VerITAS: Verifying Image Transformations at Scale

Trisha Datta, Binyi Chen, Dan Boneh
Stanford University

***Abstract*—Verifying image provenance has become an important topic, especially in the realm of news media. To address this issue, the Coalition for Content Provenance and Authenticity (C2PA) developed a standard to verify image provenance that relies on digital signatures produced by cameras. However, photos are usually edited before being published, and a signature on an original photo cannot be verified given only the published edited image. In this work, we describe VerITAS, a system that uses zero-knowledge proofs (zk-SNARKs) to prove that only certain edits have been applied to a signed photo. While past work has created image editing proofs for photos, VerITAS is the first to do so for realistically large images (30 megapixels). Our key innovation enabling this leap is the design of a new proof system that enables proving knowledge of a valid signature on a large amount of witness data. We run experiments on realistically large images that are more than an order of magnitude larger than those tested in prior work. In the case of a computationally weak signer, such as a camera, we are able to generate proofs of valid edits for a 90 MB image in under an hour, costing about $2.42 on AWS per image. In the case of a more powerful signer, we are able to generate proofs of valid edits for 90 MB images in under five minutes, costing about $0.09 on AWS per image. Either way, proof verification time is about 2 seconds in the browser. Our techniques apply broadly whenever there is a need to prove that an efficient transformation was applied correctly to a large amount of signed private data.**

1. Introduction

Verifying where and when a digital image was taken has become increasingly difficult. After Russia invaded Ukraine in February 2022, several photographs and videos [3], [4], [5] circulated online that falsely claimed to show the conflict. In one instance, a BBC program showed footage of what was supposedly the Russian invasion of Ukraine, but was actually footage of a Russian military parade rehearsal [6]. The fact that even reputable news organizations like the BBC can make these mistakes demonstrates that there is much room for improvement in current image provenance verification processes.

The Coalition for Content Provenance and Authenticity (C2PA) has developed a standard [7] to verify image provenance that relies on digital signatures. This standard proposes that cameras digitally sign every photo they take along with the photo's metadata (e.g., location, timestamp, focal length, exposure time, etc.). Leica, Sony, and Nikon have all developed cameras with such signing capabilities [8], [9]. Leica has even developed an on-camera trusted execution environment (TEE) to protect the signing key. More recently, AI companies, such as OpenAI, have also begun issuing C2PA attestations on images that they generate to ensure that they are not falsely blamed for content that they did not generate. In Section 3 we discuss the threat model that the C2PA is designed to address.

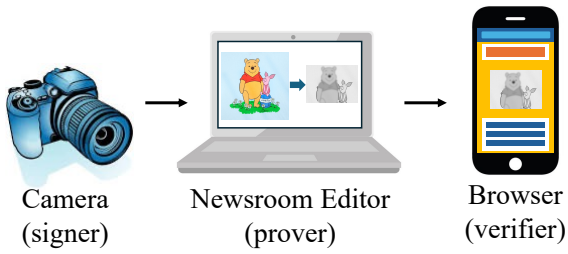

Figure 1: Image editing pipeline

Users could in theory verify the provenance of a photo in a news article by verifying the accompanying C2PA signature. However, photos are rarely published as is. Before being posted in a news story, they are often cropped, some faces and objects may be blurred to protect privacy, images are resized to save bandwidth, and in some cases they are converted to grayscale. The *Associated Press* published a list of acceptable edits [10] that do not fundamentally alter the content of the photo. See Figure 1 for an overview of the image editing pipeline.

Publishing edited photos presents a problem because the C2PA signature on the original image cannot be verified given only the edited image. To address this, the C2PA proposes that all edits be performed by a C2PA-enabled (and approved) editing application that maintains a secret signing key and signs the processed photo. These application-generated signatures will then be verified by the reader to validate the metadata of the photo. A major problem with this approach is that it changes the trust model and breaks end-to-end security. The end user must now trust the editing application and the security of its signing key. Moreover, it is not at all clear how open-source photo editing tools will be used in this context. These tools typically have no way to protect a signing key.

We therefore need a method for editing a signed photo

* This paper is the full version of our work presented in [1] and [2].

such that a news consumer who only has the published edited photo can be assured that (i) the original unedited photo was properly signed by a C2PA camera, (ii) only permissible edits, such as cropping, blurring, resizing, and grayscale, were made to the signed photo, and (iii) the metadata of the edited photo is equal to the metadata of the original photo. The scheme should preserve end-to-end security, from the camera to the user's screen, without requiring the user to trust some editing software, the article's publisher, or a third-party fact-checker. We call this property glass-to-glass security.

Our contributions. In this paper, we present VerITAS (Verifying Image Transformations At Scale), a system that uses succinct zero-knowledge arguments (zk-SNARKs) [11] to prove the provenance of edited photos. A zero-knowledge proof is a statement about a secret witness that can be verified by anyone without revealing anything about the witness other than the validity of the statement. These proofs are *complete*, meaning that verification will succeed for honestly generated proofs, and *knowledge sound*, meaning that verification will fail if the prover does not have a valid witness. These properties entail that the verifier need not trust the prover, which solves the trust problem posed by the C2PA protocol. Moreover, these proofs are *zero-knowledge*, meaning that the proof reveals nothing about elements that were removed from the original photo; this is vital when sensitive information is cropped or blurred. VerITAS uses succinct zero-knowledge arguments to enable the editor to make modifications to a captured C2PA image, and replace the signature with a zk-SNARK that the edited image was derived from a properly signed C2PA image via an authorized transformation. The resulting proofs are *succinct*, meaning that they are "short" and "fast" to verify. Proof verification can be done in the reader's browser, which could automatically detect and verify these proofs in news articles.

In 2016, Naveh and Tromer [12] implemented PhotoProof, a system for producing zero-knowledge proofs for various photo edits. While this work demonstrated the feasibility of creating zk-SNARKs for image edits, their proving time was too large to be practical. In concurrent work with ours, Kang et al. [13] developed a library that achieved a 100x speed-up over PhotoProof. However, the largest photo used in their experiments is 720p, or 900 kilopixels (KP). The pictures produced by the Sony and Leica cameras mentioned above are about 33 megapixels (MP), which is an order of magnitude larger than any photo used in previous work. VerITAS is the first system to produce ZK proofs of image edits for photos on the order of 30 MP or more.

We describe VerITAS as a protocol between a prover (a newsroom editor) and a verifier (a news consumer). Given a public edited image x (e.g., a photo in a news article) and a public editing function f, a prover convinces a verifier of the provenance of the published photo by proving that it knows a secret witness that comprises a photo w and a signature σ, such that (i) σ is a valid signature on w under a public verification key vk, and (ii) applying f to the witness photo w results in the public photo x. In other words, we need a zk-SNARK for the following instance-witness relation:

$$\mathcal{R} := \big\{ \big((\mathsf{vk}, f, x) \; ; \; (w, \sigma) \big) : \\ f(w) = x \quad \wedge \quad \text{SigVerify}(\mathsf{vk}, w, \sigma) = 1 \big\} \tag{1}$$

The witness (w, σ) provided as input to the zero-knowledge prover contains the original 30 megapixel (MP) image, which is about 90 MB, along with a signature on this image. Hence, in our settings, the prover must build a proof using an unusually large witness.

The main bottleneck in systems that use zk-SNARKs to prove simple photo edits is building a zk-SNARK for a circuit that verifies that the original image is properly signed. The difficult step is the first step of signature verification: proving that a 90 MB witness was hashed correctly with a collision resistant hash. Doing so using SHA256 is vastly inefficient because SHA256 employs many non-linear operations, which are expensive inside of a zk-SNARK circuit. Even proving knowledge of SNARK-friendly hashes like Poseidon, which is what Kang et al. [13] use, is too costly for hashing a 30 MP photo in a SNARK circuit.

Efficiently proving knowledge of a signature. VerITAS solves this problem by introducing two modes for proving knowledge of (w, σ) such that σ is a valid signature on w. One mode is designed for a computationally-limited signer (such as a camera); the other mode is designed for a more powerful signer (such an OpenAI). The former has a lightweight signing procedure but slower editing proof generation time. The latter has a more heavyweight signing procedure but a much faster editing proof generation time.

Mode 1. To accommodate a computationally limited signer (such as a camera), the VerITAS C2PA signer hashes the captured image using a lattice-based collision resistant hash to obtain a 4 KB digest. It then hashes that digest down to 32 bytes using Poseidon and signs the resulting hash value using its secret key. The benefit of the lattice hash is that it uses only linear operations over a finite field. Even for a large amount of data (such as a 90MB photo), this lattice hash is far more amenable to being proved in a SNARK circuit than other collision resistant hash algorithms. We design a custom SNARK to show that a lattice hash has been computed correctly (see Section 5). We emphasize that no prior work has been able to create hashing proofs for 30 MP images and has thus been limited to proving edits on photos an order of magnitude smaller than we are able to support.

To use our scheme, the C2PA camera would use our hash function to hash the captured image and then sign the computed hash using a standard signature scheme such as ECDSA. When editing the image, the newsroom would produce a proof that the original image is signed correctly by verifying the hash in the zk-SNARK circuit. This design may be of independent interest for anyone looking to create proofs for SNARK circuits that hash a large amount of data.

Mode 2. When the C2PA signer is computationally powerful (e.g., OpenAI), VerITAS uses a more heavyweight signing

algorithm. Here the C2PA signer first computes a polynomial commitment to the captured photo and then signs the short polynomial commitment using, say, ECDSA. Computing a polynomial commitment takes more time and memory than computing a simple hash (we give detailed numbers in the evaluation section). The benefit is that now VerITAS can greatly reduce the time to generate a proof of valid edit. In particular, we modify how the zk-SNARK prover operates so that now the SNARK circuit only needs to verify the photo edits, *but does not need to verify the signature or the polynomial commitment*. Hence, by modifying the underlying proof system we obtain a massive savings for the newsroom editor when generating a proof of a valid edit. The details are provided in Section 4.1.

Implementation. We split our implementation of VerITAS into two parts: proving knowledge of a valid signature on the original photo and proving a valid edit of the original photo. We implement the former in Rust and the latter with Plonky2 [14]. We report results for proof generation time, proof verification time, and peak memory usage. Our experiments show that for a 30 MP photo, proving knowledge of a signature using our Lattice-Poseidon hash is much faster compared to using a Poseidon hash alone. In fact, our machine could not prove knowledge of a Poseidon hash of even a 1 MP photo. If a photo is signed by a more powerful signer, as in mode 2, then an editor avoids proving knowledge of a signature altogether, and can just prove validity of the image edits, which takes less than five minutes. However, for the signer, computing a polynomial commitment of a 30 MP photo requires more memory and computational power than a camera might have, which means that cameras will likely use the lattice hash method (mode 1).

To summarize, our contributions are threefold:

- VerITAS is the first system, to our knowledge, that can produce editing proofs for 30 MP signed images (all other work has been limited to proving edits on photos over an order of magnitude smaller);
- A custom proof system for computationally weak signers that can prove that the hash of a very large amount of witness data was computed correctly;
- A custom proof system for more powerful signers that enables editors to produce editing proofs without verifying a signature on the witness in the SNARK circuit. This greatly reduces the time to produce an editing proof.

Our techniques apply more broadly than images. They apply whenever there is a need to prove that an efficient transformation was applied correctly to a large amount of signed witness data. Some examples include signed financial or health records. However, our focus in this paper in on transformations applied to signed images.

Alternate designs. While VerITAS uses zk-SNARKs to support editing signed images, a very different approach is to use redactable signatures [15], [16], or more generally, homomorphic signatures [17], [18]. Homomorphic signatures enable anyone to transform a message-signature pair (m, σ) into another message-signature pair $\big(\big(f(m), f\big), \sigma'\big)$, where σ' is a valid signature on $\big(f(m), f\big)$. In other words, σ' is a signature on the transformed message m, and a description of the transformation function f. When f is a simple redaction operation, such as cropping, this can be implemented very efficiently using redactable signatures. However, more complicated transformations, such as blurring and resizing algorithms in image processing packages, cannot be reduced to redaction. For example, VerITAS provides a proof of correct resizing using the bilinear resizing algorithm [19], a standard resizing method in Adobe Photoshop, which uses linear transformations and cannot be reduced to redaction. Many other standard edits, such as brightness, contrast adjustments, tinting, dodging, and burning, can be proven in zero knowledge as in VerITAS, but cannot be done using redactable signatures. One could try to use homomorphic signatures [17], [18], but for these image transformations, the best homomorphic signatures that do not rely on SNARKs are impractical. We also note that for resizing, the camera does not know the resizing dimensions ahead of time and therefore cannot simply pre-sign a resized image.

2. Preliminaries

We use $[n]$ to denote the set $\{0, \ldots, n-1\}$ and use $[a, b]$ to denote the set $\{a, \ldots, b\}$. We use $\mathbb{F}_q$ to denote a finite field of size q. Let $r \leftarrow\!\$\, S$ denote drawing a random value from the finite set S. We let ω denote a primitive n^{th} root of unity in $\mathbb{F}$, so that the set $\Omega := \{1, \omega, ..., \omega^{n-1}\}$ has size n. We use $Z_\Omega \in \mathbb{F}[X]$ to denote the vanishing polynomial on Ω. This Z_Ω is the lowest-degree polynomial such that $Z_\Omega(x) = 0$ for all x in Ω. It has the form $Z_\Omega(X) = X^n - 1$, which can be evaluated using at most $2\log_2 n$ field multiplications. We use $\mathbb{F}^{<d}[X]$ to denote the set of all univariate polynomials of degree less than d over the field $\mathbb{F}$.

We use bold-faced lowercase letters for vectors. For a vector $\mathbf{v} \in \mathbb{F}^m$, we denote the elements of $\mathbf{v}$ as $(v_0, ..., v_{m-1})$. We write the concatenation of two vectors as $\mathbf{v}\|\mathbf{w}$. We denote matrices with bold-faced capital letters (e.g., $\mathbf{A} \in \mathbb{F}^{n\times m}$). We denote the rows of a matrix $\mathbf{A} \in \mathbb{F}^{n\times m}$ as $\mathbf{a}_0, ..., \mathbf{a}_{n-1} \in \mathbb{F}^m$. We denote element j of row i in matrix $\mathbf{A}$ as $a_{i,j}$ (e.g., the second element in the topmost row of $\mathbf{A}$ is $a_{0,1}$). We assume access to a hash function $\mathsf{H} : \mathbb{F}^* \to \mathbb{F}$ that can take as input any (finite) number of field elements as input.

2.1. Digital Signatures

A digital signature scheme $\mathcal{S}$ is a triple of efficient algorithms $(\mathsf{KGen}, \mathsf{Sign}, \mathsf{Vf})$ such that:

- $\mathsf{KGen}(1^\lambda) \to (\mathsf{sk}, \mathsf{vk})$, where sk is the secret signing key and vk is the public verification key.
- $\mathsf{Sign}(\mathsf{sk}, m) \to \sigma$, where σ is a signature on message m.
- $\mathsf{Vf}(\mathsf{vk}, m, \sigma) \to 0/1$, where 0 implies rejection and 1 implies acceptance.

We say that a signature scheme is secure if it is existentially unforgeable under a chosen message attack [20] (see Appendix A.1 for the definition). Digital signatures in practice are implemented as a two step process: first hash the data using a collision-resistant hash and then sign the hash.

2.2. Commitment Schemes

A commitment scheme enables a party to commit to a value $x \in \mathcal{X}$ by producing a commitment string com. The commitment should be hiding and binding (see Appendix A.2 for definitions). More precisely, a commitment scheme $C = (\mathsf{setup}, \mathsf{commit})$ is a pair of PPT algorithms:

- $\mathsf{setup}(1^\lambda) \to \mathsf{pp}$, where pp are public parameters for the scheme
- $\mathsf{commit}(\mathsf{pp}, x, r) \to \mathsf{com}$, where com is a commitment to a message $x \in \mathcal{X}$ with randomness $r \in \mathcal{R}_\mathrm{C}$

To open the commitment com, the committer reveals x and r and the verifier accepts if $\mathsf{commit}(x, r) = \mathsf{com}$. In some cases the setup algorithm is trivial in which case we say that the commitment scheme is just the algorithm $\mathsf{commit}(x, r) \to \mathsf{com}$.

Polynomial commitments. A polynomial commitment scheme [21] lets a prover commit to a polynomial $f \in \mathbb{F}[X]$ of bounded degree d. Additionally, the committer can provide an evaluation proof for the committed polynomial at any point $x \in \mathbb{F}$. More precisely, a polynomial commitment scheme C is a tuple of four efficient algorithms $C = (\mathsf{setup}, \mathsf{commit}, \mathsf{open}, \mathsf{Vf})$ such that:

- $\mathsf{setup}(1^\lambda, d) \to \mathsf{pp}$, where pp are public parameters to commit to a polynomial of degree at most d.
- $\mathsf{commit}(\mathsf{pp}, f, r) \to \mathsf{com}$, where com is a commitment to a polynomial $f \in \mathbb{F}[X]$ of degree at most d using randomness $r \in \mathcal{R}_\mathrm{C}$.
- $\mathsf{open}(\mathsf{pp}, f, x, r) \to (\pi, y)$, where π is an opening proof that proves that $f(x) = y$.
- $\mathsf{Vf}(\mathsf{pp}, \mathsf{com}, x, y, \pi) \to 0/1$, where 0 implies rejection and 1 implies acceptance.

A polynomial commitment scheme must be correct, evalution binding, and optionally hiding. We defer these definitions to Appendix A.2. The KZG polynomial commitment scheme [21] is built from pairings. Another polynomial commitment scheme is built from the Fast Reed Solomon IOP of Proximity (FRI IOPP) protocol [22] using a collision resistant hash function.

2.3. zk-SNARKs: Zero-Knowledge Succinct Arguments of Knowledge

Zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs) are efficiently verifiable statements about a secret witness. A zk-SNARK Π is a tuple of efficient algorithms $\Pi = (\mathsf{setup}, \mathsf{prove}, \mathsf{Vf})$ such that, for a given instance-witness relation $\mathcal{R}$ and $(x, w) \in \mathcal{R}$:

- $\mathsf{setup}(1^\lambda) \to \mathsf{pp}$, where pp are public parameters.
- $\mathsf{prove}(\mathsf{pp}, x, w) \to \pi$, where, on input instance x and witness w, the proof π shows that $(x, w) \in \mathcal{R}$.
- $\mathsf{Vf}(\mathsf{pp}, x, \pi) \to 0/1$, where 0 implies rejection and 1 implies acceptance.

The zk-SNARK must be complete, knowledge sound, zero-knowledge, non-interactive, and succinct (see Appendix A.3 for definitions).

zk-SNARKs can be designed to prove a specific relation, or prove a general $\mathcal{NP}$ relation (e.g., Groth16 [23], PLONK [24]). We use both types in our system: we design a custom zk-SNARK for our lattice hash and use a general-purpose zk-SNARK for proving Poseidon hashes and photo transformations. Plonky2 [14] is a system that allows users to write constraints in the form of circuits that can then be proven and verified by the PLONK [24] system. We utilize PLONK to produce proofs for photo editing. For our scheme involving a computationally powerful signer, we make non-black-box use of PLONK.

2.3.1. Lookup Table Arguments. A lookup argument is a relation-specific zk-SNARK. Given a table $T \in \mathbb{F}^t$, a prover can use a lookup argument to show that all elements of some (committed) vector $\mathbf{v} \in \mathbb{F}^m$ are contained in T. Plookup [25], Baloo [26], cq [27], and Lasso [28] are state-of-the-art lookup arguments. Baloo and cq's prover complexity is sublinear in the table size t. However, compared to Plookup, their prover time grows faster in m (the dimension of $\mathbf{v}$). In our use of lookup tables we have $t \ll m$, and we therefore use a Plookup-based approach.

2.3.2. The Schwartz-Zippel Lemma. Let f be a non-zero n-variate polynomial over a finite field $\mathbb{F}$, where the total degree of f is d. The Schwartz-Zippel Lemma [29], [30] says that for random elements $\alpha_1, \ldots \alpha_\ell \leftarrow\!\$\, \mathbb{F}$ we have

$$Pr\big[f(\alpha_1, \ldots, \alpha_\ell) = 0\big] \leq d/|\mathbb{F}|$$

We will use the this lemma to prove equality of polynomials.

2.3.3. Fiat-Shamir Transform. A public-coin interactive protocol can be made non-interactive using the Fiat-Shamir transform [31], which replaces verifier challenges with hashes of the transcript up until that point. For a protocol that has special soundness, applying the Fiat-Shamir transform retains its soundness properties [32].

2.3.4. PLONK. PLONK [24] is a zk-SNARK for proving correct evaluation of an arithmetic circuit. Figure 2 shows the (simplified) PLONK system design. For our photo editing proof that requires a computationally powerful signer, we modify the permutation argument in PLONK.

Let us briefly describe PLONK and its permutation argument. To prove correct evaluation of an arithmetic circuit we first build a table representing the computation trace (the left hand side of Figure 2). Every row of the trace corresponds to a single gate in the circuit. Each gate has two input wires $\mathsf{In}_0, \mathsf{In}_1$, one output wire Out, and an associated operation Op (here, either addition or multiplication). We require that $\mathsf{Op}(\mathsf{In}_0, \mathsf{In}_1) = \mathsf{Out}$ for all gates. This is called

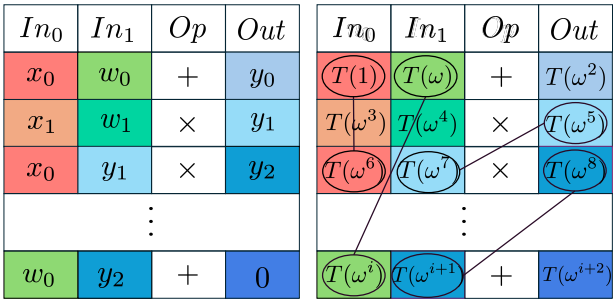

Figure 2: This figure illustrates how a circuit trace (left) is encoded as a polynomial $T(x)$ (right). The circled pairs on the right represent copy constraints (there are copy constraints between all cells of the same color); the PLONK prover must prove that the edge-connected cells have the same value and that the values of cells in each row satisfy the gate constraint. Recall that ω is a primitive n-th root of unity in $\mathbb{F}$.

the *gate constraint*. Additionally, some input/output wires over different gates may be required to share the same values. This captures the wiring structure of the circuit and is called a *copy constraint*. For example, in the left hand side of Figure 2, the cells with same color must have identical values. Copy constraints are determined by the structure of the circuit and not by the wire values assigned by the prover. We categorize the set of wire values into three types: the public instance $\mathbf{x}$, the secret witness $\mathbf{w}$, and the internal wires $\mathbf{y}$. For example, in our application, $\mathbf{x}$ is the published edited photo, $\mathbf{w}$ is the original image being signed (e.g. by the camera), and $\mathbf{y}$ is the intermediate values computed during the transformation from the original image $\mathbf{w}$ to the published image $\mathbf{x}$. In Figure 2, the public instance is $\mathbf{x} := (x_0, x_1)$, the secret witness is $\mathbf{w} := (w_0, w_1)$, and the internal wires are $\mathbf{y} := (y_0, y_1, y_2)$. Given a public instance $\mathbf{x}$, the prover needs to prove that it knows $\mathbf{w}$ and $\mathbf{y}$ that satisfy all the gate constraints and copy constraints.

PLONK uses a permutation argument to prove copy constraints. Let $T(x)$ be the polynomial that interpolates the trace values, i.e., for gate number i we have

$$T(\omega^{3i}) = \mathsf{In}_{i,0}, \quad T(\omega^{3i+1}) = \mathsf{In}_{i,1}, \quad T(\omega^{3i+2}) = \mathsf{Out}_i,$$

where $(\mathsf{In}_{i,0}, \mathsf{In}_{i,1}, \mathsf{Out}_i)$ are the wire values for gate i. Let $\tau : \Omega \to \Omega$ be a permutation such that for every copy constraint $T(s_1) = T(s_2) = \ldots = T(s_\ell)$ where $s_1, \ldots, s_\ell \in \Omega$, we have

$$\tau(s_1) = s_2, \quad \tau(s_2) = s_3, \quad \ldots, \quad \tau(s_\ell) = s_1.$$

We can represent τ as a polynomial of degree $n = |\Omega|$. It is clear that the copy constraints are satisfied if and only if $T(s) = T(\tau(s))$ for all $s \in \Omega$. The public statement in the PLONK permutation argument is a commitment to the polynomials T and τ. The argument proves that $T(s) = T(\tau(s))$ for all $s \in \Omega$. Moreover, the permutation argument supports proving permutation relations across multiple polynomials, say T_1 and T_2. This will be important in our photo editing proofs, as explained in Section 4.3.

2.4. Short Integer Solution (SIS) and Lattice Hash

The Short Integer Solution (SIS) problem [33] is defined as follows. Fix some parameters $n, m, q, b \in \mathbb{N}$ where $n < m$ and q is a prime. An instance of the problem is specified by a random matrix $\mathbf{A} \leftarrow\$ \mathbb{F}_q^{n \times m}$. To solve the given SIS instance, the adversary must find a non-zero vector $\mathbf{v} \in \mathbb{Z}^m$ such that $\mathbf{A}\mathbf{v} = 0 \pmod q$ and $\|\mathbf{v}\|_\infty \leq b$ (i.e., $\mathbf{v}$ is short). For a sufficiently large $n \in \mathbb{N}$, solving SIS is conjectured to be hard for any choice of $m, q, b \in \mathbb{N}$ whenever $q > b \cdot \text{poly}(n)$ and $m > n \log_2 q$.

We next describe a hash function whose collision resistance follows from the hardness of SIS [33], [34]. We represent the data to be hashed as a low-norm vector $\mathbf{v} \in \mathbb{Z}_q^m$. For a random matrix $\mathbf{A} \in \mathbb{F}_q^{n \times m}$, the hash function is defined as

$$H_{\mathbf{A}}(\mathbf{v}) := \mathbf{A}\mathbf{v} \pmod q \tag{2}$$

To see why this function is collision-resistant suppose towards a contradiction that there is an adversary $\mathcal{A}(\mathbf{A})$ that can find a collision for $H_{\mathbf{A}}$, when $\mathbf{A}$ is sampled as $\mathbf{A} \leftarrow\$ \mathbb{F}_q^{n \times m}$. Then $\mathcal{A}(\mathbf{A})$ will output low-norm distinct vectors $\mathbf{v}$ and $\mathbf{v}'$ in $\mathbb{Z}^m$, such that $\mathbf{A}\mathbf{v} = \mathbf{A}\mathbf{v}'$ in $\mathbb{F}_q$. But then $\mathbf{A}(\mathbf{v} - \mathbf{v}') = 0$, and since $\mathbf{v}$ and $\mathbf{v}'$ are low-norm, so is their difference. Hence, $\mathcal{A}$ can solve SIS. We stress that this shows that (2) is collision resistant only when $\mathbf{v}$ is low-norm.

3. Threat Model

In the C2PA setting, every camera is equipped with an embedded certified signing key for a secure signature scheme. The secret signing key is generated on camera and certified by a C2PA certificate authority at manufacturing time. Every time a camera takes a photo, it signs the raw RGB values of the photo's pixels and relevant metadata (e.g., location, timestamp, exposure time). We assume that the adversary has access to the camera and the camera's public key. However, the C2PA assumes that the attacker cannot extract the signing key from the camera, nor can the attacker cause the camera to sign an image that was not captured by the camera's optical hardware. In our settings, the only root of trust is the camera and its signing key. The editor of a photo is not trusted in any way. The verifier wants to ensure that the received edited photo is the result of applying an acceptable transformation [10] to a C2PA signed image.

Non threats. While C2PA is an important step towards image provenance, it is by no means a complete solution and must be combined with other defenses. Specific attacks on the C2PA are out of scope for this paper because our primary focus is on securing the image editing pipeline (Figure 1). Nevertheless, for completeness, we describe a few potential attacks on C2PA and how they might be addressed. These attacks are considered in the C2PA documentation.

First, the adversary might extract the C2PA signing key from some deployed camera. The Leica camera implements

a hardware trusted execution environment (TEE) to protect the key and make extraction harder (but not impossible). Moreover, if a key is extracted, the standard includes a revocation mechanism that alerts all verifiers to revoke a compromised C2PA certificate.

Second, an attacker might try a picture-of-picture attack: it displays an AI-generated picture on a laptop screen and takes a picture of the screen using a C2PA camera. The result is a properly signed image of a fake event. This is a known challenge for the C2PA. One way to defend against this is to require the verifying client to run a picture-of-picture detector. For example, the focal length in the signed image metadata will be that of a camera taking a picture of a screen, and that is likely to be very different from the focal length needed to take a picture of the real-world portrayed event. Several other detection strategies have been suggested by C2PA, but this may turn into a cat-and-mouse game.

Third, the list of allowed transformations by the *Associated Press* may change the semantic meaning of the image. For instance, if presented with a photo of Alice and Bob, an adversary could crop out Alice and claim that Bob was alone when the photo was taken.

Fourth, C2PA may pose a privacy risk in that the signature on a photo can identify the camera that took it. This can be mitigated by having the camera sign photos using a group signature [35], [36]. We discuss this further in Section 9.

As explained above, these attacks are out of scope for this paper. Here we operate within the threat model that C2PA is designed to defend against — which excludes the attacks mentioned above — and focus on securing the editing pipeline.

4. The Design of VerITAS

We present VerITAS as an interaction between a news organization (prover) and a web browser (verifier). Every photo displayed in a news article should be accompanied by its metadata (location, timestamp, focal length, etc.), a description of the edits that were made to the original photo, and a succinct zero knowledge proof. The public statement consists of the published edited photo (x), the edits performed to the original photo (f), and the camera's public key (vk). The secret witness is the original photo (w) and the camera's signature (σ) on the original photo w. Recall that, abtractly, our goal is to design an efficient proof system for the instance-witness relation

$$\mathcal{R} := \big\{ \big((\mathsf{vk}, f, x) \ ; \ (w, \sigma) \big) \ : \ f(w) = x \quad \wedge \quad \mathsf{Vf}(\mathsf{vk}, w, \sigma) = 1 \big\} \tag{3}$$

While (3) places vk in the public statement, we could protect the photographer's identity by moving vk and its certificate to the secret witness. To simplify the presentation we will use the relation (3) and discuss the more private variant in Section 9.

A web browser will only accept a photo's provenance if the photo x is accompanied by a triple (π, vk, f) where π is a valid ZK proof for $\mathcal{R}$, vk is a properly certified C2PA verification key, and the function f, which encodes the list of edits, is "acceptable," as defined by the Associated Press. We can thus think of VerITAS as enforcing a whitelist of allowable edits.

VerITAS relies on all the properties of a zk-SNARK. Completeness and knowledge-soundness of the zk-SNARK mean that the web browser does not need to trust the editor, preserving end-to-end security of the signature. Non-interactivity means that the news organization does not need to interact with any web browser and can instead publish a single proof along with the news article that any verifier can check. Zero-knowledge ensures that the proof does not reveal information that was cropped or blurred in the original photo. Succinctness ensures that the proof can be quickly verified by the browser within a few seconds.

We next turn to designing a proof system for the relation $\mathcal{R}$ from (3). The proof has two parts:

- First, a proof that $f(w) = x$, for an image transformation function f. We come back to building such a proof in Section 6.2.
- Second, a proof that σ is a valid signature on w. This is more complicated, and we present our approach in Section 4.1.

We also need to ensure that the secret witness w used to generate both proofs are identical. We discuss how to achieve this in Section 6.1.2.

4.1. Proving Knowledge of a Valid Signature

When verifying a signature σ on some data w, the verifier: (i) computes $h \leftarrow \mathsf{H}(w)$, where H is a collision resistant hash function, and (ii) verifies that σ is a valid signature on h. When the data being verified is large, as in the case of a photo, most of the time is spent on computing the hash h in step (i). The same is true when proving knowledge of a valid signature. The challenge is to design an efficient SNARK circuit that can verify that the hash h of a large amount of data w is computed correctly. Once the verifier has a valid hash h, proving knowledge of an ECDSA signature on h can be done using existing circuits [37].

Ideally, we would like to use a standard hash function like SHA256. Unfortunately, proving that we have honestly applied SHA256 to a 30 megapixel (MP) witness is practically infeasible. This is because SHA256 consists of mainly non-algebraic operations (e.g. logical operations), and proving non-algebraic constraints in a zk-SNARK is time-consuming. There are SNARK-friendly hash functions like Poseidon [38], but proving that we have honestly applied Poseidon to a 30 MP witness is also quite challenging in practice.

Our approach. We propose two solutions to this problem. Our first solution, presented in Section 4.2, is to design a collision resistant hash function H for which there is an especially efficient way to prove the instance-witness relation

$$\mathcal{R}_{\text{hash}} := \big\{ \big(h; w \big) \ : \ h = \mathsf{H}(w) \big\} \tag{4}$$

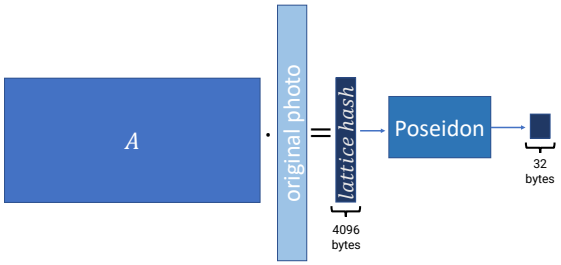

Figure 3: Our Lattice + Poseidon Hash Construction

even when w is a large string. To do so we use a composition of a lattice-based hash function (see Section 2.4) and the Poseidon hash function. Using this hash function, in the context of a SNARK, is of independent interest.

Our second solution, presented in Section 4.3, uses a polynomial commitment scheme as the collision resistant hash H. Computing this hash function on the image w takes more computing resources than in our first solution. However, once the hash value is computed, incorporating it into a SNARK proof requires no additional work. Hence, this approach is suitable when the original photo signer has enough computing power to compute a polynomial commitment to the original photo. If so, then the editor's work to produce the proof-of-valid-edit completely eliminates the expensive step of producing a proof for the relation $\mathcal{R}_{\text{hash}}$.

4.2. Lattice + Poseidon Hash Function

The hash function used in our first solution is a sequential composition of the lattice hash from Section 2.4 and a Poseidon hash. We represent the photo w being hashed as three low-norm vectors, each containing either the R, G, or B (all 8 bits long) values for every pixel. This means that if a photo has m pixels, we transform it into three vectors $\mathbf{v}_r, \mathbf{v}_g, \mathbf{v}_b$ of length m whose values are all in $[0, 255]$. We hash these three vectors separately. We set q to be the prime field of the SNARK system used to prove knowledge of the signature (e.g., a 256 bit prime for a KZG-based SNARK and a 64 bit prime for a FRI-based SNARK). We then generate a random matrix $\mathbf{A} \in \mathbb{F}_q^{n \times m}$, where in our settings m is about thirty million and n is 128. Let $\mathsf{Poseidon}(x)$ be the function that applies the Poseidon hash function defined over $\mathbb{F}_q$ to the input x. We define our hash function H on input $\mathbf{v} \in \mathbb{F}_q^m$ as:

$$\mathsf{H}(\mathbf{v}) := \mathsf{Poseidon}(\mathbf{A} \cdot \mathbf{v} \bmod q) \tag{5}$$

as shown in Figure 3. We compute the Poseidon hash of the lattice hash because $\mathbf{Av}$ is still fairly large (4096 bytes) and the Poseidon hash is much smaller (32 bytes). The function in (5) is collision resistant because the composition of two collision resistant hash functions is also collision resistant.

The point is that computing $\mathsf{H}(\mathbf{v})$ requires mostly linear operations over $\mathbb{F}_q$. In particular, the large matrix-vector product is all linear, and by choosing q to be compatible with the SNARK field, we can prove these linear operations very efficiently in a zk-SNARK. The major challenge of this approach is that the lattice-based hash is only collision-resistant when the input is a low-norm vector, so the prover must additionally prove that $\mathbf{v}$ is low-norm. In other words, we need a proof system for the following relation:

$$\begin{aligned}\mathcal{R}_{\text{VH}} := \big\{ \big((\mathbf{A}, \mathbf{h}, b)\ ;\ \mathbf{v}\big)\ :\ \mathbf{A} \in \mathbb{F}^{n \times m},\ b \in \mathbb{N},\\ \mathbf{h} = \mathsf{Poseidon}(\mathbf{A} \cdot \mathbf{v}),\quad \|\mathbf{v}\|_\infty < b\ \big\}\end{aligned} \tag{6}$$

where b is a norm bound needed for collision resistance of the lattice hash (as in Section 2.4). This relation implies that (i) the prover has honestly calculated the product of the public matrix $\mathbf{A}$ and the secret vector $\mathbf{v}$, (ii) this vector $\mathbf{v}$ is low-norm, and (iii) the prover has honestly applied Poseidon to this product. To prove (iii), we can use an available Groth16 circuit for Poseidon [39]. Proving (i) and (ii) is the focus of Section 5.

4.3. A Polynomial Commitment Hash

We next describe our second approach that is designed for a signer that has more compute power (such as the owner of a generative AI model). Again, let us encode the original image w as three vectors $\mathbf{v}_r, \mathbf{v}_g, \mathbf{v}_b \in [0, 255]^m$. For simplicity, in this section we only consider one such vector, denoted by $\mathbf{v}$, which can be thought of as any of the three vectors. In our second approach (mode 2) the signer interpolates a univariate polynomial W, of degree at most $m-1$, such that $W(\omega^i) = v_i$ for $i = 1, \ldots, m$. We denote this polynomial by $\text{poly}(w)$, namely

$$W := \text{poly}(w).$$

Next, the signer computes a succinct polynomial commitment to $W(X)$ as

$$\mathsf{com} \leftarrow \mathsf{PCS.commit}(\mathsf{pp}, W, r)$$

where r is the commitment randomness chosen by the signer. The signer signs com to obtain a signature σ on com. It sends the image w along with $(\mathsf{vk}, \mathsf{com}, \sigma, r)$ to the news editor, where vk is the signer's certified public key.

The news editor will create an edited image $x := f(w)$. Because the polynomial commitment is binding, it suffices for the editor to construct a ZK proof for the following instance-witness relation

$$\begin{aligned}\mathcal{R}_{\text{simple}} := \big\{ \big((\mathsf{pp}, f, x, \mathsf{com})\ ;\ (w, r)\big)\ :\ f(w) = x\ \wedge\\ \mathsf{com} = \mathsf{PCS.commit}(\mathsf{pp}, \text{poly}(w), r)\big\}\end{aligned}$$

A naive way to prove this is to have the SNARK circuit verify both that $f(w) = x$ and that com is a polynomial commitment to $\text{poly}(w)$. However, proving the latter would be far more expensive than proving that the editor has correctly hashed w, as we did in Section 4.2.

Instead, the key insight is that for the PLONK proof system, building a proof for $\mathcal{R}_{\text{simple}}$ is no more work for the

editor than simply proving that $f(w) = x$. This means that the SNARK circuit never needs to hash w, which greatly reduces the work for the editor. To achieve this reduction in work, VerITAS modifies PLONK's permutation argument. Let us see how.

Let $\mathcal{C}(x, w)$ be a circuit that outputs 0 iff $f(w) = x$. Remarkably, if an editor wants to build a proof for $\mathcal{R}_{\text{simple}}$, it can simply use $\mathcal{C}$ as the circuit provided to the PLONK system. However, this requires modifying the PLONK prover to indirectly prove that com is a valid commitment to poly(w). First, the editor builds the computation trace $T(X)$ for $\mathcal{C}(x, w)$ just as the standard PLONK prover would. The standard PLONK prover would then construct a proof π that $T(X)$ is a valid computation with respect to the gate constraints and copy constraints specified by $\mathcal{C}$. Our editor must additionally prove that com is a valid commitment to poly(w). Recall that some entries in $T(\Omega)$ correspond to the witness w (see Figure 4 for an example). Thus, proving that com is a valid commitment to poly$(w) = W(X)$ is equivalent to proving that the witness elements represented by entries of $W(\Omega)$ are equal to the corresponding witness entries in $T(\Omega)$. The editor can prove this equality by extending the PLONK permutation argument.

Recall that the permutation argument proves that the vector $T(\Omega)$ is equal to the vector $T(\tau(\Omega))$, where τ is a polynomial that implements a permutation of Ω. The standard PLONK prover defines τ to capture copy constraints within the circuit $\mathcal{C}$ and then uses the permutation argument to prove that the computation trace respects the circuit wiring. In our scheme, the editor extends the PLONK permutation argument to prove that all the entries in $T(\Omega)$ that correspond to a witness element are equal to the corresponding witness element in $W(\Omega)$. More specifically, it extends τ to a new permutation τ' that captures additional copy constraints between $T(X)$ and $W(X)$. The right side of Figure 4 gives an example where the black edges represent the standard PLONK copy constraints, and the thick red edges represent the extended copy constraints between $T(\Omega)$ and $W(\Omega)$. In other words, we use the PLONK permutation argument to prove copy constraints across two different polynomials: $T(X)$ and $W(X)$. The permutation argument can be adapted to this task, as was already shown in the original PLONK paper [40, §5.1].

The news editor can use this to efficiently construct a proof that $\mathcal{C}(x, w) = 0$ and that com is a valid commitment to poly(w). It uses both $T(X)$ and $W(X)$ to build a PLONK proof with respect to the permutation τ' which includes the new copy constraints shown in Figure 4. The result is a proof π that $(\mathsf{pp}, f, x, \mathsf{com})$ is a valid instance of $\mathcal{R}_{\text{simple}}$.

The final data sent to the verifier (the news reader) is (x, f) along with the proof $\pi' := (\mathsf{vk}, \mathsf{com}, \sigma, \pi)$. The verifier checks that

- $\mathsf{Vf}(\mathsf{vk}, \mathsf{com}, \sigma)$ accepts (σ is a valid sig on com), and
- π is a valid proof that $(\mathsf{pp}, f, x, \mathsf{com})$ is a $\mathcal{R}_{\text{simple}}$ instance.

We note that because the polynomial commitment is hiding, the commitment com to w reveals nothing to the verifier about w beyond the edited image x.

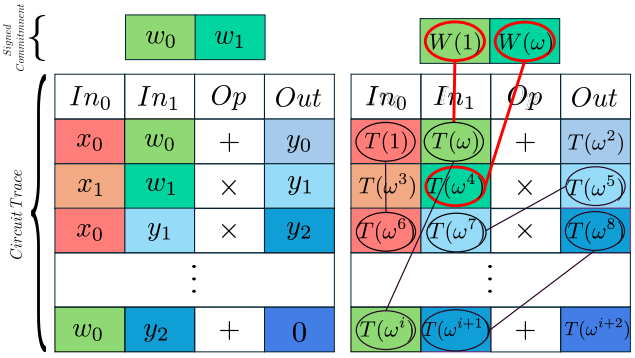

Figure 4: This figure illustrates the additional copy constraints (thick red lines) that the prover must prove in our polynomial commitment-based signing scheme. If $W(X)$ is the witness polynomial, then the prover must show that evaluations of W are the same as the corresponding witness cells in the circuit trace.

Augmenting the permutation argument to operate over two polynomial T and W in this way does not change the proving time much over simply proving that $f(w) = x$ for a public value x and witness w. This is because the additional copy constraints are by far fewer than the copy constraints required for the circuit $\mathcal{C}$.

Summary. This method results in a massive reduction in work for the editor, because the zk-SNARK circuit is now much simpler than the circuit in Section 4.2. In particular, the circuit does not need to hash the large original image w. While this saves work for the editor, it creates more work for the signer because computing a polynomial commitment to w is more costly than computing a lattice hash of w. We quantify these tradeoffs in Section 7.

5. A Proof System for the Lattice Hash

In this section, we describe the custom proof system that VerITAS uses to prove that the hash function in (5) was computed correctly. Specifically, we construct a proof system for the following instance-witness relation

$$\mathcal{R}_{\text{LH}} := \big\{ \big((\mathbf{A}, \mathbf{h}, b)\ ;\ \mathbf{v}\big)\ :\ \mathbf{A} \in \mathbb{F}^{n \times m},\ b \in \mathbb{N},\ \mathbf{h} = \mathbf{A} \cdot \mathbf{v},\quad \|\mathbf{v}\|_\infty < b \big\} \tag{7}$$

That is, the prover shows that it knows a low-norm vector $\mathbf{v} \in \mathbb{F}^m$ such that $\mathbf{h} = \mathbf{A} \cdot \mathbf{v}$.

First, to prove that $\|\mathbf{v}\|_\infty < b$ it suffices to prove that all the elements of $\mathbf{v}$ are in the set* $[0, b-1]$. We can implement this range proof with a lookup argument using the table $T := [0, b-1]$. The lookup argument proves that every element of $\mathbf{v} \in \mathbb{F}^m$ is in T. In VerITAS, we set $b = 2^8$ and $m = 30,000,000$ (30 MP), so that $b \ll m$. In these settings a simplified Plookup-based lookup argument [41]

*Technically, we need to prove that the elements are in $(-b, b)$, but since $\mathbf{v}$ only contains values in $[0, b)$, we can ignore the negative part.

minimizes prover work. In particular, the prover creates a second vector $\mathbf{u}$, which is a list of all the values in T, and a third vector $\mathbf{z} \in \mathbb{F}^{m+b}$, which is the sorted concatenation of $\mathbf{v}$ and $\mathbf{u}$. If the prover can show that $\mathbf{z}$ is a permutation of $\mathbf{v}\|\mathbf{u}$, and that the difference between consecutive elements in $\mathbf{z}$ is either 0 or 1, then it has shown that all the elements of $\mathbf{v}$ lie in T, and the range proof is complete. We explain how to do this in Section 5.3.

Second, to prove $\mathbf{h} = \mathbf{A} \cdot \mathbf{v}$ we use the classic Freivalds' algorithm [42]. That is, we use the observation that to prove that $\mathbf{h} = \mathbf{A} \cdot \mathbf{v}$ it suffices to prove that $\mathbf{r}^\intercal\mathbf{h} = (\mathbf{r}^\intercal\mathbf{A})\mathbf{v}$ holds for a random vector $\mathbf{r} \leftarrow\$ \mathbb{F}^n$ chosen by the verifier. This collapses the matrix-vector product check to simply testing that the dot-product of $(\mathbf{r}^\intercal\mathbf{A})$ with $\mathbf{v}$ is equal to the public scalar $\mathbf{r}^\intercal\mathbf{h} \in \mathbb{F}$. This can be proved via the sum-check protocol, as observed by Thaler [43]. In our case we use a univariate sum-check proof introduced in the Aurora system [44]. We give the details in Section 5.4.

5.1. Polynomial Representation of Vectors

All of our subsequent proof systems prove statements about a witness vector $\mathbf{v} \in \mathbb{F}^m$. These protocols encode the vector $\mathbf{v}$ as a polynomial, and then prove the required statement about the polynomial. To do so, let us define the polynomial encoding for a vector $\mathbf{v}$ to be the unique polynomial $v(X) \in \mathbb{F}^{<m}[X]$ where $\forall i \in [m]$, $v(\omega^i) = v_i$. This means:

$$\mathbf{v} = (v_0, ..., v_{m-1}) = \big(v(1), ..., v(\omega^{m-1})\big) \in \mathbb{F}^m.$$

We let poly : $\mathbb{F}^m \to \mathbb{F}^{<m}[X]$ be the function that maps a vector $\mathbf{v}$ to a polynomial $v(X)$. This function can be implemented with any polynomial interpolation method.

5.2. Zero, Sum, and Permutation Check Proofs

In what follows we will be using proof systems for three well-known instance-witness relations: zero check, sum check, and permutation check. As usual, let $\Omega_m := \{1, \omega, \omega^2, \ldots, \omega^{m-1}\} \subseteq \mathbb{F}$, and let $d > m$ be some degree bound.

- **The ZeroCheck relation:**

$$\mathcal{R}_{\text{ZC},m} := \big\{\big((\mathsf{pp},\mathsf{com}_u)\ ;\ (u,r)\big)\ :\ u \in \mathbb{F}^{<d}[X],\ \forall\omega \in \Omega_m : u(\omega) = 0,\ \mathsf{commit}(\mathsf{pp},u,r) = \mathsf{com}_u\big\}$$

- **The Univariate SumCheck relation:**

$$\mathcal{R}_{\text{SC},m} := \big\{\big((\mathsf{pp},\mathsf{com}_u,s)\ ;\ (u,r)\big)\ :\ u \in \mathbb{F}^{<d}[X],\ \sum_{\omega\in\Omega_m} u(\omega) = s,\ \mathsf{commit}(\mathsf{pp},u,r) = \mathsf{com}_u\big\}$$

- **The PermCheck relation:** Let $u \in \mathbb{F}^{<b}[X]$, $v \in \mathbb{F}^{<m}[X]$, and $z \in \mathbb{F}^{<b+m}[X]$ be three committed polynomials. A permutation check convinces the verifier that the vector $z(\Omega_{b+m})$ is a permutation of the vector $u(\Omega_b)\|v(\Omega_m)$. More precisely, it is a proof for the following relation

$$\begin{aligned}\mathcal{R}_{\text{PC}} :=\big\{\big((\mathsf{pp},\mathsf{com}_u,\mathsf{com}_v,\mathsf{com}_z)\ ;\ (u,v,z,r_u,r_v,r_z)\big)\ :\ & u \in \mathbb{F}^{<b}[X],\quad v \in \mathbb{F}^{<m}[X],\quad z \in \mathbb{F}^{<b+m}[X],\\ & \prod_{\alpha\in\Omega_{b+m}}(X - z(\alpha)) = \prod_{\beta\in\Omega_b}(X - u(\beta)) \prod_{\gamma\in\Omega_m}(X - v(\gamma)),\\ & \mathsf{commit}(\mathsf{pp},u,r_u) = \mathsf{com}_u,\\ & \mathsf{commit}(\mathsf{pp},v,r_v) = \mathsf{com}_v,\\ & \mathsf{commit}(\mathsf{pp},z,r_z) = \mathsf{com}_z\quad\big\}\end{aligned}$$

  The equality on the third line is an equality of univariate polynomials in the indeterminate X. The equality holds if and only if z is a permutation of $u\|v$.

A zk-SNARK for the ZeroCheck relation works by committing to the quotient polynomial $q(X) := u(X)/Z_{\Omega_m}(X)$. Aurora [44] gives a zk-SNARK for the SumCheck relation, and Plonk [24] gives zk-SNARK for the PermCheck relation. We denote these proof systems by $(\mathsf{P}_{\text{ZC}}, \mathsf{V}_{\text{ZC}})$, $(\mathsf{P}_{\text{SC}}, \mathsf{V}_{\text{SC}})$, and $(\mathsf{P}_{\text{PC}}, \mathsf{V}_{\text{PC}})$ respectively. All three proof systems produce proofs whose length is independent of the degree of the witness. We note that Haböck [45] recently gave an improved arguement for PermCheck, by replacing the product by a sum of rational functions. PermCheck can also be proved efficiently using the GKR proof system [46].

5.3. The Range Proof

Next, we explain how to prove that all elements of a vector $\mathbf{v} \in \mathbb{F}^m$ are in a given set $T := [0, b-1]$. This range proof is inspired by the Plookup lookup table protocol [41]. Define the vector $\mathbf{u}_b := (0, 1, ..., b-1)$. A range proof amounts to proving that all elements of $\mathbf{v}$ are in $\mathbf{u}_b$.

We will work with the polynomial representation of these vectors, namely $v := \text{poly}(\mathbf{v})$ and $u := \text{poly}(\mathbf{u}_b)$. The verifier has the statement $(\mathsf{pp}, \mathsf{com}_v)$. The prover has the same statement and the witness (v, r_v) such that $\mathsf{commit}(\mathsf{pp}, v, r_v) = \mathsf{com}_v$ and $v \in \mathbb{F}^{<m}[X]$. Now, the range check proof system is described in Algorithms 1 and 2.

Algorithm 1 $\mathsf{RangeCheckProver}(\mathsf{pp}, b, \mathsf{com}_v; v, r_v)$

$\mathbf{z} \leftarrow \text{poly}(\text{sort}(\mathbf{v}\|\mathbf{u}_b))$
$\mathsf{com}_z \leftarrow \mathsf{commit}(\mathsf{pp}, \text{poly}(\mathbf{z}), r_z)$
$\mathsf{com}_u \leftarrow \mathsf{commit}(\mathsf{pp}, \text{poly}(\mathbf{u}_b), 0)$
// Do the permutation check on z and u, v.
$\pi_{\text{PC}} \leftarrow \mathsf{P}_{\text{PC}}(\mathsf{pp}, \mathsf{com}_u, \mathsf{com}_v, \mathsf{com}_z; u, v, z, 0, r_v, r_z)$
// Compute a polynomial f that is zero on Ω_{m+b}
// if the gap between consecutive elements of z is either 0 or 1.
$f(X) \leftarrow \big(z(\omega X) - z(X)\big) \cdot \big(z(\omega X) - z(X) - 1\big)$
// A commitment com_f to f is implied by a commitment to z.
// Prove that f is zero on Ω_{m+b}.
$\pi_{\text{ZC}} \leftarrow \mathsf{P}_{\text{ZC,m+b}}(\mathsf{pp}, \mathsf{com}_f; f, r_f)$
Output $\pi \leftarrow (\mathsf{com}_z, \pi_{\text{PC}}, \pi_{\text{ZC}})$

Algorithm 2 $\mathsf{RangeCheckVerifier}(\mathsf{pp}, b, \mathsf{com}_v; \pi)$

parse $(\mathsf{com}_z, \pi_{\text{PC}}, \pi_{\text{ZC}}) \leftarrow \pi$
$\mathsf{com}_u \leftarrow \mathsf{commit}(\mathsf{pp}, \mathsf{poly}(\mathbf{u}_b), 0)$
accept if $\mathsf{V}_{\text{PC}}(\mathsf{pp}, \mathsf{com}_u, \mathsf{com}_v, \mathsf{com}_z; \pi_{\text{PC}})$
and $\mathsf{V}_{\text{ZC,m+b}}(\mathsf{pp}, \mathsf{com}_f; \pi_{\text{ZC}})$ both accept

The following theorem states the security property of this proof system. The proof is immediate and is omitted.

Theorem 5.1. Suppose that $(\mathsf{P}_{\text{ZC}}, \mathsf{V}_{\text{ZC}})$ and $(\mathsf{P}_{\text{PC}}, \mathsf{V}_{\text{PC}})$ are zk-SNARKs for $\mathcal{R}_{\text{ZC}}$ and $\mathcal{R}_{\text{PC}}$ respectively. Further, suppose that the polynomial commitment scheme used is secure. Then the proof system in Algorithms 1 and 2 is a zk-SNARK for the relation

$$\mathcal{R}_{\text{RP}} := \Big\{ \big((\mathsf{pp}, b, \mathsf{com}_v)\ ;\ (v, r_v)\big)\ :\ v \in \mathbb{F}^{<m}[X], \\ \forall \omega \in \Omega_m : v(\omega) \in [0, b-1] \text{ for } b \in \mathbb{N}, \\ \mathsf{com}_v = \mathsf{commit}(\mathsf{pp}, v, r_v) \quad \Big\}$$

5.4. The Lattice Hash Proof

Finally, we show a proof system that lets the prover show that, given a lattice hash $\mathbf{h} \in \mathbb{F}^n$, it knows a low-norm preimage $\mathbf{v} \in \mathbb{F}^m$. That is, we provide a proof system for the relation $\mathcal{R}_{\text{LH}}$ from (7), as required.

First, we augment the relation $\mathcal{R}_{\text{LH}}$ as follows

$$\mathcal{R}'_{\text{LH}} := \Big\{ \big((\mathbf{A}, \mathbf{h}, b, \mathsf{pp}, \mathsf{com}_v)\ ;\ (\mathbf{v}, r_v)\big)\ :\ \mathbf{A} \in \mathbb{F}^{n \times m}, \\ \mathbf{h} = \mathbf{A} \cdot \mathbf{v}, \quad \|\mathbf{v}\|_\infty < b \text{ for } b \in \mathbb{N}, \\ \mathsf{com}_v = \mathsf{commit}(\mathsf{pp}, \mathsf{poly}(\mathbf{v}), r_v) \quad \Big\} \tag{8}$$

This relation is the same as $\mathcal{R}_{\text{LH}}$ except that we force the prover to send to the verifier a commitment com_v to $\mathbf{v}$. Observe that the only difference between this relation and the relation $\mathcal{R}_{\text{RP}}$ from Theorem 5.1 is the additional constraint that $\mathbf{h} = \mathbf{A} \cdot \mathbf{v}$. We explained at the beginning of Section 5 that this constraint can be reduced to checking a single dot-product by taking a random linear combination of the rows of $\mathbf{A}$. The random linear combination is provided by the public coin of the verifier, and the protocol can be made non-interactive using the Fiat-Shamir transform. Finally, this single dot-product is exactly a univariate SumCheck relation, and can be verified by a single univariate SumCheck proof. Hence, a proof system for $\mathcal{R}'_{\text{LH}}$ uses the proof system from Section 5.3 along with a univariate SumCheck proof.

6. VerITAS Implementation Details

We implement the two components of VerITAS separately: we use the arkworks [47] Rust library to generate proofs of correct hashing (for the relation $\mathcal{R}_{\text{VH}}$ from (6)), and we use Plonky2 to generate the photo editing proofs.[†] In addition, the editor would generate a proof of knowledge of a valid ECDSA signature on the Lattice+Poseidon hash using an existing signature checking circuit [37] on top of our proof system, which only adds 45 seconds to proof generation time.

[†]Our code available at https://github.com/zk-VerITAS/VerITAS

6.1. Implementing a Proof System for $\mathcal{R}_{\text{VH}}$

To implement our proof system for the relation $\mathcal{R}_{\text{VH}}$ from (6), we use the KZG polynomial commitment [21] implementation in the arkworks Gemini library [48]. This implementation allows us to batch commit to polynomials and to batch open these commitments at multiple points. During setup, we generate a committing/verifying key pair. Our implementation uses a Groth16 proof to prove knowledge of a Poseidon hash preimage, so we also generate a Groth16 proving/verification key pair.

To prevent the prover and verifier from having to store all the elements in the (large) hashing matrix $\mathbf{A}$, we generate the entries of $\mathbf{A}$ using the upper 32 bits of a linear congruential generator [49] with a 64-bit modulus q'. For our SIS parameters, we set $n = 128$, and $b = 256$. As discussed in Section 4.2, the choice of q depends on the polynomial commitment scheme used. FRI uses a 64 bit prime, while KZG uses a 256 bit prime. If q is 64 bits, the SIS lattice estimator calculator for these parameters gives 192 bits of security [50] (bigger q values lead to more security, so the setting q to be 256 bits gives at least this much security). The prover generates the random Fiat-Shamir challenge for the permutation argument by taking the hash of the transcript thus far. Our prover proves knowledge of a lattice hash $\mathbf{Av}$ as described in Section 5 and then proves that applying a Poseidon hash to this lattice hash results in the final public hash $\mathbf{h}$ using Groth16. The verifier checks the commitment proofs and openings specified by the lattice hash proof and lastly checks the Groth16 proof.

6.1.1. Optimized $(\mathbf{r}^\intercal\mathbf{A})$ Derivation. Recall that in Section 5, we use the Freivalds' algorithm to reduce the checking of the matrix vector product $\mathbf{h} = \mathbf{A} \cdot \mathbf{v}$ to the dot-product of $(\mathbf{r}^\intercal\mathbf{A})$ with $\mathbf{v}$, where $\mathbf{r}$ is a random vector. The most time-intensive part for the verifier is to rederive $\mathbf{r}^\intercal\mathbf{A}$—the random linear combination of $\mathbf{A}$'s rows. To reduce verifying time, we implement opt-VeriTAS where we assume the existence of public trusted commitments to the rows of $\mathbf{A}$. These commitments can be generated in a preprocessing phase and used for every proof thereafter. Given the trusted (polynomial) commitments, the prover provides an opening proof for each row polynomial at a random point α. The verifier can then get the evaluation of $\mathsf{poly}(\mathbf{r}^\intercal\mathbf{A})$ at α and verify the proof. The cost is a roughly a factor of two increase in prover time, and a factor of six increase in proof size, for a roughly 800 times reduction in the verifier's time.

6.1.2. Consistency with Photo Editing Proofs. Both the SNARK circuits for the hash proof and the photo editing proof take the original photo w as part of the secret witness. However, a malicious prover might assign different values for the original photo in these two circuits. To prevent this, we leverage the fact that both proofs use polynomial

commitments. The idea is to require the prover to provide a polynomial commitment com to the original photo. Then by the technique described in Section 4.3, we can ensure that the vector committed in com is consistent with the partial witness used in both the hash and the photo editing circuits.

6.2. Photo Editing Proof Implementation

We generate the photo editing proofs using Plonky2 [14], a Rust-based general-purpose zk-SNARK system. Plonky2 lets developers specify a circuit, and use PLONK to prove that, given some public instance and private witness as input, the circuit output equals a certain value. For every edit we want to prove, we construct a circuit that applies the edit on the private witness (the original photo) and outputs the result. The verifier can then check that this output is the same as the public instance (the published edited photo).

Our cropping circuit computes the cropped photo by outputting the RGB values of the original photo in the cropped range.

Our grayscale circuit applies the standard grayscale formula used by Adobe Photoshop [51] to the RGB values of the original photo. This formula obtains the gray value gr for a pixel by taking a weighted linear combination of the RGB values: $gr = round(0.30R + 0.59G + 0.11B)$. Plonky2's circuits work over finite fields, so to support the floating point arithmetics during the grayscale calculation, we scale all pixel values by a factor of 100 and pass in the remainders of the rounding calculations to the verifier as part of the instance.

Our resizing circuit implements bilinear resizing, which is one of the standard resizing options offered in Adobe Photoshop [19]. Bilinear resizing calculates the RGB values for every pixel in the resized image by taking a weighted linear combination of the RGB values of four pixels in the original image. Just as with the grayscale circuit, we accommodate floating point arithmetic by passing remainders to the verifier. Because the resized values for R, G, and B channels are calculated independently, we can produce these proofs independently in parallel.

Our blur circuit implements a box blur, which is one of the standard blur options offered in Adobe Photoshop [52]. A box blur calculates the RGB values for a pixel at position (i, j) by averaging the RGB values of the pixels in the 3x3 "box" in the original image where pixel (i, j) is at the center. Just as in the grayscale and resizing circuits, the box blur calculation involves floating point arithmetic. However, because the purpose of blurring is to obscure information, passing remainders to the verifier could reveal private information to the verifier. Instead, we check within the circuit that the remainders are in the range $[0, 8]$.

7. Experimental Results

We report proof generation time, verification time, and proof size for the hash relation $\mathcal{R}_{\text{VH}}$ from (6) and for the image editing relations. We ran our timing experiments on randomly-generated RGB channels on a virtual machine with 131 GB of RAM and 12 CPU cores. When considering what is a reasonable amount of time to generate and verify proofs, it is important to remember that proof generation only needs to happen once per photo, while proof verification needs to be performed by every browser that accesses the article. This means that while proof generation needs to be fast, proof verification needs to be very fast. Given peak memory usage and running time, we estimate that generating a proof for the hash relation would cost about $0.99 on AWS per image for VerITAS and $2.42 for opt-VerITAS. Proofs for the editing relations would add a maximum of $0.09 per edit to the cost. Recall that the polynomial commitment hash method described in Section 4.3 only needs a proof for the editing relation; there is no need to prove the hash relation.

7.1. $\mathcal{R}_{\text{VH}}$ Proof Generation Results

Figure 5 compares the proving times for proving knowledge of a Poseidon hash and proving knowledge of our Lattice + Poseidon hash (relation $\mathcal{R}_{\text{VH}}$ from (6)) using VerITAS and opt-VerITAS. These results assume that the hashes for the RGB vectors $\mathbf{v}_r$, $\mathbf{v}_g$, and $\mathbf{v}_b$ are generated in parallel. Our results show that, for smaller image sizes (about ≤ 10 KP), it is faster to generate a proof of knowledge of a Poseidon hash than a proof of knowledge of our Lattice + Poseidon hash construction. However, for more realistically-sized images (> 10 KP), it takes much less time to generate a proof of knowledge for the Lattice + Poseidon hash construction in both VerITAS (24 min for 30 MP) and opt-VerITAS (60 min for 30 MP). In fact, when we tried to generate a proof of knowledge of a Poseidon hash for a picture of 1 MP, our machine ran out of memory and aborted the process (the 10 MP point at 34 min and 30 MP point at 103 min shown for Poseidon are projected points). With parallelism, the opt-VerITAS proving time could be cut down to 27.5 minutes (which, again, would be a once-per-image cost). Given that FRI is much faster than KZG, we estimate that VerITAS and opt-VerITAS would be much faster with FRI.

We report peak memory usage for our hash proof generation in Table 1. For 30 MP images, verification time is about 3 minutes for VerITAS (third column of Table 1) and 0.22 seconds for opt-VerITAS (last column of Table 1). Proof size is 4.6 KB for VerITAS and 28 KB for opt-VerITAS. Since these proofs are sent along with a megabyte image, these sizes are quite reasonable.

7.2. Photo Edit Proof Generation Results

To demonstrate the practicality of our Plonky2 implementations, we report setup and proof generation timing results for "realistic" image sizes. The signature-producing Sony camera mentioned earlier is a 33 MP camera. The edited photo size depends on clients. E.g., photos on *The New York Times* are resized to 2048 x 1365 pixels. Thus, in our experiments, for the editing operations that involve changing image dimensions (resizing and cropping), we report the times associated with resizing a 33 MP photo

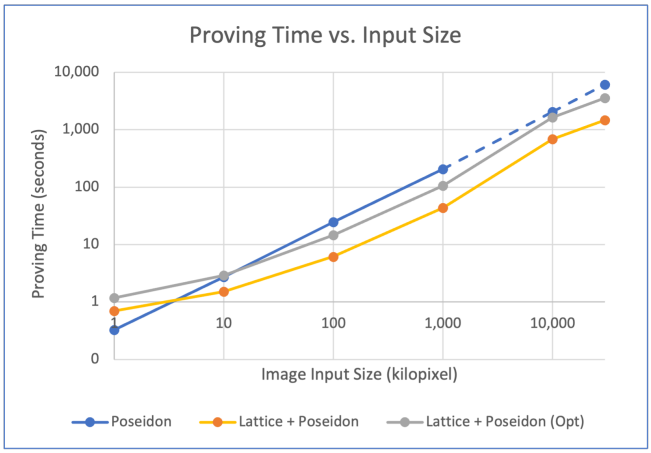

Figure 5: Graph showing proof generation time for generating a proof of a Poseidon hash and generating a proof of a Lattice + Poseidon hash (our construction) in both VerITAS and opt-VerITAS. The dashed part of the Poseidon line refers to extrapolated values for sizes that exceeded the prover's capacity.

Image Size (KiloPixel)	Peak Memory (GB)	Verify Time (sec)	Opt Peak Memory (GB)	Opt Verify Time (sec)
1	3.40	0.087	3.44	0.218
10	3.44	0.156	3.44	0.218
100	3.61	0.77	3.66	0.231
1,000	5.08	6.72	5.65	0.212
10,000	30.36	66.18	32.55	0.219
30,000	57.35	196.37	71.95	0.219

TABLE 1: Prover memory needs and Verifier time for Lattice+Poseidon hash generation for VerITAS and opt-VerITAS.

to the standard *New York Times* size. For operations that do not involve changing dimensions (grayscale conversion, blurring), we report the times associated with editing a photo of the standard *New York Times* size. For blurring, we report results for blurring 10% of the pixels.

Table 2a shows the timing results for cropping proof generation. Table 2b shows the timing results for resizing proof generation for a single color channel. Table 2c shows the timing results for grayscale proof generation. Table 2d shows the timing results for blur proof generation.

Overall, setup and proof generation take just a few minutes. Because proofs only need to be generated once by the news organization, these times are suitable for practical implementation. Verification time takes 2 seconds in a browser. Plonky2 proofs are about 100-200 KB, which is reasonable compared to edited photos on the order of 8 MB. Moreover, Plonky2 proofs can be further compressed via a constant-sized zkSNARK (e.g., Groth16 or PLONK) that proves the correctness of Plonky2 proof verification.

(a) Timing Results for Cropping

Original Size (pixels)	Reduced Size (pixels)	Setup Time (minutes)	Proof Gen Time (minutes)
6632 x 4976	2048 x 1365	2.59	2.27

(b) Timing Results for Resizing

Original Size (pixels)	Reduced Size (pixels)	Setup Time (minutes)	Proof Gen Time (minutes)
6632 x 4976	2048 x 1365	5.94	3.15

(c) Timing Results for Grayscale Conversion

Photo Size (pixels)	Setup Time (minutes)	Proof Generation Time (minutes)
2048 x 1365	1.99	1.38

(d) Timing Results for Blurring

Original Size (pixels)	Blur Region Size (pixels)	Setup Time (minutes)	Proof Gen Time (minutes)
2048 x 1365	529 x 529	10.67	5.83

TABLE 2: The time to generate a photo edit proof

Hashing Scheme	Time (s)	Memory (GB)
SHA256	1.71	0.003
Lattice (64 bit)	4.24	0.003
FRI-PCS	19.84	18.90
KZG-PCS	52.46	14.88

TABLE 3: Timing comparison for different hashing schemes of a 30 MP image. Hashing using FRI-PCS and KZG-PCS takes much longer and requires more resources than the first two hash functions.

7.3. Comparing the Two Signing Schemes

Table 3 compares how long it takes to calculate a SHA256 hash, a lattice hash, a FRI polynomial commitment (as in Plonky2), and a KZG polynomial commitment, of a 30 MP picture. We assume the photo is read in as a stream. In practice, we expect C2PA to use an FRI-based proof system, so we report lattice hash timings with a 64 bit prime q.

Table 3 shows that it takes much longer to compute a polynomial commitment of an image (using either FRI-PCS or KZG-PCS) than a simple hash (using either SHA256 or lattice hash), making it much more feasible for a computationally-limited signer like a camera to sign a lattice hash rather than a polynomial commitment. Furthermore, for a camera to calculate a KZG commitment for 30 MP photo, it would have to store an enormous reference string (SRS) that allows for commitments to polynomials of degree 3×10^7. Similarly, FRI-PCS requires access to a large amount of memory to efficiently perform FFTs. Our experiments, for instance, required about 20 GB of RAM, which is far above than what a camera would have. Thus, in practice, computationally-limited signers, such as cameras, would most likely use our lattice-based signature scheme.

8. Related Work

One cryptographic tool proposed for image authentication is perceptual hashing [53]. The goal of perceptual hashing is to design a hash function that is resilient to content-preserving manipulations but can detect malicious manipulations. While this attempts to guarantee semantic meaning in a photo, our solution aims to guarantee something more rigorous, namely to certify that only certain edits have been made to a photo.

Another potential cryptographic approach to image authentication is homomorphic signatures [17], [18], as discussed in the introduction. Homomorphic signatures that are not built from a zk-SNARK can be used for cropping (or other forms of redaction), but are too inefficient to handle more complex image edits, such as blurring and resizing.

The most closely related work to ours is that of Naveh and Tromer [12] and Kang et al. [13] mentioned in the introduction. Those works applied to images that are more than an order of magnitude smaller than image sizes from modern cameras. More recently, Della Monica et al. [54] proposed dividing a photo into N non-overlapping "tiles" and producing a proof for each tile that attests to the hash and a certain transformation on that image. Because these tiles are smaller than the entire image, the memory and time required to generate the N proofs is smaller than the memory and time required to generate one large proof. Just like Kang et al., Della Monica et al. only consider photos ≤ 900 KP, which is an order of magnitude smaller than the photos we consider. Moreover, verification time for these tiled proofs is about 3 minutes, which could be problematic for someone reading a newspaper in a browser. Another issue with this approach is that image transformations must be applied per tile rather than on the image as whole, which is the standard practice in photo editing software like Adobe Photoshop. For instance, to resize or blur a photo, tiles must be individually resized or blurred and then collated together. Consequently, these methods are unable to support standard Photoshop algorithms. We also note that very recently Dziembowski et al. [55] experimented with folding schemes for proving image edits.

9. Extensions and Conclusion

In this paper, we have discussed how to use zk-SNARKs to enable practical provenance verification for realistically large edited images in online news articles. Our system uses signing keys embedded in cameras as the origin of trust, but rather than trusting a third-party application to digitally sign edited images, we propose to use zero-knowledge proofs to prove to a news reader that an edited published photo was taken when and where the article claims it was taken. We create proofs for 30 MP images, which is the size of images produced by actual cameras equipped with embedded signing keys. The bottleneck in image editing proof systems is proving knowledge of a valid signature on the unedited photo. Our key innovations are two-fold: first, we introduce a new SNARK-friendly hashing method that reduces the hash proof generation time. We believe this SNARK-friendly method, which is a sequential composition of lattice hash with a Poseidon hash, may be of independent interest to those looking to create SNARKs that prove hashes of large amounts of data. Additionally, we introduce a polynomial commitment hash that completely eliminates the need for proving knowledge of a valid signature in the SNARK circuit. However, signing the unedited image using a polynomial commitment hash is more expensive than the lattice hash scheme.

We note that the description of VerITAS given here does not protect the identity of the signer (the photographer). Indeed, vk (and in σ in mode 2) are sent along with the ZK proof to the verifier. If the editor wants to hide the identity of the signer, then the editor could replace vk (and σ) by a public commitment com to those values, and move vk and σ to the zk-SNARK secret witness. The zk-SNARK circuit would then verify that com is a valid commitment to (vk, σ), instead of directly using those values as public inputs. This fully hides vk and σ from the verifier and protects the identity of the signer.

Finally, this paper has only discussed how to prove edits for photos, but videos are also a major source of misinformation. The main challenge with videos is that once they are edited, they are stored in a lossy compressed format. Directly applying the techniques discussed here to videos would thus require us to prove statements about video compression in a SNARK, which is challenging due to the size of a video file. Another avenue for future research is exploring different kinds of range proofs. In our work, we used a Plookup-based range proof. More recent methods, such as Lasso [28], may lead to time and memory savings for the editor.

Acknowledgments. This work was funded by NSF, DARPA, the Simons Foundation, UBRI, and NTT Research. Opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

References

[1] T. Datta and D. Boneh, "Using ZK proofs to fight disinformation," Medium, 2022, link.

[2] ——, "Using ZK proofs to fight disinformation," Real World Crypto, 2023, link.

[3] A. Coleman and S. Sardarizadeh, "Ukraine conflict: Many misleading images have been shared online," *BBC News*, 2022, link.

[4] V. Pavilonis, "Fact check: Images show mosul in 2017, kyiv one day after russian invasion began," *USA Today*, 2022, link.

[5] A. Coleman, "Ukraine conflict: Further false images shared online," *BBC News*, 2022, link.

[6] "BBC breakfast uses old footage of russian parade rehearsal to show invasion of ukraine," *Full Fact*, 2022, link.

[7] "C2PA technical specification," link.

[8] "Partnership for greater trust in digital photography: Leica and content authenticity initiative," *Leica*, 2022, link.

[9] "Sony unlocks in-camera forgery-proof technology," *Sony*, 2022, link.

[10] "Visuals," *Associated Press*, 2022, link.

[11] N. Bitansky, R. Canetti, A. Chiesa, and E. Tromer, "From extractable collision resistance to succinct non-interactive arguments of knowledge, and back again," in *ITCS 2012*, S. Goldwasser, Ed. ACM, Jan. 2012, pp. 326–349.

[12] A. Naveh and E. Tromer, "Photoproof: Cryptographic image authentication for any set of permissible transformations," in *2016 IEEE Symposium on Security and Privacy (SP)*, 2016, pp. 255–271.

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

[14] "plonky2," https://github.com/0xPolygonZero/plonky2.

[15] R. Johnson, D. Molnar, D. Song, and D. Wagner, "Homomorphic signature schemes," in *Topics in Cryptology — CT-RSA 2002*, B. Preneel, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 244–262.

[16] D. Derler, H. C. Pöhls, K. Samelin, and D. Slamanig, "A general framework for redactable signatures and new constructions," in *Information Security and Cryptology - ICISC 2015*, S. Kwon and A. Yun, Eds. Cham: Springer International Publishing, 2016, pp. 3–19.

[17] D. Boneh and D. M. Freeman, "Homomorphic signatures for polynomial functions," in *EUROCRYPT 2011*, ser. LNCS, K. G. Paterson, Ed., vol. 6632. Springer, Heidelberg, May 2011, pp. 149–168.

[18] S. Gorbunov, V. Vaikuntanathan, and D. Wichs, "Leveled fully homomorphic signatures from standard lattices," in *47th ACM STOC*, R. A. Servedio and R. Rubinfeld, Eds. ACM Press, Jun. 2015, pp. 469–477.

[19] "Image size and resolution," Adobe, 2024, link.

[20] D. Boneh and V. Shoup, *A Graduate Course in Applied Cryptography*, 2023, https://toc.cryptobook.us/book.pdf.

[21] A. Kate, G. M. Zaverucha, and I. Goldberg, "Constant-size commitments to polynomials and their applications," in *Advances in Cryptology-ASIACRYPT 2010: 16th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 5-9, 2010. Proceedings 16*. Springer, 2010, pp. 177–194.

[22] E. Ben-Sasson, I. Bentov, Y. Horesh, and M. Riabzev, "Fast reed-solomon interactive oracle proofs of proximity," in *ICALP 2018*, ser. LIPIcs, I. Chatzigiannakis, C. Kaklamanis, D. Marx, and D. Sannella, Eds., vol. 107. Schloss Dagstuhl, Jul. 2018, pp. 14:1–14:17.

[23] J. Groth, "On the size of pairing-based non-interactive arguments," Cryptology ePrint Archive, Paper 2016/260, 2016, https://eprint.iacr.org/2016/260. [Online]. Available: https://eprint.iacr.org/2016/260

[24] A. Gabizon, Z. J. Williamson, and O. Ciobotaru, "Plonk: Permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge," Cryptology ePrint Archive, Paper 2019/953, 2019, https://eprint.iacr.org/2019/953. [Online]. Available: https://eprint.iacr.org/2019/953

[25] A. Gabizon and Z. J. Williamson, "Plookup: A simplified polynomial protocol for lookup tables," Cryptology ePrint Archive, Paper 2020/315, 2020, https://eprint.iacr.org/2020/315. [Online]. Available: https://eprint.iacr.org/2020/315

[26] A. Zapico, A. Gabizon, D. Khovratovich, M. Maller, and C. Ràfols, "Baloo: Nearly optimal lookup arguments," Cryptology ePrint Archive, Paper 2022/1565, 2022, https://eprint.iacr.org/2022/1565. [Online]. Available: https://eprint.iacr.org/2022/1565

[27] L. Eagen, D. Fiore, and A. Gabizon, "CQ: Cached quotients for fast lookups," Cryptology ePrint Archive, Paper 2022/1763, 2022, https://eprint.iacr.org/2022/1763. [Online]. Available: https://eprint.iacr.org/2022/1763

[28] S. Setty, J. Thaler, and R. Wahby, "Unlocking the lookup singularity with lasso," Cryptology ePrint Archive, Paper 2023/1216, 2023, link.

[29] J. T. Schwartz, "Fast probabilistic algorithms for verification of polynomial identities," *J. ACM*, vol. 27, no. 4, p. 701–717, oct 1980. [Online]. Available: https://doi.org/10.1145/322217.322225

[30] R. Zippel, "Probabilistic algorithms for sparse polynomials," in *Proceedings of the International Symposiumon on Symbolic and Algebraic Computation*, ser. EUROSAM '79. Berlin, Heidelberg: Springer-Verlag, 1979, p. 216–226.

[31] A. Fiat and A. Shamir, "How to prove yourself: Practical solutions to identification and signature problems," in *Advances in Cryptology — CRYPTO' 86*, A. M. Odlyzko, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1987, pp. 186–194.

[32] T. Attema, S. Fehr, and M. Klooß, "Fiat-shamir transformation of multi-round interactive proofs (extended version)," *Journal of Cryptology*, vol. 36, no. 4, p. 36, Oct. 2023.

[33] M. Ajtai, "Generating hard instances of lattice problems," in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 1996, pp. 99–108.

[34] O. Goldreich, S. Goldwasser, and S. Halevi, "Collision-free hashing from lattice problems," ser. Lecture Notes in Computer Science. Springer, 2011, vol. 6650, pp. 30–39.

[35] M. Bellare, D. Micciancio, and B. Warinschi, "Foundations of group signatures: Formal definitions, simplified requirements, and a construction based on general assumptions," in *EUROCRYPT 2003*, ser. LNCS, E. Biham, Ed., vol. 2656. Springer, Heidelberg, May 2003, pp. 614–629.

[36] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," in *CRYPTO 2004*, ser. LNCS, M. Franklin, Ed., vol. 3152. Springer, Heidelberg, Aug. 2004, pp. 41–55.

[37] "circom-ecdsa circuit," link.

[38] L. Grassi, D. Khovratovich, C. Rechberger, A. Roy, and M. Schofnegger, "Poseidon: A new hash function for zero-knowledge proof systems," Cryptology ePrint Archive, Paper 2019/458, 2019, https://eprint.iacr.org/2019/458. [Online]. Available: https://eprint.iacr.org/2019/458

[39] "Poseidon circuit," link.

[40] A. Gabizon, Z. J. Williamson, and O. Ciobotaru, "PLONK: Permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge," Cryptology ePrint Archive, Report 2019/953, 2019, https://eprint.iacr.org/2019/953.

[41] A. Gabizon and Z. J. Williamson, "plookup: A simplified polynomial protocol for lookup tables," Cryptology ePrint Archive, Report 2020/315, 2020, https://eprint.iacr.org/2020/315.

[42] R. Freivalds, "Probabilistic machines can use less running time," p. 839–842, 1977.

[43] J. Thaler, "The unreasonable power of the sum-check protocol," The Art of Zero Knowledge, 2020. [Online]. Available: https://zkproof.org/2020/03/16/sum-checkprotocol/

[44] E. Ben-Sasson, A. Chiesa, M. Riabzev, N. Spooner, M. Virza, and N. P. Ward, "Aurora: Transparent succinct arguments for R1CS," in *EUROCRYPT 2019, Part I*, ser. LNCS, Y. Ishai and V. Rijmen, Eds., vol. 11476. Springer, Heidelberg, May 2019, pp. 103–128.

[45] U. Haböck, "Multivariate lookups based on logarithmic derivatives," Cryptology ePrint Archive, Report 2022/1530, 2022, https://eprint.iacr.org/2022/1530.

[46] S. Goldwasser, Y. T. Kalai, and G. N. Rothblum, "Delegating computation: interactive proofs for muggles," in *40th ACM STOC*, R. E. Ladner and C. Dwork, Eds. ACM Press, May 2008, pp. 113–122.

[47] "arkworks," https://github.com/arkworks-rs/.

[48] J. Bootle, A. Chiesa, Y. Hu, and M. Orrù, "Gemini: Elastic snarks for diverse environments," Cryptology ePrint Archive, Paper 2022/420, 2022, https://eprint.iacr.org/2022/420. [Online]. Available: https://eprint.iacr.org/2022/420

[49] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 2007.

[50] M. R. Albrecht, R. Player, and S. Scott, "On the concrete hardness of learning with errors," Cryptology ePrint Archive, Paper 2015/046, 2015, https://eprint.iacr.org/2015/046. [Online]. Available: https://eprint.iacr.org/2015/046

[51] S. Valentine, "How photoshop translates rgb color to gray," *insider*, 2018, https://insider.kelbyone.com/how-photoshop-translates-rgb-color-to-gray-by-scott-valentine/.

[52] "Blur and sharpen effects," Adobe, 2024, link.

[53] F. Ahmed, M. Y. Siyal, and V. Uddin Abbas, "A secure and robust hash-based scheme for image authentication," *Signal Process.*, vol. 90, no. 5, p. 1456–1470, may 2010, link.

[54] P. D. Monica, I. Visconti, A. Vitaletti, and M. Zecchini, "Do not trust anybody: Zk proofs for image transformations tile by tile on your laptop," Real World Crypto, 2024.

[55] S. Dziembowski, S. Ebrahimi, and P. Hassanizadeh, "VIMz: Verifiable image manipulation using folding-based zkSNARKs," Cryptology ePrint Archive, Paper 2024/1063, 2024, https://eprint.iacr.org/2024/1063. [Online]. Available: https://eprint.iacr.org/2024/1063

Appendix A. Preliminaries

This appendix contains the security definitions for the primitives defined in Section 2.

A.1. Digital Signatures

For a signature scheme SIG to be existentially unforgeable under a chosen message attack, every efficient adversary $\mathcal{A}$ with access to the verification key vk, and a signing oracle for messages of its choice, should not be able to produce a signature on a new message m^* for which $\mathcal{A}$ has not queried the signing oracle. The advantage of the adversary $\mathcal{A}$ in the corresponding security game with security parameter λ is $\mathsf{Adv}^{\mathrm{sig}}_{\mathcal{A},\mathsf{SIG}}(\lambda)$.

A.2. Commitment Schemes

We define the hiding and binding security property for commitment schemes below:

- **Hiding:** for every PPT adversary $\mathcal{A}$, there exists a negligible function $\nu(\cdot)$:

$$\Pr\left[b' \neq b : \begin{array}{l} \mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda) \\ (x_0, x_1) \leftarrow\!\$\ \mathcal{A}(\mathsf{pp}) \\ b \leftarrow\!\$\ \{0,1\} \\ r \leftarrow\!\$\ \mathcal{R}_{\mathrm{C}} \\ c \leftarrow \mathsf{commit}(\mathsf{pp}, m_b, r) \\ b' \leftarrow \mathcal{A}(c) \end{array}\right] \leq \frac{1}{2} + \nu(\lambda)$$

- **Binding:** for every PPT adversary $\mathcal{A}$ the following probability is negligible

$$\mathsf{Adv}^{\mathrm{bind}}_{\mathcal{A},C}(\lambda) := \Pr\left[\mathsf{commit}(\mathsf{pp}, x, r) = \mathsf{commit}(\mathsf{pp}, x', r')\right]$$

where $\mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda)$ and $(x, x', r, r') \leftarrow\!\$\ \mathcal{A}(\mathsf{pp})$.

We define correctness, evaluation binding, and hiding for a polynomial commitment scheme below.

- **Correctness**: for all $\lambda, d \in \mathbb{N}$, all $x \in \mathbb{F}$, and all $f \in \mathbb{F}[X]$ of degree a most d, the following probability is 1:

$$\Pr\left[\begin{array}{l} \mathsf{Vf}(pp, \mathsf{com}, \\ x, y, \pi) = 1 \end{array} : \begin{array}{l} \mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda, d) \\ r \leftarrow\!\$\ \mathcal{R}_{\mathrm{C}} \\ \mathsf{com} \leftarrow \mathsf{commit}(\mathsf{pp}, f, r) \\ (\pi, y) \leftarrow \mathsf{open}(\mathsf{pp}, f, x, r) \end{array}\right]$$

- **Evaluation Binding**: it is not possible to open a committed polynomial to two different values at one point. That is, for every PPT adversary $\mathcal{A}$ and for all $\lambda, d \in \mathbb{N}$, the following function is negligible

$$\Pr\left[\begin{array}{c} \mathsf{Vf}(\mathsf{pp}, \mathsf{com}, x, y, \pi) = 1 \\ \wedge \\ \mathsf{Vf}(\mathsf{pp}, \mathsf{com}, x, y', \pi') = 1 \\ \wedge y \neq y' \end{array} : \begin{array}{l} \mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda, d) \\ (\mathsf{com}, x, y, \pi, \\ y', \pi') \leftarrow\!\$\ \mathcal{A}(\mathsf{pp}, d) \end{array}\right]$$

Optionally, a polynomial commitment scheme can also satisfy the following property:

- **Hiding**: for every PPT adversary $\mathcal{A}$, there exists a negligible function $\nu(\cdot)$ such that:

$$\Pr\left[b' \neq b : \begin{array}{l} \mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda) \\ (f_0, f_1) \leftarrow\!\$\ \mathcal{A}(\mathsf{pp}) \\ b \leftarrow\!\$\ \{0,1\} \\ r \leftarrow\!\$\ \mathcal{R}_{\mathrm{C}} \\ c \leftarrow \mathsf{commit}(\mathsf{pp}, f_b, r) \\ b' \leftarrow \mathcal{A}(c) \end{array}\right] \leq \frac{1}{2} + \nu(\lambda)$$

A.3. zk-SNARKs

We define completeness, knowledge soundness, zero-knowledge, non-interactivity, and succinctness for a zk-SNARK below.

- **Completeness**: if $(x, w) \in \mathcal{R}$, then verification should pass. That is, for all $\lambda \in \mathbb{N}$ and all $(x, w) \in \mathcal{R}$:

$$\Pr\left[\mathsf{Vf}(\mathsf{pp}, x, \pi) = 1 \ : \begin{array}{l} \mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda) \\ \pi \leftarrow \mathsf{prove}(\mathsf{pp}, x, w) \end{array}\right] = 1$$

- **Knowledge Soundness**: if an adversary can produce a valid proof for some x, then there should be a polytime extractor that can compute a witness w such that $(x, w) \in \mathcal{R}$. That is, Π has knowledge error $\epsilon \in [0, 1]$ if for every PPT adversary $\mathcal{A} = (\mathcal{A}_0, \mathcal{A}_1)$ there exists a PPT extractor $\mathcal{E}$ such that:

$$\Pr\left[(x, w) \in \mathcal{R} \ : \begin{array}{l} \mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda) \\ (x, \mathsf{state}) \leftarrow\!\$\ \mathcal{A}_0(\mathsf{pp}) \\ w \leftarrow\!\$\ \mathcal{E}^{\mathcal{A}_1(\mathsf{state})}(\mathsf{pp}, x) \end{array}\right] \geq$$
$$\Pr\left[\mathsf{Vf}(\mathsf{pp}, x, \pi) = 1 \ : \begin{array}{l} \mathsf{pp} \leftarrow\!\$\ \mathsf{setup}(1^\lambda) \\ (x, \mathsf{state}) \leftarrow\!\$\ \mathcal{A}_0(\mathsf{pp}) \\ \pi \leftarrow\!\$\ \mathcal{A}_1(\mathsf{state}) \end{array}\right] - \epsilon$$

- **Zero-Knowledge**: We state the definition in the random oracle model where all the algorithms are oracle machine that can query an oracle $H : \mathcal{X} \to \mathcal{Y}$ for some finite sets $\mathcal{X}$ and $\mathcal{Y}$. The zk-SNARK is zero knowledge if there is a

PPT simulator $\Pi.\mathsf{sim}$ such that for all $(x, w) \in \mathcal{R}$ and all PPT adversaries $\mathcal{A}$, the following function is negligible

$$\mathsf{Adv}^{\mathrm{zk}}_{\mathcal{A},\Pi}(\lambda) := \left| \begin{array}{r} \Pr\left[\mathcal{A}^{H}\left(\mathsf{pp}, x, \mathsf{prove}^{H}(\mathsf{pp}, x, w)\right) = 1\right] - \\ \Pr\left[\mathcal{A}^{H[h]}\left(\mathsf{pp}, x, \pi\right) = 1\right] \end{array} \right|$$

where $\mathsf{pp} \leftarrow\!\!\$\ \mathsf{setup}(1^{\lambda})$ and $(\pi, h) \leftarrow\!\!\$\ \Pi.\mathsf{sim}(\mathsf{pp}, x)$. Here h is a partial function $h : \mathcal{X} \to \mathcal{Y}$ output by $\Pi.\mathsf{sim}$, and $H[h]$ refers to the oracle $H : \mathcal{X} \to \mathcal{Y}$ modified by entries in h. That is, we allow $\Pi.\mathsf{sim}$ to program the oracle H.

- **Non-interactive**: the proof is non-interactive, and a proof created by the prover can be checked by any verifier.
- **Succinct**: the proof size and verifier runtime are $o(|w|)$. The verifier can run in linear time in $|x|$.