Where on Earth? Zero-Knowledge Proof of Location from a Phone Photograph of the Moon and Stars

Pavel Shuvalov
DevNull
pavel@devnull.ae

18 November 2026

Abstract

We study whether a single photograph taken with commodity hardware, such as an unmodified phone, can serve as a privacy-preserving proof of location. The physical anchor is *lunar parallax*. The Moon is close enough to the Earth that its position against the background stars shifts by up to about one degree depending on where the observer stands, so a frame that shows the Moon among a few stars, together with its capture time, determines the camera's position: one arcsecond of measurement error corresponds to about 1.9 km on the ground. Nothing is transmitted and nothing is received; the only infrastructure is the sky, a star catalogue and an ephemeris (a published table of where the Moon is at any given time).

We make three contributions. First, we measure what phone cameras actually achieve. Our study is based on 356 captures, of which 6 RAW frames give a position, with errors from tens to under two hundred kilometres. We analyse what limits the measurement and the ranges of capture parameters in which it is feasible. Second, we turn the measurement into a zero-knowledge proof. One proof binds a signed RAW file to a Merkle commitment of its pixels; a second proof recomputes the star centroids, the camera calibration, the Moon's centre and the parallax from the committed pixels, and shows that *every* position consistent with the photograph lies inside a public region, revealing neither the position, the photograph, nor the signing device. The prover is left with essentially no choices: the camera must be the least-squares fit to a *complete*, magnitude-limited star set, which cuts the prover's influence on the result from more than a thousand kilometres to under 25 km in our measurements. Six telephoto frames from one phone at two sites are accepted, with regions between roughly 370×370 and 710×640 km, as are three further test photographs, two of them published photographs from dedicated cameras. Both proofs run on two zkVMs; on the faster one the full pipeline for a photograph, ending in a zero-knowledge proof, takes about ten minutes of GPU and CPU time, and verification takes under half a second. Third, we analyse what the proof does *not* establish. A photograph of a laptop screen displaying a rendered sky is accepted as a proof for the original site, because the proof binds pixels to a signing device and its clock, not to the physical sky. We enumerate the trust assumptions, survey which device signals could narrow the gap, and outline measures that would raise the cost of more advanced forgeries well beyond the resources of an individual.

The result is a working approach for building location proofs from commodity-phone photographs, at country-to-region granularity, and not yet a complete proof-of-location system.

1 Introduction

Many applications want to know where a person or a device is without learning more than they need: that a journalist's footage was taken in a given country, that a user is inside a jurisdiction where a service is licensed, that a sensor reading comes from a claimed area, that a participant in a network is not located where they say they are not. The two halves of this problem pull

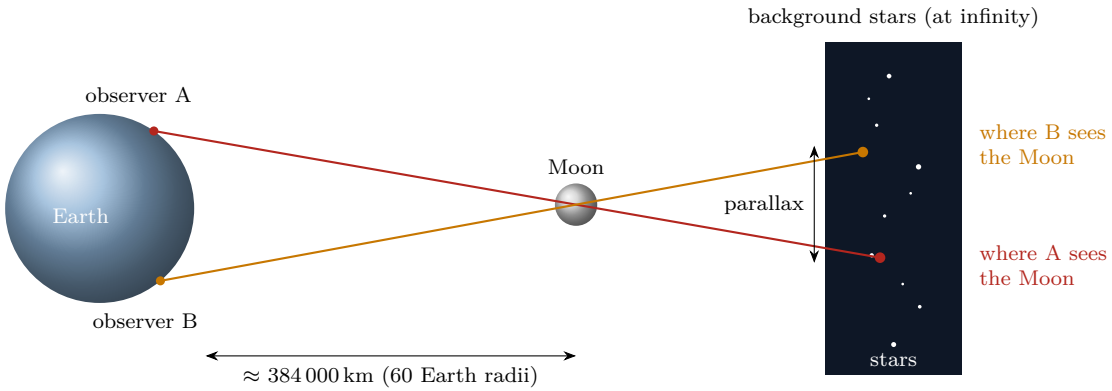


Figure 1: Lunar parallax (not to scale). Two observers see the same stars in the same directions but the Moon in front of different stars. Read backwards, the Moon's place among the stars tells an observer where on Earth they are. The true effect is at most about 1° relative to the Earth's centre.

against each other. *Authenticity* asks for evidence that cannot be fabricated without being at the claimed place; *privacy* asks that the evidence reveal no more than a coarse statement such as "inside region S".

Existing sources of location evidence obtain authenticity from infrastructure. Satellite navigation receives unauthenticated civil signals that can be spoofed with modest equipment [39, 66]. Distance-bounding and beacon-based location proofs need trusted verifiers or witnesses near the prover [12, 59, 61, 69]. Network measurements localise an address rather than a person and can be manipulated [32]. Recent cryptographic work shows how to prove *in zero knowledge* that a private coordinate lies in a public region [27], but it takes the coordinate as given: the proof is exactly as trustworthy as whatever produced the latitude and longitude.

This paper explores a source of location evidence that needs no terrestrial infrastructure at all, and asks how far it can be carried inside a zero-knowledge proof.

The idea. Stars are so far away that every observer on Earth sees them in the same directions. The Moon is not: at a distance of about 384 000 km, sixty Earth radii, its apparent position against the stars depends on where one stands (Figure 1). The displacement, the *topocentric parallax*, reaches about one degree, twice the Moon's own diameter. A photograph that contains both the Moon and stars therefore encodes the photographer's position. The stars tell us how the camera was oriented in space; the Moon's centre then gives a direction in space; an *ephemeris*, a published table of the positions of Solar System bodies over time, tells us where the Moon was at that instant; and the observer must lie where the line from the Moon, along that direction, meets the Earth's surface. One image and one timestamp give one position fix.

Navigators used the Moon against the stars for centuries, in the method of lunar distances [11]. There the Moon served as a clock for finding longitude, and parallax was a nuisance to be "cleared". With an accurate clock in every phone we can invert the roles: the time is known, and the parallax is the signal.

From a measurement to a proof. A position computed from a photograph is evidence only if the photograph is authentic, and it is private only if the photograph is not shown. We address both with a zero-knowledge proof over a photograph signed by its capturing device, using the C2PA content-credential format [20]. Few phones sign their captures today (Section 2.2), so our photographs are signed after capture with a test certificate. The verifier learns a public region S, a capture time, and a 204-byte claim; it does not see the image, the position, the camera's calibration, or which device maker signed the file.

Doing this soundly turned out to be less about arithmetic than about *removing the prover's*

choices. A photograph does not come with star positions and a Moon centre attached; these are estimated, and every free choice in the estimation — which stars to use, where to place a centroid window, how to calibrate the lens — is a lever a dishonest prover can pull. On one real frame, choosing six convenient stars and letting the lens model absorb the difference moves the "proven" position by more than 1 700 km while every star still fits to a fraction of a pixel. Most of our design is a sequence of measurements of such freedoms and of rules that remove them.

What we found.

- **The physics works on phones, coarsely.** On synthetic images the solver recovers the position to about 1 km with a telephoto field of view. On real phone photographs the six RAW frames that pass the input gates are located to between 30 and 180 km. The error is dominated by the Moon, which a phone overexposes into a glare blob when the exposure is long enough to show stars (Section 5).
- **The measurement fits in a zero-knowledge proof.** A two-stage proof in a RISC-V zkVM binds a signed 16 MB RAW file to its pixels and re-derives the whole measurement from them. Six real frames from Dubai and the Maldives are accepted; in each, the GPS ground truth lies inside the proven uncertainty shape. Three further test photographs are accepted as well: a cropped wide-camera phone frame and two photographs published by other photographers, taken with dedicated cameras. Both proofs were proved for nine photographs on SP1 and on RISC Zero. On SP1 a 16 MB phone photograph takes 1.6–1.9 minutes on one A100 GPU, and a further 39–44 s for a claim against national borders; the proof is 1.27 MB and verifies in under 0.1 s. It is not zero-knowledge until it is wrapped into a Groth16 proof of a few hundred bytes, which takes a further six minutes on a CPU and verifies in 0.4 s (Sections 6–8).
- **The proof can be fooled from outside the mathematics.** The same phone photographing a laptop screen that shows a rendered sky, with its clock set back, yields a proof for the original location. The proof establishes that a trusted camera recorded, at the signed time, an image whose Moon-against-stars geometry is consistent only with positions in S. It does not establish that the light came from the sky (Section 9).

Contributions. (1) A description and feasibility measurement of single-image lunar-parallax positioning with phone cameras. (2) A verification predicate for that measurement designed for zero-knowledge proving, in which calibration, star selection and estimator placement are all pinned by public rules, and in which the region claim covers the whole set of positions consistent with the image rather than one witnessed point. (3) An implementation on two zkVMs with measured costs on real signed RAW files. (4) A security analysis that separates what is enforced by the proof from what is assumed of the device, demonstrated by a successful screen-replay attack, with a survey of mitigations and their limits.

Scope and limitations. The evidence base is small: six accepted frames from two sessions with one phone, photographed by the author, and three further test photographs that have no device provenance. The photographs were signed after capture with a test certificate, standing in for a camera that signs RAW output natively. The accuracy supports claims at the scale of a country or a large region, not a city. The system in its present form is not secure against an attacker who controls what the camera looks at. We report these limits alongside the results because they determine what should be built next.

2 Background

The paper joins two fields that rarely meet. This section gives the minimum of astrometry that the rest of the paper needs, and then the two building blocks on the cryptographic side: zkVMs and signed photographs.

2.1 Astrometry in one page

Directions, not distances. Astrometry measures directions. A direction is a unit vector from the observer, expressed in axes that are fixed relative to the distant universe and do not rotate with the Earth (the ICRS). Angles are small: one degree is 60 arcminutes ($'$) or 3 600 arcseconds ($''$). The Moon's disc is about $31'$ across. A phone telephoto camera maps roughly one arcminute onto one pixel.

Star catalogues. The directions of bright stars are known far better than a phone can measure them. We use the Hipparcos catalogue [28, 67]; its 8 870 stars brighter than visual magnitude $V = 6.5$ (roughly the naked-eye limit; smaller V means brighter) are all we need. For our purposes a catalogue is a public table.

The camera as a measuring device. A camera maps a direction to a pixel. The map has two parts: the *attitude*, a rotation matrix R from camera axes to the sky frame, and the *intrinsics*: focal length f in pixels, principal point (c_x, c_y) and lens distortion coefficients k_1, k_2 of the standard radial model [13]. Given a few identified stars, both can be fitted by least squares; this is how spacecraft *star trackers* determine their orientation [47], and finding which star is which without prior knowledge is the well-studied "lost in space" problem [14, 46]. Once the camera is fitted, any pixel can be turned into a direction in space.

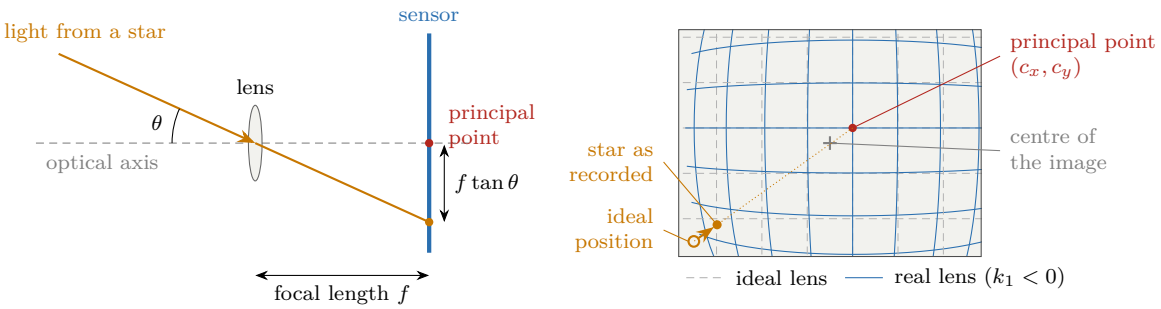
Figure 2 shows what the intrinsics mean. In an ideal camera a star at angle θ from the optical axis lands at distance $f \tan\theta$ from the principal point, the pixel where the optical axis meets the sensor. A real lens moves each image point along the line through the principal point, by a factor that grows towards the corners. Writing (x, y) for the ideal position in units of the focal length and $r^2 = x^2 + y^2$, the pixel (u, v) where the star is actually recorded is

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} c_x \\ c_y \end{pmatrix} + f\left(1 + k_1 r^2 + k_2 r^4\right) \begin{pmatrix} x \\ y \end{pmatrix}. \tag{1}$$

With $k_1 = k_2 = 0$ the lens is ideal. On a phone the principal point can sit tens of pixels from the middle of the image, and the distortion reaches several pixels at the corners; both are large compared with the fraction of a pixel the method needs, so they are fitted from the stars of each photograph.

The ephemeris. An *ephemeris* (plural *ephemerides*) is a table that gives the positions of celestial bodies as functions of time. Modern ones are computed by numerically integrating the equations of motion of the Solar System and fitting the result to decades of observations, including laser ranging to the Moon. JPL publishes the standard series; we use DE440 [54], whose error for the Moon is far below anything a camera can resolve. Like the catalogue it is a public table, evaluated at the capture time.

Small effects that are not small here. At 1.9 km per arcsecond, several textbook corrections matter: *light time* (we see the Moon where it was 1.3 s ago); *aberration* (the Earth's orbital speed tilts every apparent direction by up to $20''$); *refraction* (the atmosphere lifts objects near the horizon, by about $1.7'$ at 30° altitude [9, 58, 62]); and *phase* (the lit part of the Moon is not centred on the Moon's centre). Section 4 treats each.



(a) ideal camera, seen from the side

(b) the sensor, seen from the front

Figure 2: The camera model. (a) In an ideal camera a star at angle θ from the optical axis is imaged at distance $f \tan\theta$ from the principal point. (b) Two departures from the ideal, both exaggerated. The principal point (c_x, c_y) is not the centre of the image, and the lens moves every image point towards or away from the principal point by the factor $1 + k_1 r^2 + k_2 r^4$ of Eq. 1: a square grid (dashed) is recorded as a bowed one (solid).

2.2 zkVMs and signed photographs

zkVMs. We express the statement to be proved as an ordinary program for a zero-knowledge virtual machine, which proves correct execution of a RISC-V binary [15, 63] with a succinct proof [8, 33, 36]. This lets us write the verifier in Rust, hash megabytes of data and use IEEE floating point, which the machine emulates deterministically. Cost is counted in *cycles* (executed instructions).

Signed photographs. C2PA [20] is an industry format in which a capturing device signs a manifest containing a hash of the image file with a key certified by its manufacturer. It tells a verifier that a file came from a certified camera unmodified; it says nothing about what the camera was pointed at. This distinction is the subject of Section 9.

Device support today. Signing at capture is not yet common on phones. Several professional cameras sign their output, and a list of supporting devices is maintained in [16]. Among phones, the Google Pixel 10 signs every photograph taken with its camera application [34], and a few other Android models have some support; no iPhone produces C2PA credentials at capture. To our knowledge, no phone signs uncompressed RAW output with C2PA. Apple has instead introduced its own format, Apple Reference Image [4], in which the sensor signs the pixel data, the capture is kept as a signed DNG, and the capture time is bounded by signed timestamps. It is new, limited to the main camera of two phone models, and not yet widely supported.

The photographs in this work are therefore signed after capture with our own test certificate, standing in for a device that signs its RAW output. We explore what the approach can achieve once such signing is available on common devices, and the binding proof (Section 6.2) is the only part that depends on the signature format.

3 Overview and threat model

Parties. A *prover* holds a phone and wants to convince a *verifier* that a photograph was taken inside a public region S at a stated time. A *device maker* certifies the phone's signing key. Public data (star catalogue, ephemeris, country borders) come from scientific sources the verifier can obtain independently.

Software. The software has three parts, and we use these names throughout. The *solver* finds the answer: it identifies the stars, fits the camera and solves for the position (Section 4.3).

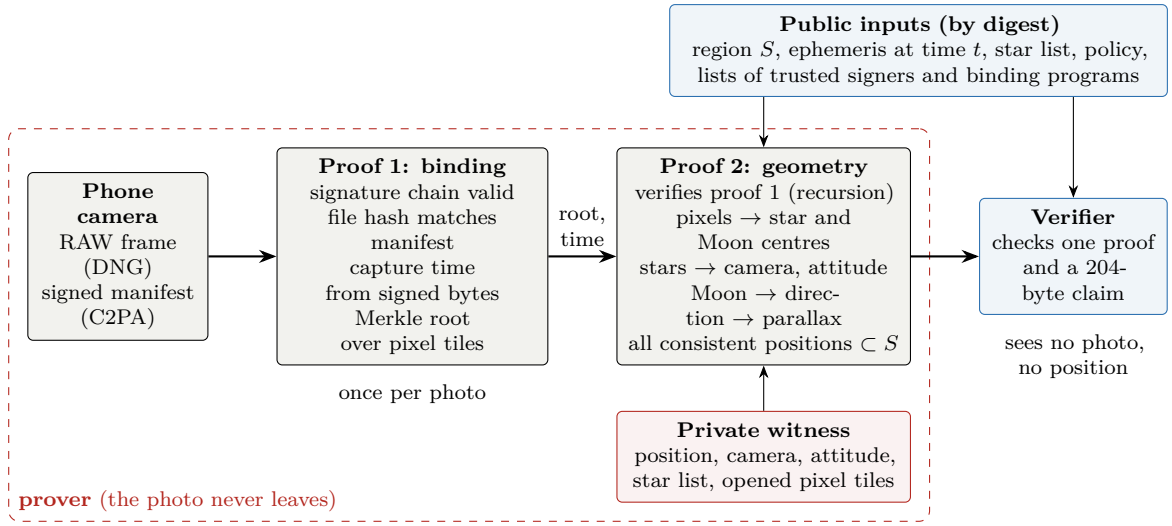


Figure 3: System overview. The prover holds a signed RAW photograph. Proof 1 is made once per photograph and commits to its pixels; proof 2 is made once per region claim. The verifier sees only the final proof and claim.

It involves search, runs on the prover's side, and is not trusted. The *prover* software takes the solver's result as a starting point, measures the photograph again under the fixed public rules of Section 6, and produces the proofs. The *verifier* checks the final proof against the public inputs.

The statement. Informally, the final proof establishes:

> *There is a RAW photograph, signed under one of the verifier's trusted roots with capture time t, such that the star centroids and the Moon's centre, computed from its pixels by fixed public rules, are consistent with an observer position on the Earth's surface; and* ***every*** *position consistent with those measurements within a public tolerance lies inside S.*

The position, the pixels, the camera calibration and the identity of the signing root within the trusted list remain private.

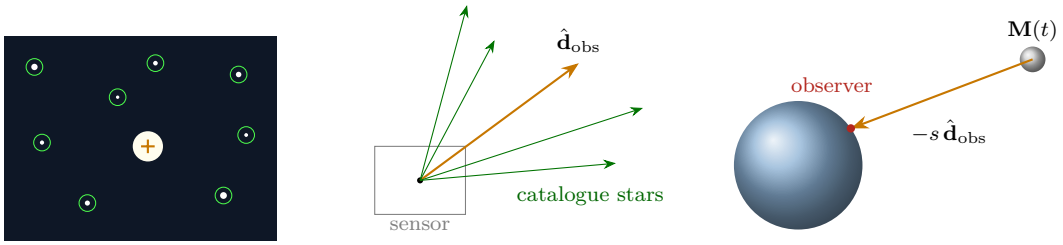
Adversary. We consider a prover who wants a proof for a region they were not in. They control everything on their side of the signature: the witness, the choice of photograph among those their device has signed, where the camera points, and (today) the phone's clock. They cannot forge the device's signature or break the proof system. Table 1 lists the assumptions; Section 9 examines each, including the ones that do not hold on current phones.

Privacy goal. The verifier should learn only that the statement holds for (S, t). In particular, two claims made from the same photograph should not be linkable through the proof, and the claim should not identify the phone's manufacturer. Section 6.8 describes what the claim contains.

Non-goals. We do not provide proof of *presence of a person*, protection against a prover who colludes with someone inside the region, or availability: the method needs a clear night sky with the Moon and enough stars in one frame.

	Assumption	Status in this work
A1	The proof systems (zkVM, hash, signatures) are sound, including the trusted setup of the Groth16 wrap.	Standard; we rely on SP1 and RISC Zero as published (Section 7).
A2	Public data are correct: ephemeris, star list, policy, borders.	Verifier recomputes every digest from its own copies.
A3	The signing key is confined to an honest capture path: signed pixels are what the sensor recorded.	Assumed. Emulated with a test certificate; no phone in our corpus signs RAW natively.
A4	The signed capture time is correct.	Out of scope here: the time is the phone's own clock. Can be addressed by binding the capture to an attested timestamp whose hash the proof pins (Section 9.3).
A5	The light that reached the sensor came from the sky.	*Not enforced*; violated by our screen-replay experiment (Section 9.2). Possible mitigations are proposed in Section 9.3 and Section 10.
A6	The public tolerance covers the measurement error of honest photographs.	Calibrated on the current corpus; more data are needed for a more precise calibration.
A7	The person making the claim is the person holding the device.	Out of scope (relay and proxy attacks).

Table 1: Trust assumptions. A1–A2 are cryptographic or publicly checkable; A3–A5 are assumptions about the physical capture and are where the system is weakest; A6 is an empirical calibration; A7 is shared with every device-based location proof.



(a) Identified stars fix the camera's orientation R and its lens model.

(b) The Moon's centre pixel becomes a direction in space.

(c) From the Moon's known position, the line along that direction meets the Earth at the observer.

Figure 4: One image, one fix. The observer is at $\mathbf{R} = \mathbf{M}(t) - s\,\hat{\mathbf{d}}_{\text{obs}}$ where this line meets the Earth ellipsoid.

4 Position from the Moon and stars

4.1 Geometry

Let $\mathbf{M}(t)$ be the Moon's position relative to the Earth's centre at capture time t, from the ephemeris, and let $\hat{\mathbf{d}}_{\text{obs}}$ be the unit direction in which the camera saw the Moon's centre. The observer lies on the line

$$\mathbf{R} \;=\; \mathbf{M}(t) - s\,\hat{\mathbf{d}}_{\text{obs}}, \qquad s > 0, \tag{2}$$

and on the Earth's surface; the nearer intersection of the two is the fix (Figure 4). Equivalently, a candidate position $\mathbf{R}$ predicts a direction $\hat{\mathbf{d}}_{\text{pred}} = (\mathbf{M} - \mathbf{R})/\|\mathbf{M} - \mathbf{R}\|$ and the solver adjusts latitude and longitude until the *Moon residual*, the angle between $\hat{\mathbf{d}}_{\text{pred}}$ and $\hat{\mathbf{d}}_{\text{obs}}$, vanishes. Height above the ellipsoid cannot be recovered this way. It matters little: over the 4.5 km of height that the proof admits, the fix moves by 0.5 km on our frames with the Moon near the zenith and by 7.5 km on those with a low Moon (Section 6.7).

Effect	Size if ignored	Treatment
Moon's phase (centre of light ≠ centre of disc)	400–600″ (∼750–1 100 km)	Analytic offset along the Sun direction (below)
Refraction, Moon at 32°	96″ absolute; 9–18″ relative to stars	Absorbed by the per-photo camera fit; altitude gate at 25°
Annual aberration	up to 20″	Applied to catalogue stars and to the predicted Moon
Lunar albedo pattern (maria)	up to 0.05 Moon radii (∼50″)	Covered by the tolerance (Section 6.6); correctable in principle
Capture time vs. EXIF stamp	half the exposure (up to 4 s, 2.6 km)	Mid-exposure time (Section 6.2)
Light time (1.3 s)	0.7″	Solved; checked as an identity in the proof
Star proper motion since catalogue epoch	up to several ″ for a few stars	Catalogue propagated to the capture epoch

Table 2: Corrections, by size, for a phone telephoto frame (58″ per pixel).

Sensitivity. The lever arm is the Moon's distance, so an angular error $\delta\theta$ displaces the fix by

$$\delta x \;\approx\; D_{\text{Moon}}\,\delta\theta \;\approx\; 1.9 \text{ km per arcsecond} \tag{3}$$

across the line of sight. One arcminute, about one pixel of a phone telephoto camera, is 110 km. Along the Moon's azimuth the footprint is stretched by $1/\sin h$ for a Moon at altitude h, because the line of sight meets the ground obliquely: at $h = 30°$ the error doubles in that direction. A clock error of one second moves the fix by 0.5–0.7 km (we measure 0.65 km/s), mostly through the Earth's rotation.

These three numbers frame everything that follows. A phone that locates the Moon's centre to a tenth of a pixel could reach about 10 km; one that misplaces it by a pixel cannot do better than 100 km.

4.2 Corrections

Table 2 lists the effects that separate the idealised geometry from a real photograph, ordered by size. Two deserve comment.

The phase offset. A phone that exposes long enough to record stars turns the Moon into a saturated blob surrounded by glare; no limb is visible to fit a circle to. What can be measured robustly is the *centre of brightness* of the blob. For a partly lit sphere that centre is displaced from the centre of the disc towards the Sun (Figure 7b). For the Lommel–Seeliger scattering law, which describes the lunar surface well, the displacement has a closed form [7, 42, 48] and is well approximated over the phases we admit by a linear law [42]

$$\delta \;=\; 0.3552\, r\, \alpha, \tag{4}$$

with r the Moon's radius in pixels and α the Sun–Moon–observer phase angle in radians. On our frames δ is 7–10 pixels, the largest correction of all by an order of magnitude: left uncorrected, it would displace the fix by roughly 750 to 1 100 km. The linear form differs from the exact one by at most 0.14 pixel (17 km) for $\alpha \le 107.5°$, a systematic bias we accept in exchange for a formula that is one multiplication inside the proof.

Refraction. The atmosphere lifts the Moon and the stars alike. The standard formula of Sæmundsson [58] (the inverse of Bennett's [9]) gives the lift ρ, in arcminutes, of an object at

true altitude h, in degrees, at standard pressure and temperature:

$$\rho(h) \;=\; 1.02\,\cot\!\Big(h + \frac{10.3}{h + 5.11}\Big). \tag{5}$$

Since the Moon is measured relative to the stars, the natural correction is the difference $\rho(h_{\text{moon}}) - \rho(h_{\text{stars}})$ between the Moon and the mean altitude of the reference stars. That turns out to be too much. Because the camera is calibrated on stars in the same frame, the fit absorbs the mean lift and also most of its gradient across the field, so what remains at the Moon is much smaller than this difference, and applying it over-corrects (4.9 km against 2.1 km on synthetic frames, where the better figure is obtained by lifting each catalogue star by ρ before the camera is fitted). The proof models no refraction at all and requires the Moon to be at least 25° above the horizon; the residual effect at that limit (about 16″) is budgeted in the tolerance.

4.3 The solver

The solver, written in Python, has six stages.

1. **Decode** the RAW file to a linear monochrome array: the sum of each 2×2 Bayer cell, at half resolution. No demosaicing, no tone curve.
2. **Detect stars** [6, 10], masking the Moon's glare, and compute sub-pixel centroids.
3. **Identify stars.** The Moon itself seeds the search: its pixel position and its predicted direction fix two of the three rotational degrees of freedom, leaving a scan over roll and plate scale.
4. **Fit the camera** (attitude and intrinsics) by least squares against catalogue directions at the capture epoch.
5. **Measure the Moon**: centre of brightness in a fixed aperture of three lunar radii, sky from a surrounding annulus, then the phase offset of Eq. 4.
6. **Solve for position** by Gauss–Newton on the Moon residual, including light time and aberration. The Skyfield library [56] evaluates the DE440 ephemeris and the Earth's orientation at the capture time.

Input gates. A frame is not solved at all unless it passes four checks, each made as early as the data allow. Before anything is decoded, the Moon must be at least 35% illuminated, which follows from the capture time alone. After star detection (stage 2), the stars must be sharp. Once the Moon is located in the frame, it must be at least 25 pixels across, and its glare must be less than three times the true disc. All four come before star identification, so no position is ever computed for a refused frame.

The gates exclude a range of capture conditions in which a solution is technically possible but unstable: the solver still returns a position, yet small changes in the input move it by hundreds or thousands of kilometres. The quantities the solver itself reports (star count, fit residual, number of detections) do not distinguish these cases from the stable ones, so the limits are placed on the input. The proof design follows the same rule: the conditions under which a photograph is accepted are public limits on the input.

5 What phones achieve

5.1 Synthetic images

A renderer produces star fields and a phased Moon at its topocentric position for a given place, time and camera, with realistic noise. On these images the error tracks the plate scale almost

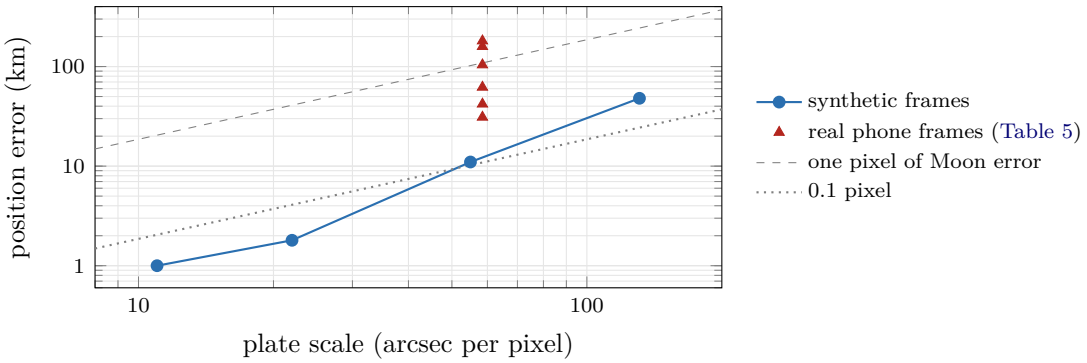


Figure 5: Accuracy against plate scale. Synthetic frames reach a few hundredths to a tenth of a pixel of Moon-centre error. Real frames sit between a quarter of a pixel and two pixels: the gap is the overexposed Moon.

linearly (Figure 5): the Moon's centre is recovered to a roughly constant fraction of a pixel, so every halving of the field of view halves the error. A 6° telephoto field gives 1 km; a 65° main-camera field gives 48 km. Repeating with different noise changes the answer by a third of its error: the solver is limited by bias, not by noise, so averaging frames helps little.

5.2 Real photographs

Between September and November 2026 we collected 356 captures in 14 sessions (at sites in the UAE, the Maldives and the EU), with a Samsung Galaxy S25 FE and iPhones, as RAW (DNG), HEIC and JPEG. GPS coordinates recorded at capture serve as ground truth.

Selection of frames. Part of the corpus was taken with an iPhone 16. This model has no telephoto camera, and we found that photographs from a wide camera very rarely produce a fix: the Moon is too small in the image, and the lens distortion over the wide field leaves larger residuals in the star fit. The results we report are therefore from the telephoto camera of the Android phone, captured as RAW (DNG), which is also the format the proof binds (Section 6.2); one wide-camera frame, cropped, is examined separately in Section 8. Six RAW captures pass the input gates and produce a position. The sessions were not tuned for a high yield: we deliberately covered a range of lunar phases and altitudes, times of night, and weather conditions, to find where the measurement is feasible. The frames that were refused are what the input gates of Section 4.3 were derived from. The frames that work are telephoto exposures on a tripod, long enough to record stars, in which the Moon is saturated but compact.

The frames that solve. The six frames come from two sites 3 000 km apart, with the Moon at 84° and 33° altitude. Their positions, computed under the rules of the proof, are reported with the evaluation in Table 5: 31–62 km from GPS ground truth at the first site and 104–181 km at the second. This is short of what the synthetic images promise by more than an order of magnitude, and Figure 6 shows why: the stars are located to a fraction of a pixel, but the Moon is a clipped blob more than twice the size of the true disc, and its centre of brightness must be corrected by seven pixels to reach the centre of the disc. The Moon measurement, not the star measurement, sets the error on every frame we examined.

What this means for a proof. A proof built on this measurement can make claims at the scale of hundreds of kilometres. The path to tighter claims is a better Moon measurement (shorter effective exposure of the Moon within the same frame, or a model of the albedo pattern), not better cryptography. We return to this in Section 6.6.

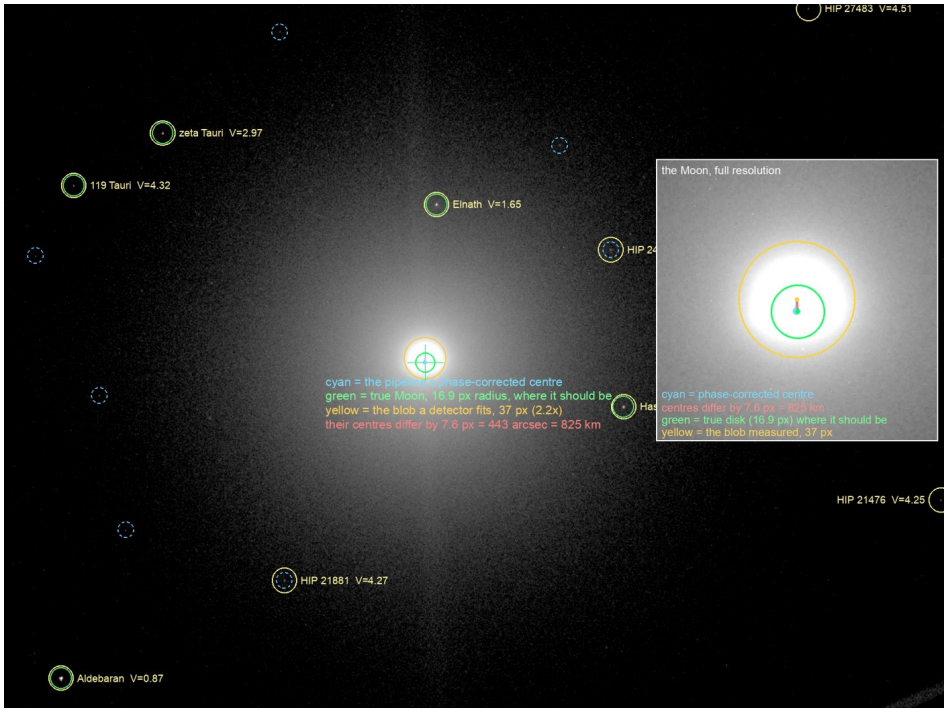


Figure 6: A frame that works (Dubai, 8 s telephoto exposure, shown stretched; annotations are the solver's own diagnostic overlay). Circles mark identified catalogue stars. The Moon is a saturated blob 2.2 times the true disc (inset: yellow, measured blob; green, true disc). Its centre of brightness sits 7.6 pixels, more than 800 km on the ground, from the centre of the disc; the phase correction of Eq. 4 removes most of that.

Instead of computing…	**…take a witness and check**
which catalogue star is which blob	the claimed star projects onto the measured centroid within $\varepsilon_{\text{star}}$
the best-fit camera and attitude	one Gauss–Newton step from the witness moves nothing
the inverse of the lens distortion	the forward model maps the witnessed point back to the pixel
the light time by iteration	$\|\mathbf{M}(t-L)-\mathbf{R}\| = cL$ holds to 1 m
the position minimising the residual	the residual at the witnessed position is $\leq 1''$
the boundary of the consistent set	the residual at each witnessed boundary point is $\geq \varepsilon_{\text{moon}}$

Table 3: Searches replaced by checks.

6 From a solver to a proof

The solver is cooperative: it assumes the photographer wants the right answer. A proof must assume the opposite. This section develops the verification predicate top-down: first the principles, then each stage.

6.1 Design principles

P1. Verify, do not search. Star identification, calibration, the position solve and the light-time iteration all happen outside the proof. The prover supplies their results as witness, and the proof checks predicates that only a correct result satisfies (Table 3).

P2. Measurements come from committed pixels. The proof never accepts a star centroid or a Moon centre from the prover. It accepts *where to look* (an integer pixel) and computes the measurement itself from pixels that are bound to the signed file.

P3. Every estimator is pinned. If an estimator has a free parameter — a window position, a threshold, an aperture centre — the prover will choose it. Each estimator is therefore fixed by public policy and made *canonical*: only one placement passes.

P4. A good fit is not evidence. With few stars and a flexible lens model, a wrong answer fits *better* than the right one (Section 6.4). The proof must constrain which stars are used and how the camera is obtained, not merely bound the residual.

P5. Prove the whole consistent set. Showing that one consistent position lies in S proves only that S comes within the tolerance of the truth. The proof must show that all consistent positions lie in S (Section 6.7).

An earlier iteration of the design targeted a hand-written arithmetic circuit, with fixed-point numbers, unit vectors in place of angles, and witnessed reciprocals and square roots so that no trigonometric function or division appears in constraints. Estimating its cost showed that the geometry would take a few thousand constraints and the pixel commitment several hundred times more. That ratio is why the implementation moved to a zkVM, where hashing a 16 MB file is routine, and the same predicate is written in Rust with IEEE doubles. The principles above are independent of that choice and the predicate remains expressible as a circuit.

6.2 Proof 1: binding pixels to a signed file

The first proof takes the signed DNG as private input and outputs a commitment to its pixels. It checks, in order: that the file contains exactly one C2PA manifest whose assertions match their hashes; that the SHA-256 hash of the file, minus the byte ranges the manifest declares excluded, equals the signed data hash; that the manifest is signed (ECDSA P-256 in a COSE envelope) by a certificate chain valid at capture time; and that every byte the proof subsequently interprets — TIFF directories, strip offsets, pixel strips, the EXIF time — lies in the hashed range. The last rule matters because C2PA signing rewrites the file and leaves a small unhashed region for the manifest itself: a parser that read from it would read attacker-controlled bytes.

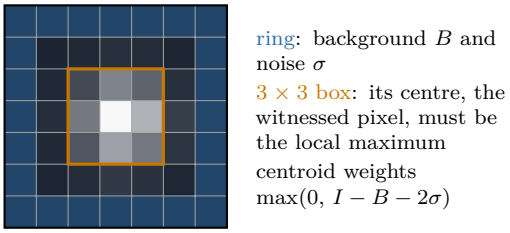
Our RAW files are uncompressed 16-bit Bayer mosaics, so pixel (x, y) sits at a fixed offset and nothing needs decoding. This is a significant simplification over JPEG or HEIC, where recovering pixels requires entropy decoding and an inverse transform inside the proof; we estimate RAW to be an order of magnitude cheaper despite the larger file.

The proof then cuts the mosaic into tiles of 32×32 raw samples (16×16 Bayer cells, all four colour sites) and outputs the root of a SHA-256 Merkle tree [51] over them. Each leaf includes its tile coordinates, so a tile cannot be presented at a different position, and leaves and inner nodes are domain-separated. All four channels are committed because the red channel alone, though it would be cheaper, records too few stars for a fix.

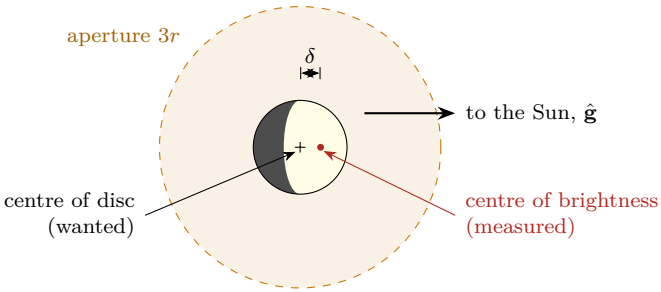
Capture time. The output includes the capture time. We found that the test phone stamps EXIF time at the *end* of the exposure (the stamp trails the shutter press by the exposure time plus about a second), whereas star and Moon positions are averages over the exposure. The binding proof therefore reads the exposure time from the signed EXIF data and outputs the mid-exposure instant. On an 8 s exposure this moves the fix by 2.6 km: irrelevant at today's accuracy, but a bias that would otherwise remain when everything else improves. The stamping convention is a firmware property, so it belongs to the binding program, of which there may be one per device family.

6.3 Pixels to measurements

Proof 2 opens only the tiles it needs (a few hundred of 7 854), checks each against the root from proof 1, and computes two kinds of measurement (Figure 7). Throughout, a pixel is one 2×2



(a) star window, 7×7 pixels



(b) Moon: phase offset $\delta = 0.3552\, r\, \alpha$

Figure 7: The two pinned estimators. (a) A star centroid is computed in a fixed window around a witnessed pixel that must be a local maximum, so the window cannot be slid. (b) The Moon's centre of brightness is computed in a fixed aperture whose centre must coincide with the result (a fixed point); the proof then steps back by δ along the projected Sun direction to the centre of the disc.

Bayer cell, the sum of its four samples; the 58″ per pixel of the telephoto camera is at this half resolution.

Stars. For each star the witness names an integer pixel P. The proof requires P to be the maximum of its 3×3 neighbourhood, and among equal maxima the first in reading order, takes the 7×7 window around it, estimates the background and its scatter from the 24 border pixels, and computes the centroid with weights $\max(0, I - B - 2\sigma)$. The local-maximum rule makes the window canonical; without the tie rule, two equally bright neighbouring pixels would both be admissible. The window size and threshold were chosen by a grid search on real frames (0.28 pixel RMS against the catalogue, worst star 0.6 pixel); larger windows were several times worse.

Moon. The witness names an aperture centre c. The proof computes the centre of brightness inside radius $3r$ about c, with the sky level taken as the median of an annulus, in exact integer arithmetic, and requires the result to equal c within 0.01 pixel. Because the centroid map is a contraction (measured factor 0.18–0.30), this fixed point is unique in practice: starting points 20–57 pixels away converge to the same centre within 0.001 pixel.

The proof also checks that *the phase law applies*. The linear law of Eq. 4 holds when the halo around the saturated blob carries about half of the centroid weight. The proof measures the share carried by the saturated plateau and requires it to be at most 0.6; on our RAW frames it is 0.48–0.59. The offset δ, the phase angle and the Sun direction in the image are all computed inside the proof from the public ephemeris and the witness's own attitude.

6.4 Stars to camera: removing the prover's freedom

The attack. Suppose the proof checks only that each claimed star projects within $\varepsilon_{\text{star}}$ of its centroid under a witnessed camera. On a real frame with 22 verified stars and an honest answer 51 km from the truth, we searched for subsets of stars and camera parameters that maximise the error while satisfying that check. With six stars the prover reaches 937 km with a fit residual of 0.09 pixel, five times *better* than the honest fit; with a looser tolerance, 1 756 km. Tightening $\varepsilon_{\text{star}}$ barely helps. A low residual indicates few constraints, not a correct answer.

First remedy: publish the camera. If the lens calibration is a public input, the camera's free parameters disappear. But this requires a trusted calibration per device model, and our two sessions with the *same phone* gave measurably different cameras: focal length differing by

0.1–0.3% and distortion coefficients by a factor of two, probably from focus position, optical stabilisation or temperature. A published camera is both a trust burden and wrong in detail.

Current design: the photo calibrates itself, with no choices. The attack needs two ingredients: a *chosen subset* of stars, and *free parameters* that absorb the error the subset induces. We remove both without publishing anything about the camera.

1. **Complete star set.** The list of all 8 870 Hipparcos stars with $V \leq 6.5$, sorted by brightness, is a public input; the proof hashes it whole. The witness declares a magnitude limit m. Using the witness's camera and attitude, the proof enumerates every list star with $V \leq m$ that projects into the image and requires the fit to use *exactly* those stars. Only public rules exclude a star: within 4 pixels of the edge, within 7 lunar radii of the Moon (inside its glare), or with a comparably bright neighbour within 6 pixels. A prover cannot drop an inconvenient star or add a convenient faint one.

2. **Least-squares stationarity.** The witnessed camera (f, c_x, c_y, k_1, k_2) and attitude must be approximately stationary for the least-squares fit to the proof's own centroids. The proof forms the residuals and their analytic Jacobian in the eight parameters, solves the normal equations, and requires that the resulting Gauss–Newton step move no star by more than 0.001 pixel. The prover cannot lean the fit in a preferred direction within the tolerance.

3. **Sanity bounds and coverage.** Focal length within a broad range, principal point within 5% of the image centre, distortion monotone over the image; at least ten stars, at least two brighter than $V = 4$, a star in every quadrant around the Moon, and the Moon within 25% of the half-diagonal from the principal point so that the lens model is interpolated, not extrapolated, where it matters. (The six telephoto frames sit at about 5%. The bound was 15% until the cropped wide-camera frame of Section 8, at 22%, was admitted; it is a heuristic resting on few frames.) The image size is taken from proof 1.

The only choice left is the limit m. An honest prover takes the faintest value at which the complete set still fits within one pixel, but the proof does not force this: any lower limit that satisfies the rules above is accepted. We measured this freedom on all nine accepted photographs by building a witness at every admissible limit, down to the ten-star minimum. The Moon's direction moves by at most $7.8''$, less than a tenth of the tolerance, which includes an allowance for it (Section 6.6). On the ground that is up to 9 km on the Dubai telephoto frames and up to 24 km on the Maldives frames, where the low Moon makes each arcsecond cost more. The largest shifts come from the smallest star sets, of 10 to 13 stars. On every photograph whose site lies inside the proven shape at the honest limit, it stays inside at every admissible limit. The freedom has gone from thousands of kilometres to tens. The proof could remove it by also checking that the next fainter limit fails, at the cost of one more least-squares check; we have not implemented this.

The price. Self-calibration needs stars. With 83–108 stars the per-photo camera reproduces a multi-frame calibration and the fix changes by under 4 km. With 13–19 stars the distortion terms are loosely determined and the Maldives fixes are 2–48 km worse than with a shared calibration. The rule also refuses frames outright when any required bright star is not measurable — hidden by cloud or haze, for instance — which is the correct outcome for soundness and a real cost for usability.

We have not run a deliberate adversarial search against the final rule set of the kind described above for the unconstrained camera; the scan over admissible limits is the only measurement so far.

6.5 Moon to direction, observer to residual

With the camera fixed, the proof converts the Moon's corrected centre to a direction without ever inverting the distortion model: the witness supplies the undistorted image point, and the proof checks that the forward model maps it to the measured centre within 10^{-6} pixel. Rotating by the attitude gives $\hat{\mathbf{d}}_{\text{obs}}$.

The observer is witnessed as a unit normal $\hat{\mathbf{n}}$ to the WGS84 ellipsoid and a height h, which must lie between -0.5 and 4 km, from which the proof computes the Earth-fixed position and rotates it to the inertial frame with the public Earth-orientation matrix. The Moon's position at the retarded time comes from a second-order Taylor expansion of the ephemeris about the capture epoch t, whose position, velocity and acceleration are public:

$$\mathbf{M}(t - L) \;=\; \mathbf{M}(t) \;-\; \dot{\mathbf{M}}(t)\, L \;+\; \tfrac{1}{2}\,\ddot{\mathbf{M}}(t)\, L^2. \tag{6}$$

The three vectors are computed outside the proof from DE440 by central differences over ± 60 s, and are part of the public ephemeris record that the verifier recomputes (assumption A2). The epoch of that record must equal the capture time from proof 1 to the millisecond. The light time $L \approx 1.3$ s is witnessed and checked as the identity $\|\mathbf{M}(t - L) - \mathbf{R}\| = cL$ to one metre. The first neglected term, $\frac{1}{6}\dddot{\mathbf{M}}L^3$, is of order nanometres. Aberration is applied to the predicted direction, as it was to the stars. The proof requires the Moon to be at least 25° above the observer's horizon and finally checks

$$\angle(\hat{\mathbf{d}}_{\text{obs}},\, \hat{\mathbf{d}}_{\text{pred}}(\hat{\mathbf{n}})) \;\le\; 1''. \tag{7}$$

The witnessed position is thus the fix itself, the centre of the set of positions that agree with the photograph within $\varepsilon_{\text{moon}}$ (Section 6.7), and not an arbitrary point of that set. An honest prover's residual is below $0.01''$. The proof also requires both rotation matrices to be proper (orthonormal with determinant $+1$). The determinant is not a formality: a mirrored attitude reproduces every star and flips the sign of the parallax.

6.6 How large must the tolerance be?

The tolerance $\varepsilon_{\text{moon}}$ is the central parameter. Too small, and honest photographs are rejected or, worse, the proven set excludes the true position. Too large, and the region that can be proven is uselessly big. It has to cover the error of an honest measurement plus whatever freedom the prover retains. The latter is at most $7.8''$ (Section 6.4), so the tolerance is mostly the honest error budget.

We derive that budget from a forward model of how the measured centre of brightness departs from Eq. 4 on a realistic Moon. The model renders the Moon for a given instant and observer with its true orientation (libration) [5], an albedo map from lunar orbiter imagery, a Lommel–Seeliger or rougher Hapke-type scattering law [37], a point-spread function with a glare halo, and sensor clipping; it then applies the proof's estimator. Over 1 706 epochs spanning 2025–2035 and 22 combinations of optical and photometric parameters, the worst error grows slowly with phase and is bounded by $0.053 + 0.009\,\alpha$ lunar radii (Figure 8). The dominant term is the albedo pattern: the dark maria pull the centre of light by about 5% of the radius even at full Moon. We therefore set, as public policy,

$$\varepsilon_{\text{moon}} \;=\; 32'' \;+\; \text{SD} \cdot (0.053 + 0.009\,\alpha), \tag{8}$$

with SD the Moon's angular semidiameter and α the phase angle, both already available inside the proof. The $32''$ floor is a budget for what is not in the Moon model, not a derived bound. It has three parts: star-fit noise propagated to the Moon's position (99th percentile $18.5''$ at ten stars), residual refraction at the altitude limit (about $16''$), which together we count as $25''$, and the $7.8''$ by which the prover's choice of magnitude limit can move the Moon. Whether the

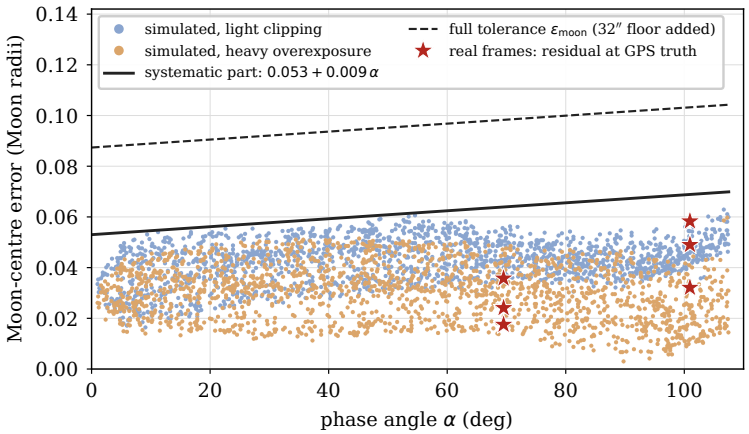


Figure 8: The tolerance law. Dots: worst simulated Moon-centre error per epoch over a decade, for lightly clipped and heavily overexposed Moons. Solid line: the adopted bound on the systematic part. Dashed: full tolerance including the 32″ floor. Stars: the residual at GPS ground truth on the six real phone frames, which includes all error sources.

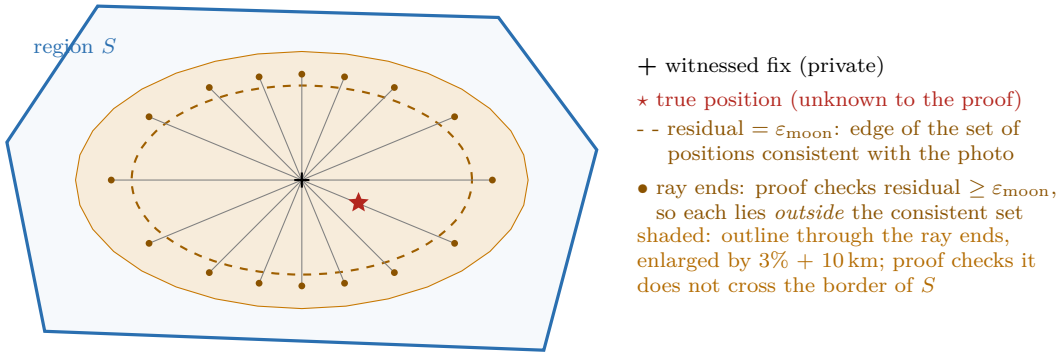


Figure 9: The uncertainty shape (schematic; the proof uses 32 rays). The prover chooses the ray lengths, but a ray can only be too long, never too short. The witnessed fix must be the centre of the consistent set; with that rule the shaded shape covered the consistent set in every case we tested.

budget is adequate is an empirical question: the margin between the residual at the site and the tolerance is at least 41″ on every accepted phone frame. On our frames Eq. 8 gives 91–96″.

On the real frames the residual at the GPS position is 17–52″.

Because the albedo term is deterministic — it depends on the epoch and a public map — it could be computed and subtracted, which would remove most of the 0.05-radius term and bring $\varepsilon_{\text{moon}}$ towards 60″. This is the most direct route to smaller regions and is future work.

6.7 From one position to a region

A proof that the witnessed position $\hat{\mathbf{n}}$ lies in S would be nearly worthless: a prover standing up to $\varepsilon_{\text{moon}} \times 1.9\,\text{km}$ outside S could witness a point just inside. What must lie in S is the whole set of positions whose predicted Moon direction agrees with the photograph within $\varepsilon_{\text{moon}}$. That set is an elongated oval on the Earth's surface, stretched along the Moon's azimuth by $1/\sin h$.

Bounding it with a linearised ellipse underestimates it by up to 28% on a low-Moon frame, so the proof samples it instead (Figure 9). The witness supplies 32 ray lengths from $\hat{\mathbf{n}}$ at equally spaced bearings. At the end of each ray the proof evaluates the Moon residual and requires it to be *at least* $\varepsilon_{\text{moon}}$: the ray end is outside the consistent set. Each sector between adjacent rays is given the longer of the two lengths, enlarged by 3% (three times the largest bulge of the true contour between rays that we measured at 32 rays; with 8 rays the bulge reaches 36%) plus

a fixed 10 km. Understating a ray is caught by the residual check.

Rays at 32 bearings do not by themselves guarantee that the enlarged shape covers the consistent set, so we tested it: on the nine photographs and on 141 synthetic geometries with the Moon between 25° and 90°, we compared the tightest shape the proof accepts with the edge of the consistent set at 3 600 bearings. When $\hat{\mathbf{n}}$ could be any point of the set, as in an earlier version of the policy, a shape built from a point near its edge left up to 5.6 km of it uncovered, at the lowest Moon the proof admits. This is why $\hat{\mathbf{n}}$ must be the centre (Section 6.5). With that rule nothing was left uncovered. The prover also chooses the height, between −0.5 and 4 km, and the shape is built at that height; the fix moves with it by up to 0.5 km with the Moon overhead and 7.5 km on the Maldives frames. Over all admitted heights together, the shape still covered the consistent set in every case, and the worst case used 4.9 of the 10 km margin. These are tests on a finite set of geometries, not a proof of coverage.

The region S is a spherical polygon: a set of include and exclude rings whose edges are great-circle arcs. The proof checks that $\hat{\mathbf{n}}$ is inside and that no ring edge intersects any enlarged sector. The test is exact: in the gnomonic projection about $\hat{\mathbf{n}}$, great circles are straight lines and spherical caps are discs, so each edge–sector pair reduces to clipping a segment against a wedge. Regions can be simple boxes, fitted polygons, or political. Our political regions are of the form "country X or the sea": everything except the land of every other country, from public map data [52]. Such a region contains all sea, not only the waters of X, because we did not use maritime boundary data.

6.8 The public claim and what it reveals

The final proof's public output is 204 bytes: the capture time, and digests of the region, the ephemeris record for that time, the policy (every tolerance and estimator parameter), the star list, the list of trusted signing roots, the list of trusted binding programs, and the number of stars used. The verifier recomputes each digest from the values it considers correct and checks the proof against the geometry program's verification key. What the claim hides also depends on the proof system being zero-knowledge, which is not true of every proof type we measured (Section 7).

Several things are deliberately absent. There is no hash of the photograph or of its pixel commitment, so two claims from one photograph are not linkable through the proof. The signing root is not named: the proof shows that the root is *in* the verifier's trusted list and outputs only the list's digest, so the claim does not reveal the device maker; the anonymity set is the list. The same holds for the binding program, which would otherwise reveal the RAW format and thereby the device family. Nothing about the camera is public. An earlier version output a hash of the camera parameters; since those are now fitted from the photo, the hash was both unverifiable and a stable identifier of the photograph.

Residual leakage. The claim is not information-free beyond "inside S", and applications should account for the following. The capture time is public to the millisecond; if the same photograph, or its metadata, is ever published elsewhere, the claim is linkable to it. This is not inherent in the method. The time could be kept private altogether, or revealed only coarsely, as an interval of seconds, minutes or hours that the proof shows the capture time to lie in. We do not do this yet, because the time must first be bound to an attested time source (Section 9.3); hiding or truncating it is future work. The star count leaks a little about sky conditions and field. More importantly, *the choice of S can leak the position.* A region derived from the fix and its error bound, as in several of our experiments, reveals the fix as its centre. Our recommendation is not to construct S this way, and to use fixed shapes defined independently of any claim: a country, a licensing zone, a service area. This is also how the question arises in practice. A service asks whether the user is inside a region that the service has defined; the claim does not

			SP1 (s)			RISC Zero (s)		
Photo	**File**	**Stars / tiles**	**P1**	**P2**	**P2$_{\text{nat}}$**	**P1**	**P2**	**P2$_{\text{nat}}$**
042106	16.0 MB RAW	13 / 282	98	17	42	458	183	755
042142	16.0 MB RAW	83 / 388	85	17	44	453	238	823
042204	16.0 MB RAW	108 / 429	91	19	44	456	251	819
193759	16.0 MB RAW	13 / 217	81	14	41	450	161	734
193811	16.0 MB RAW	19 / 260	81	14	39	453	170	746
193831	16.0 MB RAW	15 / 232	86	14	43	453	164	746
0416crop	12.3 MB*	15 / 83	66	8	39	355	77	663
hyades	7.7 MB*	30 / 835	43	33	61	222	509	1 063
moonjupiter	12.0 MB*	70 / 568	67	24	47	345	341	853
Proof size				1.27 MB			224 kB	
Verification				0.06–0.08 s			0.02 s	
Peak GPU memory				25.9 GB			13.9 GB	

Table 4: Proving time in seconds on one A100, for the proof alone. P1: binding proof. P2: geometry proof with a box or octagon as the region, including the verification of P1. P2$_{\text{nat}}$: the same with a region defined by national borders (82 000–93 000 vertices). "Tiles" are the 32×32 pixel tiles the geometry proof opens. The RISC Zero P1 times are from the earlier run (see text). *Not a camera RAW file: a HEIC or JPEG rewritten as an uncompressed DNG (Section 8).

come with a region of its own. With a fixed S the prover checks locally whether the uncertainty shape fits, and publishes a proof only if it does.

7 Implementation and cost

Both proofs are Rust programs whose core logic compiles unchanged for native execution and for the zkVM guest; the test suite (54 tests for the geometry alone) runs natively, and the guest is checked to reach the same verdict on every test frame. Arithmetic is IEEE double precision, emulated deterministically in the guest; tolerances inside the proof (10^{-6} pixel, 1 m, 10^{-10} for orthonormality) are many orders above rounding error. A refused proof is a guest that panics: the proof system will neither verify nor recurse over an execution that did not exit cleanly.

Both proofs were implemented for SP1 [63] (version 6.10.0) and RISC Zero [15] (version 3.1.0, which fixes soundness bugs in the 3.0 releases). Proof 2 verifies proof 1 inside the zkVM: by recursion in SP1, and in RISC Zero by taking the binding receipt as an assumption that the final receipt resolves. Photographs were signed with `c2patool` [21] under a test certificate chain. Measurements used a rented server with one NVIDIA A100 (40 GB) and 16 CPU cores.

We ran nine test photographs (Section 8) through proof 1, through proof 2 with a simple region (a box or an octagon), and through proof 2 again with a region defined by national borders. On each zkVM that is 27 proofs. Every proof verified, and its public output equals the claim computed natively. The geometry proofs were made again after the last change of the policy (Appendix A). The binding program did not change, so on RISC Zero the binding proofs of the earlier run were kept and the new geometry proofs verify them.

Where the cost goes. Table 4 gives the measurements. On SP1 a 16 MB phone photograph takes 94–116 s (1.6–1.9 minutes) for proof 1 and one claim with a simple region. Proof 1 is the larger part, and its time follows the file size (43 s for 7.7 MB, 81–98 s for 16 MB): it hashes the file twice, once for the C2PA data hash and once for the tile tree. Proof 1 is made once per photograph; each additional region claim costs only proof 2. The time of proof 2 follows the number of tiles opened, from 8 s for 83 tiles to 33 s for 835. Within proof 2, the star list hash and the completeness scan account for 12–17 M cycles and tile openings for about 10 M on the

phone frames; the geometry itself is negligible, consistent with the early circuit estimate.

The two zkVMs. RISC Zero is slower on this workload: about 5 times for proof 1, 10–15 times for proof 2, and 17–19 times for proof 2 with national borders. Its proof is 5.7 times smaller, it verifies about three times faster, and it needs about half the GPU memory. Its times follow its cycle counts: 139 M cycles for proof 1 on a 16 MB file and 28–195 M for proof 2. Proving on CPU, though possible, takes hours.

Constant-size and zero-knowledge proofs. For on-chain verification both zkVMs can wrap their proof into a Groth16 proof [36], whose size does not depend on the computation. Groth16 requires a trusted setup; we use the circuits and keys that each project publishes. We wrapped the proof 2 of all nine photographs on each zkVM, without proving them again, and every wrapped proof verifies. The SP1 proof becomes 356 bytes. The wrap ran on the CPU of the proving server and took 345–363 s with 30 GB of memory; the wrapped proof verifies in 0.38 s. The RISC Zero proof becomes 256 bytes; its wraps took 89–126 s on a CPU and verify in under 0.01 s. The wrap is a fixed cost per proof and does not depend on the photograph. We have not measured it on a GPU for either zkVM.

The wrap also matters for privacy. According to SP1's documentation [64], its STARK proofs, which include the 1.27 MB proof of Table 4, do not currently have the zero-knowledge property; the Groth16 wrap does. On SP1, a proof that is meant to hide the photograph and the position therefore has to be the wrapped one, and the wrap is part of its cost. RISC Zero states that its receipts are zero-knowledge; we use its succinct receipt, which is produced by recursion. We rely on these properties as the two projects state them and have not analysed them ourselves.

National borders. A region defined by real borders adds 23–31 s to proof 2 on SP1, and 199–225 M cycles on RISC Zero, because every one of its 82 000–93 000 vertices is deserialised and hashed although only nearby rings matter. Committing the rings in a Merkle tree and opening only those within reach would remove most of this.

8 Evaluation on real photographs

Scope. We ran every RAW frame in the corpus that could conceivably carry a proof through the whole chain: prover-side star identification, self-calibration, signing with the test certificate, witness generation, and the bound verification predicate, executed natively. We started from 356 photographs of our own. We also examined 489 publicly available photographs of the Moon among stars, to look for frames that give a stable fix.

Accepted phone frames. Six telephoto frames are accepted (Table 5, first group). In all six, the GPS ground truth lies inside the proven shape: no accepted proof excludes the true position. All six were proved end to end on both zkVMs (Table 4). One of them (042106) was never used during development; with its own complete star set it lands at 62 km.

The two sites differ. The Dubai frames, with the Moon near the zenith, are fixed to 31–62 km. The Maldives frames are fixed to 104–181 km, which is consistent with the lower Moon (the $1/\sin(\text{altitude})$ stretch is 1.8 at 33°) and with a lens calibrated on only 13–19 stars. The proven shapes reflect the same difference: their radius is 168 km in Dubai and about 370 km in the Maldives.

Photographs from other cameras. Three further photographs test whether the predicate, built on one phone's telephoto frames, accepts anything else (Table 5, second group). None of

Frame	Camera, site	Stars (to V)	$\varepsilon_{\text{moon}}$	Residual at site	Fix error	Shape radius	Region proved
Galaxy S25 FE telephoto, RAW; GPS ground truth							
042106	Dubai, 1 s	13 (4.66)	95″	35″	62 km	168 km	UAE, Oman, Iran
042142	Dubai, 4 s	83 (6.10)	95″	17″	31 km	168 km	UAE, Oman, Iran
042204	Dubai, 8 s	108 (6.29)	95″	24″	42 km	168 km	UAE, Oman, Iran
193759	Maldives, 1/6 s	13 (4.54)	94″	44″	159 km	370 km	Maldives
193811	Maldives, 1/3 s	19 (4.85)	94″	52″	181 km	373 km	Maldives
193831	Maldives, 1/3 s	15 (4.64)	94″	29″	104 km	365 km	Maldives
Other cameras; converted files without device provenance							
0416crop	iPhone 16 wide, Dubai	15 (3.40)	95″	53″	94 km	168 km	GCC
hyades	Canon 60D, 135 mm, Alberta	30 (6.50)	96″	43″	92 km	272 km	Canada
moonjupiter	Canon 20D, 78 mm, Rome	70 (6.50)	91″	108″	317 km	300 km	EU

Table 5: Photographs accepted by the proof's predicate. "Residual at site" is the Moon residual at the ground-truth position (GPS for the first group, the photographer's stated site for hyades and moonjupiter); it must be below $\varepsilon_{\text{moon}}$ for the site to lie inside the proven shape, which fails only for moonjupiter. "Fix error" is the distance from the witnessed position to that site and is not revealed by the proof. "Shape radius" is the longest ray of the uncertainty shape. A national region is the land of the named countries together with all sea.

them is an original RAW file. Each was rewritten as an uncompressed DNG and signed with the test certificate, so they exercise the geometry and the cost of the proofs, not the binding to a device.

A cropped wide-camera frame (0416crop). The full frame, from the 26 mm main camera of an iPhone 16, is refused: a building covers a fifth of it and hides bright stars that the complete-set rule requires. Cropping the building away, and the same width on the opposite side, leaves a frame that is accepted with 15 stars, 94 km from GPS, with the GPS position inside the shape. Two reservations apply. Admitting it required relaxing the off-axis bound from 15% to 25% (Section 6.4), and its plate scale (137″ per pixel) is one at which the tolerance law has not been calibrated. And a crop of a genuinely signed file would break its signature; doing this honestly needs an editing manifest and a proof that the pixels are a sub-rectangle of the original, as in the work on signed-image edits [22, 53].

Two published photographs. "Waning Moon in the Hyades" by Alan Dyer [24] (Canon EOS 60D, 135 mm, 0.4 s, 2014) and "The Moon, Jupiter and Aldebaran" by Gianluca Masi [50] (Canon EOS 20D, 78 mm, 4 s, 2012) are the first photographs through the proof that are neither from a phone nor our own. We used the published JPEG files and report only numbers derived from them. Neither file carries a position, so the "site" is what the photographer states, and the time is the camera's own unsynchronised clock.

The first is accepted with 30 stars. Its fix is 92 km from the photographer's site in Alberta, which lies inside the proven shape, and the shape lies inside Canada.

The second is accepted with 70 stars and a good star fit, yet its fix is in the Tyrrhenian Sea, 317 km west of Rome, and Rome is *outside* the proven shape (108″ against a tolerance of 91″). The likely cause is the timestamp: the camera clock carries no time zone, and its error is unknown; shifting it by five minutes brings the fix to about 100 km from Rome. We keep this photograph as an illustration of assumption A4, not as an accuracy measurement. The predicate accepted, although the reported site lay outside the tolerance, because the proof trusts the signed time exactly as it would on a phone.

Refused frames. Of our own photographs, 32 RAW frames reached the proof and 26 of them are refused: 8 are short Moon exposures with no real stars; 4 are swamped by glare seven times the Moon's size; 2 have no plate solution; and 12 are main-camera frames (23 mm equivalent) that fail the complete-set rule, because some required bright star has no measurable

flux — several sit in haze or thin cloud. These refusals are the rule working as intended. As a diagnostic, we let the prover drop the unmeasurable stars on those twelve frames: three then fit on 6–8 stars, and two of the three would have placed the true position *outside* the accepted set (residuals at GPS of 229″ and 343″ against a tolerance of about 95″). The tolerance law was calibrated at telephoto scale and does not transfer to a lens 2.5 times coarser.

How large are the provable regions? Figure 10 shows five of the proven shapes. The region is much larger than the fix error, and necessarily so: it must contain every position the policy would accept, which is set by $\varepsilon_{\text{moon}}$, not by how well a particular photograph happened to do. At 95″ the half-width across the line of sight is $95 \times 1.9 \approx 180$ km at the Moon's mean distance. With the Moon overhead (Dubai) the observer is an Earth radius nearer to it; the shape is nearly a disc of radius 168 km and fits a 370×370 km box. With the Moon at 33° (Maldives) it stretches to 370 km along the Moon's azimuth and needs a polygon of about 710×640 km.

Political regions. The last column of Table 5 lists the region defined by national borders that was proved for each photograph on both zkVMs. Each region is the land of the named countries together with all sea (Section 6.7), so a proof shows that the photographer was not on the land of any other country; it does not show that they were within the maritime borders of the named ones. The three Maldives frames each prove "the Maldives or the sea". The three Dubai telephoto frames prove "UAE, Oman, Iran or the sea" but not "UAE or the sea" or the same for the member states of the Gulf Cooperation Council (GCC): the shape reaches Iran's coast, 136–141 km from the site. A GCC claim from this site needs $\varepsilon_{\text{moon}} \leq 69''$ and a UAE claim $\leq 41''$; the law gives 95″. The union of the three countries is not a meaningful political region; we use it only because it is the smallest set of countries that contains the shape. The cropped wide-camera frame, taken at the same site, does prove "GCC or the sea": its fix happens to fall further from the Iranian coast, so its shape stays inside. The hyades photograph proves "Canada or the sea". The moonjupiter photograph proves "EU or the sea" but neither "EU land" nor "Italy or the sea", since its shape lies largely over the sea and reaches Corsica. Whether a claim is provable thus depends on where in the region the prover stands and on where the fix happens to fall, a property any uncertainty-aware location proof shares.

How the policy relates to these frames. The policy was not fixed before the evaluation. The input gates were derived from the frames that fail, the tolerance law was checked on five of the six accepted phone frames, and the off-axis bound was widened to admit the cropped frame. Only 042106 was first examined after the rules were set. The results therefore show that the approach works end to end on real data. They do not yet measure how often an honest photograph is accepted, or how often the true position falls outside the proven shape. Measuring both needs a policy frozen in advance and a new set of photographs from several devices.

9 Security analysis

We separate three questions: what a dishonest prover can do *within* the proof's rules; what they can do by violating the capture assumptions A3–A5; and what the proof means once those are accounted for.

9.1 Attacks the proof rejects

Table 6 lists manipulations we tested against the final predicate on the real frames; each is refused. Most were found the hard way, as described in Section 6.

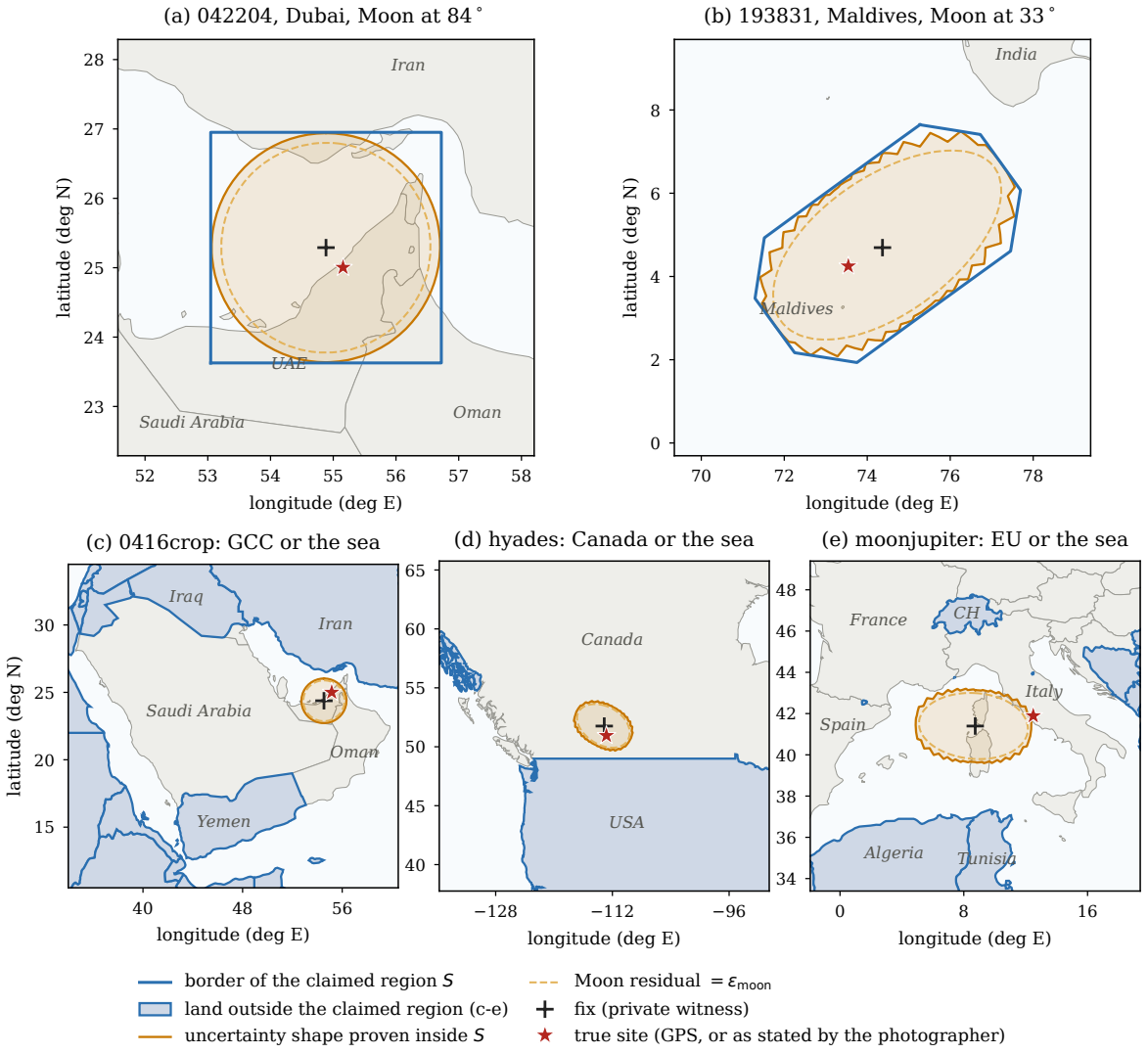


Figure 10: Proven uncertainty shapes for five accepted photographs. The verifier sees only the region. Top: two phone frames with the simple regions used in testing. (a) With the Moon near the zenith the shape is nearly circular. It crosses into Iran and Oman, which is why a "UAE" or "GCC" claim is not provable from this frame. (b) A low Moon stretches the shape along the Moon's azimuth. Bottom: the three photographs from other cameras, each with the national region proved for it; the region is everything except the darker land, and the blue line is its border. (c) The whole GCC; the shape stays clear of Iran. (d) The shape lies inside Canada. (e) The shape lies over the sea, Corsica and Sardinia.

Manipulation	Stopped by
Altered, foreign or withheld pixel tiles	Merkle opening against proof 1's root
Invented, misidentified or brightened star	centroid computed from committed pixels; residual bound; star list hash
Sliding a star window to shift its centroid	witnessed pixel must be the local maximum
Moving the Moon aperture	fixed-point condition
Choosing a convenient Sun direction	computed in the proof
Dropping inconvenient stars / adding convenient ones	complete star set
Tilting attitude or lens within the star tolerance	least-squares stationarity
Mirrored attitude (flips the parallax)	determinant +1
Declaring a smaller image to exempt edge stars	image size from proof 1
Wrong ephemeris epoch or light time	epoch bound to signed time; light-time identity
Placing the fix at the edge of the consistent set	residual at the witnessed position must be $\leq 1''$
Understating the uncertainty shape	residual at ray ends must be $\geq \varepsilon_{\text{moon}}$
Untrusted signer or binding program	membership in the verifier's lists
Not-a-number in place of a measured or witnessed value	every input must be finite; every comparison rejects NaN

Table 6: Manipulations tested against the final predicate (all refused).

Frame	Stars (to V)	Star RMS	Fix vs GPS	Residual at GPS	Verdict
real sky, 042142	83 (6.10)	0.24 px	31 km	$17.3''$	accepted
screen, shot 1	107 (6.29)	0.32 px	64 km	$36.5''$	accepted
screen, shot 2	107 (6.29)	0.33 px	72 km	$41.2''$	accepted
screen, shot 3	138 (6.50)	0.32 px	68 km	$38.5''$	accepted

Table 7: Screen replay. A photographed laptop screen is accepted by the proof's predicate for the location of the original capture. The tolerance $\varepsilon_{\text{moon}}$ is $95''$ for all four frames.

9.2 An attack the proof accepts: screen replay

The proof binds its conclusion to pixels signed by a trusted device. Figure 11 shows the part of the chain this leaves uncovered. We tested it directly.

Experiment. The same Galaxy S25 FE, on a tripod, photographed a high-brightness laptop display showing a rendered version of the sky as it appeared from Dubai during one of our real captures. The EXIF capture time of the resulting RAW file was set to that of the original before signing, which emulates a phone whose clock has been set back; the file was then run through the unmodified prover. Three shots were **accepted** as claims on a 370 km box around Dubai, with 107–138 stars each, fix errors of 64–72 km and the original GPS position inside the proven shape (Table 7). The forged frames satisfy the complete-star-set rule to a *fainter* magnitude than the real frame. The rule that the ephemeris time equals the signed capture time does not hinder the attack: the signed time of a forged shot is the middle of its 1/6 s exposure, which is 1.9 s after the middle of the original 4 s exposure, and in that time the Moon moves about $1''$ against the stars.

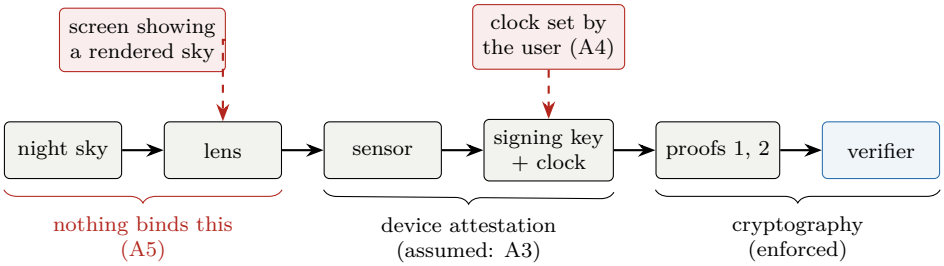


Figure 11: Where the binding holds. The proofs bind the verifier's conclusion to the signed pixels and the signed time. Whether the pixels depict the sky, and whether the time is true, are outside the proof.

What the attacker needs. The phone is unmodified and the pixels are what its sensor recorded. As everywhere in this work, the file was then signed with our test certificate, not by the phone; the attack does not depend on this, since a camera that signed its own captures would sign the same pixels. The attacker needs (i) the ability to set the phone's clock (emulated here), (ii) a display, and (iii) a one-time calibration: photographing a dot grid to learn the mapping from screen to sensor, including the lens distortion at the close focus distance, which differs from its distortion at infinity. With that, each catalogue star is rendered where the real sky would have put it on the sensor. For the Moon we warped the pixels of the original photograph; an attacker without an original would have to synthesise the glare blob.

It took 68 failed shots to get there. Each failure traced to a rendering or placement detail — bright stars predicted onto the screen bezel, the camera off the projection axis, uncalibrated near-focus distortion — and none to a property the attacker cannot control. We regard none of the checks that caught the early attempts as a security boundary.

Could the proof tell the difference? In this particular forgery, yes, in several ways: the stars are about 600 times too bright for their exposure settings (a photometric zero-point check), have no colour, sit on a black background without sky glow, and have footprints three times too large. Every one of these is a choice the attacker made for convenience and can change: dim the screen or add a neutral filter, render colours and a background, sharpen the stars. Forensic checks on pixels raise the cost of an attack and are worth adding; they are not a boundary either. This is the familiar situation of recapture detection [17, 65], with the difference that a star field is a far simpler scene to reproduce than a natural image.

A forged Moon under real stars. The stars look the same from every site, and only the Moon's place among them carries the position. An attacker therefore does not need to forge the sky, only the Moon. The proof measures the centre of brightness of an overexposed blob and does not check that the blob is the Moon. If a second bright source is in the field, such as a lamp or a reflection in a sheet of glass, the prover can name it as the Moon. The stars are real, so their colour, brightness and sharpness and the sky background are those of a real photograph, and the forensic cues that expose our screen replay are absent. We have not tested this attack.

What the proof establishes. Stated carefully:

> *A device trusted by the verifier recorded, at the time it signed, an image whose Moon-against-stars geometry is consistent only with positions inside S.*

Without a trusted time source this does not even establish *when*; with one, it still does not establish that the image is of the sky.

Attack	What it needs today	Mitigation, and what remains
Replay of a past sky (A4)	Set the phone clock; show a recorded or rendered sky.	Trusted time inside the signed manifest (timestamp authority [1], authenticated network time [44]), or a verifier nonce in the manifest. Remains: rendering the *current* sky for a remote site, within the $\sim$270 s that $\varepsilon_{\text{moon}}$ tolerates.
Screen at close range (A5)	Display, calibration, tripod.	Capture app holds focus in its long-range setting and signs the focus state; the proof's sharp-star requirement does the rest. Remains: screen beyond the hyperfocal distance (15–20 m for the telephoto, hence several metres wide), or a close-up lens.
Screen at distance / projector (A5)	Large screen at tens of metres.	Simultaneous capture with a second camera on the phone (parallax between lenses is zero only at infinity); signed gravity vector checked against the fitted attitude and claimed site. Remains: collimator.
Forged Moon under real stars (A5)	A lamp or a reflection in the field of a real photograph of the stars.	Proof 1 already reads every pixel; it can also commit to the list of tiles that contain saturated pixels, and proof 2 can require that all of them belong to the one blob it measures. This needs no further tile openings.
Collimator (A5)	Screen at the focus of a lens larger than the phone's camera baseline: light arrives from optical infinity, as in star-tracker test benches [57].	No optical depth cue distinguishes it. An interactive capture that also tests sky coverage, sideways motion and agreement with the inertial sensors (see text). Raises the cost well beyond an individual's resources; out of scope here.
Frame injection (A3)	Compromised camera stack or extracted key.	Hardware-backed keys with remote attestation [2], verified boot state, revocation. Standard, and outside our control.
Relay / proxy (A7)	An accomplice in S takes the photograph, or sells one.	Bind the signing key to an identity or session; verifier nonce. Inherent to device-based evidence.
Honest error beyond $\varepsilon_{\text{moon}}$ (A6)	Nothing: an unlucky honest photo.	Larger calibration corpus per lens; conservative bounds on focal length; albedo model. A too-small $\varepsilon_{\text{moon}}$ harms honest users (rejection) and soundness (true position outside the shape).

Table 8: Attacks on the capture assumptions, and how far each mitigation goes.

9.3 Trust assumptions, attacks and mitigations

Table 8 goes through the assumptions of Table 1 from the attacker's side.

Time first. Trusted time is the cheapest and most valuable addition. With the capture time attested, an attacker can no longer reuse a recording of the sky from the target location; they must compute what the sky looks like *now* from there and display it. Our rendering pipeline already draws stars from the catalogue, so this is a modest step for stars; the Moon's glare blob would have to be synthesised convincingly. A verifier nonce bound into the signed

manifest gives freshness without a time authority. The proof already takes the time from signed bytes, so supporting a countersigned timestamp is a change to the binding program only. The moonjupiter photograph of Section 8 shows that this matters for honest users too: a camera clock a few minutes off yields a valid proof whose shape excludes the true position.

Then optics. A screen at arm's length is the cheap attack, and it can be excluded by focus: the phone's telephoto lens has a hyperfocal distance of roughly 15–20 m, so a display nearer than that cannot be sharp when focus is held at long range, and the proof already demands sharp stars. This requires the focus state to be *signed*, which leads to the next section.

Advanced fakes, and raising the bar. The sky can be faked with more advanced equipment than a laptop screen. A collimator presents a displayed image at optical infinity; laboratory "star simulators" built on this principle are the standard way to test star trackers on the ground [57]. Against a collimator the focus test of the previous paragraph shows only that the light arrives from optical infinity.

The real sky nevertheless has three properties that a fake has to reproduce *at the same time*. It is everywhere: there are stars in every direction, not only around the Moon. It is infinitely far: moving the camera sideways shifts nothing. And it is fixed in space with no delay: every rotation that the phone's gyroscope reports appears in the image at the same instant. The apparatus we can think of each fails one of the three. A collimator covers a few degrees of sky. A projection dome is at a finite distance and shows parallax. A display that is attached to the phone and follows its motion lags behind the gyroscope.

This suggests an interactive capture, run by an attested application (Section 10), that tests the three properties together. The application obtains a fresh nonce with a signed time and derives from it a sequence of requests: point at a clear part of the sky, then at another, then at the Moon; move the phone sideways; all while it compares the camera stream with the inertial sensors. The final photograph is taken in the same session and must continue what the stream showed. We expect such a protocol to raise the cost of a forgery from a laptop and a tripod to purpose-built optics with millisecond tracking, well beyond the resources of an individual. It does not make forgery impossible, and it relies on the integrity of the device and of the application. We mention these measures as directions; their design and evaluation are outside the scope of this paper.

10 Towards binding to the sky

Which signals could a phone sign today that bear on Table 8?

Files carry almost nothing. We examined the metadata of the 356 captures of our corpus and of public image sets. None of the files from the test phone, RAW or JPEG, records focus distance, lens position, or inertial data. iPhone files (120) carry undocumented maker-note fields related to focus; on 52 night frames from one iPhone these values cluster by what autofocus locked onto. Even on genuine frames of the Moon and sky the reading is not the nominal infinity value but a point near the far end of the scale, so a usable rule is "focus within the long-range part of the scale", not "focus must read infinity". Dedicated cameras (Canon, Nikon) record focus distance in metres. In short, no current phone file carries a signed measurement of where the light came from.

The camera API has what the files lack. On Android, the Camera2 interface [3] reports per-frame focus distance in dioptres with a declared calibration level, lens state, the sensor timestamp at exposure start, optical-stabilisation samples, per-lens intrinsic calibration and

Approach	Evidence is hard to fake because...	Needs nearby infrastructure?	Main weakness
Civil GNSS	(it is not: signals are unauthenticated)	no	spoofing, replay
Authenticated GNSS	navigation messages are signed	no	signal replay with delay
Distance bounding, beacons	radio signals cannot beat light	yes, verifiers	deployment; relay
Witness-based proofs	co-located peers vouch	yes, peers	collusion
Network delay	packets cannot beat light	landmarks	coarse; delay manipulation
Attested sensor reading	hardware signs its GPS fix	no	inherits GNSS spoofing
ZK over a coordinate	(assumed input)	n/a	proves a property of an unverified input
This work	the Moon's parallax cannot be changed	**no**	camera can be shown a fake sky; coarse

Table 9: Location evidence by its physical anchor.

distortion, and, through the sensor interface, accelerometer and gyroscope samples on the same clock. The reported focus distance is the commanded lens position, not a measurement of the scene, so its use is to *force* long-range focus and let image sharpness decide. The setting at which the sky is sharp need not be the nominal infinity value: lens calibration and temperature move it, so the admissible range has to be measured per lens.

An attested capture application (design sketch). With public APIs on a stock, locked phone one can: capture RAW through Camera2 and write an uncompressed DNG, the format the binding proof already reads; hold a signing key in the hardware keystore with an attestation chain to the platform vendor's root, which certifies the key's hardware protection, the application's identity and the verified-boot state [2]; and embed, in the signed C2PA manifest, an assertion with focus state and calibration level, sensor timestamp, lens calibration, inertial samples, a verifier nonce and optionally a frame from a second camera. The proofs would then additionally check the attestation chain and apply policy to those fields. This moves trust from the camera vendor's imaging pipeline to an application process on an attested operating system, the model of earlier trusted-sensor proposals [60]; injecting frames would require a system-level exploit on a device that still passes attestation. We have not built it.

11 Related work

We organise prior work by the question a verifier would ask: *what makes the location evidence hard to fake?* Table 9 gives the one-line answers; the paragraphs below give context.

Satellite navigation and its authentication. A phone's own position estimate is the obvious evidence, and the easiest to forge: civil GNSS signals are unauthenticated, and portable spoofers have been demonstrated for well over a decade [39, 66]. Signed navigation messages [29] authenticate the data but not the signal's arrival time. A scheme that has a device attest its GNSS reading [60] certifies that the receiver reported a position, not that the position is true. Our method shares this structure — an honest device measuring a possibly dishonest world

— but the "signal" is passive starlight, which cannot be jammed or spoofed at a distance; the attacker has to get between the sky and the lens.

Location proofs from infrastructure and witnesses. Distance-bounding protocols [12] upper-bound the distance to a verifier by timing; location-verification schemes build on them with several verifiers [18, 61]. Location-proof architectures have access points or co-located users issue signed statements [59, 69], and decentralised variants replace the trusted issuer with incentives [30]; [68] surveys the area. All of these need cooperating parties near the prover. Delay-based internet geolocation needs only landmarks but localises a network endpoint and can be misled [32]. The appeal of a celestial anchor is that it is present, identical and free everywhere on Earth; its drawbacks are weather, night-time, and the resolution reported here.

Zero-knowledge location. The closest cryptographic work is ZKLP [27], which builds SNARK circuits for IEEE floating-point arithmetic and uses them to prove that a private coordinate lies in a public region, with applications to private proximity testing. ZKLP and this paper are complementary: ZKLP assumes an authentic coordinate and proves a private predicate about it efficiently; we address where an authentic coordinate could come from, and our region test over an uncertainty shape would sit where ZKLP's point-in-region test does. Our arithmetic uses a zkVM's emulated floating point; purpose-built floating-point proof systems [27, 31] would likely be much cheaper for the geometry, though the geometry is not where our cost is.

Proofs about signed photographs. A line of work proves that a published image was derived from a camera-signed original by permitted edits (crop, resize, redact) without revealing the original [22, 25, 35, 43, 53], or that it was correctly compressed [23] or cropped [55]. We borrow the setting — a signature on the original, a proof about a derived statement — but the derived statement is a measurement, not an edited image, and we open only a few hundred small windows of the photograph rather than transforming all of it. Working from uncompressed RAW removes the image decoder from the proof. The C2PA ecosystem [20] provides the signature format; to our knowledge it has no standard pixel-level commitment, which is why proof 1 exists.

Celestial navigation and star trackers. Finding position from the sky is as old as navigation. Classical methods measure star altitudes above the horizon or, in modern automated forms, relative to the direction of gravity [40, 41]; both need a vertical reference, which a phone photograph of the sky lacks and which an attacker could tilt. The Moon's parallax needs no vertical: the reference is the star field itself. Spacecraft optical navigation measures the direction to a nearby body against stars for the same reason [19], and that literature is the source of the centre-of-brightness correction [42, 48]. Amateur photographs have been used to triangulate events on the Moon through parallax [70], the inverse of our problem. The closest precedent for the measurement itself is a student project by Holmes [38], which estimated an observer's position from a single wide-field photograph of the Moon and two identified stars. One method recovered the longitude to within 3°, about 260 km, and a second the latitude to within 1.4° while losing the longitude; the report names the glare of the overexposed Moon as a source of error, which is also what limits our measurement. Calibrating a camera from the star field alone is standard for star trackers [26, 49]; our contribution there is making the calibration *canonical* so that a prover cannot steer it. Image-based geolocation in computer vision has used the Sun and sky appearance [45], at much coarser accuracy and without an adversary.

Recapture detection. Detecting that an image is a photograph of a screen is a studied forensic problem [17, 65]. Those methods are classifiers on image statistics; Section 9.2 explains why we treat such signals as cost-raising but not as a security boundary in this setting.

12 Limitations and future work

Evidence. Everything rests on six accepted frames from two sessions with one phone model, and three further photographs without device provenance. The acceptance rate for an ordinary user is unknown and likely low: the method needs a tripod, a telephoto lens, a RAW exposure long enough to record stars, a Moon between roughly 35% and full illumination at least 25° up, and a sky clear enough that every bright star in the field is visible.

Accuracy. Regions are hundreds of kilometres across. The Moon measurement dominates; the plausible improvements are an albedo-aware centre of brightness (Section 6.6), a capture mode that records the Moon unsaturated within the same frame (a dual-gain or bracketed readout bound in one signature), and longer focal lengths. On the three Maldives frames the Moon residual along the Sun direction (29–52″) is several times what the Moon model predicts; understanding it needs more data at different altitudes and phases. The prover can still choose the magnitude limit, which moves the fix by up to 24 km on our frames; the proof does not yet enforce the faintest limit (Section 6.4).

Coverage of devices. The tolerance law has been calibrated at one plate scale; each lens class needs its own calibration of Eq. 8. The one wide-camera frame we examined in detail suggests what such lenses need: a third radial distortion term (with k_1, k_2 alone, 13 bright stars all fit within a pixel and the fix is still 174 km out), and a rule for foreground objects that hide required stars. Neither is in the proof; both should wait for more than one frame. Only uncompressed DNG is supported.

Capture trust. As analysed in Section 9: time is not attested; nothing ties the image to the sky; signing was emulated. The attested capture application of Section 10 is the next engineering step; the interactive capture outlined in Section 9.3, which addresses more advanced forgeries, is a direction beyond it.

Cost and deployment. About ten minutes of GPU and CPU time is acceptable for occasional claims and too slow for interactive use; hashing the file twice dominates the GPU part and could be reduced, or avoided entirely if the device signed a pixel commitment natively. Constant-size proofs were produced for all nine photographs, but no on-chain verifier has been deployed, and on-device proving has not been attempted.

13 Conclusion

The Moon's place among the stars is a location signal that is free, global and cannot be altered at its source. We have shown that a current phone can read it to within 200 km from a single RAW frame; that the reading can be re-derived inside a zero-knowledge proof from signed pixels with almost no latitude left to the prover; and that such a proof, today, certifies a statement about an image and a device, not yet about the world. The first two results suggest the approach is worth pursuing as one ingredient of proof-of-location systems, especially where no terrestrial infrastructure can be trusted. The third defines the work that remains: an attested time, a capture path that signs what the optics and inertial sensors were doing, and capture protocols that make a forged sky expensive to build.

Acknowledgements

I thank the friends who helped to build the test corpus by photographing the Moon under many different sky conditions. I also thank the members of the Cloudy Nights forum for sharing details

about their photographs.

References

[1] C. Adams, P. Cain, D. Pinkas, and R. Zuccherato. Internet X.509 public key infrastructure time-stamp protocol (TSP). RFC 3161, Aug. 2001. URL https://www.rfc-editor.org/rfc/rfc3161.

[2] Android Developers. Verify hardware-backed key pairs with key attestation. Online documentation, . URL https://developer.android.com/privacy-and-security/security-key-attestation. Visited 2026-11.

[3] Android Developers. CaptureResult (android.hardware.camera2 API reference). Online documentation, . URL https://developer.android.com/reference/android/hardware/camera2/CaptureResult. Visited 2026-11.

[4] Apple Security Engineering and Architecture and Apple Camera & Photos. Apple Reference Image: A new approach for verified photography. Apple Security Research blog, Sept. 2026. URL https://security.apple.com/blog/apple-reference-image/. Visited 2026-11.

[5] B. A. Archinal, M. F. A'Hearn, E. Bowell, A. Conrad, G. J. Consolmagno, R. Courtin, T. Fukushima, D. Hestroffer, J. L. Hilton, G. A. Krasinsky, G. Neumann, J. Oberst, P. K. Seidelmann, P. Stooke, D. J. Tholen, P. C. Thomas, and I. P. Williams. Report of the IAU Working Group on Cartographic Coordinates and Rotational Elements: 2009. *Celestial Mechanics and Dynamical Astronomy*, 109 (2):101–135, 2011. doi: 10.1007/s10569-010-9320-4.

[6] K. Barbary. SEP: Source Extractor as a library. *The Journal of Open Source Software*, 1(6):58, 2016. doi: 10.21105/joss.00058.

[7] Y. V. Batrakov, Y. A. Chernetenko, G. K. Gorel, and L. A. Gudkova. Hipparcos catalogue orientation as obtained from observations of minor planets. *Astronomy and Astrophysics*, 352:703–711, 1999. URL https://ui.adsabs.harvard.edu/abs/1999A&A...352..703B.

[8] E. Ben-Sasson, I. Bentov, Y. Horesh, and M. Riabzev. Scalable, transparent, and post-quantum secure computational integrity. IACR Cryptology ePrint Archive, Paper 2018/046, 2018. URL https://eprint.iacr.org/2018/046.

[9] G. G. Bennett. The calculation of astronomical refraction in marine navigation. *Journal of Navigation*, 35(2):255–259, 1982. doi: 10.1017/S0373463300022037.

[10] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement Series*, 117(2):393–404, 1996. doi: 10.1051/aas:1996164.

[11] N. Bowditch. *The American Practical Navigator: An Epitome of Navigation*. Pub. No. 9. National Geospatial-Intelligence Agency, Springfield, VA, USA, 2024 edition, 2024. URL https://msi.nga.mil/Publications/APN.

[12] S. Brands and D. Chaum. Distance-bounding protocols. In T. Helleseth, editor, *Advances in Cryptology – EUROCRYPT '93*, volume 765 of *Lecture Notes in Computer Science*, pages 344–359. Springer, 1994. doi: 10.1007/3-540-48285-7_30.

[13] D. C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3):444–462, 1966.

[14] J. Brown and K. Stubis. TETRA: Star identification with hash tables. In *Proceedings of the 31st Annual AIAA/USU Conference on Small Satellites*, Logan, UT, USA, 2017. URL https://digitalcommons.usu.edu/smallsat/2017/all2017/124/.

[15] J. Bruestle, P. Gafni, and the RISC Zero Team. RISC Zero zkVM: Scalable, transparent arguments of RISC-V integrity. White paper (draft), RISC Zero, Aug. 2023. URL https://dev.risczero.com/proof-system-in-detail.pdf.

[16] C2PA Viewer. Devices supporting C2PA content credentials. Web page, updated 2026-09-18, 2026. URL https://c2paviewer.com/supported-devices. Visited 2026-11.

[17] H. Cao and A. C. Kot. Identification of recaptured photographs on LCD screens. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1790–1793, 2010. doi: 10.1109/ICASSP.2010.5495419.

[18] S. Čapkun and J.-P. Hubaux. Secure positioning in wireless networks. *IEEE Journal on Selected Areas in Communications*, 24(2):221–232, 2006. doi: 10.1109/JSAC.2005.861380.

[19] J. A. Christian. A tutorial on horizon-based optical navigation and attitude determination with space imaging systems. *IEEE Access*, 9:19819–19853, 2021. doi: 10.1109/ACCESS.2021.3051914.

[20] Coalition for Content Provenance and Authenticity. Content credentials: C2PA technical specification, version 2.4. Specification, Apr. 2026. URL https://spec.c2pa.org/specifications/specifications/2.4/specs/C2PA_Specification.html. Visited 2026-11.

[21] Content Authenticity Initiative. c2pa-rs: Rust SDK for the core C2PA specification. Software repository. URL https://github.com/contentauth/c2pa-rs. Visited 2026-11.

[22] T. Datta, B. Chen, and D. Boneh. VerITAS: Verifying image transformations at scale. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 4606–4623, 2025. doi: 10.1109/SP61157.2025.00097. URL https://eprint.iacr.org/2024/1066. Full version: IACR Cryptology ePrint Archive 2024/1066.

[23] S. Dittmer, S. Lu, K. Model, and J. Near. ZK-JPEG: Zero-knowledge image editing and compression. IACR Cryptology ePrint Archive, Paper 2026/2039, 2026. URL https://eprint.iacr.org/2026/2039. To appear at SCN 2026 according to the ePrint record.

[24] A. Dyer. Dawn conjunction of Venus and Jupiter. The Amazing Sky (web log), Aug. 2014. URL http://amazingsky.net/2014/08/18/dawn-conjunction-of-venus-and-jupiter/. Contains the photograph "Waning Moon in the Hyades near Aldebaran". Visited 2026-11.

[25] S. Dziembowski, S. Ebrahimi, and P. Hassanizadeh. VIMz: Private proofs of image manipulation using folding-based zkSNARKs. *Proceedings on Privacy Enhancing Technologies*, 2025(2):344–362, 2025. doi: 10.56553/popets-2025-0065. IACR Cryptology ePrint Archive 2024/1063.

[26] J. Enright, I. Jovanovic, and B. Vaz. Autonomous recalibration of star trackers. *IEEE Sensors Journal*, 18(18):7708–7720, 2018. doi: 10.1109/JSEN.2018.2857621.

[27] J. Ernstberger, C. Zhang, L. Ciprian, P. Jovanovic, and S. Steinhorst. Zero-knowledge location privacy via accurate floating-point SNARKs. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 3440–3459, 2025. doi: 10.1109/SP61157.2025.00057. URL https://eprint.iacr.org/2024/1842. Full version: IACR Cryptology ePrint Archive 2024/1842; arXiv:2404.14983.

[28] ESA. *The Hipparcos and Tycho Catalogues*. ESA SP-1200. European Space Agency, 1997. URL https://www.cosmos.esa.int/web/hipparcos/catalogues. 17 volumes.

[29] I. Fernández-Hernández, V. Rijmen, G. Seco-Granados, J. Simon, I. Rodríguez, and J. D. Calle. A navigation message authentication proposal for the Galileo open service. *NAVIGATION: Journal of the Institute of Navigation*, 63(1):85–102, 2016. doi: 10.1002/navi.125.

[30] Foamspace Corp. FOAM whitepaper. White paper, 2018. URL https://foam.space/publicAssets/FOAM_Whitepaper.pdf. Visited 2026-11.

[31] S. Garg, A. Jain, Z. Jin, and Y. Zhang. Succinct zero knowledge for floating point computations. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*, pages 1203–1216. ACM, 2022. doi: 10.1145/3548606.3560653.

[32] P. Gill, Y. Ganjali, B. Wong, and D. Lie. Dude, where's that IP? Circumventing measurement-based IP geolocation. In *19th USENIX Security Symposium (USENIX Security 10)*. USENIX Association, 2010. URL https://www.usenix.org/conference/usenixsecurity10/dude-wheres-ip-circumventing-measurement-based-ip-geolocation.

[33] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989. doi: 10.1137/0218012.

[34] Google. How Pixel and Android are bringing a new level of trust to your images with C2PA Content Credentials. Google Online Security Blog, Sept. 2025. URL https://security.googleblog.com/2025/09/pixel-android-trusted-images-c2pa-content-credentials.html. Visited 2026-11.

[35] G. Greiner, T. Mowery, and P. Soni. HyperVerITAS: Verifying image transformations at scale on Boolean hypercubes. *Proceedings on Privacy Enhancing Technologies*, 2026(2):58–75, 2026. doi: 10.56553/popets-2026-0036. IACR Cryptology ePrint Archive 2026/641.

[36] J. Groth. On the size of pairing-based non-interactive arguments. In M. Fischlin and J.-S. Coron, editors, *Advances in Cryptology – EUROCRYPT 2016*, volume 9666 of *Lecture Notes in Computer Science*, pages 305–326. Springer, 2016. doi: 10.1007/978-3-662-49896-5_11.

[37] B. Hapke. Bidirectional reflectance spectroscopy: 3. Correction for macroscopic roughness. *Icarus*, 59(1):41–59, 1984. doi: 10.1016/0019-1035(84)90054-X.

[38] M. Holmes. Global position determination from observed relative position of celestial bodies. Senior project, Aerospace Engineering Department, California Polytechnic State University, San Luis Obispo, June 2010. https://digitalcommons.calpoly.edu/aerosp/13.

[39] T. E. Humphreys, B. M. Ledvina, M. L. Psiaki, B. W. O'Hanlon, and P. M. Kintner, Jr. Assessing the spoofing threat: Development of a portable GPS civilian spoofer. In *Proceedings of the 21st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2008)*, pages 2314–2325, Savannah, GA, USA, Sept. 2008. URL https://www.ion.org/publications/abstract.cfm?articleID=8132.

[40] M. Izadmehr and M. Khakian Ghomi. Design and construction of a portable high resolution positioner using star patterns. *Astrophysics and Space Science*, 364:75, 2019. doi: 10.1007/s10509-019-3565-5.

[41] M. Izadmehr and M. Khakian Ghomi. An accuracy of better than 200 m in positioning using a portable star tracker. *New Astronomy*, 74:101272, 2020. doi: 10.1016/j.newast.2019.04.004.

[42] M. Kaasalainen and P. Tanga. Photocentre offset in ultraprecise astrometry: Implications for barycentre determination and asteroid modelling. *Astronomy and Astrophysics*, 416(1):367–373, 2004. doi: 10.1051/0004-6361:20031711.

[43] D. Kang, T. Hashimoto, I. Stoica, and Y. Sun. ZK-IMG: Attested images via zero-knowledge proofs to fight disinformation. arXiv:2211.04775, 2022.

[44] W. Ladd and M. Dansarie. Roughtime: A protocol for rough time synchronization. RFC 10049 (Experimental), Oct. 2026. URL https://www.rfc-editor.org/rfc/rfc10049.

[45] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What do the sun and the sky tell us about the camera? *International Journal of Computer Vision*, 88(1):24–51, 2010. doi: 10.1007/s11263-009-0291-4.

[46] D. Lang, D. W. Hogg, K. Mierle, M. Blanton, and S. Roweis. Astrometry.net: Blind astrometric calibration of arbitrary astronomical images. *The Astronomical Journal*, 139(5):1782–1800, 2010. doi: 10.1088/0004-6256/139/5/1782.

[47] C. C. Liebe. Accuracy performance of star trackers – a tutorial. *IEEE Transactions on Aerospace and Electronic Systems*, 38(2):587–599, 2002. doi: 10.1109/TAES.2002.1008988.

[48] L. Lindegren. Meridian observations of planets with a photoelectric multislit micrometer. *Astronomy and Astrophysics*, 57:55–72, 1977. URL https://ui.adsabs.harvard.edu/abs/1977A&A....57...55L.

[49] H.-b. Liu, J.-c. Tan, H. Jia, X.-j. Li, J.-q. Wang, and J.-k. Yang. Autonomous on-orbit calibration of a star tracker camera. *Optical Engineering*, 50(2):023604, 2011. doi: 10.1117/1.3542039.

[50] G. Masi. The Moon, Jupiter and Aldebaran: 8 Sept. 2012. The Virtual Telescope Project, Sept. 2012. URL https://www.virtualtelescope.eu/2012/09/08/the-moon-jupiter-and-aldebaran-8-sept-2012/. Visited 2026-11.

[51] R. C. Merkle. A digital signature based on a conventional encryption function. In C. Pomerance, editor, *Advances in Cryptology – CRYPTO '87*, volume 293 of *Lecture Notes in Computer Science*, pages 369–378. Springer, 1988. doi: 10.1007/3-540-48184-2_32.

[52] Natural Earth. Natural Earth: Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales. Public-domain map dataset. URL https://www.naturalearthdata.com/. Visited 2026-11.

[53] A. Naveh and E. Tromer. PhotoProof: Cryptographic image authentication for any set of permissible transformations. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 255–271, 2016. doi: 10.1109/SP.2016.23.

[54] R. S. Park, W. M. Folkner, J. G. Williams, and D. H. Boggs. The JPL planetary and lunar ephemerides DE440 and DE441. *The Astronomical Journal*, 161(3):105, 2021. doi: 10.3847/1538-3881/abd414.

[55] P. Perazzo, M. Mattei, G. Anastasi, M. Avvenuti, G. Dini, G. Lettieri, and C. Vallati. JPEGs just got snipped: Croppable signatures against deepfake images. In *2025 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6, 2025. doi: 10.1109/IJCNN64981.2025.11227387.

[56] B. Rhodes. Skyfield: High precision research-grade positions for planets and Earth satellites generator. Astrophysics Source Code Library, record ascl:1907.024, 2019. URL https://ascl.net/1907.024.

[57] G. Rufino and A. Moccia. Laboratory test system for performance evaluation of advanced star sensors. *Journal of Guidance, Control, and Dynamics*, 25(2):200–208, 2002. doi: 10.2514/2.4888.

[58] Þ. Sæmundsson. Astronomical refraction. *Sky and Telescope*, 72:70, 1986. URL https://ui.adsabs.harvard.edu/abs/1986S&T....72...70S.

[59] S. Saroiu and A. Wolman. Enabling new mobile applications with location proofs. In *Proceedings of the 10th Workshop on Mobile Computing Systems and Applications (HotMobile '09)*, pages 1–6. ACM, 2009. doi: 10.1145/1514411.1514414.

[60] S. Saroiu and A. Wolman. I am a sensor, and I approve this message. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems and Applications (HotMobile '10)*, pages 37–42. ACM, 2010. doi: 10.1145/1734583.1734593.

[61] N. Sastry, U. Shankar, and D. Wagner. Secure verification of location claims. In *Proceedings of the 2nd ACM Workshop on Wireless Security (WiSe '03)*, pages 1–10. ACM, 2003. doi: 10.1145/941311.941313.

[62] R. C. Stone. An accurate method for computing atmospheric refraction. *Publications of the Astronomical Society of the Pacific*, 108:1051–1058, 1996. doi: 10.1086/133831.

[63] Succinct Labs. SP1: A zero-knowledge virtual machine that proves the correct execution of programs compiled for the RISC-V architecture. Software repository and documentation, . URL https://github.com/succinctlabs/sp1. Documentation: https://docs.succinct.xyz/docs/sp1/introduction. Visited 2026-11.

[64] Succinct Labs. SP1 documentation: Security model. https://docs.succinct.xyz/docs/sp1/security/security-model, . Accessed November 2026.

[65] T. Thongkamwitoon, H. Muammar, and P.-L. Dragotti. An image recapture detection algorithm based on learning dictionaries of edge profiles. *IEEE Transactions on Information Forensics and Security*, 10(5):953–968, 2015. doi: 10.1109/TIFS.2015.2392566.

[66] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, and S. Čapkun. On the requirements for successful GPS spoofing attacks. In *Proceedings of the 18th ACM Conference on Computer and Communications Security (CCS '11)*, pages 75–86. ACM, 2011. doi: 10.1145/2046707.2046719.

[67] F. van Leeuwen. Validation of the new Hipparcos reduction. *Astronomy and Astrophysics*, 474(2):653–664, 2007. doi: 10.1051/0004-6361:20078357.

[68] F. Zafar, A. Khan, A. Anjum, C. Maple, and M. A. Shah. Location proof systems for smart Internet of Things: Requirements, taxonomy, and comparative analysis. *Electronics*, 9(11):1776, 2020. doi: 10.3390/electronics9111776.

[69] Z. Zhu and G. Cao. APPLAUS: A privacy-preserving location proof updating system for location-based services. In *2011 Proceedings IEEE INFOCOM*, pages 1889–1897, 2011. doi: 10.1109/INFCOM.2011.5934991.

[70] J. I. Zuluaga, M. Tangmatitham, P. Cuartas-Restrepo, J. Ospina, F. Pichardo, S. A. López, K. Peña, and J. M. Gaviria-Posada. Location, orbit, and energy of a meteoroid impacting the Moon during the lunar eclipse of 2019 January 21. *Monthly Notices of the Royal Astronomical Society*, 492(1): 1432–1449, 2020. doi: 10.1093/mnras/stz3531.

A Policy parameters

Table 10 lists the public policy. Its digest is part of every claim; changing any value changes the claim. This is the fifth version of the policy, and all proofs reported in the paper were made with it. It differs from the version used during most of the development in five rules: the tolerance floor (32″, was 25″), the residual at the witnessed position (new), the observer height (was up to 9 km), the epoch skew (was up to 1 s) and the star-peak tie rule (new).

Parameter	**Value**	**Basis**
Moon tolerance floor	32″	star-fit noise at 10 stars; refraction at 25°; choice of magnitude limit
Moon tolerance, systematic	SD $(0.053 + 0.009\alpha)$	simulated envelope, 2025–2035
Star tolerance $\varepsilon_{\text{star}}$	1 px	all kept stars on real frames < 1 px
Least-squares step	0.001 px	removes attitude slack
Minimum stars; bright stars	10; 2 with $V \leq 4$	subset-attack measurements
Quadrants around the Moon	all four	no extrapolated attitude
Radial spread of stars	≥ 300 px	lever arm for rotation
Moon off-axis	≤ 0.25 half-diagonal	lens model interpolated at the Moon
Moon altitude	$\geq 25°$	stands in for a refraction model
Phase angle	$\leq 107.5°$	validity of the linear phase law
Focal length; principal point	1 000–10 000 px; ±5%	sanity bounds only
Star exclusion: edge; Moon; crowding	4 px; $7r$; 6 px, 2.5 mag	estimator window; glare; blending
Star window; threshold	7×7; $B + 2\sigma$	grid search on real frames
Star peak	first 3×3 maximum	in reading order; one window per star
Moon aperture; sky annulus	$3r$; 1.6–2.2 apertures	as in the solver
Moon fixed-point tolerance	0.01 px	contraction factor 0.18–0.30
Plateau share for the linear phase law	≤ 0.6	RAW frames measure 0.48–0.59; provisional
Uncertainty shape	32 rays, +3%, +10 km	measured contour bulge $\leq 0.97\%$
Residual at the witnessed position	$\leq 1''$	the fix is the centre of the consistent set
Observer height	−0.5 to 4 km	the prover chooses it; covered by the margin
Epoch skew	0	the ephemeris epoch is the capture time

Table 10: Public policy of the geometry proof.

B Contents of the claim

Field	Bytes	Meaning
binding-program list digest	32	set of trusted proof-1 programs; which one was used is hidden
trusted-root list digest	32	set of trusted signing roots; which one signed is hidden
capture time (ms)	8	mid-exposure time from the signed file
region digest	32	the claimed region S
ephemeris digest	32	Moon, Earth, Sun state and Earth orientation at that time
policy digest	32	Table 10
star list digest	32	Hipparcos, $V \leq 6.5$
number of stars	4	stars in the complete set
total	204	

Table 11: Public output of the final proof.