognition Based on Light CNN+LSTM Model. *Cyborg and Bionic Systems.* Vol 2021. 2021. DOI:10.34133/2021/9794610
14. Silva A. D., Perera M. V., Wickramasinghe K., Naim A. M., Dulantha Lalitharatne, Kappel S. L. Real-Time Hand Gesture Recognition Using Temporal Muscle Activation Maps of Multi-Channel Semg Signals, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Barcelona, Spain, 2020, P. 1299–1303, DOI: 10.1109/ICASSP40776.2020.9054227
15. Asif A.R., Waris A., Gilani S.O., Jamil M., Ashraf H., Shafique M., Niazi I.K. Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG. *Sensors* 2020, № 20, 1642 p. DOI: https://doi.org/10.3390/s20061642
16. Gopal P., Gesta A., Mohebbi A. A Systematic Study on Electromyography-Based Hand Gesture Recognition for Assistive Robots Using Deep Learning and Machine Learning Models. *Sensors* 2022, № 22, 3650 p. DOI: https://doi.org/10.3390/s22103650
17. Yu G., Deng Z., Bao Z., Zhang Y., He B. Gesture Classification in Electromyography Signals for Real-Time Prosthetic Hand Control Using a Convolutional Neural Network-Enhanced Channel Attention Model. *Bioengineering* 2023, № 10, 1324 p. DOI: https://doi.org/10.3390/bioengineering10111324

References

1. Guo, B., Ma, Y., Yang, J., Wang, Z., Zhang, X. (2020), "Lw-CNN-Based Myoelectric Signal Recognition and Real-Time Control of Robotic Arm for Upper-Limb Rehabilitation". *Computational Intelligence and Neuroscience*, 2020, 8846021 p. DOI: https://doi.org/10.1155/2020/8846021
2. Wan, Y., Han, Z., Zhong, J., & Chen, G. (2018), "Pattern recognition and bionic manipulator driving by surface electromyography signals using convolutional neural network". *Advances in Mechanical Engineering*, Vol. 10(10), P. 1–11. DOI: https://doi.org/10.1177/1729881418802138
3. Bao, T., Zaidi, S. A. R., Xie, S., Yang, P., Zhang, Z. (2019), "A CNN-LSTM Hybrid Framework for Wrist Kinematics Estimation Using Surface Electromyography". *IEEE Transactions on Instrumentation and Measurement.* 82 p. 2019. DOI: https://doi.org/10.1109/TIM.2020.3036654
4. "Real-time Bionic Arm Control Via CNN-based EMG Recognition". available at: https://www.hackster.io/emgarm/real-time-bionic-arm-control-via-cnn-based-emg-recognition-b013d3
5. Meng, Q., Yue, Y., Li, S., Yu, H. (2022), "Electromyogram-based motion compensation control for the upper limb rehabilitation robot in active training". *Mechanical Sciences,* 13(2), P. 675–685. DOI: https://doi.org/10.5194/ms-13-675-2022
6. Atzori, M., Gijsberts, A., Castellini, C., Caputo, B., Mittaz Hager, A.G., Elsig, S., Giatsidis, G., Bassetto, F., Müller, H. (2014), "Electromyography data for non-invasive naturally-controlled robotic hand prostheses". *Scientific Data*, № 1, 140053 p. DOI: https://doi.org/10.1038/sdata.2014.53
7. Banzi, M., Shiloh, M. (2014), "Getting Started with Arduino: The Open Source Electronics Prototyping Platform. Maker Media", available at: https://www.sarcnet.org/files/Getting%20Started%20With%20Arduino.pdf
8. Patel, S., Patel, V. (2018), "A study on normalization techniques in data mining". *International Journal of Computer Applications*, № 179(14), P. 1–4. DOI: https://doi.org/10.5120/ijca2018917759
9. Bhandari, P. (2020), "The Standard Normal Distribution | Calculator, Examples & Uses. Scribbr". available at: https://www.scribbr.com/statistics/standard-normal-distribution/
10. He, K., Zhang, X., Ren, S., Sun, J. (2016), "Deep residual learning for image recognition". *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, P. 770-778. DOI: https://doi.org/10.1109/CVPR.2016.90
11. Kingma, D. P., Ba, J. (2014), "Adam: A method for stochastic optimization". I*n Proceedings of the International Conference on Learning Representations (ICLR)*. DOI: https://doi.org/10.48550/arXiv.1412.6980
12. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D., Steiner, B., Zheng, X. (2016), "TensorFlow: A system for large-scale machine learning". *In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, P. 265-283. available at: https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi
13. Dianchun, Bai, Tie, Liu, Xinghua, Han, Hongyu, Yi. (2021), "Application Research on Optimization Algorithm of sEMG Gesture Recognition Based on Light CNN+LSTM". *Cyborg and Bionic Systems.* Vol 2021. 2021. DOI:10.34133/2021/9794610
14. Silva, A.D., Perera, M.V., Wickramasinghe, K., Naim, A.M., Dulantha, Lalitharatne, Kappel, S. L. (2020), "Real-Time Hand Gesture Recognition Using Temporal Muscle Activation Maps of Multi-Channel Semg Signals," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Barcelona, Spain, 2020, P. 1299-1303, DOI: 10.1109/ICASSP40776.2020.9054227
15. Asif, A.R.; Waris, A.; Gilani, S.O.; Jamil, M.; Ashraf, H.; Shafique, M.; Niazi, I.K. (2020), "Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG". *Sensors.* № 20, 1642 p. DOI: https://doi.org/10.3390/s20061642

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2025. No. 3 (33)*

16. Gopal, P.; Gesta, A.; Mohebbi, A. (2022), "A Systematic Study on Electromyography-Based Hand Gesture Recognition for Assistive Robots Using Deep Learning and Machine Learning Models". *Sensors*. № 22, 3650 p. DOI: https://doi.org/10.3390/s22103650
17. Yu, G.; Deng, Z.; Bao, Z.; Zhang, Y.; He, B. (2023), "Gesture Classification in Electromyography Signals for Real-Time Prosthetic Hand Control Using a Convolutional Neural Network-Enhanced Channel Attention Model". *Bioengineering*. № 10, 1324 p. https://doi.org/10.3390/bioengineering10111324

Надійшла (Received) 03.01.2025

Відомості про авторів / About the Authors

Пастушенко Антон Олександрович — аспірант кафедри біомедичної інженерії та опто-електронних систем, Вінницького національного технічного університету, м. Вінниця, Україна. e-mail: hate6death88@gmail.com; ORCID ID: https://orcid.org/0009-0006-7028-4071

Коваль Леонід Григорович — кандидат технічних наук, доцент, завідувач кафедри біомедичної інженерії та оптико-електронних систем, Вінницький національний технічний університет, Вінниця, Україна. e-mail: koval.l@vntu.edu.ua; ORCID ID: https://orcid.org/0000-0001-9887-2605

Pastushenko Anton — a graduate student of the Department of Biomedical Engineering and Opto-Electronic Systems, Vinnytsia National Technical University, Vinnytsia, Ukraine.

Koval Leonid — Candidate of Technical Sciences, Associate Professor, Head of the Department of Biomedical Engineering and Optoelectronic Systems, Vinnytsia National Technical University, Vinnytsia, Ukraine.

CONTROL OF A ROBOT MANIPULATOR USING SURFACE ELECTROMYOGRAPHY SIGNALS

The subject of the study is the methods of controlling a robotic manipulator based on surface electromyography (EMG) signals using individualized normalization and classification of gestures by a convolutional neural network (CNN). **The aim of the work** is to create an effective control system that combines pre-normalization of EMG signals and deep learning to increase the accuracy of gesture recognition and stability in real time. **For this purpose**, the following were used: signal collection using the Myo Armband, pre-processing (min-max and zero-mean), signal conversion into images through a sliding window, CNN training (ResNet, Adam optimization) and comparison with SVM and Random Forest models. **The results** obtained showed a classification accuracy of 97.27% in the test environment and 91.71% in real time. The CNN model outperformed traditional methods by 18–19%, and zero-mean increased the accuracy by 2.34% compared to min-max normalization. The system remained stable even with variations in the bracelet position due to individual normalization. **Conclusions**: the proposed system demonstrates high accuracy, reliability and adaptability in real time. The scientific novelty lies in the combination of individualized normalization of EMG signals with ResNet, which provides stability and exceeds the accuracy of traditional algorithms. In the future, it is planned to expand the set of gestures, study more complex conditions and optimize neural networks for embedded systems.

Keywords: electromyography; gesture control; robotic manipulator; convolutional neural network; ResNet; data normalization; bionic prostheses; real-time control.

Бібліографічні описи / Bibliographic descriptions

Пастушенко А. О., Коваль Л. Г. Керування роботом-маніпулятором за допомогою сигналів поверхневої електроміографії. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 3 (33). С. 203–212. DOI: https://doi.org/10.30837/2522-9818.2025.3.203

Pastushenko, A., Grigorovich, K. (2025), "Control of a robot manipulator using surface electromyography signals", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (33), P. 203–212. DOI: https://doi.org/10.30837/2522-9818.2025.3.203

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 3 (33)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

АЛФАВІТНИЙ ПОКАЖЧИК

ALPHABETICAL INDEX

НАУКОВЕ ВИДАННЯ

Сучасний стан наукових досліджень та технологій в промисловості

Щоквартальний науковий журнал

№ 3 (33), 2025

Відповідальний секретар журналу *І.Г. Перова*
Відповідальний за випуск *А.А. Коваленко*
Відповідальний за ліцензування *В.В. Косенко*
Редактор *Л.В. Кузьміна*
Комп'ютерна верстка *Л.Ю. Свєтайло*

Формат 60×84/8. Умов. друк. арк. 23,25.
Тираж 150 прим.

Віддруковано з готових оригінал-макетів
в типографії ФОП Андреєв К.В.
Єдиний державний реєстр юридичних осіб
та фізичних осіб-підприємців.
Запис №24800170000045020 від 30.05.2003.

61157, Харків, вул. Акад. Богомольця, 9, кв. 50,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com

SCIENTIFIC PUBLICATION

Innovative technologies and scientific solutions for industries

Quarterly scientific journal

No. 3 (33), 2025

Responsible secretary of journal *I. Perova*
Responsible for release *A. Kovalenko*
Responsible for licensing *V. Kosenko*
Editor *L. Kuzmina*
Computer layout *L. Svietailo*

Format 60×84/8. Conventional printed sheets 23,25.
Edition of 150 copies.

Printed from ready-made original layouts
in the printing house
of Individual Entrepreneur Andreev K.V.
Unified State Register of Legal Entities
and Individual Entrepreneurs.
Entry No. 24800170000045020 of 30.05.2003.

fl. 50, 9, Acad. Bogomolets Str., Kharkiv, 61157,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com