

15. Digital security transformation: change that works. URL: <https://ohoronapraci.kiev.ua/article/news/cifrova-transformacia-bezpeki-zmini-aki-pracuut> (accessed June 20, 2026).
16. WINWIN. Global innovation strategy of Ukraine. URL: <https://winwin.gov.ua/> (accessed June 20, 2026).
17. Cybersecurity. URL: <https://amintegrator.com/cybersecurity/> (accessed June 20, 2026).
18. Yakovenko Y. The importance of cybersecurity in shaping the digital economy. URL: https://www.researchgate.net/publication/383193213_Znacenna_kiberbezpeki_u_for_muvanni_cifrovoi_ekonomiki (accessed June 20, 2026).
19. Tarasovsky Yu. Ukraine has approved the Strategy for the Digital Development of Innovation until 2030. URL: <https://forbes.ua/news/ukraina-zatverdila-strategiyu-tsifrovogo-rozvitku-innovatsiy-do-2030-roku-shcho-vona-peredbachae-14012025-26270> (accessed June 20, 2026).
20. Science park of additive technologies. URL: <https://www.amspace.com.ua/26270> (accessed June 20, 2026).
21. Directive on Network and Information Security. URL: https://nis2.org.ua/?gad_source=1&gad_campaignid=23017071562&gbraid=0AAAA-BBThWcdWCFXVbeUIPC10mJmbhR0zl&gclid=EAIaIQobChMI1dvk4NDv1AMV_5qDBx1opjDYEAAAYASAAEgLBKfD_BwE (accessed June 20, 2026).
22. Enhanced cybersecurity rules in the European Union. URL: <https://www.h-x.technology.ua/services/nis-2-cybersecurity-directive-ua> (accessed June 20, 2026).

2.4. Protection of Critical Information Infrastructure and Cyber Risk Management: European Practices and Opportunities for Implementation in Ukraine

Olena DVIRNA,

Serhii NABOKA

National University "Yuri Kondratyuk Poltava Polytechnic"

The rapid development of cloud technologies and the widespread adoption of distributed, hybrid, and multi-cloud environments have increased the relevance of efficient computing resource management. Dynamic workload variations, resource heterogeneity, and the need to ensure a specified level of quality of service require the simultaneous consideration of multiple criteria, including execution time, resource usage cost, and energy consumption. Therefore, the resource allocation problem in cloud systems is inherently multi-objective, and the application of evolutionary methods, which enable the generation of a set of Pareto-optimal solutions, represents a promising approach to its solution (Chen J., Du T., Xiao G., 2021; Malti A.N., Hakem M., Benmammar B., 2022).

At the same time, optimization based exclusively on performance and economic efficiency indicators does not fully meet the current requirements for cloud systems. Distributed architectures, containerization, microservices, application programming interfaces, and remote access introduce additional cybersecurity risks. The major threats to cloud environments include identity and access management issues, insecure APIs, misconfigurations, security architecture deficiencies, and software supply chain risks (Cloud Security Alliance, 2022). This necessitates the consideration of the level of cyber risk directly when selecting a cloud resource allocation strategy.

One of the approaches to ensuring the security of modern cloud environments is the Zero Trust concept, according to which access to resources should not be based on implicit trust associated with the network or physical location of a user or resource (Rose S. et al., 2020). In cloud-native and multi-cloud environments, these principles involve the verification of subjects and resources and the application of context-dependent access policies (Chandramouli R., Butcher Z., 2023). Accordingly, trust and security characteristics should be integrated into decision-making procedures for the placement and reallocation of computing resources.

This approach is especially important in the context of the European policy on digital resilience. ENISA pays significant attention to cybersecurity risks related to cloud services and to the development of mechanisms that increase trust in these services (ENISA, 2023). At the same time, EU Directive 2022/2555 (NIS2) extends

cybersecurity risk management requirements to cloud computing service providers and data centre operators (European Parliament and Council of the European Union, 2022). For Ukraine, in the context of European integration, this creates a need to align cloud infrastructure management approaches with European principles of cybersecurity and digital resilience.

Therefore, multi-objective optimization of resource allocation is an important task in which security is considered not only as a set of additional constraints but also as a separate optimization criterion. In general, this problem can be represented by the following vector objective function:

$$F(x) = (T(x), C(x), E(x), R(x)) \rightarrow \min, \quad (2.14)$$

where x is a variant of computing resource allocation; $T(x)$ is the task execution time or an overall measure of time efficiency; $C(x)$ is the cost of using cloud resources; $E(x)$ is energy consumption; and $R(x)$ is an overall assessment of cyber risk. This formulation makes it possible to move from searching for one optimal solution to creating a set of compromise solutions. The final solution can be selected depending on the current workload, performance requirements, economic limitations, and the acceptable level of risk.

The aim of the research is to develop an approach to multi-objective resource allocation optimization in cloud computing systems by including cyber risk as an optimization criterion and considering Zero Trust principles and European approaches to digital resilience. To achieve this aim, it is necessary to formalize the system of optimization criteria, define the role of cyber risk in the resource allocation model, study the possibilities of using multi-objective evolutionary methods to find Pareto-optimal solutions, and develop recommendations for applying the proposed approach in accordance with current European requirements for cloud infrastructure security.

The efficiency of cloud computing systems largely depends on how well the available resources are allocated among users' tasks. Virtualization makes it possible to separate physical processor, network, and storage resources from their hardware and

represent them as a set of virtual resources that can be provided according to the current needs of users and applications. As cloud platforms become larger, resource allocation becomes more complex because of the differences between physical nodes, dynamic workloads, different application performance requirements, and the need to meet defined quality-of-service requirements. Therefore, resource allocation and task scheduling are among the main tasks of cloud infrastructure management (Chen J., Du T., Xiao G., 2021; Zhang H., Wu Y., Sun Z., 2022).

In general, a cloud computing infrastructure can be represented as a set of physical nodes $H = \{h_1, h_2, \dots, h_m\}$, each of which is characterized by available processing power CPU_i , RAM_i capacity, disk space S_i , network bandwidth B_i , and other parameters. Virtual machines or containers are deployed on these physical nodes and form the logical resource layer $V = \{v_1, v_2, \dots, v_n\}$. A set of tasks must then be allocated to these virtual resources $J = \{j_1, j_2, \dots, j_k\}$.

Thus, the optimization problem is to determine an allocation of tasks to available resources that meets resource and technological constraints and, at the same time, provides the best values of the defined criteria. Even the problem of scheduling a set of tasks on heterogeneous resources is generally a complex combinatorial problem. This significantly limits the use of exact methods when the size of the system increases (Malti A.N., Hakem M., Benmammar B., 2022).

It is important to distinguish between the optimization of virtual machine or container placement and task scheduling. In the first case, the goal is to determine which physical nodes should host the virtualized resources, taking into account their CPU, RAM, storage, and network requirements. In the second case, computational tasks are assigned to the available virtual resources.

In practice, these processes are closely connected. Inefficient placement of VMs or containers can lead to an uneven workload of physical servers, while inefficient task scheduling can increase execution time and energy consumption and may cause violations of SLA requirements. For this reason, modern models increasingly consider resource allocation and task scheduling as interconnected levels of a single optimization problem.

The main criteria for cloud resource optimization are execution time, cost, energy consumption, and latency. Let these criteria be denoted as $T(x)$, $C(x)$, $E(x)$ and $L(x)$, where x represents a specific resource allocation solution. The criterion $T(x)$ may represent the total task execution time or the makespan, which is the time when the last task is completed: $T(x) = \max_{j \in J} t_j^{finish}$.

Minimizing $T(x)$ improves the overall performance of the system. However, it does not guarantee minimum cost or balanced resource utilization. Studies of modern scheduling algorithms confirm that makespan remains one of the main criteria for evaluating the efficiency of task allocation in cloud environments (Malti A.N., Hakem M., Benmammar B., 2022; Mangalampalli S., Karri G.R., Kose U., 2023).

The cost criterion $C(x)$ represents the cost of using computing resources. In a simplified form, it can be defined as $C(x) = \sum_{i=1}^n p_i t_i$, where p_i is the cost per unit of time for using the i -th resource, and t_i is the duration of its use. For public and multi-cloud platforms, this criterion is especially important because providers offer different prices, virtual machine types, and payment models. Cost minimization may require the use of less powerful resources, which can increase task execution time. Therefore, there is a natural conflict between the $T(x)$ and $C(x)$ criteria.

The energy consumption criterion $E(x)$ represents the amount of energy required to operate the computing resources involved. Its value depends on the number of active physical servers, their utilization level, task execution time, and power management mechanisms. In modern research, energy efficiency is considered not only as a way to reduce the operating costs of data centres but also as a separate measure of scheduling algorithm efficiency. In particular, the EHEFT-R algorithm focuses on task execution efficiency, QoS, and energy consumption at the same time (Zhang H., Wu Y., Sun Z., 2022).

The criterion $L(x)$ represents the delay in data processing or transmission. This criterion is especially important for interactive services, real-time systems, and distributed applications, where the physical location of resources and the current

network conditions directly affect the quality of service. Therefore, the general optimization problem can be represented as

$$\min F(x) = (T(x), C(x), E(x), L(x)), \quad (2.15)$$

subject to a set of constraints on the available processing, memory, network, and other resources.

Depending on the type of input parameters, cloud resource optimization models can be divided into static and dynamic models. A static model assumes that the sets of tasks and resources, as well as their main characteristics, are known before the optimization process starts and do not change significantly during the considered time period. Under these conditions, the optimization algorithm can create a complete resource allocation plan before the tasks are executed. The advantage of this approach is the possibility of performing a more detailed global search. However, its efficiency decreases when the workload changes rapidly.

In dynamic models, user requests arrive at different times, while the availability and characteristics of resources may change during system operation. Therefore, resource allocation should be adjusted according to the current state of the cloud infrastructure. Dynamic behaviour is one of the main characteristics of real cloud environments: heterogeneous tasks come from different sources, while their number, size, and computing requirements change over time (Mangalampalli S., Karri G.R., Kose U., 2023). This creates a need for adaptive algorithms that can perform repeated optimization or quickly adjust an existing solution.

Exact, heuristic, metaheuristic, and evolutionary methods are used to solve resource allocation problems. Exact mathematical programming methods can provide an optimal solution when the model is correctly defined and the problem size is acceptable. However, the combinatorial growth of the solution space makes these methods difficult to use in large cloud systems. For this reason, algorithms that do not guarantee a mathematically exact global optimum but can provide high-quality solutions within an acceptable time are of significant practical interest.

Heuristic methods use rules developed according to the structure of a specific problem. These methods include algorithms that select the most or least loaded resource, priority-based task allocation rules, and different list-scheduling approaches. Their main advantages are simple implementation and relatively low computational costs. For example, algorithms from the HEFT family use task ranking and estimated completion times to select suitable resources. In EHEFT-R, additional priority rules are used to improve the initial solution, while a modified HEFT mechanism is applied to find a more efficient schedule (Zhang H., Wu Y., Sun Z., 2022).

A disadvantage of heuristic methods is their strong dependence on the specific features of the problem. A rule that is effective for one type of workload may produce worse results when the task structure or resource characteristics change. In addition, the local focus of many heuristic methods limits their ability to explore a large space of alternative resource allocation solutions.

Metaheuristic methods are more universal because they define a general search strategy and can be adapted to different optimization problems. In cloud computing, genetic algorithms (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Whale Optimization Algorithm (WOA), Flower Pollination Algorithm (FPA), and other population-based and nature-inspired approaches are used. For example, a method based on flower pollination behaviour has been proposed for multi-objective task scheduling. In this method, the Pareto principle is combined with the TOPSIS method. Makespan, execution cost, and task allocation reliability are considered as optimization criteria (Malti A.N., Hakem M., Benmammar B., 2022).

Evolutionary algorithms have a special place among metaheuristic approaches. They work with a population of possible solutions instead of a single solution and use selection, crossover, and mutation mechanisms. This makes it possible to explore different areas of the solution space and reduces the dependence of the final result on a single initial point. This feature is especially important for cloud optimization problems because task and resource allocation has a discrete and combinatorial nature.

The use of evolutionary methods is especially promising for multi-objective optimization. The traditional approach is to combine several criteria into a single scalar objective function

$$F(x) = w_1T(x) + w_2C(x) + w_3E(x) + w_4L(x), \quad (2.16)$$

where $w_1, \dots, w_4$ are weighting coefficients. However, the choice of weights is subjective, and changes in user priorities or the system state require the weights to be defined again. In addition, combining several criteria into one value may hide information about possible trade-offs.

An alternative approach is to use the principle of Pareto optimality. A solution x_1 dominates a solution x_2 , if it is not worse than x_2 for all criteria and is better for at least one criterion. The set of non-dominated solutions forms the Pareto front, where each point represents a specific trade-off between the criteria. This approach does not require a single set of weights to be defined in advance. Instead, it provides a set of alternative solutions that can be selected according to the current system requirements. In particular, Chen, Du, and Xiao proposed a multi-objective cloud resource allocation model and the modified RAA-PI-NSGAI algorithm. This algorithm is designed to improve the quality and distribution of the obtained solutions and to speed up the search process (Chen J., Du T., Xiao G., 2021).

The potential of the multi-objective approach is also confirmed by the fact that reliability and trust characteristics are gradually being added to traditional optimization criteria. For example, in the model proposed by Malti, Hakem, and Benmammar, task-to-resource mapping reliability is optimized together with makespan and cost (Malti A.N., Hakem M., Benmammar B., 2022). In the study by Mangalampalli, Karri, and Kose, the multi-objective scheduler considers task and VM priorities as well as trust indicators. These indicators are defined through the availability of virtual resources, successful task execution, and task processing time efficiency. The proposed approach also aims to reduce makespan and energy consumption (Mangalampalli S., Karri G.R., Kose U., 2023).

A general comparison of the main approaches is presented in Table 2.4.

Table 2.4

Comparative Characteristics of Approaches to Cloud Resource Optimization

Approach	Advantages	Limitations	Suitability for Dynamic Cloud Systems
Exact methods	Ability to obtain an optimal solution	High computational complexity as the problem size increases	Low/Medium
Heuristic methods	Fast operation, simple implementation	Dependence on specific rules and problem structure	Medium/High
Metaheuristic methods	More global search, ability to work with complex nonlinear models	Dependence on parameters, no guarantee of finding the global optimum	High
Evolutionary methods	Ability to work with discrete solutions and explore a large space of alternatives	High computational cost	High
Multi-objective evolutionary methods	Formation of the Pareto front, simultaneous consideration of conflicting criteria	Difficulty in selecting the final solution and increased computational cost	High

Source: compiled by the authors based on (Chen J., Du T., Xiao G., 2021; Malti A.N., Hakem M., Benmammar B., 2022; Zhang H., Wu Y., Sun Z., 2022; Mangalampalli S., Karri G.R., Kose U., 2023).

The comparison shows that multi-objective metaheuristic and evolutionary approaches are the most promising for scalable cloud systems with heterogeneous

resources and dynamic workloads. These approaches make it possible to consider conflicting criteria and create a set of alternative resource allocation solutions without the need to define a single trade-off in advance. At the same time, recent studies show a gradual expansion of optimization criteria: together with execution time, cost, and energy consumption, reliability, availability, and trust are also considered.

However, even these models are mainly focused on performance and quality of service. A trust indicator based on availability, success rate, or SLA is not the same as a cyber risk assessment. Cyber risk should also consider the probability of threats, vulnerability characteristics, access context, and the potential impact of a security incident. Therefore, the next stage in the development of cloud resource optimization models is to include security directly in the system of decision-making criteria. This creates the need to study cyber risk as a separate factor in cloud infrastructure optimization and to define the corresponding criterion $R(x)$, which is the subject of the next section.

The spread of cloud technologies is accompanied not only by increased computing capabilities but also by a larger attack surface. Resource virtualization, shared use of physical infrastructure by different users, containers and microservices, application programming interfaces (APIs), remote access, and automated orchestration tools create additional sources of cyber risk. Therefore, the efficiency of a specific cloud system configuration cannot be evaluated only by execution time, cost, energy consumption, and latency. Decisions about the placement of computing tasks and services directly affect their level of security.

The most significant threats to cloud systems include weak identity and access management mechanisms, configuration errors, insecure APIs, weaknesses in security architecture, software supply chain risks, data leakage or loss, and attacks aimed at disrupting service availability (Cloud Security Alliance, 2022; ENISA, 2024). A specific feature of cloud environments is the interdependence of their components. A threat affecting one node, account, or service may also affect other components of the system. Therefore, the level of security depends not only on the characteristics of individual resources but also on how they interact and on the configuration of the infrastructure as a whole.

Another important characteristic is the dynamic nature of cyber risk. The state of cloud infrastructure changes over time because of scaling, migration of virtual machines and containers, the arrival of new tasks, changes in network routes, access policies, and workloads. At the same time, external conditions may also change: new vulnerabilities may appear, the intensity of attacks may increase or decrease, and the level of trust in specific components may change. ENISA reports consider threat assessment and resilience as parts of a continuous cybersecurity management process, which is especially important for digital infrastructure and cloud services (ENISA, 2023; ENISA, 2024).

For a formal description of these factors, it is useful to apply the concept of cyber risk as a characteristic that combines the probability of a specific threat and the possible consequences of the related incident. For a specific threat q a simplified risk assessment can be represented as $r_q = p_q \cdot I_q$, where p_q is the estimated probability of the q -th threat, and I_q is the estimated potential impact of this threat on the system. In a cloud environment, these parameters may depend on the resource configuration. Therefore, it is reasonable to write $r_q(x) = p_q(x) \cdot I_q(x)$, where x represents a specific task and resource allocation solution. This representation makes cyber risk different from a fixed external parameter: a change in x can affect not only system performance but also its level of security.

For example, two allocation options for the same task may have similar execution time and cost but may differ significantly in their level of risk. Placing a critical service on a node with a less secure access configuration, insufficient segmentation, or higher exposure to an external network may be more cost-effective but may also create a higher cyber risk. Similarly, placing several critical services on one physical node may reduce energy consumption through resource consolidation, but at the same time increase the possible consequences if this node is compromised. Therefore, conflicts may arise between economic efficiency, energy efficiency, and security, which is typical of multi-objective optimization problems.

Considering these relationships is the basis of the **security-aware resource allocation** approach, in which security parameters are directly integrated into the

resource scheduling and allocation process. Unlike the traditional approach, where a configuration is first optimized for performance and security mechanisms are applied afterwards, the security-aware approach evaluates security during the process of creating the solution. Research on security-aware scheduling in heterogeneous hybrid clouds shows that security requirements can be included in the resource selection process together with time constraints (*Security-aware task scheduling with deadline constraints on heterogeneous hybrid clouds*, 2021). A similar trend can be found in trust-aware models, where resource trust characteristics are considered during multi-objective task scheduling (Mangalampalli S., Karri G.R., Kose U., 2023).

For a set of threats $Q = \{q_1, q_2, \dots, q_s\}$ the overall risk of a specific resource allocation solution can be defined as $R(x) = \sum_{q=1}^s w_q p_q(x) I_q(x)$, where w_q is the importance coefficient of the corresponding threat. The weighting coefficients can be defined according to the type of data, the criticality of the service, SLA requirements, or the category of the information resource. This model makes it possible to compare alternative cloud infrastructure configurations quantitatively according to their potential level of cyber risk.

At the same time, in a dynamic cloud environment, it is important to consider changes in threat intensity over time. For this purpose, a function $\lambda_q(t)$ can be introduced to represent the intensity or relevance of the q -th threat at time t . Then, the risk assessment over the time interval $[0, T]$ can be represented as

$$R(x, T) = \sum_{q=1}^s w_q \int_0^T \lambda_q(t) p_q(x, t) I_q(x, t) dt.$$

This representation makes it possible to consider not only the static characteristics of resources but also the current state of the cloud system and changes in the security situation. As a result, a resource allocation solution that is acceptable at time t_1 may no longer be optimal at time t_2 , This creates the need for adaptive resource reallocation.

Therefore, cyber risk should be considered as a separate optimization criterion rather than only as a constraint on solution feasibility. If the security requirement is defined only as $R(x) \leq R_{max}$, all solutions that meet the defined threshold are considered equal from the security point of view. However, including $R(x)$ in the vector of optimization criteria makes it possible to evaluate the trade-off between risk reduction and other system characteristics. Therefore, the basic multi-objective resource allocation model (2.15) should be extended to a security-aware formulation

$$\min F(x) = (T(x), C(x), E(x), L(x), R(x)) \quad (2.17)$$

subject to the established system of resource, technological, and security constraints.

The proposed extension makes it possible to create Pareto-optimal configurations that represent different trade-offs between performance, cost, energy efficiency, latency, and security. At the same time, minimizing $R(x)$ does not mean that the most secure but potentially very expensive solution should always be selected. The task is to determine a resource allocation solution in which the level of cyber risk is balanced with other requirements of the cloud system. Further formalization of this approach requires the identification of factors that influence $R(x)$, including identity management, access control, segmentation, and continuous trust verification. In modern cloud-native environments, these mechanisms are directly related to the Zero Trust concept. The integration of Zero Trust into cloud resource management is discussed below.

Traditional information security models are mainly based on the concept of a protected network perimeter. Resources and users located inside a corporate network usually receive a higher level of trust than external users and systems. However, this model is not sufficient for modern cloud systems because the boundaries of corporate infrastructure are becoming less clear. Computing resources can be located in both private and public clouds, applications can operate as a set of microservices, and users and administrators can access them from different networks and devices. Under these

conditions, being located within a specific network segment cannot be considered a sufficient reason for trust.

An alternative to the perimeter-based approach is the Zero Trust concept. According to this concept, trust should not be automatically given to a user or resource only because of its physical or network location. Zero Trust Architecture requires the verification of the user and device before establishing an access session to a corporate resource. Authentication and authorization are considered separate functions that must be performed before access is granted (Rose S. et al., 2020). Thus, the basic principle of Zero Trust can be summarized as “never trust automatically, always verify.”

One of the main components of Zero Trust is identity-based access control, which means that access decisions are made based on the verified identity of the subject. In a cloud environment, identity becomes one of the key elements of security because a user, service, or software component can interact with resources regardless of their physical location. Access decisions should consider not only user credentials but also the context of the request, including device characteristics, resource type, level of criticality, current security status, and other parameters. For cloud-native applications, NIST considers the implementation of Zero Trust using identity-tier and network-tier policies, which makes it possible to apply granular access policies in multi-cloud environments (Chandramouli R., Butcher Z., 2023).

The second principle is least privilege, according to which a user, service, or process receives only the minimum set of permissions required to perform its functions. From the point of view of cloud infrastructure management, this principle is important not only for user accounts. It also applies to interactions between virtual machines, containers, microservices, and application programming interfaces. Limiting unnecessary privileges reduces the possibility of lateral movement by an attacker if one component is compromised and, therefore, reduces the potential impact of a security incident.

An important mechanism for implementing Zero Trust is microsegmentation, in which the infrastructure is divided into logically isolated segments and communication between them is controlled according to defined policies. In a cloud system, this makes it possible to isolate individual VMs, containers, applications, or groups of resources

even when they are located on shared physical infrastructure. In cloud-native environments, such policies can be supported by service mesh technologies, which provide control over communication between microservices and support attribute-based access control (Chandramouli R., Butcher Z., Chetal A., 2021). Microsegmentation reduces the potential area of attack spread and helps to limit the consequences of a compromised component.

Another important principle is continuous verification, which means continuous checking of the conditions under which access was granted. In a dynamic cloud system, successful authentication at one moment does not guarantee that the level of risk will remain the same during the whole session. Changes in the state of a device, location or type of access, user behaviour, threat level, or resource configuration may require additional verification or changes to the access policy. Therefore, Zero Trust represents a transition from static control to continuous assessment of the interaction context.

These principles are especially important for cloud-native and multi-cloud environments. An application can be distributed across several cloud platforms, while its individual components can be dynamically created, scaled, moved, or removed. NIST SP 800-207A directly adapts the Zero Trust model to cloud-native applications in multi-cloud environments and considers granular resource access policies regardless of network location (Chandramouli R., Butcher Z., 2023). In such an architecture, security control should be connected not so much with a fixed IP address or physical server but with the identity and characteristics of the resource itself.

In the study by H. V. Holovko, O. A. Dvirna, and S. V. Naboka, Zero Trust is considered from the perspective of its practical implementation under the limited resources of small and medium-sized businesses. The authors emphasize the need to adapt the model to the real capabilities of companies that may have limited budgets, a small number of IT specialists, and outdated infrastructure. The proposed approach is not based on simplifying the Zero Trust concept itself but on its step-by-step implementation according to the available resources of the organization (Holovko H.V., Dvirna O.A., Naboka S.V., 2026). This idea is also important for cloud resource optimization: the level of security cannot be considered separately from the financial and computing resources required to provide it.

From the optimization perspective, it is especially important that the implementation of Zero Trust can change the characteristics of a specific cloud system configuration. Additional authentication, policy verification, encryption of communication between services, monitoring, and the use of service mesh components require additional computing and network resources. Microsegmentation, in turn, may introduce restrictions on possible placement options for interconnected services. Therefore, a higher level of security may require additional costs and may also affect execution time and latency.

On the other hand, the choice of a resource allocation solution affects the possibility of implementing Zero Trust principles. Let each resource v_i have a set of security characteristics $Z_i = \{I_i, P_i, M_i, V_i\}$, where I_i is the level of identity and authentication support; P_i is the level of implementation of the least privilege principle; M_i is the microsegmentation characteristic; and V_i is the ability to provide continuous verification and access control. Therefore, placing the same task on resources v_i and v_j may result in different levels of cyber risk even if these resources have the same CPU, RAM, and other computing characteristics.

This relationship should be included in the risk function introduced in the previous section:

$$R(x) = R(x, Z(x), Q(t)), \quad (2.18)$$

where x is a resource allocation solution; $Z(x)$ is the set of Zero Trust characteristics of the resources used in configuration x ; $Q(t)$ represents the current threat state. In this case, changing the placement of a task may change $R(x)$, even if the functional characteristics of the task remain unchanged.

For example, let task j_k be executed on one of two resources v_1 and v_2 . The first resource provides lower cost and shorter execution time but has a lower level of segmentation and access control. The second resource requires higher costs but provides MFA, granular access control, microsegmentation, and continuous

monitoring. Therefore, if only the $T(x)$ and $C(x)$ criteria are considered, selecting v_1 may be optimal:

$$T(v_1) < T(v_2), C(v_1) < C(v_2), \quad (2.19)$$

However, after including the security criterion, the following relationship may be observed:

$$R(v_1) > R(v_2). \quad (2.20)$$

Therefore, there is no single “best” resource. Instead, a trade-off appears between performance, economic efficiency, and cyber risk.

Considering this, Zero Trust principles should be integrated not only into the security architecture but also directly into the decision-making process for cloud resource allocation. For critical tasks, the algorithm can exclude resources that do not meet the minimum requirements for identity verification, privilege management, or segmentation. For other tasks, the level of Zero Trust implementation can affect the value of $R(x)$, and, therefore, the position of the corresponding solution on the Pareto front. As a result, security-aware optimization combines two mechanisms: strict security constraints for unacceptable configurations and minimization of residual cyber risk for the set of feasible solutions.

Thus, Zero Trust provides a conceptual basis for moving from traditional cloud resource optimization to a risk-oriented management model, where security is considered a property of a specific resource configuration. Identity-based access control, least privilege, microsegmentation, and continuous verification can be formalized as parameters that affect the value of $R(x)$, while the implementation of these mechanisms requires additional resources. This explains the need to find trade-off solutions between $T(x)$, $C(x)$, $E(x)$, $L(x)$ and $R(x)$. Such a formulation provides a basis for further mathematical formalization of the multi-objective cloud resource allocation problem, taking into account cyber risk and Zero Trust principles.

The development of cloud computing in the European Union is accompanied by the formation of a comprehensive approach to cybersecurity, where technical security mechanisms are combined with risk management, certification, regulatory requirements, and digital resilience. A special role in this process belongs to the European Union Agency for Cybersecurity (ENISA), whose activities are aimed at improving the common level of cybersecurity across the EU. For cloud technologies, this approach is especially important because a disruption in the operation of one provider may affect a large number of organizations, information systems, and digital services at the same time.

In the Cloud Cybersecurity Market Analysis, ENISA considers cloud cybersecurity as an important separate segment of the European cybersecurity market. Among the main problems of cloud environments, special attention is given to misconfigurations, insecure APIs, and the lack of qualified specialists. Secure cloud architectures, edge computing, and fog computing are identified as promising areas for future development (ENISA, 2023). This confirms that security should not be considered only as an additional element of cloud infrastructure, but as one of the main factors in its design, configuration, and optimization.

The importance of a risk-oriented approach is increasing because of changes in the nature of cyber threats. According to the ENISA Threat Landscape 2024, the main threats include attacks against availability, ransomware, threats against data, malware, social engineering, and supply chain attacks. Attacks against availability were among the leading categories of the threats analysed (ENISA, 2024). This is directly relevant to cloud systems because availability is both an indicator of service quality and an important part of security. Therefore, resource allocation decisions should consider not only the performance of individual nodes but also the possible consequences of their unavailability or compromise.

An important part of the European approach is the development of cybersecurity certification mechanisms for cloud services. ENISA developed the candidate European Cybersecurity Certification Scheme for Cloud Services (EUCS) within the European cybersecurity certification framework. EUCS is aimed at assessing the cybersecurity of cloud services and developing a common approach to confirming their level of

security (ENISA, 2020). In the context of this research, it is important that the certification approach creates a basis for formalizing the security characteristics of cloud resources, which can be used when evaluating alternative resource allocation configurations.

The Directive (EU) 2022/2555 (NIS2) has a significant regulatory impact on cyber risk management in the EU. Its provisions directly cover cloud computing service providers and data centre service providers. NIS2 is based on the need for a high level of harmonization of requirements for these organizations because of the cross-border nature of their activities (European Parliament and Council of the European Union, 2022).

Importantly, NIS2 focuses not on the implementation of separate and isolated security measures but on systematic cyber risk management. These measures should include incident prevention, business continuity, crisis management, supply chain security, security in the acquisition, development, and maintenance of network and information systems, assessment of the effectiveness of cybersecurity measures, access management, and other areas. Special attention is also given to risks related to suppliers and third-party services (European Parliament and Council of the European Union, 2022).

For cloud infrastructure optimization, this means that security parameters may be related not only to a specific physical or virtual resource but also to the provider, the supply chain, and the service deployment model. For example, two resources with the same CPU, RAM, cost, and performance characteristics may differ in the provider's level of security, implemented access control mechanisms, fault tolerance, backup capabilities, and characteristics of related suppliers. Therefore, when defining the cyber risk function $\backslash(R(x)\backslash)$, it is reasonable to consider not only the technical characteristics of a specific resource but also the characteristics of the environment in which it operates.

In general, for the i -th cloud resource, a vector of security parameters can be defined as

$$S_i = (s_i^{id}, s_i^{acc}, s_i^{seg}, s_i^{mon}, s_i^{res}, s_i^{sup}), \quad (2.21)$$

where s_i^{id} represents identity and authentication mechanisms; s_i^{acc} represents access management; s_i^{seg} is the level of segmentation; s_i^{mon} represents monitoring capabilities; s_i^{res} describes the resilience and recovery capabilities of the resource; and s_i^{sup} represents risks related to the provider and the supply chain. The overall risk of a configuration can then be defined as a function of not only the resource allocation x , but also the characteristics of the resources involved:

$$R(x) = R(x, S_1, S_2, \dots, S_n, Q(t)), \quad (2.22)$$

where $Q(t)$ represents the current state of cyber threats at time t .

This approach is directly related to the concept of digital resilience. For a cloud system, resilience does not mean the complete elimination of all threats, which is practically impossible. Instead, it means the ability to maintain critical functions, adapt to negative events, and recover after an incident. Therefore, resource allocation optimization can focus not only on reducing the probability of an incident but also on limiting its possible consequences. Resource redundancy, distribution of critical components across independent nodes, avoidance of a single point of failure, and the ability to quickly migrate workloads may increase cost or energy consumption, but at the same time they can improve the resilience of the system.

For Ukraine, the implementation of European approaches to cloud cybersecurity should be considered not only as a matter of regulatory alignment but also as a basis for developing practical methods of digital infrastructure management. A promising direction is the use of the NIS2 risk-oriented model, ENISA approaches to threat analysis and cloud cybersecurity, and the principles included in the candidate EUCS scheme to define requirements for national cloud platforms, government information systems, enterprises, and critical infrastructure. At the same time, it is especially important to move from static compliance checks to continuous risk assessment and adaptive resource management.

Thus, the European approach demonstrates a transition from considering cybersecurity as a set of separate technical measures to a comprehensive model of risk management and digital resilience. For the problem of multi-objective cloud resource optimization, this provides a basis for including NIS2 requirements, ENISA recommendations, Zero Trust characteristics, and cloud service security parameters in the formation of the $R(x)$ criterion. As a result, a configuration can be selected not only according to performance, cost, energy consumption, and latency, but also by considering the ability of the system to resist cyber threats and maintain its functionality during security incidents. This integration of European cyber risk management principles with mathematical optimization methods provides the basis for the formalization of the multi-objective cloud resource allocation model in the next section.

Effective cloud infrastructure management requires the formalization of the process of allocating computing tasks among available resources while considering performance, cost, energy consumption, and security at the same time. Unlike traditional models, where security characteristics are mainly defined as additional constraints, the proposed formulation considers cyber risk as an independent optimization criterion. This makes it possible to consider possible trade-offs between task execution speed, economic efficiency, energy consumption, and the security level of a specific cloud system configuration.

Let a set of computing tasks be given as

$$J = \{j_1, j_2, \dots, j_n\}, \quad (2.23)$$

and let the set of available cloud resources be

$$V = \{v_1, v_2, \dots, v_m\}. \quad (2.24)$$

Each task j_i is characterized by a set of computing resource requirements:

$$d_i = (cpu_i, ram_i, bw_i, s_i), \quad (2.25)$$

where cpu_i is the required processing capacity; ram_i is the required amount of RAM; bw_i is the required network bandwidth; and s_i represents additional service requirements, including acceptable latency, availability level, or other SLA parameters.

Each resource v_j is characterized by a vector of available capacities

$$r_j = (CPU_j, RAM_j, BW_j), \quad (2.26)$$

as well as by cost, energy consumption, performance, and security parameters. To describe the resource allocation process, a binary variable is introduced:

$$x_{ij} = \begin{cases} 1, & \text{if task } j_i \text{ is assigned to resource } v_j, \\ 0, & \text{otherwise.} \end{cases} \quad (2.27)$$

Then, each task must be assigned to exactly one resource:

$$\sum_{j=1}^m x_{ij} = 1, i = 1, \dots, n. \quad (2.28)$$

In addition, the constraints related to available resources must be satisfied. For processing capacity:

$$\sum_{i=1}^n cpu_i x_{ij} \leq CPU_j, j = 1, \dots, m. \quad (2.29)$$

For RAM:

$$\sum_{i=1}^n ram_i x_{ij} \leq RAM_j, j = 1, \dots, m. \quad (2.30)$$

For network bandwidth:

$$\sum_{i=1}^n bw_i x_{ij} \leq BW_j, j = 1, \dots, m. \quad (2.31)$$

Thus, a configuration $\backslash(x\backslash)$ is considered feasible only if all tasks are allocated among the resources and the available processing, memory, and network capacities are not exceeded.

A separate group includes constraints related to the quality of service. If the maximum acceptable latency L_i^{max} is defined for task j_i then the following condition must be satisfied:

$$L_i(x) \leq L_i^{max}. \quad (2.32)$$

Similarly, if the SLA defines a minimum level of availability A_i^{min} , the corresponding availability constraint must be satisfied.

This group may also include constraints related to the geographical location of data, the type of cloud provider, the permitted data storage region, or requirements for using resources with a specific level of security certification.

The first optimization criterion is task execution time. For task j_i , assigned to resource v_j , the execution time can be represented as

$$t_{ij} = \frac{w_i}{p_j} + l_{ij}, \quad (2.33)$$

where w_i is the computational workload of the task; p_j is the performance of resource and l_{ij} is the additional network or communication delay.

The overall makespan of the system is then defined as

$$T(x) = \max_{i=1, \dots, n} t_i^{finish}. \quad (2.34)$$

Minimizing $T(x)$ is aimed at reducing the overall workload processing time and improving the performance of the cloud system.

The second criterion is the cost of resource usage. If c_j is the cost per unit of time for using resource v_j , the total cost can be defined as

$$C(x) = \sum_{i=1}^n \sum_{j=1}^m c_j t_{ij} x_{ij}. \quad (2.35)$$

The function $C(x)$ takes into account that different resources may have different pricing models and that a more powerful resource is not necessarily the most cost-effective option.

The third criterion is energy consumption. In a simplified form, it can be represented as

$$E(x) = \sum_{j=1}^m P_j(u_j) \cdot \tau_j, \quad (2.36)$$

where $P_j(u_j)$ is the power consumed by resource v_j at utilization level u_j ; τ_j is the duration of its active use. This formulation makes it possible to consider the effect of resource consolidation: reducing the number of active nodes can decrease energy consumption, but at the same time it may affect system performance and the level of risk.

The fourth criterion is the overall cyber risk $R(x)$. For each resource, parameters can be defined to describe its level of security, including the implementation level of identity mechanisms, access control, microsegmentation, monitoring, backup, and resilience. Let each threat $q \in Q$ be characterized by the probability of its occurrence $p_q(x, t)$ and its potential impact $I_q(x, t)$. Then, the current cyber risk can be represented as

$$R(x, t) = \sum_{q=1}^s w_q p_q(x, t) I_q(x, t), \quad (2.37)$$

where w_q is the weighting coefficient that represents the importance of the corresponding threat.

Since the cloud environment is dynamic, it is also reasonable to consider changes in threat intensity over time. For this purpose, a function $\lambda_q(t)$ is introduced, which represents the intensity, relevance, or relative probability of the q -th threat at time t . Then, the overall risk over the time interval $[0, T]$ can be defined as

$$R(x) = \sum_{q=1}^s w_q \int_0^T \lambda_q(t) p_q(x, t) I_q(x, t) dt. \quad (2.38)$$

The function $\lambda_q(t)$ makes it possible to consider changes in the external security situation, such as the appearance of new vulnerabilities, an increase in attack intensity, or a change in the level of trust in a specific group of resources. If the threat environment is relatively stable, $\lambda_q(t)$ can be considered constant. In dynamic scenarios, it can be determined based on monitoring data, event logs, or external threat indicators.

In this formulation, cyber risk depends not only on the external environment but also on the resource allocation solution x . For example, placing several critical tasks on the same node may reduce cost and energy consumption but increase the potential consequences if this node is compromised. Similarly, using a resource with a lower level of segmentation or less strict access policies may improve $T(x)$ or $C(x)$, but increase $R(x)$. Therefore, cyber risk should be considered not only as a threshold constraint

$$R(x) \leq R_{max}, \quad (2.39)$$

but also as a separate optimization criterion.

As a result, the problem takes the form of a multi-objective optimization problem:

$$\min F(x) = (T(x), C(x), E(x), R(x)) \quad (2.40)$$

subject to the system of resource, SLA, and security constraints.

If necessary, latency $L(x)$ can also be added to the vector of criteria. However, in the further model, it is reasonable to consider it either as a component of or as an SLA constraint. This makes it possible to avoid an excessive number of interdependent criteria and keep the model interpretable.

Because the objective functions are conflicting, in the general case there is no single solution that simultaneously minimizes $T(x)$, $C(x)$, $E(x)$ and $R(x)$. Therefore, a natural approach to solving this problem is to use the principle of Pareto optimality.

Let $x^{(1)}$ and $x^{(2)}$ be two feasible solutions. Solution $x^{(1)}$ dominates $x^{(2)}$, if $f_k(x^{(1)}) \leq f_k(x^{(2)})$ for all criteria k , and a strict inequality holds for at least one criterion. If there is no other feasible solution that dominates a given solution, this solution is Pareto-optimal.

The set of such solutions can be represented as $P^* = \{x \in X : y \in X, F(y) \prec F(x)\}$, where X is the set of feasible configurations, and the symbol $\prec$ represents Pareto dominance.

The formation of the Pareto front makes it possible to obtain not one solution but a set of trade-off solutions. One solution may provide the minimum cost with a slightly higher level of risk, another may provide the minimum cyber risk at the cost of additional expenses, while a third may provide a balance between performance and energy consumption. This is especially important for cloud systems, where priorities may change depending on the type of workload, task criticality, and the current level of cyber threats.

Thus, the proposed mathematical model combines the classical parameters of cloud resource allocation with a dynamic cyber risk criterion. Its key feature is that $R(x)$ depends not only on the external threat environment but also on the resource configuration itself. This makes it possible to directly consider Zero Trust parameters,

resilience, and provider security characteristics during the optimization process. To find a set of Pareto-optimal solutions, it is reasonable to use a multi-objective evolutionary algorithm, particularly NSGA-II, adapted to consider dynamic cyber risk and security-aware constraints.

To solve the formulated multi-objective cloud resource allocation problem, it is reasonable to use the NSGA-II (Non-dominated Sorting Genetic Algorithm II) evolutionary algorithm. Its selection is based on the discrete nature of the problem, the large number of possible resource allocation configurations, and the presence of several conflicting criteria. Unlike approaches that combine several criteria into a single weighted objective function, NSGA-II makes it possible to generate a set of non-dominated solutions within one optimization process and analyse the trade-offs between execution time $T(x)$, cost $C(x)$, energy consumption $E(x)$, and cyber risk $R(x)$. The effectiveness of NSGA-II modifications for multi-objective resource allocation problems in cloud environments is supported by recent research (Chen J., Du T., Xiao G., 2021).

A specific feature of the proposed approach is the combination of NSGA-II with a security-aware solution evaluation mechanism and adaptation to changes in cyber risk. This makes it possible not only to find trade-off resource configurations but also to adjust the search process when the state of the threat environment changes. Thus, NSGA-II is used as the basic multi-objective search mechanism, while additional procedures for risk assessment, security constraint checking, and adaptive parameter adjustment form its modified version.

A solution to the problem is encoded as a chromosome $X = (x_1, x_2, \dots, x_n)$, where n is the number of computing tasks, and the value $x_i = j$ means that task j_i is assigned to cloud resource v_j , $j \in \{1, \dots, m\}$. This representation is compact and directly corresponds to the matrix of binary variables x_{ij} introduced in the mathematical model. For example, the chromosome $X = (2, 1, 4, 2, 3)$ v_2 , the second task on v_1 , the third task on v_4 , and the fifth task on v_3 .

At the first stage, the initial population is generated $P_0 = \{X_1, X_2, \dots, X_N\}$, where N is the predefined population size. The simplest approach is to generate assignments randomly. However, a significant number of such solutions may violate CPU, RAM, or bandwidth constraints. Therefore, it is reasonable to use a hybrid approach to initial population generation: some individuals are generated randomly to ensure diversity, while others are created using heuristic rules. In particular, tasks can be assigned to resources with sufficient remaining capacity, minimum predicted execution time, or an acceptable level of cyber risk. This approach combines the global search capabilities of the evolutionary algorithm with the advantages of heuristic generation of initial solutions.

After each individual is generated, its feasibility is checked:

$$\sum_i cpu_i x_{ij} \leq CPU_j, \quad \sum_i ram_i x_{ij} \leq RAM_j, \quad \sum_i bw_i x_{ij} \leq BW_j. \quad (2.41)$$

In addition, SLA requirements and defined critical security requirements are checked. If a solution violates resource constraints, a repair procedure can be applied. This procedure moves individual tasks to resources with sufficient remaining capacity. This makes it possible to avoid the accumulation of a large number of infeasible configurations in the population.

For each feasible individual X the vector of criteria is calculated $F(X) = (T(X), C(X), E(X), R(X))$.

The values of $T(X)$, $C(X)$ and $E(X)$ are determined according to the mathematical model presented in the previous section. The value of $R(x)$ is evaluated by considering the characteristics of the resources, Zero Trust mechanisms, and the current state of the threat environment. Thus, two solutions with similar performance characteristics may receive significantly different evaluations because of different levels of cyber risk.

To consider the dynamic nature of cyber risk, the threat intensity function $\lambda_q(t)$ is used. For the current time interval, the risk value is defined as $R(X, t) = \sum_{q=1}^s w_q \lambda_q(t) p_q(X, t) I_q(X, t)$.

The parameters $p_q(X, t)$ and $I_q(X, t)$ depend on the resource configuration. In particular, they can be adjusted depending on the availability of multi-factor authentication, privilege levels, microsegmentation, monitoring, backup, and other Zero Trust characteristics. Accordingly, a resource with a higher level of security reduces the risk component of the solution, but its use may increase $T(X)$, $C(X)$ or $E(X)$.

For critical tasks, it is reasonable to introduce an additional minimum acceptable security level. Let Z_j represent the overall level of Zero Trust implementation for resource v_j and let Z_i^{min} represent the minimum security requirement for task i . Then, the following condition must be satisfied $x_{ij} = 1 \Rightarrow Z_j \geq Z_i^{min}$.

Thus, security is considered in the proposed method in two ways: solutions that do not meet the security requirements are rejected as infeasible, while the residual cyber risk $R(x)$ is minimized among feasible configurations. This mechanism makes it possible to avoid situations in which the algorithm selects a low-cost and high-performance solution that is unacceptable from a security perspective.

The next stage of NSGA-II is non-dominated sorting. The population is divided into a sequence of fronts:

$$F_1, F_2, \dots, F_k, \quad (2.42)$$

where F_1 contains solutions that are not dominated by any other solution in the current population. After these solutions are removed, F_2 is identified, followed by F_3 and so on. Individuals from the first front receive the highest priority during selection. In the proposed formulation, dominance is determined simultaneously according to four criteria.

Thus, solution X_a dominates X_b if

$$\begin{aligned} T(X_a) &\leq T(X_b), C(X_a) \leq C(X_b), \\ E(X_a) &\leq E(X_b), R(X_a) \leq R(X_b), \end{aligned} \quad (2.43)$$

and at least one of these inequalities is strict.

To maintain the diversity of solutions within the same front, crowding distance is used. For each criterion, the individuals are sorted according to the corresponding value. Then, for each internal point, the normalized distance between neighbouring solutions is calculated as:

$$CD_i = \sum_{k=1}^K \frac{f_k^{i+1} - f_k^{i-1}}{f_k^{max} - f_k^{min}}, \quad (2.44)$$

where $K = 4$ is the number of criteria. Preference is given to solutions with a larger crowding distance, which prevents the population from becoming concentrated in only one area of the Pareto front and ensures the representation of different trade-offs between performance, cost, energy consumption, and cyber risk.

New solutions are generated using crossover and mutation operators. During the crossover of two parent chromosomes, some task assignments are inherited by the first offspring from one parent, while the remaining assignments are inherited from the other parent. After the crossover operation, resource and security constraints are checked. Mutation involves changing the assigned resource for one or more tasks:

$$x_i = j \rightarrow x_i = j', j' \neq j. \quad (2.45)$$

The mutation operator generates new configurations and reduces the probability of premature convergence of the algorithm.

To improve efficiency, it is proposed to use an adaptive mutation mechanism related to changes in cyber risk. Let $\bar{\lambda}(t)$ is the aggregated intensity of current threats.

The mutation probability can be defined as:

$$p_m(t) = p_m^{\min} + (p_m^{\max} - p_m^{\min}) \frac{\bar{\lambda}(t)}{\lambda_{\max}}, \quad (2.46)$$

where $p_m^{\min}$ and $p_m^{\max}$ are the minimum and maximum mutation probabilities. When the security situation is stable, the algorithm operates with a lower intensity of changes, preserving the efficient configurations that have already been found. When the intensity of threats increases $p_m(t)$ increases, the mutation probability also increases, which expands the search and improves the probability of quickly finding an alternative and more secure resource allocation solution.

The hybridization of the method also includes the use of a local security-aware heuristic for some of the obtained solutions. If a specific task makes a significant contribution to $R(x)$, the algorithm analyses alternative resources that meet the task requirements and performs local reassignment if it reduces the risk without an unacceptable deterioration of the other criteria. Thus, the global evolutionary search of NSGA-II is combined with targeted local improvement of resource configurations.

The general pseudocode of the proposed adaptive hybrid method is as follows.

Input:

J – set of tasks;

V – set of cloud resources;

N – population size;

$G_{\max}$ – maximum number of generations;

$Q(t)$ – current characteristics of cyber threats.

1. Generate the initial population $P0$:

generate some solutions randomly;

generate some solutions using heuristic rules.

2. For each $X \in P0$:

check CPU, RAM, bandwidth, and SLA constraints;

check security/Zero Trust constraints;

if necessary, apply the repair procedure;

calculate $T(X)$, $C(X)$, $E(X)$, $R(X, t)$.

3. For $g = 1, \dots, G_{\max}$:

update $\lambda_q(t)$ and the risk parameters;
 perform non-dominated sorting;
 calculate the crowding distance;
 perform tournament selection of parents;
 apply crossover;
 determine the adaptive mutation probability $p_m(t)$;
 apply mutation;
 check the feasibility of offspring and apply the repair procedure;
 apply security-aware local improvement to selected solutions;
 calculate $T(X)$, $C(X)$, $E(X)$, $R(X, t)$ for the new individuals;
 combine the parent and offspring populations;
 perform non-dominated sorting of the combined population;
 generate the next population P_g according to rank and crowding distance.

4. Return the set of non-dominated solutions P^* .

Output:

the Pareto set of trade-off cloud resource allocation configurations.

Thus, the proposed method is an adaptive security-aware modification of NSGA-II, in which the classical Pareto optimization mechanism is extended by three components: dynamic cyber risk assessment using $\lambda_q(t)$, security constraints on the feasible placement of critical tasks, and a local heuristic for resource reassignment. Adaptive adjustment of the evolutionary search parameters makes it possible to respond to changes in threat intensity without the need to completely revise the mathematical model.

The result of the algorithm is not a single configuration but a set of Pareto-optimal solutions $P^* = \{X^{(1)}, X^{(2)}, \dots, X^{(k)}\}$, from which a solution can be selected according to the current operating conditions of the system. Under a normal threat level, priority may be given to cost and performance. However, when $\lambda(t)$ increases, it becomes reasonable to move to configurations with a lower $R(x)$, even if they have slightly higher costs or execution time. This feature provides a connection between

multi-objective resource optimization, Zero Trust principles, and adaptive cyber risk management in a dynamic cloud environment.

To evaluate the proposed adaptive security-aware approach to multi-objective cloud resource allocation optimization, it is reasonable to consider several simulation scenarios with different workload levels and cyber threat intensities. The basic scenario represents the normal operating mode of the system with a stable workload and a low value of $\lambda(t)$. The second scenario includes an increase in the number of tasks and the utilization of computing resources, while the third scenario considers an increase in cyber threat intensity. In addition, a combined scenario can be considered, where both the workload and the level of cyber risk increase at the same time. This approach makes it possible to evaluate the ability of the algorithm to adapt resource allocation to changes in both computing and security conditions.

The comparison between traditional and security-aware optimization should be performed using the same sets of tasks and resources. In the traditional case, the solution is generated according to the criteria $F_{trad}(x) = (T(x), C(x), E(x))$, while the proposed approach uses the extended vector $F_{sec}(x) = (T(x), C(x), E(x), R(x))$.

The main difference between these formulations becomes clear when a resource with a low cost or high performance is characterized by increased cyber risk. A traditional algorithm does not have a separate criterion for distinguishing such configurations according to their security level. In contrast, security-aware optimization makes it possible to exclude solutions that do not meet the defined security constraints and to consider $R(x)$ when forming the set of non-dominated solutions.

The result of multi-objective optimization should be analysed using the Pareto front, where each point represents a feasible resource allocation configuration. Moving from three to four criteria expands the space of possible trade-offs. For example, solution x_1 may provide the minimum cost $C(x)$, but have a higher cyber risk $R(x)$; solution x_2 may provide the minimum execution time $T(x)$ but have higher energy consumption; and solution x_3 may provide lower cyber risk by using more secure but more expensive resources. Therefore, the Pareto front makes it possible to avoid

defining a single fixed relationship between the weights of the criteria in advance and to select a configuration according to the current state of the cloud system.

The behaviour of the model when the function $\lambda(t)$ changes is especially important. When the threat intensity is low, priority may be given to solutions with lower $T(x)$, $C(x)$ and $E(x)$. An increase in $\lambda(t)$ increases the contribution of the corresponding threats to $R(x)$. As a result, some configurations that were previously efficient may lose their advantage. The algorithm can then move to another area of the Pareto front, where resources with a higher level of Zero Trust implementation, better segmentation, access control, and monitoring are used. In this way, resource allocation can adapt to changes in the security situation.

The practical application of the proposed approach is especially appropriate for hybrid and multi-cloud systems, where it is possible to choose between resources with different levels of performance, cost, and security. For non-critical tasks under a low threat level, cost-effective configurations can be used. For tasks related to critical services or sensitive data, it is reasonable to define minimum security and Zero Trust requirements and select only resources that meet these requirements. If $\lambda(t)$ increases, the system should provide a repeated evaluation of the Pareto set and, if necessary, workload migration.

Thus, the practical advantage of security-aware optimization is not the unconditional minimization of cyber risk at the cost of other system characteristics, but the possibility of making a controlled trade-off between cost, performance, energy consumption, and cyber risk. The proposed approach can be used as a component of decision support systems for cloud infrastructure management, where current monitoring data and threat assessments are used to dynamically update $R(x)$ and select the most appropriate resource configuration. This provides a connection between mathematical optimization, Zero Trust principles, and risk-oriented management of the digital resilience of cloud systems.

REFERENCES FOR SUBSECTION 2.4

1. Chen J., Du T., Xiao G. A multi-objective optimization for resource allocation of emergent demands in cloud computing. *Journal of Cloud Computing*. 2021. Vol. 10. Art. 20. DOI: 10.1186/s13677-021-00237-7.
2. Shi F., Lin J. Virtual Machine Resource Allocation Optimization in Cloud Computing Based on Multiobjective Genetic Algorithm. *Computational Intelligence and Neuroscience*. 2022. Vol. 2022. Art. 7873131. DOI: 10.1155/2022/7873131.
3. Malti A. N., Hakem M., Benmammar B. Multi-objective task scheduling in cloud computing. *Concurrency and Computation: Practice and Experience*. 2022. Vol. 34, № 25. Art. e7252. DOI: 10.1002/cpe.7252.
4. Zhang H., Wu Y., Sun Z. EHEFT-R: multi-objective task scheduling scheme in cloud computing. *Complex & Intelligent Systems*. 2022. Vol. 8. P. 4475–4482. DOI: 10.1007/s40747-021-00479-7.
5. Latreche I., Slatnia S., Kazar O., Batouche M. A third generation genetic algorithm NSGA-III for task scheduling in cloud computing. *Journal of King Saud University – Computer and Information Sciences*. 2022. Vol. 34, No. 9. P. 7515–7529. DOI: 10.1016/j.jksuci.2022.03.017.
6. Sissodia R., Rauthan M. M. S., Naithani K. A Survey on Multi-Objective Tasks and Workflow Scheduling Algorithms in Cloud Computing. *International Journal of Cloud Applications and Computing*. 2022. Vol. 12, № 1. P. 1–16. DOI: 10.4018/IJCAC.297100.
7. Mangalampalli S., Karri G. R., Kose U. Multi objective trust aware task scheduling algorithm in cloud computing using whale optimization. *Journal of King Saud University – Computer and Information Sciences*. 2023. Vol. 35. P. 791–809. DOI: 10.1016/j.jksuci.2023.01.016.
8. Luo Z.-Y., Chen Y.-N., Liu X.-T. Research on Sparrow Search Optimization Algorithm for multi-objective task scheduling in cloud computing environment. *Journal of Intelligent & Fuzzy Systems*. 2023. Vol. 45, № 6. DOI: 10.3233/JIFS-232527.
9. Wang X., Lou H., Dong Z., Yu C., Lu R. Decomposition-based multi-objective evolutionary algorithm for virtual machine and task joint scheduling of cloud

computing in data space. *Swarm and Evolutionary Computation*. 2023. Vol. 77. Art. 101230. DOI: 10.1016/j.swevo.2023.101230.

10. Prathiba S., Sankar S. Energy-efficient resource allocation in cloud infrastructure using L3F-MGA and E-ANFIS. *Measurement: Sensors*. 2024. Vol. 31. Art. 100965. DOI: 10.1016/j.measen.2023.100965.

11. Çağlar İ., Altılar D. T. Look-ahead energy efficient VM allocation approach for data centers. *Journal of Cloud Computing*. 2022. Vol. 11. Art. 11. DOI: 10.1186/s13677-022-00281-x.

12. Liu J., Liu X. Energy-efficient allocation for multiple tasks in mobile edge computing. *Journal of Cloud Computing*. 2022. Vol. 11. Art. 71. DOI: 10.1186/s13677-022-00342-1.

13. Wang B., Wang C., Huang W., Song Y., Qin X. Security-aware task scheduling with deadline constraints on heterogeneous hybrid clouds. *Journal of Parallel and Distributed Computing*. 2021. Vol. 153. P. 15–28. DOI: 10.1016/j.jpdc.2021.03.003.

14. Wang B., Liu F., Lin W., Ma Z., Xu D. Energy-efficient collaborative optimization for VM scheduling in cloud computing. *Computer Networks*. 2021. Vol. 201. Art. 108565. DOI: 10.1016/j.comnet.2021.108565.

15. Saxena D., Singh A. K. A proactive autoscaling and energy-efficient VM allocation framework using online multi-resource neural network for cloud data center. *Neurocomputing*. 2021. Vol. 426. P. 248–264. DOI: 10.1016/j.neucom.2020.08.076.

16. Rose S., Borchert O., Mitchell S., Connelly S. Zero Trust Architecture. NIST Special Publication 800-207. Gaithersburg : National Institute of Standards and Technology, 2020. DOI: 10.6028/NIST.SP.800-207.

17. Chandramouli R., Butcher Z. A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Cloud Environments. NIST Special Publication 800-207A. Gaithersburg : National Institute of Standards and Technology, 2023. DOI: 10.6028/NIST.SP.800-207A.

18. Chandramouli R., Butcher Z., Chetal A. Attribute-based Access Control for Microservices-based Applications Using a Service Mesh. NIST Special Publication

800-204B. Gaithersburg : National Institute of Standards and Technology, 2021. DOI: 10.6028/NIST.SP.800-204B.

19. Chandramouli R. Implementation of DevSecOps for a Microservices-based Application with Service Mesh. NIST Special Publication 800-204C. Gaithersburg : National Institute of Standards and Technology, 2022. DOI: 10.6028/NIST.SP.800-204C.

20. Chandramouli R., Kautz F., Torres-Arias S. Strategies for the Integration of Software Supply Chain Security in DevSecOps CI/CD Pipelines. NIST Special Publication 800-204D. Gaithersburg : National Institute of Standards and Technology, 2024. DOI: 10.6028/NIST.SP.800-204D.

21. Chandramouli R., Hales W. A Data Protection Approach for Cloud-Native Applications. NISTIR 8505. Gaithersburg : National Institute of Standards and Technology, 2024. DOI: 10.6028/NIST.IR.8505.

22. Chandramouli R., Butcher Z., Callaghan J. Service Mesh Proxy Models for Cloud-Native Applications. NIST Special Publication 800-233. Gaithersburg : National Institute of Standards and Technology, 2024. DOI: 10.6028/NIST.SP.800-233.

23. Sehgal N. K., Bhatt P. C. P., Acken J. M. Cloud Computing with Security: Concepts and Practices. 2nd ed. Cham : Springer, 2020. 355 p. DOI: 10.1007/978-3-030-24612-9.

24. Calles M. A. Serverless Security: Understand, Assess, and Implement Secure and Reliable Applications in AWS, Microsoft Azure, and Google Cloud. Berkeley : Apress, 2020. 347 p. DOI: 10.1007/978-1-4842-6100-2.

25. Cloud Security Alliance. Security Guidance for Critical Areas of Focus in Cloud Computing v5. Cloud Security Alliance, 2024.

26. Cloud Security Alliance. Top Threats to Cloud Computing: Pandemic Eleven. Cloud Security Alliance, 2022.

27. European Union Agency for Cybersecurity. Cloud Cybersecurity Market Analysis. Athens : ENISA, 2023.

28. European Union Agency for Cybersecurity. EUCS – Cloud Services Scheme. Draft Candidate Scheme v1.0. ENISA, 2020. ENISA, 2020.

29. European Union Agency for Cybersecurity. ENISA Threat Landscape 2024. Athens : ENISA, 2024.
30. European Union Agency for Cybersecurity. NIS Investments 2024. Athens : ENISA, 2024.
31. Dvirna O. A., Naboka S. V. Cloud Service Performance Optimization: Methods and Their Effectiveness. Control, Navigation and Communication Systems. 2026. No. 1 (83). P. 62–68. DOI: 10.26906/SUNZ.2026.1.062.
32. Holovko H. V., Dvirna O. A., Naboka S. V. Zero Trust for Small and Medium-Sized Businesses. Taurida Scientific Herald. Series: Technical Sciences. 2026. Issue 1, Part 1. P. 84–92. DOI: 10.32782/tnv-tech.2026.1.1.9.
33. Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive). Official Journal of the European Union. 2022. L 333. 27.12.2022. P. 80–152.