12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

# A MULTI-OBJECTIVE APPROACH TO OPTIMIZING VIRTUALIZED MILITARY DATA CENTER INFRASTRUCTURE WITH GENETIC ALGORITHMS

NIKOLA ZOGOVIĆ
Mihajlo Pupin Institute, University of Belgrade, Belgrade, Republic of Serbia
nikola.zogovic@pupin.rs, ORCID: 0000-0002-9052-6422

SAVA STANIŠIĆ
Faculty of Technical Sciences, University of Kragujevac, Čačak, Republic of Serbia
sava.stanisic@vs.rs, ORCID: 0009-0002-3118-0537

BORISLAV ĐORĐEVIĆ
Mihajlo Pupin Institute, University of Belgrade, Belgrade, Republic of Serbia
borislav.djordjevic@pupin.rs, ORCID: 0000-0002-6145-4490

***Abstract:*** *Selecting cost-efficient, highly available, and reliable virtualized infrastructure is a critical challenge for military data centers. This paper extends a prior two-objective optimization framework by introducing reliability as a third objective alongside cost minimization and availability maximization. The framework uses the Non-dominated Sorting Genetic Algorithm II (NSGA-II) to model trade-offs among three conflicting objectives within a virtualized military data center, where services are deployed as virtual machines across physical nodes with heterogeneous failure characteristics. Five VM types with distinct cost, availability, and failure rate profiles are defined to model realistic military infrastructure. Experimental results with the pymoo library demonstrate that the three-objective approach yields 119 Pareto-optimal solutions spanning $0.18–$4.94/h in cost, 98.0–100.0000% in availability, and 23.7–99.9996% in reliability over a 720-hour horizon. Critically, at matched cost points, the three-objective approach achieves up to 20.5 percentage points higher reliability than the two-objective baseline by diversifying VM type selection rather than stacking identical low-cost instances.*

***Keywords:*** *multi-objective optimization, NSGA-II, Pareto optimality, military data center, virtualization, availability, reliability*

## 1. INTRODUCTION

Modern military operations depend on robust digital infrastructure to support command and control (C2), intelligence, surveillance, and reconnaissance (ISR), logistics management, and secure communications [1, 2]. These services are increasingly deployed on virtualized platforms within military data centers, both in garrison facilities and in field-deployable configurations. Unlike commercial cloud environments where downtime results in financial loss, failures in military IT infrastructure can compromise mission execution, situational awareness, and force protection [3].

Military data centers operate under unique constraints. They are often isolated from public networks, must function in austere or contested environments, and face rigid procurement budgets. The selection of virtual machine (VM) types and their allocation across physical servers is a critical planning decision that directly affects operational readiness. From the requirement to ensure uninterrupted support for command, control, and intelligence functions, availability and reliability arise as fundamental system objectives. Availability, the probability that a system is operational, determines acceptable outage duration [4]. Reliability, the probability of continuous failure-free operation over time T, is equally critical, as even brief C2 or ISR interruptions during operations have tactical consequences [5].

Traditional single-objective optimization approaches fall short in addressing this multi-faceted problem. The authors proposed a two-objective NSGA-II framework in [6] to balance cost and availability. However, it did not explicitly optimize reliability, instead assuming that higher availability implies higher reliability, an assumption that fails when VM types have heterogeneous failure characteristics.

This paper extends [6] by incorporating reliability as a third objective and adapting the optimization framework to military data centers. The results show that optimizing only cost and availability systematically favors homogeneous VM deployments, typically using multiple low-cost instances to maximize availability while reducing reliability because of their higher failure probability. Similar to a military force relying on a single weapon system, such configurations are efficient but vulnerable to type-specific weaknesses. Adding reliability

**Corresponding author:** N. Zogović;
E-mail: nikola.zogovic@pupin.rs
DOI: 10.5937/OTEH26079Z

as a third objective diversifies VM portfolios by failure-rate profile, providing a computational analogue of defense in depth. This optimization increases modeled reliability by up to 20.5 percentage points at the same cost.

This paper makes three contributions: (1) it introduces reliability as a third optimization objective for virtualized military data centers; (2) it defines a five-tier VM model capturing the heterogeneous failure characteristics of military-grade hardware; and (3) it demonstrates that VM type diversification, rather than redundant replication of identical instances, improves reliability. Unlike prior multi-objective cloud studies that optimize cost and availability or task scheduling without reliability [6–10], the proposed formulation explicitly captures the cost–availability–reliability trade-off, providing actionable guidance for infrastructure planners.

## 2. RELATED WORK

Multi-objective optimization (MOO), particularly using NSGA-II, has been widely applied to cloud infrastructure optimization. Tan et al. [7] applied NSGA-II to microservice allocation in containerized clouds, while Verma et al. [8] identified it as a state-of-the-art method for Pareto-based combinatorial optimization.

In military computing, virtualization is a key enabler of deployable C2 systems and tactical data centers. Fault-tolerant VM placement for redundant server clusters was studied in [3], [9] presents NSGA-II application to cloud task scheduling, while [10] presents heterogeneous IoT–fog–cloud environments. More recent work [11,12] has extended multi-objective, QoS-aware fault-tolerant VM placement to data-center-scale systems.

In previous work [6], the authors proposed a two-objective NSGA-II framework balancing cost and availability, and Stanišić [13] investigated Docker file-system performance for containerized workloads. However, existing studies do not explicitly optimize reliability, often consider only a limited range of VM types, and largely overlook VM type diversification as a mechanism for improving system resilience. This paper addresses these gaps by introducing reliability as a third optimization objective and modeling heterogeneous VM failure characteristics.

## 3. METHODOLOGY

The optimization framework uses NSGA-II [14] to simultaneously minimize cost and maximize availability and reliability for a virtualized military data center.

### 3.1. Problem Formulation and Military VM Tier Model

The decision vector $x = [n_1, n_2, n_3, n_4, n_5]$ represents the number of VM instances of each type to deploy, where $n_i \in \{0, 1, \ldots, 10\}$. The uniform upper bound of 10 is a scenario-specific capacity proxy based on data center resource constraints (compute, memory, space, and power). For operational use, it must be replaced by tier-specific limits derived from the actual host inventory. Five VM types are defined to model the range of hardware typically available in military data centers, from commercial off-the-shelf (COTS) servers used for administrative workloads to hardened, fault-tolerant platforms designed for mission-critical C2 systems. The specifications are shown in Table 1.

**Table 1.** Military Data Center VM Tier Specifications

| VM Tier | Cost ($/h) | Avail. | $\lambda$ (/h) | Typical Use |
|---|---|---|---|---|
| T1: COTS Basic | $0.18 | 0.980 | 0.002 | Admin, logs |
| T2: COTS Enhanced | $0.37 | 0.990 | 0.001 | Logistics |
| T3: Hardened | $0.52 | 0.995 | 0.0005 | SIGINT proc. |
| T4: Mission-Crit. | $0.78 | 0.999 | 0.0002 | ISR feeds |
| T5: Fault-Tolerant | $1.15 | 0.9995 | 0.0001 | C2 systems |

The tier model reflects real-world military procurement trade-offs. Tier 1 (COTS Basic) represents commodity servers suitable for non-critical workloads such as administrative systems and log aggregation. Tiers 3–5 represent progressively hardened platforms with redundant power supplies, ECC memory, and hot-swappable components, designed for signal intelligence processing, ISR data feeds, and command and control systems respectively. The failure rate $\lambda$ decreases by an order of magnitude from T1 to T5, reflecting the engineering investment in higher-tier hardware.

The failure rate parameters are derived from representative mean time between failures (MTBF) values: T1 corresponds to MTBF ≈ 500h (commodity COTS hardware), T2 to MTBF ≈ 1000h (enterprise-grade with ECC RAM), T3 to MTBF ≈ 2000h (ruggedized platforms, e.g., MIL-STD-810 compliant), T4 to MTBF ≈ 5000h (mission-critical with redundant PSU and cooling), and T5 to MTBF ≈ 10000h (fault-tolerant systems with hot-standby components). Availability values are derived from failure rates assuming a mean time to repair (MTTR) of 1 h (T1–T2) and 0.25 h (T3–T5), reflecting the faster repair of modular, hot-swappable hardware in higher tiers.

### 3.2. Objective Functions

Three objectives are defined. All objectives are formulated as minimization for NSGA-II compatibility.

**Objective 1: Cost Minimization.** The total infrastructure cost is the sum of per-instance costs:

$$C(x) = \sum_i n_i \cdot c_i \quad (1)$$

where $c_i$ is the hourly cost of VM type $i$, given in column 2 of Table 1. The optimization minimizes $f_1 = C(x)$.

**Objective 2: Availability Maximization.** Service availability is modeled using the classical parallel-system

availability model, assuming independent VM failures. The probability that at least one VM remains operational is

$$A(x) = 1 - \prod_i (1 - a_i)^{n_i} \quad (2)$$

where $a_i$ is the per-instance availability of VM type $i$ (Table 1). The optimization minimizes $f_2 = 1 - A(x)$.

**Objective 3: Reliability Maximization.** System reliability over the time horizon $T$ is modeled using the classical exponential lifetime model and parallel-system reliability assumption, yielding the probability that at least one VM remains operational:

$$R(x) = 1 - \prod_i (1 - e^{-\lambda_i \cdot T})^{n_i} \quad (3)$$

where $\lambda_i$ is the failure rate of VM type i (Table 1), and $T = 720\text{h}$ (30 days). The time horizon is mission-dependent and reflects the operation or deployment duration (Table 3). The optimization minimizes $f_3 = 1 - R(x)$. Unlike availability, reliability depends on the failure rate $\lambda_i$ rather than the availability $a_i$; consequently, a VM type may exhibit high availability through rapid recovery yet low reliability due to frequent failures, or vice versa.

Equations (2) and (3) assume independent VM failures and omit common-cause failures from shared infrastructure. Thus, the calculated availability and reliability may be optimistic when instances share a physical fault domain. Incorporating common-cause failures and fault-domain placement is left for future work.

### 3.3. Constraints and Problem Statement

The optimization is subject to a budgetary constraint $C(x) \leq B$, where $B$ is the predefined operational budget. Two scenarios model distinct military contexts: $B = \$5/h$ represents a field-deployable tactical data center with limited power and hardware, while $B = \$15/h$ represents a garrison facility with greater infrastructure capacity. The per-type instance count is bounded as described in Section 3.1; no explicit floor is imposed on availability or reliability, so the reported Pareto fronts span the full feasible range of these objectives.

Combining the three objectives with the budget and instance-count constraints, the complete optimization problem is stated as:

$$x* = argmin_x \left(f_1(x), f_2(x), f_3(x)\right), \; s.t. \; C(x) \leq B, \; n_i \in \{0, \ldots, 10\} \quad (4)$$

Each component $n_i$ takes one of 11 integer values, so the decision space comprises $11^5 = 161{,}051$ candidate configurations, over which NSGA-II searches for the Pareto-optimal set.

### 3.4. Algorithm Configuration

NSGA-II is configured with a population size of 120, 300 generations, simulated binary crossover (SBX; probability 0.9, $\eta$ = 15), polynomial mutation (probability 0.2, $\eta$ = 20), and integer-rounding repair. For the algorithm comparison in Section 4.6, NSGA-III is configured with the same population size, generation count, and genetic operators, using Das–Dennis reference directions to guide selection. The implementation uses the pymoo library (v0.6.1). Hypervolume convergence is tracked to verify algorithm convergence.

## 4. EXPERIMENTAL RESULTS

### 4.1. Two-Objective Baseline

The two-objective optimization (cost vs. availability), the problem originally defined in our prior work [6], under the \$5/h budget produces 10 Pareto-optimal solutions. Fig. 1 shows the Pareto front.

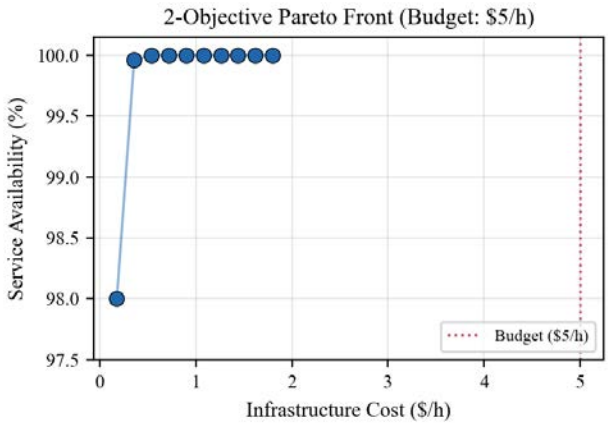



**Figure 1.** 2-Objective Pareto Front (Budget: \$5/h)

A critical observation is that all solutions consist exclusively of GP-Low instances (the cheapest VM type). The algorithm maximizes availability by stacking identical low-cost VMs, ranging from 1 instance (\$0.18/h, 98.0%) to 10 instances (\$1.80/h, 100.0%). This monoculture strategy is optimal for availability but ignores reliability implications.

### 4.2. Three-Objective Optimization

The three-objective optimization, defined by the objectives in Eqs. (1)–(3) and the constraints of Section 3.3, under the $\$5/h$ budget yields 119 Pareto-optimal solutions spanning cost $\$0.18 – \$4.94/\text{h}$, availability $98.0 – 100.0000\%$, and reliability $23.7 – 99.9996\%$ (T = 720h). Fig. 2 shows the 3D Pareto front.

The Pareto surface reveals the three-way trade-off. Low-cost solutions (bottom-left) sacrifice both availability and reliability. As cost increases, both metrics improve, but the rate of improvement differs, with reliability showing steeper gains at moderate cost levels due to the exponential nature of the failure rate model.

Under the relaxed budget of $\$15/\text{h}$, the Pareto front extends to 119 solutions with reliability reaching 100.00% (Fig. 3).

### 4.3. Key Finding: Reliability Improvement over the Two-Objective Baseline through VM Diversification

The central finding of this study is revealed by comparing 2-objective and 3-objective solutions at matched cost points. Fig. 4 presents this comparison and Fig. 5 shows the cost-reliability projection for both budget scenarios.

The mechanism is clear: the 2-objective approach, optimizing only cost and availability, consistently selects homogeneous configurations of the cheapest VM tier (T1: COTS Basic, $\lambda = 0.002/\text{h}$). While stacking many T1 instances achieves near-perfect availability through redundancy, the high per-instance failure rate degrades reliability over the 720h (30-day) evaluation horizon, a duration representative of a typical military deployment rotation or exercise cycle.

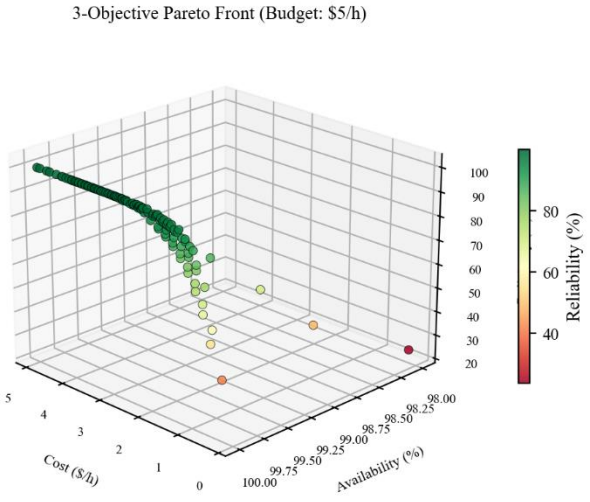



**Figure 2.** 3-Objective Pareto Front (Budget: $5/h)

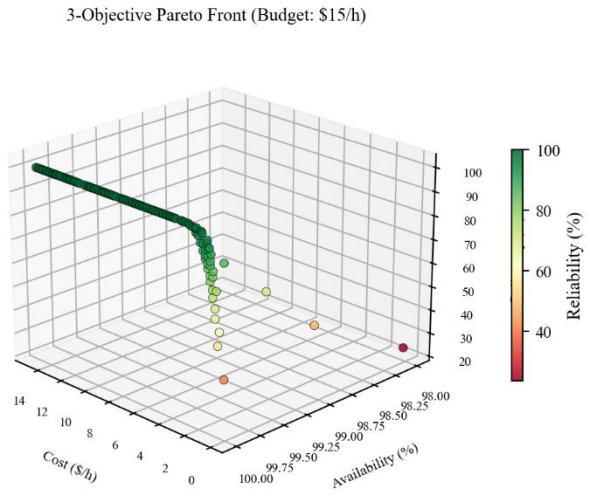



**Figure 3.** 3-Objective Pareto Front (Budget: $15/h)

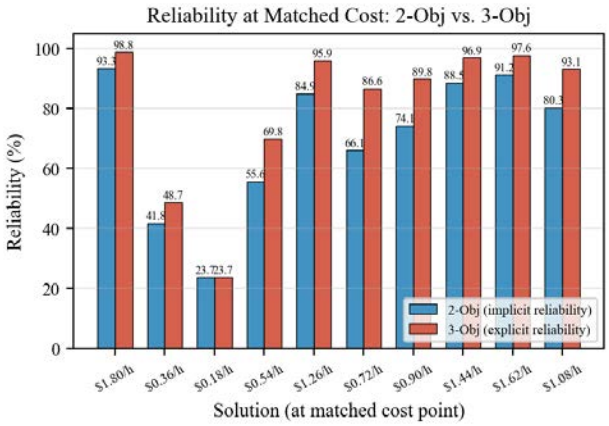



**Figure 4.** Reliability Comparison at Matched Cost Points

The 3-objective approach discovers that mixing VM tiers, particularly by including T3: Hardened ($\lambda = 0.0005/\text{h}$) and T4: Mission-Critical ($\lambda = 0.0002/\text{h}$) instances, dramatically improves reliability. For example, at $\$0.72/\text{h}$, the 2-objective solution deploys 4×T1 instances and achieves 66.1% reliability, while the 3-objective solution deploys 1×T4 instance and achieves 86.6%, a relative gain of 20.5 percentage points over the two-objective baseline at the same cost. In military terms, this is the difference between a C2 system that is expected to fail multiple times during a 30-day deployment and one that has an 87% probability of continuous operation.

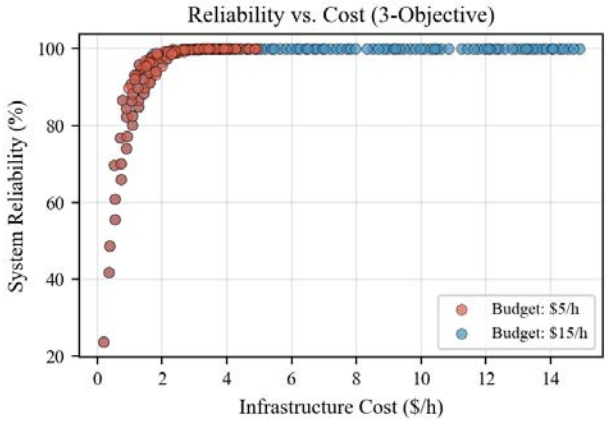



**Figure 5.** Reliability vs. Cost Projection

**Table 2.** Reliability Improvement: 2-Obj vs. 3-Obj at Matched Cost

| Cost ($/h) | 2-Obj Rel.(%) | 3-Obj Rel.(%) | Δ Rel.(pp) |
|---|---|---|---|
| $0.54 | 55.6 | 69.8 | +14.2 |
| $0.72 | 66.1 | 86.6 | +20.5 |
| $1.08 | 80.3 | 93.1 | +12.9 |
| $1.26 | 84.9 | 95.9 | +11.0 |
| $1.44 | 88.5 | 96.9 | +8.4 |

This result is a computational manifestation of the defense-in-depth principle: heterogeneous systems are more resilient than homogeneous ones because they are not susceptible to common-mode failures. A vulnerability or hardware defect affecting T1 COTS servers would

simultaneously compromise all 4 instances in the 2-objective solution, while the 3-objective solution's T4 mission-critical hardware operates on a fundamentally different failure profile.

### 4.4. VM Type Allocation Patterns

Fig. 6 shows how VM type allocation varies across the Pareto front.

Low-cost Pareto solutions use T1 COTS instances exclusively, mirroring the 2-objective behavior. As budget increases, the optimizer progressively introduces higher-tier hardware. At moderate cost levels ($\$0.52 - \$1.30/h$), the algorithm discovers mixed configurations combining T1 and T3/T4 instances, providing both the redundancy benefits of multiple instances and the low failure rates of hardened hardware. At higher budgets, T4 Mission-Critical and T3 Hardened instances dominate, reflecting the military planning principle that mission-critical services should run on mission-critical hardware when resources permit.

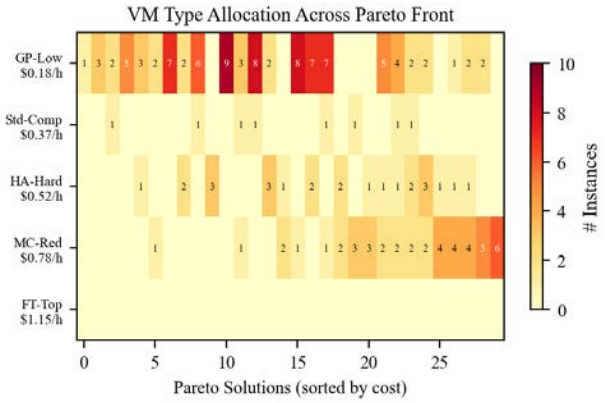



**Figure 6.** VM Type Allocation Across Pareto Front

### 4.5. Sensitivity Analysis: Effect of Time Horizon

A key question for military planners is how the operational time horizon affects the value of three-objective optimization. Figure 7 shows the reliability gap between the two- and three-objective formulations across four operationally relevant time horizons.

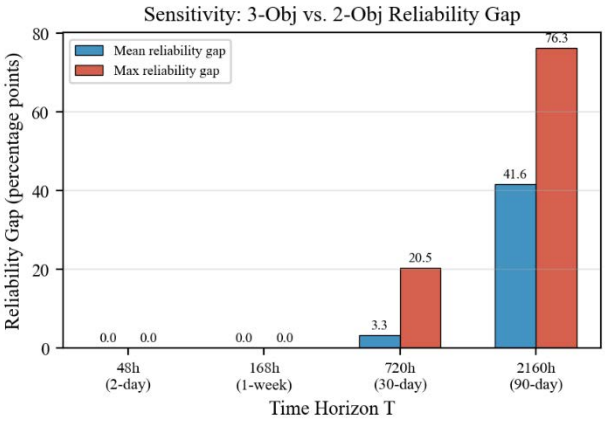



**Figure 7.** Reliability Gap vs. Time Horizon

The results reveal a clear pattern. For short operations ($T \leq 168$h), all VM types remain sufficiently reliable, and the two-objective formulation performs as well as the three-objective approach. As the time horizon increases to 720h (30-day deployment), a reliability gap emerges, reaching up to 20.5 percentage points. At $T = 2160h$ (90-day rotation), the gap becomes substantial, with an average of 41.6 and a maximum of 76.3 percentage points. These results indicate that while explicit reliability optimization provides little benefit for short missions, it becomes essential for sustained deployments exceeding one week.

### 4.6. NSGA-II vs. NSGA-III comparison

Fig. 8 compares the convergence of NSGA-II and NSGA-III on the three-objective optimization problem. NSGA-II produces 119 Pareto-optimal solutions with a final hypervolume of 14.45, whereas NSGA-III produces 12 solutions with a hypervolume of 14.26. NSGA-II performs better because NSGA-III is designed for many-objective optimization (four or more objectives), where reference directions are needed to maintain diversity. In contrast, the present three-objective problem has a small, bounded integer search space, allowing NSGA-II's crowding-distance mechanism to preserve diversity more effectively. These results justify the use of NSGA-II for military data center optimization with integer VM allocation variables.

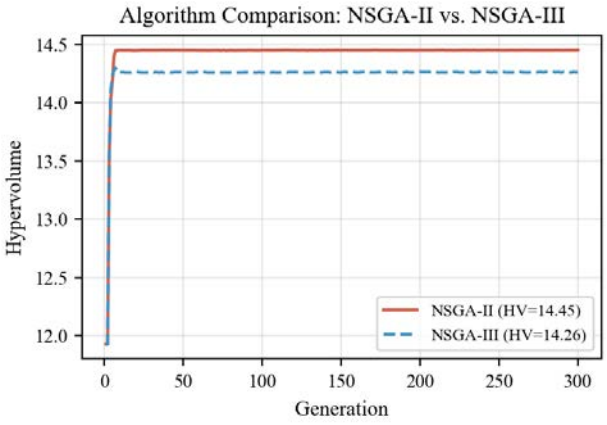



**Figure 8.** NSGA-II vs. NSGA-III Convergence

Hypervolume analysis further shows that both algorithms reach 99% of their final hypervolume within the first few tens of generations. The three-objective formulation converges as quickly as the two-objective formulation, indicating that adding reliability as a third objective incurs negligible computational overhead.

### 4.7. Full Objective Space and Objective-Level Constraints

Sections 4.1–4.2 present only Pareto-optimal solutions. To characterize the search space more completely, all $11^5$ = 161,051 VM configurations were exhaustively evaluated using Eqs. (1)–(3). Fig. 9 shows the resulting objective space projected onto the cost–reliability plane for the tactical budget (B = \$5/h), under which 2,503 configurations satisfy the budget constraint. Reliability

ranges from 23.7% (a single T1 COTS instance) to 99.9996% (highly redundant mixed configurations), demonstrating that the budget constraint alone permits configurations with widely different mission suitability.

The formulation in Section 3.3 imposes no minimum reliability, so all budget-feasible configurations are included in the Pareto analysis. In practice, however, missions often require a minimum reliability. Applying an objective-level constraint, $R(x) \geq R_{min}$, removes the lower portion of the objective space. For a three-nines requirement ($R_{min} = 99.9\%$ over the 720h horizon, dashed line in Fig. 9), 625 of the 2,503 budget-feasible configurations (25.0%) are eliminated, leaving 1,878 feasible configurations. A four-nines requirement ($R_{min} = 99.99\%$) excludes 1,582 configurations (63.2%). Most eliminated configurations are the low-cost, low-redundancy monocultures identified in Section 4.3.

These results show that realistic mission-reliability requirements eliminate many of the cheapest solutions, leaving diversified, higher-reliability VM portfolios as the feasible design choices. When mission reliability is specified a priori, the threshold can be incorporated directly as an additional optimization constraint.

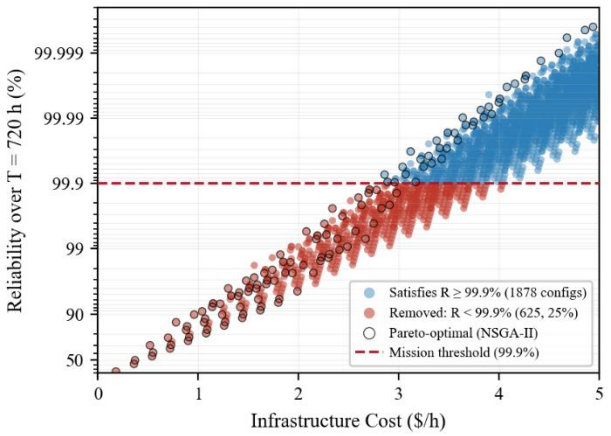



**Figure 9.** Full objective space for B = $5/h

## 5. CONCLUSION

This study extends the two-objective cloud optimization framework in [6] by introducing reliability as a third objective for virtualized military data centers. The results show that optimizing only cost and availability systematically favors homogeneous COTS-tier VM deployments, which achieve high availability through redundancy but poor long-term reliability. Incorporating reliability fundamentally changes the optimal solutions, with diversified VM portfolios combining COTS, hardened, and mission-critical hardware improving reliability by up to 20.5 percentage points at the same cost.

The findings provide three practical recommendations. First, budget-constrained tactical data centers should deploy mixed T1–T4 VM configurations rather than replicate identical low-cost instances. Second, higher-budget garrison facilities should reserve T3–T5 hardware for mission-critical workloads while using T1–T2 for non-critical services. Third, the resulting Pareto front provides a decision-support tool for balancing cost, availability, and reliability according to mission priorities.

Future work will extend the framework by incorporating adversarial failure models, workload-specific reliability targets, anti-affinity constraints for physical fault-domain separation, complete analysis of the objective space, and AI-based adaptive reconfiguration [15] for dynamic optimization under changing mission conditions.

## References

[1] MALHOTRA, R., BANSAL, A., KESSENTINI, M.: A systematic literature review on maintenance of software containers, ACM Computing Surveys, 56(8), pp. 1–38, 2024.

[2] SMITH, W. et al.: Cloud computing in tactical environments, Proc. MILCOM 2017 – IEEE Military Communications Conference, pp. 882–887, 2017.

[3] MACHIDA, F., KAWATO, M., MAENO, Y.: Redundant virtual machine placement for fault-tolerant consolidated server clusters, IEEE/IFIP NOMS, pp. 32–39, 2010.

[4] NABI, M., TOEROE, M., KHENDEK, F.: Availability in the cloud: State of the art, J. Network and Computer Applications, 60, pp. 54–67, 2016.

[5] TRIVEDI, K.S., KIM, D.S., GHOSH, R.: Resilience in computer systems and networks, Proc. IEEE/ACM ICCAD, pp. 74–80, 2009.

[6] STANIŠIĆ, S., ZOGOVIĆ, N.: A Multi-Objective Approach to Optimizing Cloud Infrastructure with Genetic Algorithms, Proc. 12th IcETRAN, pp. 1–4, 2025.

[7] TAN, B., MA, H., MEI, Y.: A NSGA-II-based Approach for Multi-objective Micro-service Allocation in Container-based Clouds, Proc. IEEE/ACM CCGRID, pp. 282–289, 2020.

[8] VERMA, S., PANT, M., SNASEL, V.: A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems, IEEE Access, 9, pp. 56745–56777, 2021.

[9] ARORA, M., KUMAR, V., DAVE, M.: Task Scheduling in Cloud Infrastructure using Optimization Technique Genetic Algorithm, Proc. 4th WorldS4, pp. 788–793, 2020.

[10] MOKNI, I., YASSA, S.: A multi-objective approach for optimizing IoT applications offloading in fog-cloud environments with NSGA-II, The Journal of Supercomputing, 80, pp. 27034–27072, 2024.

[11] ZHANG, W., CHEN, X., JIANG, J.: A multi-objective optimization method of initial virtual machine fault-tolerant placement for star topological data centers of cloud systems, Tsinghua Science and Technology, 26(1), pp. 95–111, 2021.

[12] BALAJI, K. et al.: Virtual Machine Fault Tolerance Placement in Data Centers using Multi-Objective Optimization, Proc. 8th CSITSS, pp. 1–6, 2024.

[13] STANIŠIĆ, et al.: Performance Optimization of File Systems for Docker Containers, Proc. Int. Scientific Conf. Sinteza 2025, pp. 117–127, 2025.

[14] SHI, L., DUAN, S.: Design and Implementation of Improved Multi-objective Genetic Algorithm Based on Uniform Distribution, DIONE 2023, LNICST, vol. 515, Springer, 2025.

[15] STANIŠIĆ, S., LUKOVIĆ, V., BELOTIĆ, B.: Automation of Monitoring and Optimization of Docker Containers Using Artificial Intelligence, Proc. 24th Int. Symp. INFOTEH-JAHORINA, pp. 1–4, 2025.