From Global Policies to Local Strategies: Multi-Objective Optimization of Resource-Specific Handover Policies -Supplementary Material-

Lukas Kirchdorfer[1,2][0000−0003−4713−9328], Artemis Doumeni[3], Han van der Aa[4][0000−0002−4200−4937], and Hugo A. López[3][0000−0001−5162−7936]

[1] SAP Signavio, Walldorf, Germany
lukas.kirchdorfer@sap.com
[2] Data and Web Science Group, University of Mannheim, Germany
[3] DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark
hulo@dtu.dk
[4] Faculty of Computer Science, University of Vienna, Vienna, Austria
han.van.der.aa@univie.ac.at

1 Introduction

This document provides supplementary material accompanying our main paper. Specifically, we provide more details on the experimental setup in Section 2 and more results in Section 3.

2 Experimental Setup

Metrics. We employ two types of metrics. The first type measures and compares the overall process performance under different handover policies. For this purpose, we compute the total process cost in \$ and the waiting time (WT) in hours. The second type of metric is intended to assess the overall quality of the discovered Pareto fronts, estimating how good each mutation variant is at searching for a diverse and high-quality set of optimal solutions. For this, we rely on four common metrics, which all make use of P_{Ref} as the joint reference Pareto front of all four mutation variants, and P_{Approx} as the approximated Pareto front discovered by one variant:

- **Hyperarea Ratio (HR)**: The Hyperarea (HA) is considered the most relevant and widely used measure to compare algorithms in the evolutionary community, and measures the area in the objective space that is dominated by a Pareto front relative to a worst-case reference point, which we set as the maximum cost and time among all the solutions explored. We then compute $HR = \frac{HA(P_{Approx})}{HA(P_{Ref})}$, resulting in a value between 0 and 1. A value close to 1 indicates a higher-quality front, with a score of 1 signifying that the approximated front dominates the same solution space as the reference front.

- **Purity**: Purity is computed as $\frac{|P_{Approx} \cap P_{Ref}|}{|P_{Approx}|}$ and measures the ratio of solutions in P_{Approx} included in P_{Ref}. A Purity score of 1 indicates excellent convergence, as it means the algorithm is highly effective at finding solutions that are part of the best-known front.
- **Averaged Hausdorff Distance (AHD)**: This metric provides a formal measure of the distance between the approximated front and the reference front. It is calculated as the maximum of two average distances: the average distance from each point in P_{Approx} to its nearest neighbor in P_{Ref}, and vice versa. A lower value for AHD signifies better convergence.
- **Delta Spread (Δ)**: Δ measures the spread and distribution of solutions along the Pareto front and is given by:

$$\Delta = \frac{d_0 + d_n + \sum_{i=1}^{n-1} |d_i - d'|}{d_0 + d_n + (n-1)d'}$$

  where d_i, $0 \leq i \leq n = |P_{Approx}|$, denotes the Euclidean distance between consecutive solutions. d_0 and d_n represent the distances between the extreme solutions in P_{Ref} and the corresponding extreme solutions in P_{Approx}. The term d' is the average of all d_i. A lower Δ value indicates a better spread and distribution of solutions, and thus a higher-quality approximation of the Pareto front.

3 Additional Results

Mutation variants. To further assess the quality of the solutions discovered by each variant, we analyze the corresponding Pareto fronts in Table 1. While no single variant consistently dominates across all four metrics, the *Guided* and *Hybrid* approaches stand out. The *Guided* variant performs best in terms of *Purity*, demonstrating its effectiveness in identifying dominant solutions, whereas the *Hybrid* variant achieves the lowest *AHD* values, indicating high convergence toward the reference front. Both variants also achieve similar *HR* values, often close to 1, meaning that their Pareto fronts cover nearly the same area in the objective space as the global reference front. The *Greedy* variant performs best on the *Delta Spread* metric, suggesting a slightly more uniform distribution of solutions, while the *Random* variant consistently produces the lowest-quality fronts across all measures.

Moreover, we provide detailed relative improvement scores of the four mutation variants compared to the as-is process in Table 2.

Runtimes. All experiments were run on an Apple M3 Pro (12-core CPU, 18GB RAM). We report runtimes in minutes in Table 3, comparing our optimization approach to the baselines. Note that the baselines all require the same runtime. While our approach adds a significant runtime factor to the heuristic baselines, we think that this is still acceptable given the substantial performance gains. Note that the runtime of our approach could further be reduced by setting the number of simulations T to a smaller value (e.g., $T = 1$ instead of $T = 3$).

Table 1: Comparison of mutation variants in terms of Pareto front quality.

Metric	Variant	LoanApp	P2P	C1000	C2000	ACR	BPI12W	BPI17W
HR ↑	Random	0.94	0.57	0.52	0.74	0.95	0.75	0.30
	Greedy	0.95	0.64	**0.93**	0.74	0.88	0.84	0.63
	Guided	**0.97**	**0.97**	0.60	0.73	0.97	**1.00**	0.38
	Hybrid	0.88	0.70	0.58	**0.87**	**0.98**	0.85	**1.00**
Purity ↑	Random	0.25	0.00	0.00	0.29	0.04	0.00	0.00
	Greedy	0.33	0.00	**1.00**	0.00	0.00	0.00	0.00
	Guided	**0.73**	**0.91**	0.00	**0.75**	**0.50**	**0.88**	0.00
	Hybrid	0.14	0.36	0.57	0.67	0.38	0.18	**1.00**
AHD ↓	Random	536.90	3537.55	4022.27	3511.11	1583.03	5642.47	14923.90
	Greedy	555.80	4380.22	1694.68	4134.00	8053.22	4539.06	11073.50
	Guided	**506.67**	5114.57	5973.75	3985.89	1996.45	**543.95**	9934.43
	Hybrid	790.21	**2325.92**	**1498.86**	**3216.44**	**599.75**	2971.95	**0.00**
Δ ↓	Random	0.83	**0.75**	0.96	0.88	0.94	0.80	0.89
	Greedy	**0.74**	0.91	**0.59**	**0.84**	1.17	**0.73**	**0.83**
	Guided	0.78	0.82	0.81	0.85	0.84	0.87	0.90
	Hybrid	1.06	0.88	0.61	0.96	**0.69**	0.85	0.94

Furthermore, the population size N and the number of iterations G significantly impact runtime.

Table 2: Relative improvement in cost and WT per mutation variant compared to the As-Is baselines.

Log	Random		Greedy		Guided		Hybrid	
	Cost	WT	Cost	WT	Cost	WT	Cost	WT
LoanApp	22.10%	18.61%	22.82%	13.79%	20.19%	18.26%	22.24%	6.26%
P2P	28.93%	80.16%	31.42%	77.03%	54.74%	78.86%	32.30%	84.21%
C1000	35.88%	51.83%	38.88%	55.35%	33.67%	52.36%	40.32%	40.68%
C2000	30.55%	34.68%	32.39%	28.39%	34.67%	26.00%	33.14%	32.32%
ACR	68.04%	79.14%	51.92%	85.81%	62.40%	90.60%	55.00%	88.61%
BPI12W	25.38%	0.75%	27.40%	23.75%	27.80%	41.38%	21.99%	35.75%
BPI17W	23.32%	96.30%	25.04%	97.81%	24.71%	96.54%	26.53%	97.69%
Average	33.46%	51.64%	32.84%	54.56%	**36.88%**	**57.71%**	33.07%	55.07%

Table 3: Runtimes in minutes for all 7 logs.

	LoanApp	P2P	C1000	C2000	ACR	BPI12W	BPI17W	Average
Baselines	0.28	0.35	1.78	3.58	1.72	2.17	10.83	2.96
Ours	9.33	16.03	70.67	142.60	41.83	67.31	379.17	103.85