Does the Source of Carrier Image Affect Steganographic Detectability?

Full Midway Proposal

Abdul Moiz Akbar | Malo Coquin | Daria Gjonbalaj | Nico Muller-Spath
Jimena Naravaez del Cid | David Wicker | Nikolas Zouros

Department of Advanced Computing Sciences
Maastricht University

Project 2.2 | February 2026

Agenda

1. Motivation and problem statement
2. Research questions and hypotheses
3. Chosen approaches
4. Experiments and validation
5. Prototype status
6. Related work positioning
7. Relation to curriculum
8. Planning and passing requirements

Motivation and Problem Statement

- Image steganography detectability depends on carrier statistics, not only embedding logic (Petitcolas et al., 1999; Cheddad et al., 2010; Fridrich & Kodovsky, 2012).
- Most steganalysis benchmarks assume camera photos with familiar noise/compression traces.
- Modern generators (Stable Diffusion, StyleGAN3) produce photorealistic images from different processes (Rombach et al., 2022; Karras et al., 2021).
- Synthetic images have measurable statistical fingerprints that may alter detector behavior (Wang et al., 2020; Corvi et al., 2023).

Central Problem

Do steganalysis methods designed and validated on photographs remain effective when the carrier is ML-generated?

Closest prior work

De et al. (2022) show AI-generated image steganography, but not a controlled real-vs-ML comparison with standardized LSB/DCT detectors.

Why This Study Matters

Security

- If synthetic carriers are harder to detect, attackers gain an easy evasion path.
- If easier, defenders get a concrete screening advantage.

Scientific gap

- Interaction between generative-model distributions and embedding distortion is largely unexplored.
- Need controlled experiments isolating carrier origin.

Practical relevance

- AI images are now common in social and communication channels.
- Practitioners need evidence on retraining/adaptation requirements.

Study scope: 2 x 2 x 3 x 2 factorial design over 1,000 images, with CPU-feasible execution in 7 weeks.

Research Questions

RQ1: Carrier origin effect

Does carrier origin (real vs ML-generated) change detectability under identical embedding settings?

RQ2: Payload sensitivity

Does increasing payload size change the real-vs-ML detectability gap?

RQ3: Embedding-method interaction

Does method choice (LSB vs DCT) change how carrier origin influences detectability?

RQ4: Encryption effect

Does AES-256-CBC payload encryption change detectability, and does this differ by carrier type?

Hypotheses and Decision Criteria

H	Claim	Primary test
H1	Carrier origin affects steganalysis AUC.	Wilcoxon + effect size
H2	Real-vs-ML AUC gap increases with payload.	Spearman trend + ANOVA interaction
H3	Carrier origin effect depends on embedding method.	Two-way ANOVA interaction
H4	Encryption changes AUC and may interact with origin.	Wilcoxon by condition + interaction check

Common statistical settings

Primary metric is ROC-AUC with Bonferroni-adjusted alpha (0.05/6 = 0.0083), plus 95% confidence intervals and Cohen's d.

Chosen Approach: Factorial Design and Conditions

Design matrix

- **Carrier type (2):** Real, ML-generated
- **Embedding (2):** Spatial LSB, frequency-domain DCT-QIM
- **Payload (3):** Low, Medium, High
- **Detectors (2 main):** RS Analysis, SRM+FLD

Sample size

- 500 real images
- 500 ML-generated images
- 1,000 total carriers
- Full condition coverage

Controlled variable principle

Carrier origin is treated as the central independent variable; image size/format, embedding pipelines, payload levels, and detector protocols are standardized across conditions.

Datasets and Preprocessing

Real photographs (500)

- RAISE: 250 images (RAW-derived forensic-quality baseline)
- COCO: 150 images
- Flickr30k: 100 images

Sources: Dang-Nguyen et al. (2015), Lin et al. (2014), Young et al. (2014)

ML-generated images (500)

- Stable Diffusion v2.1: 250 images
- StyleGAN3: 250 images
- Prompts aligned to COCO/Flickr semantics

Sources: Rombach et al. (2022), Karras et al. (2021)

Normalization and quality gate

All images normalized to 512x512 RGB 8-bit PNG. BRISQUE <= 50 filter for generated outputs to exclude low-quality artifacts.

Embedding Methods

LSB substitution (spatial)

- PRNG-keyed pixel/channel selection
- k = 1 for low and medium payload
- k = 2 for high payload
- Optional AES-256-CBC payload encryption

DCT-QIM (frequency)

- 8x8 block DCT per channel
- Mid-frequency zigzag coefficients (10–54)
- QIM embedding:

$$C_i' = \Delta \cdot \mathrm{round}(C_i/\Delta) \pm \Delta/4$$

Method rationale

LSB and DCT represent the two canonical embedding domains, allowing direct tests of method-origin interactions under controlled payload settings.

Payload and Encryption Conditions

Level	Approx. bpp	LSB setting	Purpose
Low	≈ 0.08	$k = 1$ sparse mask	Near-threshold detectability
Medium	≈ 0.16	$k = 1$ denser mask	Baseline operating point
High	≈ 0.32	$k = 2$	Stress-test detector sensitivity

Encryption condition

Each payload level is tested in:

- Plain payload mode
- AES-256-CBC pre-encrypted mode

Why include encryption?

It isolates whether message-bit structure contributes to detectability beyond carrier-level distortion.

Steganalysis Detectors and Validation Metrics

Detector set

- **RS Analysis** (training-free statistical baseline)
- **SRM+FLD** (feature-based classical ML detector)
- χ^2 **attack** as supplementary LSB check

References: Fridrich et al. (2001), Westfeld and Pfitzmann (1999), Fridrich and Kodovsky (2012)

Primary metrics

- ROC-AUC (primary)
- Accuracy at Youden's J
- Equal Error Rate
- FPR at 5% FNR

Statistical analysis plan

Two-way ANOVA (carrier x method; payload covariate), Wilcoxon pairwise tests, effect sizes, and Bonferroni-adjusted significance threshold.

Experiment Plan I (RQ1 and RQ2)

Exp. 1 – Carrier origin effect (RQ1)

- Apply RS and SRM+FLD across all payload levels and methods.
- Compare AUC for real vs ML-generated carriers.
- Decide with Wilcoxon significance and effect size.

Exp. 2 – Payload sensitivity (RQ2)

- Track real-vs-ML AUC gap across Low, Medium, High payload.
- Test monotonic trend (Spearman) and carrier x payload interaction.
- Decide whether payload amplifies origin-dependent detectability.

Experiment Plan II (RQ3 and RQ4)

Exp. 3 – Method interaction (RQ3)

- Run two-way ANOVA on SRM AUC with carrier origin and method factors.
- Confirm whether the detectability gap depends on LSB vs DCT.

Exp. 4 – Encryption effect (RQ4)

- Compare plain vs AES-encrypted payload AUC per carrier and method.
- Test whether encryption effect differs by carrier origin.

Interpretation policy

Null findings remain valid outcomes; all major results reported with confidence intervals and effect sizes.

Prototype Status

Vertical prototype (depth)

Isolated implementation and verification of:

- LSB embed/extract (BER = 0 on test set)
- DCT-QIM embed/extract (lossless payload recovery)
- RS Analysis sanity checks vs expected behavior
- SRM feature extraction + baseline classifier

Horizontal prototype (breadth)

Integrated end-to-end run on 50 images:

- 25 real + 25 ML-generated
- Medium LSB payload condition
- Pipeline interfaces validated before full scale

Readiness: Vertical and horizontal checks reduce implementation risk before full 1,000-image execution.

Related Work Landscape

Generative steganography

Prior work often embeds messages during generation itself, including GAN and diffusion pipelines (Hu et al., 2023; Liu et al., 2024; Duan et al., 2020).

AI-generated carriers

De et al. (2022) show feasibility with bespoke probabilistic coupling, but without controlled real-vs-ML detectability comparisons.

Cross-domain and synthetic-image forensics

Existing studies cover camera-domain shifts and real-vs-synthetic discrimination (Wang et al., 2020; Corvi et al., 2023), but not payload detectability across carrier origin.

Related Work: Our Positioning

What this proposal adds

- Controlled real-vs-ML carrier comparison
- Standardized LSB and DCT embedding
- Explicit steganalysis outcome reporting
- Payload and encryption interaction analysis

Methodological choice

- Classical SRM+FLD over deep detectors
- Better interpretability for cross-domain behavior
- Feasible CPU runtime for full study scale

References: Fridrich and Kodovsky (2012), Luo et al. (2024)

Relation to Curriculum

Cryptography and Steganography

LSB and DCT-QIM embedding, payload encryption (AES-256-CBC), and detectability-focused reasoning.

Machine Learning

Feature-based steganalysis with SRM representations and FLD-style linear classification.

Research Methods

Factorial design, hypothesis testing, ANOVA/Wilcoxon analysis, effect sizes, and controlled significance correction.

Algorithm Design and Implementation

Block-wise transforms, coefficient-level embedding logic, and reproducible experiment orchestration in Python.

Planning and Milestones

Phase windows

- **Phase 2:** 30 Mar – 15 May 2026
- **Phase 3:** 25 May – 12 Jun 2026

Milestones

- M1 (end Wk2): dataset ready
- M2 (end Wk3): embedding pipelines verified
- M3 (end Wk5): experiments complete
- M4 (end Wk7): report submitted

Planned outputs

- Curated real and ML-generated datasets
- Verified LSB/DCT + AES pipelines
- Detector runs and statistical analysis
- Final deliverables: slides, poster, paper

Execution strategy

Parallel workstreams for data, embedding, and analysis with fixed scope after Week 2 to protect timeline and avoid design drift.

Minimal Passing Requirements

Product minimum

- Functional LSB and DCT pipelines
- Plain and AES-encrypted payload modes
- RS and SRM+FLD evaluated on all 1,000 images
- Full condition coverage for carrier/method/payload

Validation minimum

- RQ1 and RQ4 answered with AUC-based significance tests
- 95% confidence intervals reported
- Effect sizes reported for practical significance
- Null results interpreted explicitly

Definition of done

Minimum deliverable is a reproducible, statistically transparent conclusion on carrier-origin and encryption effects, even if findings are null.

Thank you

Questions and Discussion

Midway proposal document: `docs/proposals/midway_proposal.tex`