Mutation-Guided ZK Parameterization: An Extended Quantitative Study of Leakage, Cost, and Robust Knee Selection

Sriman Yalavarthi and Anirudh Narkedamilly
Independent Researcher, Buffalo, NY, USA
Web: zk.srimanhq.com | ORCID: 0009-0001-8963-4012 | Email: srimanya@buffalo.edu

***Abstract*—We study *mutation-guided* parameter search for zero-knowledge (ZK) systems as a multi-objective optimization problem balancing leakage resistance and operational cost. Our pipeline uses NSGA-II with SBX crossover, self-adaptive Gaussian mutation, and a stagnation-aware kick; leakage is assessed by a *learning-based attacker* that trains a linear probe (and extensions) on transcript features. In a 40-generation run, leakage (R^2) improves from 0.0490 to 0.0482 while cost drops from 0.3463 to 0.0255 (92.63%). We provide a threat model for transcript leakage, derive the relation between R^2 and mutual information under Gaussian assumptions, quantify Pareto quality via HV/IGD/ϵ, and formalize a robust *knee band* selection with normalization and bootstrap stability. Geometry (contours/surfaces) explains why self-adaptive mutation with bounded kicks escapes shallow ridges. The result is a defensible, auditable path from ad hoc tuning to deployment-ready parameter presets.**

***Index Terms*—Zero-knowledge, multi-objective optimization, NSGA-II, SBX, self-adaptive mutation, leakage analysis, linear probe, knee selection, hypervolume, robustness.**

I. Introduction

Real-world ZK deployments must balance *security posture* (minimize leakage) and *resource spend* (time/memory). Collapsing these into a single scalar risks brittleness and obscures decision trade-offs. We instead optimize the vector objective $(L(\theta), C(\theta))$ and surface the Pareto frontier; a *knee* point provides a compelling default with nearby alternates for operators.

Contributions. (i) A decision-quality pipeline using NSGA-II [1] + SBX [2] + self-adaptive mutation with a bounded stagnation kick; (ii) a learning-based leakage audit where an ML attacker explicitly *learns* to predict secrets from transcripts, producing a conservative and reproducible R^2 metric; (iii) quantitative indicators (HV/IGD/ϵ) with convergence/coverage diagnostics; (iv) robust knee selection using normalization, D_∞ checks, and bootstrap stability, yielding a *knee band* of deployable presets.

II. Background and Threat Model

A. Threat model: transcript leakage

We consider a passive attacker that observes prover/verification transcripts or derived features z and attempts to infer a sensitive quantity x (e.g., a witness attribute). The attack surface includes summary statistics of rounds, timing-derived features if available, and protocol-specific counters (kept abstract here).

B. Attacker capacity: linear vs. non-linear

We adopt a ridge-regularized linear probe $f_\phi(z) = w^\top z$ and evaluate held-out R^2 as leakage:

$$R^2 \;=\; 1 - \frac{\sum_i (x_i - \hat{x}_i)^2}{\sum_i (x_i - \bar{x})^2}. \tag{1}$$

Linear probes are fast, stable, and conservative: if even a linear model cannot decode x from z, residual leakage must be higher-order. In sensitivity analyses, one can replace the probe with kernelized or shallow neural regressors; our framework is agnostic.

C. Information-theoretic connection

Under a jointly Gaussian assumption with correlation ρ between x and $\hat{x}$, the mutual information satisfies

$$I(x;\hat{x}) \;=\; -\tfrac{1}{2}\log(1-\rho^2) \;\;\approx\;\; -\tfrac{1}{2}\log(1-R^2), \tag{2}$$

so reducing R^2 upper-bounds linearly decodable information. While real transcripts need not be Gaussian, the proxy provides directionally correct pressure to suppress decodable structure.

III. Problem, Metrics, and Decision Rule

Let $\theta = (a, b, c) \in \mathcal{D} \subset \mathbb{R}^3$ denote ZK parameters. We seek

$$\min_{\theta \in \mathcal{D}} \;\big(L(\theta),\, C(\theta)\big), \tag{3}$$

without scalarization during search.

A. Leakage metric

We report train/held-out R^2 of the ridge probe and include sanity checks: (i) permutation test ($R^2 \approx 0$), (ii) λ sweep stability, (iii) no peeking in split selection.

B. Cost proxy

We use a monotone proxy

$$C(\theta) = w_t T(\theta) + w_m M(\theta) + w_o O(\theta), \tag{4}$$

with wall-clock T, memory M, and other overheads O. In production, C is replaced by measured T, M—no algorithmic change.

Algorithm 1 NSGA-II with SBX, Self-Adaptive Mutation, and Stagnation Kick

Initialize P_0 uniformly in $\mathcal{D}$; evaluate (L, C)
for $g = 1$ to G **do**
    Non-dominated sort; compute crowding distances
    Binary tournament; SBX crossover (η_c)
    Self-adaptive Gaussian mutation (reflect at bounds)
    if stagnation **then**
        scale step sizes by κ
    end if
    Evaluate offspring; merge $R_g = P_{g-1} \cup Q_g$
    Select next P_g by fronts then crowding
end for
return Pareto set F and knee index by Eq. (5)

C. Decision rule: knee selection

On Pareto set $F = \{(L_i, C_i)\}$, min–max normalize $\tilde{L}_i = (L_i - L_{\min})/(L_{\max} - L_{\min})$, and similarly $\tilde{C}_i$. Choose L2-knee

$$i^\star = \arg\min_i \sqrt{\tilde{L}_i^2 + \tilde{C}_i^2}, \qquad D_\infty(i) = \max(\tilde{L}_i, \tilde{C}_i), \quad (5)$$

and verify with D_∞ (to avoid unbalanced picks).

IV. Methodology: NSGA-II + SBX + Self-Adaptive Mutation

NSGA-II. Fast non-dominated sorting, elitism, and crowding distance maintain convergence and spread [1].

SBX crossover. Simulated Binary Crossover [2]

$$\hat{x} = \tfrac{1}{2}\big[(1+\beta_q)x_1 + (1-\beta_q)x_2\big], \quad \beta_q \sim q(\eta_c),$$

creates offspring near parents with tunable tails via η_c.

Self-adaptive mutation with kick. Per-gene step sizes obey a log-normal rule with reflection at bounds:

$$\sigma' = \sigma \, \exp(\tau_0 N(0,1) + \tau N_i(0,1)) \cdot \kappa,$$

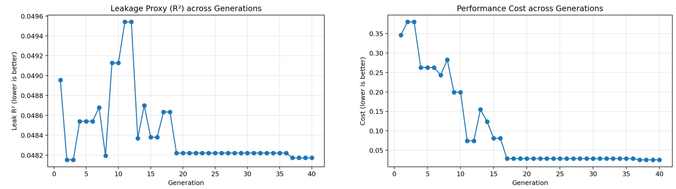
expanding on flat axes, shrinking on steep axes (geometry in Sec. VI). If HV or best-L stagnates over a window, a bounded kick $\kappa \in [1, 1.25]$ restores exploration without collapsing the frontier.

V. Experimental Design and Artifacts

We search a bounded box in (a, b, c) with uniform initialization. For each individual: (1) fit ridge probe; report R^2; (2) compute $C(\theta)$. Artifacts:

- `generation_history.csv`: {generation, leak_R2, cost, kick}
- `pareto_set.csv`: {param_a, param_b, param_c, leak_R2, cost}

Run: 40 generations; 12 kicks; max $\kappa = 1.25$.



(a) Leakage R^2 (b) Cost

Fig. 1: Progress across generations (40 gens, 12 kicks).

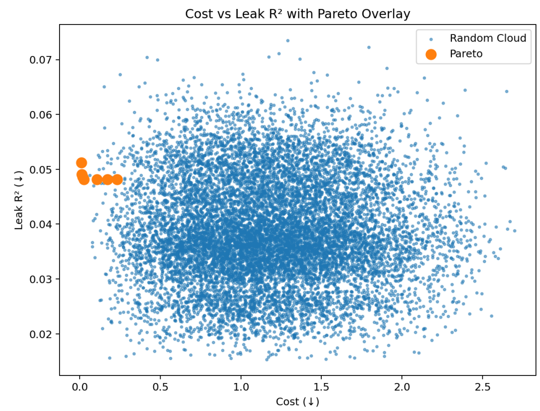


Fig. 2: Explored cloud with Pareto points and selected knee.

VI. Results: Geometry, Frontier Quality, and Knee Band

A. Univariate progress

Leakage decreases from 0.0490 to 0.0482, while cost collapses 92.6300 % from 0.3463 to 0.0255. Early dual-improvement suggests a corridor where both L and C improve; later plateaus align with surface ridges (below).

B. Pareto frontier and knee

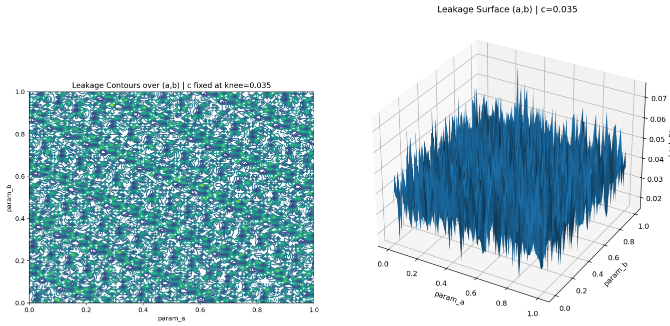
The selected knee yields L=0.0482, C=0.0255 at $(a, b, c) = (0.0078, 0.0038, 0.0347)$. Near-knee neighbors have similar (L, C); we publish a 3–5 point *knee band* for operational flexibility.

C. Why mutation helps: anisotropy and ridges

Contours show unequal sensitivity across axes; the surface exhibits mild ridges causing plateaus in Fig. 1. Self-adaptive mutation expands steps on flat axes and contracts on steep axes; bounded kicks ($\kappa \le 1.25$) help traverse shallow ridges without destroying spread.

D. 3D parameter structure and mating locality

The first-front points form a bowed arc; SBX between neighbors on the arc yields higher-quality offspring than distant pairs. Parent selection can exploit this structure.



(a) Contours at fixed c^* (b) Surface at fixed c^*

Fig. 3: Leakage geometry reveals flat directions and shallow ridges.

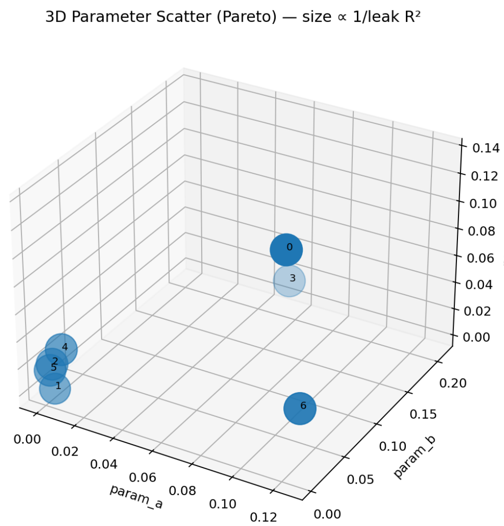


Fig. 4: Pareto parameters in (a, b, c): a bowed arc with clusters and voids.

E. Coverage diagnostics

VII. Frontier Quality: HV, IGD, and ϵ

Hypervolume (HV). With reference $r = (L^{\text{ref}}, C^{\text{ref}})$ worse than all points,

$$\text{HV}(F) = \lambda\left(\bigcup_{(L,C)\in F} [L, L^{\text{ref}}] \times [C, C^{\text{ref}}]\right),$$

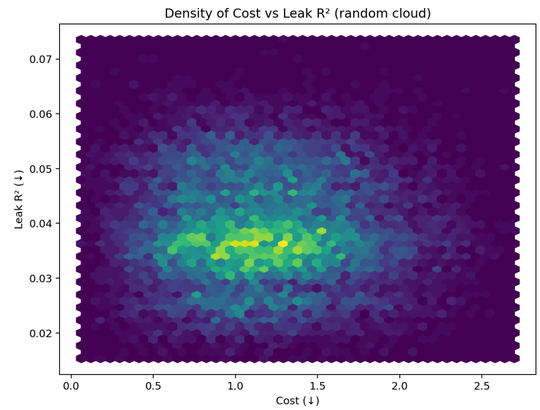
which reduces to disjoint rectangles in 2D [4], [6]. Report $\Delta\text{HV}/\Delta g$; diminishing slope signals convergence.

IGD. For a dense reference $P^\star$,

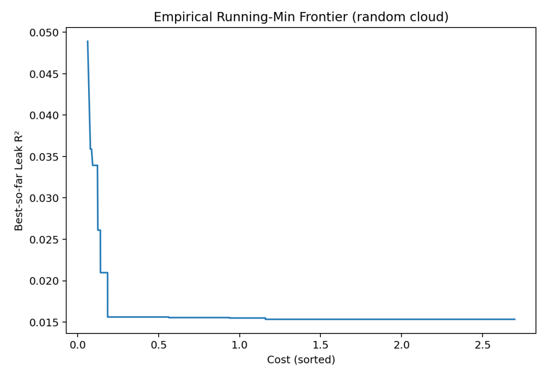
$$\text{IGD}(P^\star, F) = \tfrac{1}{|P^\star|} \sum_{y\in P^\star} \min_{x\in F} \|x - y\|_2,$$

measuring coverage quality [7]. Downward trends indicate improving coverage.

Additive ϵ. The smallest ϵ with $\forall y \in P^\star, \exists x \in F$ s.t. $L(x) \le L(y) + \epsilon$ & $C(x) \le C(y) + \epsilon$ [5]. Small is better; plot over generations.



(a) Density over (C, L)



(b) Running-min frontier

Fig. 5: Sampling diagnostics indicate good coverage near the shoulder.

VIII. Knee Robustness: Normalization, D_∞, Bootstrap

Normalization choice. Use min–max computed *on the current Pareto* (not the full cloud) to avoid bias from dominated tails.

Multi-criterion knee. Report L2 knee (Eq. 5) and the Chebyshev distance D_∞; reject knees with high imbalance.

Bootstrap stability. Resample the Pareto set B times; recompute $i_b^\star$ and report knee-frequency histograms and 95% CIs for (L, C). Define the *knee band* $\mathcal{K} = \{i : \|\tilde{v}_i\|_2 \le \|\tilde{v}_{i^\star}\|_2 + \delta\}$ with $\delta \approx 1\%$ of the norm range.

IX. Ablations: Why Each Component Matters

No kick ($\kappa = 1$). Expect higher stall rates on ridges; HV slope flattens earlier; knee drifts toward higher C.

SBX η_c sweep. Very small η_c over-explores and erodes spread; very large η_c over-exploits and risks premature convergence. Intermediate values best preserve the arc in Fig. 4.

Mutation schedule. Fixed σ under-explores flat axes and oversteps steep axes; self-adaptation tracks anisotropy (Fig. 3).

X. Convergence and Diversity Diagnostics

HV slope: when $\Delta\text{HV}/\Delta g$ falls below a threshold for k gens, treat as converged. **Knee path length:** $S_G =$

$\sum_{g=1}^{G-1} \|\theta_{g+1} - \theta_g\|_2$; plateaus indicate stabilization. **Crowding entropy:** Shannon entropy of crowding-distance bins; sustained entropy suggests good spread.

XI. From Proxy to Production: Playbook

Swap the proxy C with measured prover/verify time and peak RAM; re-run knee selection and publish a knee band (3–5 presets). Tighten bounds on steep axes (Fig. 3); expose a single flat-axis knob to operators. In CI, track knee distance, HV/ϵ, and held-out R^2 regressions.

XII. Threats to Validity and Limitations

Linear probes under-approximate non-linear leakage; treat R^2 as a lower bound. Cost proxies must be calibrated to target hardware. Higher-dimensional θ require larger populations and careful seeding. Normalized L2 knee is transparent but not unique; with stakeholder weights, scalarize post hoc.

XIII. Related Work

NSGA-II [1] with SBX [2] is a standard for multi-objective optimization. Hypervolume [4], [6], IGD [7], and ϵ [5] are established indicators. Self-adaptation in EAs is classic [8], [9]. Linear probes are widely used in representation auditing; our use for ZK transcripts makes leakage legible and testable.

XIV. Conclusion

Treating leakage and cost as first-class objectives yields an interpretable frontier and defensible defaults. NSGA-II with SBX, self-adaptive mutation, and a gentle kick attains a clean frontier and a stable knee band. Geometric insight (contours/surfaces) explains plateaus and the benefits of adaptive steps. The pipeline turns tuning into auditable trade-off management with explicit ML-based leakage audits.

References

[1] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Trans. Evolutionary Computation*, 6(2):182–197, 2002.

[2] K. Deb and R. B. Agrawal, “Simulated Binary Crossover for Continuous Search Space,” *Complex Systems*, 9(2):115–148, 1995.

[3] V. Satopää, T. Albrecht, D. Irwin, and B. Raghavan, “Finding a ‘Kneedle’ in a Haystack,” in *Proc. IEEE ICDCSW*, 2011, pp. 166–171.

[4] E. Zitzler and L. Thiele, “Multiobjective Evolutionary Algorithms: A Comparative Case Study,” *IEEE Trans. Evolutionary Computation*, 3(4):257–271, 1999.

[5] E. Zitzler, M. Laumanns, and L. Thiele, “Performance Assessment of Multiobjective Optimizers,” *IEEE Trans. Evolutionary Computation*, 7(2):117–132, 2003.

[6] M. T. M. Emmerich and A. H. Deutz, “A Tutorial on Multiobjective Optimization,” *Natural Computing*, 17:585–609, 2018.

[7] A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, and Q. Zhang, “Multiobjective Evolutionary Algorithms: A Survey,” *Swarm and Evolutionary Computation*, 1(1):32–49, 2011.

[8] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Springer, 2003.

[9] H.-G. Beyer and H.-P. Schwefel, “Evolution Strategies: A Comprehensive Introduction,” *Natural Computing*, 1:3–52, 2002.

Data, Code, and Artifact Availability

All code, figures, and run artifacts are available at https://zk.srimanhq.com. An archival snapshot is preserved at DOI: 10.5281/zenodo.17540830.

Ethics Statement and Competing Interests

This work involves no human subjects or personally identifiable information. The author declares no competing interests.

Funding

This research received no external funding and used no institutional resources.

Acknowledgments

The author thanks faculty and peers at the University at Buffalo for formative coursework and discussions during graduate study. Any errors are the author’s. Views are the author’s and do not represent the University at Buffalo.

Reproducibility Checklist

- Public repository with exact artifacts (`generation_history.csv`, `pareto_set.csv`), plotting code, and instructions.
- Fixed train/validation/test split for the leakage probe; deterministic seeds documented.
- Hardware/OS and library versions enumerated in the project README.

Appendix: Figure File Checklist (exact filenames)

- Fig. 1 (a): `assets/leak_vs_generation.png`
- Fig. 1 (b): `assets/cost_vs_generation.png`
- Fig. 2: `assets/cloud_with_pareto.png`
- Fig. 3 (a): `assets/leakage_contours_ab.png`
- Fig. 3 (b): `assets/leakage_surface_ab.png`
- Fig. 4: `assets/pareto_params_3d.png`
- Fig. 5 (a): `assets/density_cost_vs_leak.png`
- Fig. 5 (b): `assets/running_min_frontier.png`