

Redaction Precision Hardening Report

A baseline-versus-hardened comparison under deliberately wrong masks, planted PDF leaks, multiple raster scales, five controlled documents, and one real online government form.

VERDICT	PASS - HARDENED PATH CLOSED EVERY TESTED GAP
BOUNDARY	This benchmark uses known regions and synthetic corruption. It does not prove automatic detector recall or resistance to unknown future attacks.
CLEAN OUTPUTS	42 / 42 recovery cells passed across 6 files
HOSTILE GATES	6 / 6 planted leaks caught; 140 / 140 scale checks passed
RUN ID	HEAD 31a4c7c66a8b / STATUS 6e9d3f3fd5e4 / DIRTY

95.18% 4PX COVERAGE	151,533 4PX LEAKED PIXELS	100.00% 12PX COVERAGE	0 12PX LEAKED PIXELS	100.00% VERIFIER SENSITIVITY
------------------------	------------------------------	--------------------------	-------------------------	---------------------------------

Deliberately Wrong Masks

Two documents were re-redacted across nine mask states and two safety margins. The approved boxes were shifted in every cardinal direction, shifted diagonally in both directions, shrunk, or both shrunk and shifted. Every artifact was regenerated from the current deterministic CLI core; no cached redaction output was accepted.

FIXTURE	MASK ERROR	4PX COVERAGE	4PX LEAKS	12PX COVERAGE	12PX LEAKS
online-w9	diagonal-down-right-8	93.307%	11,688	100.000%	0
online-w9	diagonal-up-left-8	91.069%	15,596	100.000%	0
online-w9	exact	100.000%	0	100.000%	0
online-w9	shift-down-6	96.893%	5,426	100.000%	0
online-w9	shift-left-6	99.684%	552	100.000%	0
online-w9	shift-right-6	99.738%	458	100.000%	0
online-w9	shift-up-6	95.823%	7,294	100.000%	0
online-w9	shrink-6	92.188%	13,642	100.000%	0
online-w9	shrink-and-shift-5	87.869%	21,184	100.000%	0
text-pii	diagonal-down-right-8	93.698%	13,784	100.000%	0
text-pii	diagonal-up-left-8	93.710%	13,759	100.000%	0
text-pii	exact	100.000%	0	100.000%	0
text-pii	shift-down-6	97.222%	6,076	100.000%	0
text-pii	shift-left-6	99.605%	864	100.000%	0
text-pii	shift-right-6	99.605%	864	100.000%	0
text-pii	shift-up-6	97.222%	6,076	100.000%	0
text-pii	shrink-6	93.698%	13,784	100.000%	0
text-pii	shrink-and-shift-5	90.634%	20,486	100.000%	0

Tradeoff: average redacted page area increased from 2.59% to 3.32%. That extra area is the measured cost of moving perturbed target coverage from 95.18% to 100.00%.

Baseline Leak Versus Hardened Coverage

SOURCE CROP

Before you begin. For guidance related to the purpose of Form W-9, see the instructions.

1 Name of entity/individual. An entry is required. (For a sole proprietor, enter the name on line 2.)

2 Business name/disregarded entity name, if different from above.

3a Check the appropriate box for federal tax classification of the entity/individual. Only one of the following seven boxes.

☐ Individual/sole proprietor ☐ C corporation ☐ S corporation

☐ LLC. Enter the tax classification (C = C corporation, S = S corporation). **Note:** Check the "LLC" box above and, in the entry space, enter classification of the LLC, unless it is a disregarded entity. A disregarded entity is a sole proprietorship, partnership, trust, or estate that is not a separate legal entity. Enter the tax classification of its owner.

☐ Other (see instructions)

3b If on line 3a you checked "Partnership" or "Trust/estate," or checked and you are providing this form to a partnership, trust, or estate in this box if you have any foreign partners, owners, or beneficiaries. See the instructions.

5 Address (number, street, and apt. or suite no.). See instructions.

6 City, state, and ZIP code

7 List account number(s) here (optional)

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the backup withholding. For individuals, this is generally your social security number. For partnerships, trusts, or estates, it is your employer identification number (EIN). If you do not have a TIN, later.

Note: If the account is in more than one name, see the instructions for Number To Give the Requester for guidelines on whose number to enter.

Part II Certification

Under penalties of perjury, I certify that:

1. The number shown on this form is my correct taxpayer identification number.

2. I am not subject to backup withholding because (a) I am exempt from backup withholding, or (b) I am subject to backup withholding as a result of a determination by the IRS that I am no longer subject to backup withholding; and

3. I am a U.S. citizen or other U.S. person (defined below); and

4. The FATCA code(s) entered on this form (if any) indicating that I am a U.S. person.

Certification instructions. You must cross out item 2 above if you have a TIN because you have failed to report all interest and dividends on your tax return.

BASELINE 4PX

Before you begin. For guidance related to the purpose of Form W-9, see the instructions.

1 Name of entity/individual. An entry is required. (For a sole proprietor, enter the name on line 2.)

2 Business name/disregarded entity name, if different from above.

3a Check the appropriate box for federal tax classification of the entity/individual. Only one of the following seven boxes.

☐ Individual/sole proprietor ☐ C corporation ☐ S corporation

☐ LLC. Enter the tax classification (C = C corporation, S = S corporation). **Note:** Check the "LLC" box above and, in the entry space, enter classification of the LLC, unless it is a disregarded entity. A disregarded entity is a sole proprietorship, partnership, trust, or estate that is not a separate legal entity. Enter the tax classification of its owner.

☐ Other (see instructions)

3b If on line 3a you checked "Partnership" or "Trust/estate," or checked and you are providing this form to a partnership, trust, or estate in this box if you have any foreign partners, owners, or beneficiaries. See the instructions.

5 Address (number, street, and apt. or suite no.). See instructions.

6 City, state, and ZIP code

7 List account number(s) here (optional)

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the backup withholding. For individuals, this is generally your social security number. For partnerships, trusts, or estates, it is your employer identification number (EIN). If you do not have a TIN, later.

Note: If the account is in more than one name, see the instructions for Number To Give the Requester for guidelines on whose number to enter.

Part II Certification

Under penalties of perjury, I certify that:

1. The number shown on this form is my correct taxpayer identification number.

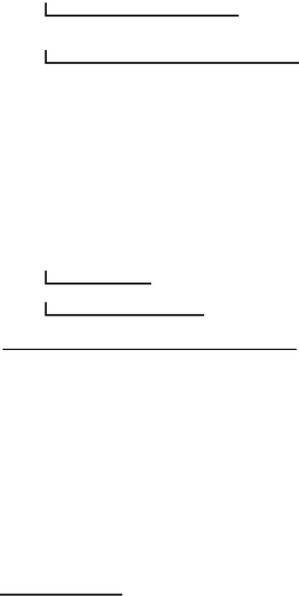
2. I am not subject to backup withholding because (a) I am exempt from backup withholding, or (b) I am subject to backup withholding as a result of a determination by the IRS that I am no longer subject to backup withholding; and

3. I am a U.S. citizen or other U.S. person (defined below); and

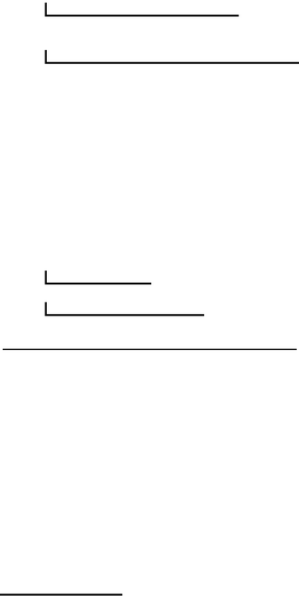
4. The FATCA code(s) entered on this form (if any) indicating that I am a U.S. person.

Certification instructions. You must cross out item 2 above if you have a TIN because you have failed to report all interest and dividends on your tax return.

4PX LEAK MAP



12PX LEAK MAP



BLACK IN LEAK MAP = TARGET PIXEL NOT OVERWRITTEN. 4PX: 21,184; 12PX: 0.

The visual uses the worst measured perturbed case. In each leak map, black means a ground-truth target pixel was not overwritten. The baseline map exposes the failed pixels directly; the hardened map is empty. This is a binary pixel comparison, not a subjective visual score.

Planted Leaks Must Produce Red Cells

INJECTED FAILURE	EXPECTED GATE	CAUGHT	ACTUAL DETECTOR
metadata	metadata-residue	YES	metadata-residue
text-layer	text-extraction	YES	text-extraction, layer-structure
attachment	metadata-residue	YES	metadata-residue
thumbnail	metadata-residue	YES	metadata-residue
trailing-history	metadata-residue	YES	metadata-residue
pixel-leak	entropy	YES	entropy

Sensitivity result: 6 of 6 planted leaks were rejected. The controls injected metadata, a selectable text layer, an attachment, an embedded thumbnail, trailing revision-like bytes, and visible source pixels inside a redaction region.

The Bars Survive Re-rasterization

DPI	REGION CHECKS	PASSED	LEAKED PIXELS
96	35	35	0
150	35	35	0
300	35	35	0
450	35	35	0

Across six clean PDFs, 140 target-region recovery checks were repeated at 96, 150, 300, 450 DPI. Every original target pixel rendered black or quantization-equivalent near-black (RGB channels 0-2); total leaked pixels: 0. These are correlated re-rasterizations, not statistically independent documents. At 96 DPI Poppler maps some exact black samples to RGB 1,1,1, so exact-byte equality alone would create a false failure without revealing source signal.

What Improved and What Remains

MEASURE	BASELINE	HARDENED	INTERPRETATION
Perturbed target coverage	95.18%	100.00%	Safety margin absorbs bounded reviewer-box error.
Perturbed leaked pixels	151,533	0	Any nonzero value is a redaction failure.
Planted-leak sensitivity	not measured	100.00%	The verifier now proves it can turn known corruption red.
Multi-scale checks	not measured	140	Re-rasterization did not reopen any target region.
Redacted page area	2.59%	3.32%	Precision cost is explicit rather than hidden.
Evidence provenance	cached artifacts allowed	fresh run + hashes	Generator and core source hashes are bound to run 6e9d3f3fd5e4.

The hardened path is materially stronger than the baseline in this benchmark: it closes every measured boundary leak, catches every planted corruption, and remains stable across multiple raster scales. This does not establish detector precision or recall. Those require a labeled candidate-generation corpus with hard negatives, which is reported separately from overwrite correctness.