Ministry of Education and Science of Ukraine

KHARKIV NATIONAL UNIVERSITY OF RADIO ELECTRONICS
RIGA NORDIC UNIVERSITY
NATIONAL AVIATION UNIVERSITY
INSTITUTE FOR INFORMATION RECORDING
NATIONAL UNIVERSITY "LVIV POLYTECHNIC"
NATIONAL DEFENSE UNIVERSITY OF AZERBAIJAN REPUBLIC
IVAN KOZHEDUB KHARKIV NATIONAL AIR FORCE UNIVERSITY

Eighth International
Scientific and Technical Conference

«COMPUTER AND INFORMATION SYSTEMS AND TECHNOLOGIES»

October 09-10

Kharkiv 2025

Eighth International Scientific and Technical Conference «COMPUTER AND INFORMATION SYSTEMS AND TECHNOLOGIES». Kharkiv: NURE. 2025. – 125 p.

This publication is prepared by
Department of Electronic Computers
KHARKIV NATIONAL UNIVERSITY OF RADIO ELECTRONICS (NURE)

61166, Ukraine,
14 Nauky Avenue, Kharkiv
tel: +38 (057) 702-13-54
E-mail: info@csitic.com
Online ISSN: 2710-463X

Adaptive Methods for Managing Distributed Computing with Built-in Self-Healing Mechanisms

Volk Maksym Oleksandrovich,
Buhrii Andrii Mykolaiovych,
Zaihraev Dmytro Serhiyovych,
Panchenko Yelisei Oleksandrovych,
Zveriev Pylyp Viktorovych

Kharkiv National University of Radio Electronics, 14 Nauky Ave, Kharkiv UA-61166, Ukraine, maksym.volk@nure.ua

Abstract. *Cloud computing systems increasingly face challenges of heterogeneity, scalability, and reliability. This paper presents an integrated approach to distributed computing management with embedded self-recovery mechanisms in heterogeneous cloud environments. The architecture combines adaptive scheduling, predictive analytics, and machine learning methods (LSTM, Random Forest, reinforcement learning) to forecast failures and optimize recovery scenarios. Experimental evaluation demonstrates improved fault prediction accuracy (12–15%), reduction of false alarms (25–32%), faster incident response (from 11–17 minutes to 5–9), and a 17% decrease in operational costs. The results confirm the technical and economic feasibility of the proposed approach for mission-critical cloud infrastructures.*

Keywords*: cloud computing; distributed systems; self-recovery; predictive analytics; machine learning; resource management, heterogeneous environments, service availability.*

I. Introduction and Problem Statement

Over the last two decades, cloud computing has evolved into one of the most influential paradigms in information technology, radically changing how organizations store, process, and analyze data. The shift from centralized data centers to distributed, dynamically scalable cloud infrastructures has provided unprecedented levels of flexibility, cost-efficiency, and service availability. As enterprises increasingly depend on cloud platforms to run mission-critical applications, expectations regarding reliability, security, and adaptability continue to rise.

A defining characteristic of modern cloud environments is heterogeneity. Unlike the early days of cloud computing, where infrastructures were relatively homogeneous and mainly CPU-based, contemporary platforms integrate a variety of hardware accelerators (GPUs, TPUs, FPGAs), virtualization technologies, operating systems, and communication protocols [1]. This diversity opens new opportunities for performance optimization, particularly for AI and data-intensive workloads, but simultaneously complicates orchestration, scheduling, and reliability management. Managing distributed processes across such heterogeneous ecosystems requires advanced approaches capable of dynamically adapting to fluctuating workloads and unforeseen system events.

Conventional management techniques – including static task allocation, heuristic scheduling, and centralized planners – were effective in more uniform and stable infrastructures but now face significant limitations. They often fail to respond adequately to rapid changes in workload, provide insufficient fault tolerance, and lack predictive mechanisms to anticipate system failures. As a result, service-level agreements (SLAs) are harder to guarantee, downtime increases, and operational costs rise due to inefficiencies in resource utilization [2].

In parallel, the concept of self-recovery has emerged as a key enabler of resilient cloud services. Self-recovery mechanisms, such as automatic restart of failed services, redeployment of corrupted containers, or migration of tasks to healthy nodes, aim to restore operability with minimal human intervention. However, these mechanisms are typically implemented independently from scheduling and orchestration modules [3-4].

The situation becomes even more complex in multi-cloud and hybrid deployments, where resources from several providers must be orchestrated. Variations in platforms, APIs, and security policies make it difficult to establish unified recovery workflows. Interoperability gaps not only reduce efficiency but also introduce new security vulnerabilities, as recovery procedures themselves can be exploited if not properly coordinated [5].

Another challenge is the rise of AI- and ML-driven workloads. Such applications are highly latency-sensitive and often depend on specialized hardware, such as GPU clusters or high-speed interconnects. Failure of these components can lead to severe performance degradation. Traditional recovery mechanisms, which do not integrate closely with workload schedulers, are often too slow or incapable of reallocating tasks efficiently in the presence of hardware-specific dependencies.

Given these challenges, the core research problem can be formulated as follows: current cloud management frameworks lack integrated mechanisms that unify resource orchestration, workload scheduling, and self-recovery into a single adaptive system. Without such integration, cloud infrastructures struggle to simultaneously maintain high performance, cost efficiency, and resilience against failures.

The objective of this research is to develop and validate an integrated approach that combines distributed computation management with predictive self-recovery in heterogeneous cloud environments.

II. Problem Solution and Results

The proposed solution introduces a unified control architecture that integrates distributed computation management with advanced self-recovery techniques in heterogeneous cloud infrastructures. Unlike conventional approaches that treat scheduling and fault recovery as separate domains, this method establishes a closed feedback cycle where monitoring, predictive analytics, adaptive planning, execution, and recovery verification interact continuously.

At the core of the approach is a layered architecture. The monitoring layer provides both periodic and event-driven telemetry acquisition, ensuring that anomalies can be captured

with minimal latency while avoiding excessive overhead. The analytics and prediction layer applies hybrid AI/ML models:

The proposed methodological framework integrates distributed task management with automated self-recovery into a closed-loop control cycle: monitoring → anomaly detection → prediction → scheduling → execution → recovery validation. This architecture ensures continuous adaptation of policies based on predictive models.

LSTM recurrent neural networks analyze load dynamics and forecast potential bottlenecks. Random Forest classifiers identify categories of failures, allowing context-aware recovery. Reinforcement Learning (RL) agents support decision-making for adaptive scheduling, learning optimal strategies under changing workloads.

The scheduling layer integrates classical heuristics (e.g., Weighted Round Robin, Genetic Scheduling) with RL-driven decision-making to balance responsiveness with robustness. The execution and orchestration layer leverages Kubernetes, OpenStack, and container orchestration mechanisms to implement task migration and resource reallocation. Finally, the self-recovery layer incorporates automated restart, container redeployment, activation of replicas, and dynamic load migration. A built-in verification module evaluates the outcome and updates predictive models to improve future decision-making.

Two distinct scenarios are supported. Predictive prevention: when anomaly forecasts exceed a defined probability threshold, the scheduler proactively migrates critical tasks, preventing cascading failures. Reactive recovery: in case of abrupt failure, backup containers or virtual machines are deployed instantly, while the failed node undergoes automated diagnostics before rejoining the pool. This dual strategy combines preventive resilience with rapid reactivity, reducing both downtime and wasted resources.

Validation was carried out using hybrid SaaS environments comprising 200 virtual machines and over 400 containers across heterogeneous clusters. Testbeds simulated varying workload intensities, including AI/ML inference tasks, real-time transactional services, and mixed I/O-bound processes. Telemetry included CPU/GPU utilization, memory load, latency, and error rates.

The evaluation revealed several key improvements. Fault prediction: accuracy increased by 12–15% compared with threshold-based systems, enabling early intervention. False alarms: reduced by 25–32%, lowering unnecessary task migrations and restarts. Incident response: mean response time shortened from 11–17 minutes to 5–9 minutes. Availability: mean time to recovery (MTTR) decreased from 2.4 hours to under 1 hour, improving availability from 99.77% to 99.94%.

Resource utilization: predictive scheduling lowered average CPU usage from 67% to 58% and RAM from 77% to 68%, while effective utilization ratios rose to 80%. Economic benefits: operational expenditures were reduced by 17% due to optimized resource allocation and lower redundancy requirements.

Benchmarking against conventional systems highlighted several advantages. Traditional systems rely heavily on static redundancy, which increases cost but not adaptability.

The approach is particularly relevant for environments requiring continuous service delivery, such as financial transaction systems, large-scale AI training platforms, and mission-critical healthcare applications. In such domains, the ability to predict failures, reallocate workloads dynamically, and recover rapidly provides not only technical resilience but also tangible economic value.

III. Conclusions

This research proposed and validated an integrated methodology for managing distributed computations with embedded self-recovery in heterogeneous cloud infrastructures. The approach distinguishes itself by unifying two traditionally separate areas—scheduling and fault recovery—into a closed-loop cycle enhanced by predictive analytics and machine learning.

The architecture demonstrated significant technical improvements. Fault prediction accuracy increased by up to 15% over threshold methods, while false positives were reduced by up to 32%. Mean time to recovery dropped by more than half, leading to measurable gains in service availability. Resource utilization was optimized, reducing redundancy without compromising resilience. These technical benefits translated into a 17% decrease in operational expenditures and an increase of 25–32% in integral efficiency indices.

Beyond measurable results, the proposed solution establishes a framework adaptable to future developments in cloud computing. The modular design allows the incorporation of additional AI models, specialized monitoring agents, or advanced orchestration tools. Its adaptability makes it suitable for multi-cloud and hybrid deployments, where interoperability and resilience are critical. The findings confirm not only the feasibility but also the necessity of integrated management in modern heterogeneous environments. The proposed unified architecture addresses these challenges by ensuring real-time adaptability, fault tolerance, and economic efficiency.

Future work should focus on extending predictive capabilities to cross-cloud orchestration, incorporating security-aware recovery mechanisms, and exploring bio-inspired or federated learning techniques to further reduce latency and improve decision accuracy. These directions will strengthen the path toward fully autonomous, self-healing cloud infrastructures capable of sustaining mission-critical operations under dynamic and uncertain conditions.

REFERENCES

[1] C. Ji, H. Luo, "Cloud-based AI systems: leveraging large language models for intelligent fault detection and autonomous self-healing," 2025 IEEE 7th International Conference on Communications, Information System and Computer Engineering (CISCE). Guangzhou, China, 2025, P. 226-229, DOI: 10.1109/CISCE65916.2025.11065204.

[2] R. Vankayalapati, C. Pandugula, "AI-powered self-healing cloud infrastructures: A paradigm for autonomous fault recovery," Migration Letters Vol.19, No.6. 2022. P. 1173-1187. DOI: 10.2139/ssrn.5052024

[3] Volk M., Kozina O., Buhrii A., Osiievskyi S., Kozin M., Volk D., Diachenko D., Turinskyi Y, "Devising a method for data consistency at replication in multicloud systems," Eastern-European Journal of Enterprise Technologies, 2025, Vol.4 №2 (136), 14-22. DOI: doi.org/10.15587/1729-4061.2025.332189

[4] O. Mamchych, M. Volk, "Smartphone Based Computing Cloud and Energy Efficiency," 12th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece. 2022. pp.1-5, DOI: 10.1109/DESSERT58054.2022.10018740

[5] S. Harshal, P. Jay, "Self-Healing AI: Leveraging Cloud Computing for Autonomous Software Recovery," International Journal of Intelligence and Self-Automated Engineering. Vol.10. No. 3s. – 2022. – pp 341-3

CONTENT

Scientific publication

«COMPUTER AND INFORMATION SYSTEMS AND TECHNOLOGIES»

Responsible for the release:	RUBAN Igor, KOVALENKO Andriy, MARTOVYTSKYI Vitalii
Computer layout:	KOVALENKO Andriy, MARTOVYTSKYI Vitalii

Collection materials are published in the author's version without editing

Approved by the Scientific and Technical Council of Kharkiv National University of Radio Electronics № 5/5 23.04.2021