

# A security analysis of the Zcash Sapling Protocol

Ariel Gabizon

Daira Hopwood

Zcash

## 1 Introduction

The purpose of this note is to show that the Sapling protocol, that will be used for Zcash shielded (private) transactions as of the Sapling network upgrade, satisfies certain security properties. This document is not completely self contained and while reading it we recommend referring often to the Zcash protocol spec[3] for full details of the Sapling protocol.

A notable property of the protocol is a separation of proving and signing authority. A “delegated spender/prover” creates transactions with the help of a proof authorizing key (or just proving key for short), but the transaction is not valid until it is signed by the signer with the spending key, that roughly corresponds to the secret key when thinking of the proving key as a public key.

We informally describe the four properties we prove.

1. **Non-malleability:** The delegated spender, after receiving a set of signatures on transactions of his choice, should not be able to create a new valid transaction, containing a nullifier appearing in one of the old transactions (overlapping nullifiers intuitively correspond to transactions from the same spending key). The way non-malleability is defined and proved here is inspired from the Zerocash paper[1].
2. **Indistinguishability:** An adversary should not be able to find two tuples (input notes, output notes) that are consistent in public data - meaning mainly that the amount going in or out of the shielded pool is the same, such that it is possible to distinguish from seeing the transaction which tuple it corresponds to.
3. **Balance:** An adversary should not be able to construct a valid ledger (even when having full control of transactions inserted) such that the total amount coming out of the shielded pool is larger than what came in.
4. **Spendability:** An adversary should not be able to send the honest party a note that was successfully received, but cannot be later spent (such an attack on [1] was found by Zooko Wilcox and coined “Faerie Gold” in [3]).

Before getting into the Sapling protocol and these properties, we begin with preliminary definitions and results regarding signature schemes.

## 2 Signature schemes

When we say an algorithm  $\mathcal{A}$  is *efficient*, we mean it runs in time  $\text{poly}(\lambda)$  for the “security parameter”  $\lambda$ .

**Definition 2.1.** Let  $\mathbb{G}$  be a group of prime order  $r$ . A signature scheme  $\mathcal{S}$  over  $\mathbb{G}$  in the random oracle model consists of algorithms  $\mathcal{S} = (\text{sign}, \text{verifySig}, \mathcal{S} = (\mathcal{S}_{\text{sign}}, \mathcal{S}_{\mathcal{R}}))$  where  $\text{sign}, \text{verifySig}$  are oracle machines with access to an oracle  $\mathcal{R}$  taking as input arbitrary strings and returning uniform elements of  $\mathbb{F}_r$ . Such that the following holds.

- The set of public/verification keys  $\{\text{pk}\}$  is  $\mathbb{G}$ , and the set of private keys  $\{\text{sk}\}$  is  $\mathbb{F}_r$ .
- For  $\text{sk} \in \mathbb{F}_r$ , the verification key of  $\text{sk}$  is  $\text{pk} = \text{sk} \cdot g$  for a fixed generator  $g \in \mathbb{G}$ .
- We have the following “zero-knowledge” property: Fix any efficient  $\mathcal{A}$ . Suppose that  $\mathcal{A}$  interacts with  $\mathcal{S}$  with two types of queries
  1. Queries  $x$ , for an arbitrary string  $x$  that are answered according to  $\mathcal{S}_{\mathcal{R}}$ .
  2. Queries  $(\text{pk}, \mathbf{m})$ , answered according to  $\mathcal{S}_{\text{sign}}$ .

Let  $\pi_1$  be the distribution of the sequence of queries and replies to  $\mathcal{A}$ . Let  $\pi_2$  be the distribution of the sequence of queries and replies to  $\mathcal{A}$  when

1.  $\mathcal{R}$  takes the place of  $\mathcal{S}_1$
2.  $\text{sign}^{\mathcal{R}}(\text{sk}, \mathbf{m})$  is returned instead of  $\mathcal{S}_2(\text{pk}, \mathbf{m})$  where  $\text{sk}$  is the secret key corresponding to  $\text{pk}$ .

Then the distance between  $\pi_1$  and  $\pi_2$  is  $\text{negl}(\lambda)$ .

We say  $\mathcal{S}$  is *unforgeable w.r.t key randomization* if the following holds. Fix any efficient  $\mathcal{A}$ . A party  $\mathcal{O}$  chooses uniform  $\text{sk} \in \mathbb{F}_r$  and sends  $\text{pk} = \text{sk} \cdot g$  to  $\mathcal{A}$ .  $\mathcal{O}$  also initializes an empty set  $T$ .  $\mathcal{A}$  adaptively makes  $\text{poly}(\lambda)$  queries of the form  $(\alpha, \mathbf{m})$ .  $\mathcal{O}$  replies with  $\sigma := \text{sign}(\text{pk} + \alpha \cdot g, \mathbf{m})$  and adds  $(\alpha, \mathbf{m}, \sigma)$  to  $T$ .

Finally  $\mathcal{A}$  outputs  $(\alpha^*, \mathbf{m}^*, \sigma^*)$ . Let  $\text{pk}^* := \text{pk} + \alpha^* \cdot g$ . Then the probability that

1.  $\text{verifySig}(\text{pk}^*, \mathbf{m}^*, \sigma^*)$ , and
2.  $(\alpha^*, \mathbf{m}^*, \sigma^*) \notin T$

is  $\text{negl}(\lambda)$ .

We assume our group  $\mathbb{G}$  has a hard DL problem; meaning that for any efficient  $\mathcal{A}$ , given uniform  $g, \text{sk} \cdot g \in \mathbb{G}$  the probability of outputting  $\text{sk}$  is  $\text{negl}(\lambda)$ .

We define the non-malleable version of Schnorr’s signature scheme:

**Schnorr:**

**Parameters:** Group  $\mathbb{G}$  of prime order  $s$ . Non-zero  $g \in \mathbb{G}$ .

**Signing:** Given message  $\mathbf{m}$  and  $\text{sk}$ ,

- Choose random  $a \in \mathbb{F}_r$  and let  $R := a \cdot g$
- Compute  $c := \mathcal{R}(R, \text{pk}, \mathbf{m})$
- Let  $u := a + c \cdot \text{sk}$ .
- Define  $\text{sign}^{\mathcal{R}}(\text{sk}, \mathbf{m}) := (R, u)$ .

**Verifying:** Given  $\text{pk}, \mathbf{m}, \sigma = (R, u)$ ,  $\text{verifySig}^{\mathcal{R}}(\text{pk}, \mathbf{m}, \sigma)$  accepts iff:

- Computing  $c := \mathcal{R}(R, \text{pk}, \mathbf{m})$ ; we have  $u \cdot g = R + c \cdot \text{pk}$ .

**Simulating:**

- $\mathcal{S}_{\mathcal{R}}(\mathbf{x})$  checks if  $\mathbf{x}$  has been queried before; if so answers consistently, otherwise answers uniformly in  $\mathbb{F}_r$  and records the answer.
- $\mathcal{S}_{\text{sign}}(\text{pk}, \mathbf{m})$ : Choose uniform  $c, u \in \mathbb{F}_r$ . Define  $R := u \cdot g - c \cdot \text{pk}$  and  $\mathbf{x} := (R, \text{pk}, \mathbf{m})$ . Check if  $\mathcal{S}_{\mathcal{R}}(\mathbf{x})$  has been defined. If so, abort. Otherwise define  $\mathcal{S}_{\mathcal{R}}(\mathbf{x}) = c$  and return  $(R, u)$ .

**Remark 2.2.** At times when we wish to change the parameter  $g$  we work with from default to an element  $h$ , we will use it in the subscript, e.g.  $\text{sign}_h^{\mathcal{R}}(\text{sk}, \mathbf{m})$ .

We refer by  $\text{Schnorr}' = (\text{sign}', \text{verifySig}')$  to the Schnorr scheme where  $\text{pk}$  is omitted from the computation of  $c$ .

**Theorem 2.3.** Schnorr is non-forgable w.r.t randomization.

*Proof.* Similarly to [2], we reduce to the non-forgability of standard Schnorr (where the public key is not part of the signature & without randomization) that was proven in [4].

Suppose we are given  $\mathcal{A}$  interacting with  $\mathcal{O}$  as described above, and finally outputting  $(\alpha^*, \mathbf{m}^*, \sigma^*)$ . We construct  $\mathcal{A}'$  that interacts with  $\mathcal{O}'$  which is a “standard” Schnorr oracle.

That is:

1.  $\mathcal{O}'$  begins by choosing a uniform  $\text{sk} \in \mathbb{F}_r$
2.  $\mathcal{O}'$  computes  $\text{pk} = \text{sk} \cdot g$  and sends  $\text{pk}$  to  $\mathcal{A}'$ .  $\mathcal{O}'$  initializes an empty set  $T'$ .
3.  $\mathcal{A}'$  sends queries  $\mathbf{m}$  to  $\mathcal{O}'$  and receives replies  $\sigma = \text{sign}'_{\text{sk}}(\mathbf{m})$ .  $\mathcal{O}'$  adds  $(\mathbf{m}, \sigma)$  to  $T'$ .
4. After all queries  $\mathcal{A}'$  outputs  $(\mathbf{m}^*, \sigma^*)$ .

$\mathcal{A}'$  wins if

- $\text{verifySig}'(\text{pk}, \mathbf{m}^*, \sigma^*)$ , and
- $(\mathbf{m}^*, \sigma^*) \notin T'$

$\mathcal{A}'$  will simulate  $(\mathcal{A})$ 's interaction with  $\mathcal{O}$  using  $\mathcal{O}'$ : Given a query  $(\alpha, \mathbf{m})$  of  $\mathcal{A}$ ,  $\mathcal{A}'$  queries  $\mathcal{O}'$  with  $\mathbf{m}' := (\mathbf{pk} + \alpha \cdot g, \mathbf{m})$ , to receive reply  $\sigma' = (R, u')$  - *this is a Schnorr'-signature of  $\mathbf{m}'$  with  $\mathbf{sk}$ , and we now convert this to a Schnorr-signature of  $\mathbf{m}$  with  $\mathbf{sk} + \alpha$* . Let  $c := \mathcal{R}(R, \mathbf{m}') = \mathcal{R}(R, \mathbf{pk} + \alpha \cdot g, \mathbf{m})$ . It sends  $\sigma := (R, u := u' + c\alpha)$  to  $\mathcal{A}$ .

We have

$$u \cdot g = u' \cdot g + c\alpha \cdot g = R + c \cdot \mathbf{pk} + c\alpha \cdot g = R + c \cdot (\mathbf{pk} + \alpha \cdot g).$$

So we have  $\text{verifySig}(\mathbf{pk} + \alpha \cdot g, \mathbf{m}, \sigma)$ . Also  $R$  is uniformly distributed, thus  $\mathcal{A}'$  is answering  $(\mathcal{A})$ 's queries with the same distribution  $\mathcal{O}$  would have.

Note that the mapping  $F(\alpha, \mathbf{m}, \sigma) := (\mathbf{m}', \sigma')$  where  $\mathbf{m}' := (\mathbf{pk} + \alpha \cdot g, \mathbf{m})$ ,  $\sigma' := (R, u - c\alpha)$  is injective.

Let  $T$  be the set of tuples  $(\alpha, \mathbf{m}, \sigma)$  such that  $\mathcal{A}$  queried  $(\alpha, \mathbf{m})$  and  $\mathcal{A}'$  answered  $\sigma$ . We have  $T' = \{F(x)\}_{x \in T}$ .

When  $\mathcal{A}$  finally outputs  $x^* = (\alpha^*, \mathbf{m}^*, \sigma^*)$ ;  $\mathcal{A}'$  outputs  $F(x^*)$ . As  $F$  is injective  $x^* \notin T$  implies  $F(x^*) \notin T'$ .

Denote  $(m', \sigma') := F(x^*)$ . From [4]'s results on unforgeability of Schnorr', the probability that

- $\text{verifySig}'(\mathbf{pk}, \mathbf{m}', \sigma')$ , and
- $(\mathbf{m}', \sigma') \notin T'$

is  $\text{negl}(\lambda)$ . Noting that  $\text{verifySig}'(\mathbf{pk}, \mathbf{m}', \sigma') \equiv \text{verifySig}(\mathbf{pk} + \alpha \cdot g, \mathbf{m}^*, \sigma^*)$ , this means that the probability that

- $\text{verifySig}(\mathbf{pk} + \alpha \cdot g, \mathbf{m}^*, \sigma^*)$ , and
- $x^* \notin T$

is  $\text{negl}(\lambda)$ . This is exactly what we had to prove.  $\square$

**Invertible group samplers** We assume that for our group  $\mathbb{G}$  we have an efficient randomized procedure **sample** that produces a group element in  $\mathbb{G}$  that is  $\text{negl}(\lambda)$ -close to uniform, such that there is an efficient deterministic algorithm **invert** that given the output of **sample**, produces w.p.  $1/\text{poly}(\lambda)$  over the randomness of **sample**, the randomness  $r$  used in that execution.

Note that when  $\mathbb{G}$  is an elliptic curve group over  $\mathbb{F}_r$  such a pair (**sample**, **invert**) is having **sample** try  $\lambda$  iterations of: Choose random  $x \in \mathbb{F}_r$ , check if there exists some  $(x, y) \in \mathbb{G}$ , if so output one of the two such elements randomly; and otherwise try another random  $x \in \mathbb{F}_r$ .

**invert**, given  $(x, y) \in \mathbb{G}$ , will output  $(x, \text{sign}(y))$ , which will be correct in the case that the first iteration of **sample** produced a good  $x$ , which happens w.p. approximately half.

We also need that Schnorr is a proof of knowledge of discrete log. For this, we state the following theorem that is almost implicit in [4]

**Theorem 2.4** (Extractability of Schnorr). *There is an algorithm  $\xi$  with the following property. Fix any efficient  $\mathcal{A}$  and group element  $\mathbf{g} \in \mathbb{G}$ . Suppose that  $\mathcal{A}$  produces w.p.  $\gamma$   $(\mathbf{pk}, \mathbf{m}, \sigma)$  such that  $\text{verifySig}_g^{\mathcal{R}}(\mathbf{pk}, \mathbf{m}, \sigma)$ . Then, given the internal randomness used by  $\mathcal{A}$  in the run,  $\xi$  produces w.p.  $\gamma/2$  over  $(\mathcal{A})$ 's randomness and its own randomness  $s \in \mathbb{F}_r$  such that  $\mathbf{pk} = s \cdot \mathbf{g}$ . Furthermore,  $\xi$ 's running time will be  $P(\lambda, 1/\gamma)$  where  $P$  is a polynomial depending on the running time of  $\mathcal{A}$ .*

### 3 Description of Sapling

#### 3.1 Basic components

##### Functions, and their requirements:

We do not explicitly state function domains and ranges; see the spec for more details. Whenever discussing a function in the properties below, we always think of an infinite sequence of functions indexed by the security parameter  $\lambda$ .

1. For any fixed values  $g, pk, v$ , and for any  $\epsilon \geq 0$ ,  $\mathbf{NC}(g, pk, v, rcm)$  is  $\epsilon$ -close to uniform when  $rcm$  is  $\epsilon$ -close to uniform.
2.  $\mathbf{NC}$  is collision resistant - i.e. the probability of finding  $note, note'$  such that  $\mathbf{NC}(note) = \mathbf{NC}(note')$  is  $\text{negl}(\lambda)$ .<sup>1</sup>
3. For any fixed  $v$  and any  $\epsilon \geq 0$ ,  $\mathbf{VC}(v, rcv)$  is  $\epsilon$ -close to uniform whenever  $rcv$  is  $\epsilon$ -close to uniform.
4.  $\mathbf{VC}$  is collision-resistant.
5. **sighash** is collision-resistant.
6. **IVK** is collision-resistant.
7. **NF** is collision resistant (see another requirement for the indistinguishability property in Section 5).

**Generators of  $\mathbb{G}$**  We assume we are given generators  $g_{sig}, g_n, g_r, g_v$  that were sampled in a way that except w.p  $\text{negl}(\lambda)$  an efficient  $\mathcal{A}$  cannot discover the discrete log relation between any two of them.

##### Statements:

OUT( $cv, cm, epk$ ): I know  $note = (g, pk, v, rcm), rcv, esk$  such that

1.  $cm = \mathbf{NC}(note)$ .
2.  $cv = \mathbf{VC}(v, rcv)$ .
3.  $epk = esk \cdot g$ .
4.  $g$  has order greater than eight.

SPEND( $rt, cv, nf, rk$ ): I know  $path, pos, note = (g, pk, v, rcm), cm, rcv, \alpha, ak, nsk$  such that

1.  $cm = \mathbf{NC}(note)$ .
2. Either  $v = 0$  (“dummy note”); or  $path$  is a merkle path from  $cm$  at position  $pos$  to  $rt$ .

---

<sup>1</sup>A caveat here is that this is true when the  $rcm$  parameter is thought of as a field element; in the actual circuit it is received as a string of bits where some elements of  $\mathbb{F}_r$  have multiple representations; inspection of the proof shows that it suffices that CR w.r.t  $rcm$  as a field element; same story with  $rcv$  in  $\mathbf{VC}$ .

3.  $rk = ak + \alpha \cdot g_{sig}$ .
4. Setting  $nk := nsk \cdot g_n$ ,  $ivk := \mathbf{IVK}(ak, nk)$ ; we have  $pk = ivk \cdot g$ .
5.  $nf = \mathbf{NF}(nk, cm, pos)$

## Components

A *note* is a tuple  $note = (g, pk, v, rcm)$  where

1.  $g, pk \in \mathbb{G}$ .
2.  $v, rcm \in \mathbb{F}_r$ .
3.  $v \leq \text{MAX}$ .

An *output base*  $output = (g, pk, v)$  is the same as a note excluding the  $rcm$  component.

**Remark 3.1.** *It is convenient for us to define a note with  $g$  rather than its GH-preimage  $d$  as in the spec, as this is what's given as input to the circuits; there are minor non-exploitable issues with this, see e.g. <https://github.com/zcash/zcash/issues/3490>.*

For  $ivk \in \mathbb{F}_r$  we say *note belongs to  $ivk$*  if  $pk = ivk \cdot g$ .

An *input base*, usually denoted  $input$ , will consist of the values required to make an input in a Sapling transaction, except the spending key; namely  $input = (note, path, pos, pak)$  where

- $note$  is a note
- $path$  is a path in a merkle tree beginning from a leaf of value  $cm = \mathbf{NC}(note)$ .
- $pos$  is the position of  $cm$  amongst the leaves of the Merkle tree ( $pos$  is redundant here as it can be derived from  $path$ , but convenient).
- $pak$  is a proving key to make SNARK spend proofs about the note.

We say  $input$  is *consistent with  $rt$*  if  $path$  ends at  $rt$ .

A *transaction input*, usually denoted  $inp$ , is the final form in which an input appears in a transaction;  $inp$  consists of

1. A value commitment  $cv$ .
2. A nullifier  $nf$ .
3. A Merkle root  $rt$  of the tree containing the used note.
4. A public key  $rk$  that is (allegedly) a randomized version of the spent note's proving key  $ak$ .
5. A SNARK proof  $\pi$  for the statement  $\text{SPEND}(rt, cv, nf, rk)$ .

### 3.2 Methods

We use the convention that  $\ell$  denotes the number of inputs in a transaction, and  $s$  the number of outputs.

**makeinp**(rt, input = (note, path, pos, pak), rcv,  $\alpha$ )  
 where input is an input base consistent with rt.

1.  $\text{cm} = \mathbf{NC}(\text{note})$
2.  $\text{nf} = \mathbf{NF}(\text{nk}, \text{note}, \text{pos})$
3. Define  $\text{rk} := \text{ak} + \alpha \cdot \mathbf{g}_{\text{sig}}$ ,  $\text{cv} := \mathbf{v} \cdot \mathbf{g}_{\mathbf{v}} + \text{rcv} \cdot \mathbf{g}_{\mathbf{r}}$ .
4. Let  $\pi = \pi_{\text{spend}}(\text{cv}, \text{rt}, \text{nf}, \text{rk}; \text{note}, \text{pak}, \alpha, \text{path}, \text{pos})$ .
5. Output  $\text{inp} = (\text{cv}, \text{rt}, \text{nf}, \text{rk}, \pi)$ .

**makeout** (note = (g, pk, v, rcm), rcv),

1. Choose random  $\text{esk} \in \mathbb{F}_r$ .
2. Let  $\text{cv} := \mathbf{VC}(\mathbf{v}, \text{rcv}) = \mathbf{v} \cdot \mathbf{g}_{\mathbf{v}} + \text{rcv} \cdot \mathbf{g}_{\mathbf{r}}$ .
3. Let  $\text{note} = (\mathbf{g}, \text{pk}, \mathbf{v}, \text{rcm})$  and  $\text{cm} := \mathbf{NC}(\text{note})$ .
4. Let  $\text{epk} = \text{esk} \cdot \mathbf{g}$ .
5. Let  $\text{enc} = \mathbf{ENC}_{\mathbf{KDF}(\text{esk} \cdot \text{pk}, \text{epk})}(\text{note})$
6. Let  $\pi = \pi_{\text{output}}(\text{epk}, \text{cm}, \text{cv}; \text{note}, \text{rcv}, \text{esk})$ .
7. Output  $(\text{cv}, \text{cm}, \text{epk}, \pi, \text{enc})$

**makerandomizedout** (note = (g, pk, v), rcv),

1. Choose random  $\text{esk}, \text{rcm} \in \mathbb{F}_r$ .
2. Let  $\text{cv} := \mathbf{VC}(\mathbf{v}, \text{rcv}) = \mathbf{v} \cdot \mathbf{g}_{\mathbf{v}} + \text{rcv} \cdot \mathbf{g}_{\mathbf{r}}$ .
3. Let  $\text{note} = (\mathbf{g}, \text{pk}, \mathbf{v}, \text{rcm})$  and  $\text{cm} := \mathbf{NC}(\text{note})$ .
4. Let  $\text{epk} = \text{esk} \cdot \mathbf{g}$ .
5. Let  $\text{enc} = \mathbf{ENC}_{\mathbf{KDF}(\text{esk} \cdot \text{pk}, \text{epk})}(\text{note})$
6. Let  $\pi = \pi_{\text{output}}(\text{epk}, \text{cm}, \text{cv}; \text{note}, \text{rcv}, \text{esk})$ .
7. Output  $(\text{cv}, \text{cm}, \text{epk}, \pi, \text{enc})$

**bindval** ( $\text{raw}_{\text{tx}} = (\overrightarrow{\text{inp}}, \overrightarrow{\text{out}}, \mathbf{v}^{\text{bal}}), \overrightarrow{\text{rcv}}$ )

1. Let  $r := \sum_{i=1}^{\ell} \text{rcv}_i - \sum_{i=\ell+1}^{\ell+s} \text{rcv}_i$
2. Let  $S := \sum_{i=1}^{\ell} \text{cv}_i - \sum_{i=\ell+1}^{\ell+s} \text{cv}_i - \mathbf{v}^{\text{bal}} \cdot \mathbf{g}_{\mathbf{v}}$

3. Let  $\sigma_{\text{bind}} := \text{sign}_{g_r}(r, \text{sighash}(\text{raw}_{\text{tx}}))$ .

4. Output  $\text{pre-tx} := (\text{raw}_{\text{tx}}, \sigma_{\text{bind}})$ .

**signtx**( $\text{pre-tx} = (\text{raw}_{\text{tx}}, \sigma_{\text{bind}})$ ,  $\vec{\text{ask}}$ ,  $\vec{\alpha}$ )

1. For each  $i \in [\ell]$ , let  $\sigma_i := \text{sign}_{g_{\text{sig}}}(\text{ask}_i + \alpha_i, \text{sighash}(\text{raw}_{\text{tx}}))$

2. Let  $\vec{\sigma} := (\sigma_1, \dots, \sigma_\ell)$ .

3. Output  $(\text{raw}_{\text{tx}}, \vec{\sigma})$ .

Given  $(\text{rt}, v^{\text{bal}})$  we say  $(\vec{\text{input}}, \vec{\text{output}})$  is *consistent* with  $\text{rt}, v^{\text{bal}}$ , if

- for each  $j \in [\ell]$   $\text{input}_j$  is consistent with  $\text{rt}$ , i.e.  $\text{pak}_j$  is from  $\text{NC}(\text{note}_j)$  to  $\text{rt}$ ,
- $\sum_{j=1}^{\ell} v_j - \sum_{j=\ell+1}^{\ell+s} v_j = v^{\text{bal}}$ .
- the positions  $\{\text{pos}_j\}_{j \in [\ell]}$  are all distinct.

and

**makerandomizedtx** ( $\text{rt}, v^{\text{bal}}, \vec{\text{input}}, \vec{\text{output}}$ )

where  $\text{input}_j = (\text{note}_j, \text{pak}_j, \text{path}_j, \text{pos}_j)$ ,  $\text{output}_j = (g_j, \text{pk}_j, v_j)$

1. Choose random  $\vec{\text{rcv}} \in \mathbb{F}_r^s$ .

2. For  $j \in [\ell]$ ,  $\text{inp}_j = \text{makeinp}(\text{rt}, \text{input}_j, \text{rcv}_j)$

3. For  $j \in [s]$ ,  $\text{out}_j = \text{makeout}(\text{output}_j, \text{rcv}_j)$

4.  $\text{pre-tx} = \text{bindval}(\vec{\text{inp}}, \vec{\text{out}}, v^{\text{bal}})$ .

5. Choose random  $\vec{\alpha} \in \mathbb{F}_r^\ell$ .

6. Output  $\text{tx} = \text{signtx}(\text{pre-tx}, \text{ask}, \vec{\alpha})$

**maketx** ( $\vec{\text{input}}, \vec{\text{output}}, \vec{\text{rcv}}, \text{ask}, \text{pak}$ ) where  $\text{input}_j = (v_j, \text{note}_j, \text{pak}_j, \text{path}_j, \text{pos}_j)$ ,  $\text{output}_j = (g_j, \text{pk}_j, v_j, \text{rcm}_j)$

1. Choose random  $\vec{\alpha} \in \mathbb{F}_r^\ell$ .

2. For  $j \in [\ell]$ ,  $\text{inp}_j = \text{makeinp}(\text{input}_j, \text{rcv}_j, \alpha_j, \text{pak})$

3. For  $j \in [s]$ ,  $\text{out}_j = \text{makeout}(\text{output}_j, \text{rcv}_j)$

4. Let  $v^{\text{bal}} := \sum_{i=1}^{\ell} v_i - \sum_{j=\ell+1}^{\ell+s} v_j$ .

5.  $\text{pre-tx} = \text{bindval}(\vec{\text{inp}}, \vec{\text{out}}, v^{\text{bal}}, \vec{\text{rcv}})$ .

6. Let  $\text{tx} = \text{signtx}(\text{pre-tx}, \vec{\alpha}, \text{ask})$

**verify-tx**( $L, \text{tx}$ )



1. Suppose that  $\text{tx} = (\text{raw}_{\text{tx}}, \vec{\sigma})$ .
2. For each  $\text{inp}_i = (\text{rt}, \text{cv}, \text{nf}, \text{rk}, \pi) \in \vec{\text{inp}}(\text{tx})$ ,
  - Check that  $\text{nf} \notin \text{nf}(\text{L}) \cup \{\text{nf}(\text{inp}_1), \dots, \text{nf}(\text{inp}_{i-1})\}$ .
  - Check that  $\text{spendverify}(\text{rt}, \text{cv}, \text{nf}, \text{rk}; \pi)$ .
  - Check that  $\text{verifySig}_{\text{g}_{\text{sig}}}^{\mathcal{R}}(\text{rk}, \text{sighash}(\text{raw}_{\text{tx}}), \sigma_i)$
3. For each  $\text{out} = (\text{cv}, \text{cm}, \text{epk}, \pi, \text{enc}) \in \vec{\text{out}}(\text{tx})$ , check that  $\text{outverify}(\text{cv}, \text{cm}, \text{epk}; \pi)$
4. Let  $S := \sum_{i=1}^{\ell} \text{cv}_i - \sum_{i=\ell+1}^{\ell+s} \text{cv}_i - v^{\text{bal}} \cdot \text{g}_v$ .
5. Check that  $\text{verifySig}_{\text{g}_r}^{\mathcal{R}}(S, \text{sighash}(\text{raw}_{\text{tx}}), \sigma_{\text{bind}})$ .

## 4 Non-Malleability of Sapling w.r.t. delegated spenders

We make the simplifying assumption when modelling non-malleability in this writeup; that *there is only one spending key* ( $\text{ask}, \text{nsk}$ ) *of the honest signer involved, and all addresses are diversified addresses derived from this spending key.*

### Modelling the adversary:

We wish to show that the delegated spender cannot create any new transactions of her own. We model this claim with the following non-malleability game: We model the honest signer as an oracle  $\mathcal{O}$  that  $\mathcal{A}$  interacts with as follows.

$\mathcal{O}$  begins by choosing a new spending key  $(\text{ask}, \text{nsk}) \leftarrow \mathcal{K}$  and sending the corresponding proof authorizing key  $\text{pak} = (\text{ak}, \text{nsk})$  to  $\mathcal{A}$ . Where  $\text{ak} = \text{ask} \cdot \text{g}_{\text{sig}}$ .

Afterwards,  $\mathcal{A}$  can make **sign-all-inputs** queries to  $\mathcal{O}$ , which intuitively correspond to asking for signatures on transactions whose inputs have spending key  $(\text{ask}, \text{nsk})$  (though see remark).

### Sign-all-inputs queries

1.  $\mathcal{A}$  sends  $(\text{pre-tx} = (\text{raw}_{\text{tx}}, \sigma_{\text{bind}}), \vec{\alpha})$  to  $\mathcal{O}$ . Where  $\text{raw}_{\text{tx}} = (\vec{\text{inp}}, \vec{\text{out}}, v^{\text{bal}})$
2.  $\mathcal{O}$  checks if  $\text{spendverify}(\text{pub}_i, \pi_i)$  holds for each  $i \in [\ell]$  and otherwise aborts.
3.  $\mathcal{O}$  computes for  $i \in [\ell]$ ,  $\sigma_i = \text{sign}_{\text{g}_{\text{sig}}}(\text{ask} + \alpha_i, \text{sighash}(\text{raw}_{\text{tx}}))$ .
4. Let  $\vec{\sigma} := (\sigma_1, \dots, \sigma_{\ell})$ .  $\mathcal{O}$  return  $\text{tx} := (\text{raw}_{\text{tx}}, \sigma_{\text{bind}}, \vec{\sigma})$ .

**Remark 4.1.** *The second item is another way of saying we assume  $\mathcal{A}$  can only ask  $\mathcal{O}$  for signatures of transactions with legitimate spend proofs. Otherwise the proof currently fails as we need to be able to extract the witness from each input.*

**Terminology:** We refer below to a transaction  $\text{tx}$  as  $\text{tx} = (\text{raw}_{\text{tx}}, \sigma_{\text{bind}}, \vec{\sigma})$ , where  $\vec{\sigma}$  contains the  $\ell$  input signatures and  $\sigma_{\text{bind}}$  is as described above in **maketx** that are added during **sign-all-inputs** and the signature  $\sigma_{\text{bind}}$  added in the last step of **maketx**.

Non-malleability says,  $\mathcal{A}$  should not be able to create a new valid transaction with inputs belonging to  $\mathcal{O}$ , even after seeing transactions of its choice with inputs of  $\mathcal{O}$ . New will mean that the  $\text{raw}_{\text{tx}}$  part will be new. (If we had changed the signature scheme to sign in order and have each signature sign the previous ones we could have required that  $\text{tx}$  including the signature part must be different from all previous transactions).

The way we formalize “transaction with inputs of  $\mathcal{O}$ ” is that the transaction created by  $\mathcal{A}$  contains overlapping nullifiers with the transactions signed previously by  $\mathcal{O}$ ; precisely transactions that are outputs of **sign-all-inputs** queries.

**Remark 4.2.** *A somewhat odd thing about the construction with the delegated spender, is that valid transactions signed by  $\mathcal{O}$ , do not exactly correspond to transactions whose inputs  $\mathcal{O}$  knows the spending key of. We can only say  $\mathcal{O}$  and  $\mathcal{A}$  together know the spending key. For example, given  $(\text{ak}, \text{nsk})$ ,  $\mathcal{A}$  can choose random  $s \in \mathbb{F}_r$ , set  $\text{ak}' := \text{ak} + s \cdot \mathbf{g}_{\text{sig}}$ . Now when  $\mathcal{A}$  wants to sign an input in address  $\text{ak}'$ , i.e. with some randomized key  $\text{rk} = \text{ak}' + \alpha \mathbf{g}_{\text{sig}} = \text{ak} + (s + \alpha) \cdot \mathbf{g}_{\text{sig}}$ , it can give  $\mathcal{O}$  the randomization  $\alpha' = s + \alpha$ .*

*A way to avoid these oddities is to have  $\mathcal{O}$  only sign transactions where he recognizes the nullifiers as belonging to a note of his. For our purposes here, we get a stronger result without this restriction by showing non-malleability holds when  $\mathcal{O}$  signs any transaction.*

**Some more terminology** Given a validly formatted transaction  $\text{tx} = ((\vec{\text{inp}}, \vec{\text{out}}, \text{v}^{\text{bal}}), \sigma_{\text{bind}}, \vec{\sigma})$ , we define

- $\text{nf}(\text{tx})$  to be the set of nullifiers appearing in one of its inputs; so  $\text{nf}(\text{tx}) := \{\text{nf}(\text{inp})\}_{\text{inp} \in \vec{\text{inp}}}$ .
- $\text{rk}(\text{tx})$  the set of randomized public keys appearing in inputs of  $\text{tx}$ , so  $\text{rk}(\text{tx}) := \{\text{rk}(\text{inp})\}_{\text{inp} \in \vec{\text{inp}}}$ .
- $\text{raw}(\text{tx}) := (\vec{\text{inp}}, \vec{\text{out}}, \text{v}^{\text{bal}})$ . For a set  $T$  of validly formed transactions we define  $\text{raw}(T) := \{\text{raw}(\text{tx})\}_{\text{tx} \in T}$ .

**Claim 4.3** (Non-malleability w.r.t delegated spenders). *Fix any efficient  $\mathcal{A}$  interacting with  $\mathcal{O}$  as described above. Let  $T = \{\text{tx}'\}$  be the set of transactions that are replies of  $\mathcal{O}$  to  $\mathcal{A}$ ’s **sign-all-inputs** queries. The probability that  $\mathcal{A}$  manages to output a ledger  $L$  and transaction  $\text{tx}$  such that*

1.  $\text{verify-tx}(L, \text{tx}) = \text{acc}$ ,
2.  $\text{raw}(\text{tx})$  is not a prefix of an element of  $T$ .
3.  $\text{nf}(\text{tx}) \cap \text{nf}(\text{tx}') \neq \emptyset$  for some  $\text{tx}' \in T$ .

*is  $\text{negl}(\lambda)$ .*

*Proof.* Let  $\mathcal{A}$  be an algorithm that after interacting with  $\mathcal{O}$  as described above outputs  $L, \text{tx}$ . Let  $\epsilon$  be the probability that  $L, \text{tx}$  satisfy the above, and assume for contradiction  $\epsilon = 1/\text{poly}(\lambda)$ .

We construct  $\mathcal{A}'$  that receives a randomized forgery challenge for Schnorr as described in Definition 2.1, and with probability  $\epsilon - \text{negl}(\lambda)$  either

- outputs a collision of **sighash**
- outputs a collision of **NF**,
- outputs a collision of **IVK**,
- Constructs a signature forgery for Schnorr w.r.t randomization.

Then, from CR of **sighash**, **NF**, **NC**, **IVK** and Theorem 2.3 the claim follows.  $\mathcal{A}'$  works as follows:

1.  $\mathcal{A}'$  will receive a challenge  $\mathbf{ak}^*$  for the signature scheme Schnorr sampled by the procedure **sample** described in Section 2.
2.  $\mathcal{A}'$  chooses random  $\mathbf{nsk} \in \mathbb{G}$  and sends to  $\mathcal{A}$  the proof authorizing key  $\mathbf{pak} = (\mathbf{nsk}, \mathbf{ak})$
3. When  $\mathcal{A}$  makes a **sign-all-inputs** query  $(\mathbf{raw}_{\mathbf{tx}}, \vec{\alpha})$   $\mathcal{A}'$  first checks that the proofs in  $\mathbf{raw}_{\mathbf{tx}}$  are valid (as  $\mathcal{O}$  does in the description of **sign-all-inputs** queries) and then answers with  $\vec{\sigma}$  where  $\sigma_i := \mathcal{S}_{\text{sign}}(\mathbf{ak} + \alpha_i \cdot \mathbf{g}_{\text{sig}}, \mathbf{m})$ . If during invocations to  $\mathcal{S}_{\text{sign}}$ ,  $\mathcal{S}_{\mathcal{R}}$  is queried on a point on which  $\mathcal{A}$  queried  $\mathcal{R}$ ,  $\mathcal{A}'$  aborts. (Note that the point queried by  $\mathcal{S}_{\mathcal{R}}$  is  $(R, \mathbf{rk}, \mathbf{m})$  for a uniform  $R$  chosen only during the execution of  $\mathcal{S}_{\text{sign}}$ , so the probability such a point was already queried is  $\text{negl}(\lambda)$ .)
4. When  $\mathcal{A}'$  makes a query to  $\mathcal{R}$ ,  $\mathcal{A}$  answers according to  $\mathcal{R}$  unless the query has already been answered according to  $\mathcal{S}_{\mathcal{R}}$  during invocations of  $\mathcal{S}_{\text{sign}}$  in **sign-all-inputs** queries; in which case  $\mathcal{A}'$  answers according to  $\mathcal{S}_{\mathcal{R}}$ . (This doesn't change the distribution of  $\mathcal{R}$  from the perspective of  $\mathcal{A}$ .)
5. When  $\mathcal{A}$  outputs  $L, \mathbf{tx}$ :  $\mathcal{A}'$  checks that it indeed satisfies the challenge - that is  $\text{verify-tx}(L, \mathbf{tx})$ ;  $\mathbf{tx}$  contains an input  $\mathbf{inp}$  with  $\mathbf{nf} = \mathbf{nf}(\mathbf{inp})$  being equal to  $\mathbf{nf}(\mathbf{inp}')$  for some  $\mathbf{inp}' \in \mathbf{tx}'$  for some  $\mathbf{tx}' \in T$ ; appearing in one of the **sign-all-inputs** queries of  $\mathcal{A}$ ; and  $\mathbf{raw}_{\mathbf{tx}} \notin \mathbf{raw}(T)$ . If not,  $\mathcal{A}'$  aborts.
6.  $\mathcal{A}'$  checks if  $\mathbf{sighash}(\mathbf{raw}_{\mathbf{tx}}) = \mathbf{sighash}(\mathbf{raw}_{\mathbf{tx}}'')$  for some  $\mathbf{tx}'' \in T$  with  $\mathbf{raw}_{\mathbf{tx}} \neq \mathbf{raw}_{\mathbf{tx}}''$ . If so it outputs  $(\mathbf{raw}_{\mathbf{tx}}, \mathbf{raw}_{\mathbf{tx}}'')$  as a collision of **sighash**.

*Explanation of where we are so far:* Denote by  $\mathbf{rk}$  and  $\sigma$  the public key and signature in  $\mathbf{inp}$ . Let  $\mathbf{m} := \mathbf{sighash}(\mathbf{raw}_{\mathbf{tx}})$ .  $\sigma$  is a valid signature for message  $\mathbf{m}$  and public key  $\mathbf{rk}$ , and  $\mathbf{m}$  was never signed in reply to the **sign-all-inputs** queries by  $\mathcal{O}$ . To obtain a forgery w.r.t randomization for the challenge  $\mathbf{ak}^*$ , what is left is to find the  $\alpha^*$  such that  $\mathbf{rk} = \mathbf{ak}^* + \alpha^* \cdot \mathbf{g}_{\text{sig}}$ . The purpose of the next steps is to obtain such  $\alpha^*$  or a collision of one of our CRH functions.

7. Otherwise, denote by  $B$  the algorithm consisting of execution of all parties up to this point outputting  $\mathbf{tx}$  and  $\mathbf{tx}'$ . Note that  $B$ 's randomness consists<sup>2</sup> of that of  $\mathcal{A}, \mathcal{A}', \mathcal{R}$  used up to this point and the randomness of **sample**. Let  $\xi$  be the extractor guaranteed to exist for  $B$  for the input  $\mathbf{inp}$  in  $\mathbf{tx}$ . Recall that  $\xi$  requires  $B$ 's internal randomness to produce a SNARK witness.  $\mathcal{A}'$  can give  $\xi$  the randomness of  $\mathcal{A}', \mathcal{A}, \mathcal{R}$  used up to this point, but instead

---

<sup>2</sup>We have not defined collision-resistant functions too formally. To be more accurate we assume all CRH functions are "public" in the sense that their seed is just a random string, and this randomness is also one of the inputs to  $B$ .

of using the actual randomness of **sample** as input to  $\xi$ ,  $\mathcal{A}'$  uses the **invert** method to obtain this randomness correctly from **ak** with  $1/\text{poly}(\lambda)$  probability. Given this input, with probability  $1/\text{poly}(\lambda) - \text{negl}(\lambda) = 1/\text{poly}(\lambda)$ ,  $\xi$  outputs for the input **inp** in **tx** a witness  $w = (\text{note}, \text{pak} = (\text{ak}, \text{nsk}), \alpha, \text{path}, \text{pos})$ . Similarly there is an extractor  $\xi'$  for the input **inp'** in **tx'** giving us a witness  $w' = (\text{note}', \text{pak}' = (\text{ak}', \text{nsk}'), \alpha', \text{path}', \text{pos}')$ . If  $\xi$  or  $\xi'$  fails  $\mathcal{A}'$  aborts (note that the probability of both succeeding is  $1/\text{poly}(\lambda)$ ).

8. Let  $\text{nk} := \text{nsk} \cdot \mathbf{g}_n$ ,  $\text{nk}' := \text{nsk}' \cdot \mathbf{g}_n$ . We have

$$\mathbf{NF}(\text{nk}, \text{note}, \text{pos}) = \mathbf{NF}(\text{nk}', \text{note}', \text{pos}') = \text{nf}.$$

If  $\text{nk} \neq \text{nk}'$ ,  $\text{note} \neq \text{note}'$  or  $\text{pos} \neq \text{pos}'$ ,  $\mathcal{A}'$  outputs  $(\text{nk}, \text{note}, \text{pos}), (\text{nk}', \text{note}', \text{pos}')$  as a collision of **NF**.

9. Otherwise we have  $\text{note} = \text{note}' = (\mathbf{g}, \text{pk}, \mathbf{v}, \text{rcm})$ . Defining  $\text{ivk} := \mathbf{IVK}(\text{ak}, \text{nk})$ ,  $\text{ivk}' := \mathbf{IVK}(\text{ak}', \text{nk})$ , we have  $\text{pk} = \text{ivk} \cdot \mathbf{g} = \text{ivk}' \cdot \mathbf{g}$ . Thus,  $\text{ivk} = \text{ivk}'$ . (Important here that  $\text{ivk}$  representation is unique and it is cause dfn of **IVK** has  $\text{mod } 2^{\ell_{\text{ivk}}=251}$ .) If  $\text{ak} \neq \text{ak}'$ ,  $\mathcal{A}'$  outputs  $(\text{ak}, \text{nk}), (\text{ak}', \text{nk}')$  as a collision of **IVK**.
10. Otherwise, we have  $\text{ak} = \text{ak}'$ . Now,  $\mathcal{A}'$  knows  $\alpha^*$  such that  $\text{rk}' = \text{ak}^* + \alpha^* \cdot \mathbf{g}_{\text{sig}}$ , where  $\text{ak}^*$  was the forgery challenge from  $\mathcal{O}$  (as  $\mathcal{A}$  used  $(\alpha^*, \text{signhash}(\text{raw}_{\text{tx}}))$  in the **sign-all-inputs** query for **tx'** for input **inp'**). And also  $\text{rk}' = \text{ak}' + \alpha' \cdot \mathbf{g}_{\text{sig}}$ . So  $\text{ak} = \text{ak}' = \text{ak}^* + (\alpha^* - \alpha') \cdot \mathbf{g}_{\text{sig}}$ . And  $\text{rk} = \text{ak}^* + (\alpha^* - \alpha' + \alpha) \cdot \mathbf{g}_{\text{sig}}$ . Thus, in this case  $\mathcal{A}'$  outputs  $(\alpha^* - \alpha' + \alpha, \text{signhash}(\text{raw}_{\text{tx}}), \sigma)$  as a signature forgery w.r.t randomization of  $\text{ak}^*$ .

□

## 5 Indistinguishability w.r.t outside adversaries

For a sequence of random variables  $X_1, \dots, X_n$  it will be convenient in this section to denote  $X_{<i} := (X_1, \dots, X_{i-1})$ . Let us say that random variables  $X, Y$  are  $\gamma$ -independent if for any events  $A, B$

$$|\Pr(X \in A \wedge Y \in B) - \Pr(X \in A) \cdot \Pr(Y \in B)| \leq \gamma.$$

We recall that the *statistical distance* between  $X$  and  $Y$  is the maximum over all events  $T$  of

$$|\Pr(X \in T) - \Pr(Y \in T)|.$$

We say  $X, Y$  are  $\gamma$ -close if they have statistical distance at most  $\gamma$ .

A calculation proves

**Claim 5.1.** *Suppose  $X = (X_1, X_2), Y = (Y_1, Y_2)$  are such that*

- $X_1$  and  $Y_1$  are on the same range, are  $\gamma_1$ -independent and  $\gamma_1$ -close.
- e.w.p  $\gamma_2$  over the value  $(x_1, y_1)$  of  $(X_1, Y_1)$ ,  $(X|X_1 = x_1), (Y|Y_1 = y_1)$  are  $\gamma_3$ -independent.
- e.w.p  $\gamma_2$  over the value  $x_1$  of  $X_1$ , we have that  $(X|X_1 = x_1), (Y|Y_1 = x_1)$  are  $\gamma_3$ -close.

*Then  $X, Y$  are  $\gamma_1 + \gamma_2 + \gamma_3$ -independent and  $\gamma_1 + \gamma_2 + \gamma_3$ -close.*

Induction then shows that

**Claim 5.2.** *Suppose  $t = \text{poly}(\lambda)$ . Suppose random variables  $X = (X_1, \dots, X_t), Y = (Y_1, \dots, Y_t)$  are such that for any  $i \in [n]$ ,*

- *e.w.p  $\text{negl}(\lambda)$  over the value  $(x, y)$  of  $(X_{<i}, Y_{<i})$ ,  
 $(X_i | X_{<i} = x)$  and  $(Y_i | Y_{<i} = y)$  are  $\text{negl}(\lambda)$ -independent; and*
- *e.w.p  $\text{negl}(\lambda)$  over the value  $x$  of  $X_{<i}$ ,  $(X_i | X_{<i} = x)$  and  $(Y_i | Y_{<i} = x)$  are  $\text{negl}(\lambda)$ -close.*

*Then  $X, Y$  are  $\text{negl}(\lambda)$ -independent and  $\text{negl}(\lambda)$ -close.*

Below we use  $\mathcal{R}_{\text{sig}}$  to denote the random oracle used by the signature algorithm.

**Theorem 5.3.** *Assume that*

1. **NF**(nk, **NC**(note), pos) =  $\mathcal{R}(\text{nk}, \mathbf{MPH}(\text{note}, \text{pos}))$  where  $\mathcal{R}$  is a random oracle and **MPH** is a collision-resistant function<sup>3</sup>
2. **KDF** and  $\mathcal{R}_{\text{sig}}$  are also random oracles.
3. **ENC**<sub>K</sub>(m) produces a uniform output when  $K$  is uniform and  $m$  is fixed.
4. The SNARK we are using is witness indistinguishable - i.e. the proof distribution depends only on the public input and not on the witness.

*Then, the probability of an efficient  $\mathcal{A}$  finding  $\text{rt}, v^{\text{bal}}, \overrightarrow{\text{input}}, \overrightarrow{\text{output}}, \overrightarrow{\text{input}'}, \overrightarrow{\text{output}'}$  such that*

- $|\overrightarrow{\text{input}}| = |\overrightarrow{\text{input}'}| = \ell, |\overrightarrow{\text{output}}| = |\overrightarrow{\text{output}'}| = s.$
- *The positioned notes in  $\overrightarrow{\text{input}}$  and  $\overrightarrow{\text{input}'}$  are all distinct.*
- $(\overrightarrow{\text{input}}, \overrightarrow{\text{output}})$  and  $(\overrightarrow{\text{input}'}, \overrightarrow{\text{output}'})$  are both consistent with  $\text{rt}, v^{\text{bal}}.$
- *The distributions of the random variables  $D := \mathbf{makerandomizedtx}(\text{rt}, v^{\text{bal}}, \overrightarrow{\text{input}}, \overrightarrow{\text{output}})$  and  
 $D' := \mathbf{makerandomizedtx}(\text{rt}, v^{\text{bal}}, \overrightarrow{\text{input}'}, \overrightarrow{\text{output}'})$ , over the randomness of the oracles  $\mathcal{R}, \mathbf{KDF}$  and  $\mathcal{R}_{\text{sig}}$ , and the inner randomness of the signer, SNARK prover and the **makerandomizedtx** method, are not  $\text{negl}(\lambda)$ -close and  $\text{negl}(\lambda)$ -independent*

*is  $\text{negl}(\lambda)$ .*

*Proof.* Let us denote by  $(\overrightarrow{\text{inp}}, \overrightarrow{\text{out}}, \sigma_{\text{bind}}, \vec{\sigma})$  the output of **makerandomizedtx**(rt,  $v^{\text{bal}}, \overrightarrow{\text{input}}, \overrightarrow{\text{output}}$ ) and by  $(\overrightarrow{\text{inp}'}, \overrightarrow{\text{out}'}, \sigma'_{\text{bind}}, \vec{\sigma}')$  the output of **makerandomizedtx**(rt,  $v^{\text{bal}}, \overrightarrow{\text{input}'}, \overrightarrow{\text{output}'})$  when using independent inner randomness, but joint randomness for the oracles  $\mathcal{R}, \mathcal{R}_{\text{sig}}, \mathbf{KDF}$ .

We will consider  $D$  and  $D'$  as sequences of random variables  $D = (X_1, \dots, X_m), D' = (Y_1, \dots, Y_m)$ , and show that for every  $i \in [m]$  they satisfy the conditions of Claim 5.2.

We begin with the inputs. Letting, for  $i \in [\ell]$ ,  $X_i = \text{inp}_i, Y_i = \text{inp}'_i$ , the following claim shows those conditions hold for the first  $i \in [\ell]$ .

<sup>3</sup>The requirement here may seem a bit odd; it models the fact that **NC**(note) is a pedersen hash which is combined in **NF** with a pos-multiple of an independent group generator, followed by an application of BLAKE-2 on the result prefixed with nk. In particular, BLAKE-2 takes the place of  $\mathcal{R}$  in the implementation.

**Claim 5.4.** *E.w.p  $\text{negl}(\lambda)$  over the randomness of  $\mathcal{A}$ , for each  $i \in [\ell]$   $\text{inp}_i, \text{inp}'_i$  are identically distributed and independent given any fixing of  $\text{inp}_{<i}, \text{inp}'_{<i}$ .*

*Proof.* We show first that e.w.p.  $\text{negl}(\lambda)$  over the randomness of  $\mathcal{A}$ ,  $\text{inp}_i, \text{inp}'_i$  are independent conditioned on any fixing of  $\text{inp}_{<i}, \text{inp}'_{<i}$ .  $\text{inp}_1, \dots, \text{inp}_\ell, \text{inp}'_1, \dots, \text{inp}'_\ell$  are results of invocations of **makeinp** with independent randomness  $\text{rcv}, \alpha$  and independent randomness of the SNARK prover. Inspection shows the only opportunity for dependence amongst any two of them, even after conditioning on the value of the others, is having the random oracle  $\mathcal{R}$  queried at the same point during the invocations.  $\mathcal{R}$  is queried for the computation of **NF**; so this only happens if

$$(\text{nk}_i, \text{MPH}(\text{note}_i, \text{pos}_i)) = (\text{nk}'_i, \text{MPH}(\text{note}'_i, \text{pos}'_i)).$$

This implies  $\text{MPH}(\text{note}_i, \text{pos}_i) = \text{MPH}(\text{note}'_i, \text{pos}'_i)$ , but  $\mathcal{A}$  will only find such a collision w.p  $\text{negl}(\lambda)$ . When this doesn't happen  $\text{inp}_i$  and  $\text{inp}'_i$  are independent also given any fixing of the previous inputs.

Now to show they are identically distributed given a fixing of  $\text{inp}_{<i}, \text{inp}'_{<i}$ .

Suppose  $\text{inp}_i = (\text{nf}, \text{rt}, \text{rk}, \text{cv}, \pi)$ , and  $\text{inp}'_i = (\text{nf}', \text{rt}', \text{rk}', \text{cv}', \pi')$ . We show each element is identically distributed conditioned on any fixing of the previous ones.

- $\text{nf} = \mathcal{R}(q)$  and  $\text{nf}' = \mathcal{R}(q')$  where  $q = (\text{nk}, \text{MPH}(\text{note}, \text{pos})), q' = (\text{nk}', \text{MPH}(\text{note}', \text{pos}'))$ . These are both uniform *unless* one of the queries  $q, q'$  was already made to  $\mathcal{R}$  in a previous invocation; which would mean  $\{(\text{note}^*, \text{pos}^*)\}_{(\text{note}^*, \text{pos}^*) \in \text{inp}_{<i+1} \cup \text{inp}'_{<i+1}}$  contains a collision of **MPH** which  $\mathcal{A}$  can find only w.p  $\text{negl}(\lambda)$ .
- $\text{rt} = \text{rt}'$ .
- $\text{rk} = \text{ak} + \alpha \cdot \text{g}, \text{rk}' = \text{ak}' + \alpha' \cdot \text{g}$ . Are both uniform in  $\mathbb{G}$  because of the uniform choice of  $\alpha, \alpha'$  in **makerandomizedtx**.
- $\text{cv} = \text{v} \cdot \text{g}_\text{v} + \text{rcv} \cdot \text{g}_\text{r}, \text{cv}' = \text{v}' \cdot \text{g}'_\text{v} + \text{rcv}' \cdot \text{g}'_\text{r}$ . Are both uniform in  $\mathbb{G}$  because of the uniform choices of  $\text{rcv}, \text{rcv}' \in \mathbb{F}_r$  in the executions of **makerandomizedtx**.
- $\pi, \pi'$  - When  $(\text{nf}, \text{rt}, \text{rk}, \text{cv}) = (\text{nf}', \text{rt}', \text{rk}', \text{cv}')$ , it follows from the witness indistinguishability of the SNARK that  $\pi$  and  $\pi'$  are identically distributed. They are independent for any fixing of the previous values, as given this fixing the value of  $\pi, \pi'$  depends only on the inner randomness of the SNARK prover.

□

We proceed with the elements of the the outputs. It will be convenient now to view each element in  $\text{out}_j, \text{out}'_j$  as separate random variables  $X_i, Y_i$ , and show that

1. E.w.p  $\text{negl}(\lambda)$  over the fixing of  $X_{<i}$ , they are identically distributed given this fixing of both  $X_{<i}$  and  $Y_{<i}$ .
2. E.w.p  $\text{negl}(\lambda)$  over the fixing of  $X_{<i}, Y_{<i}$  they are independent given the fixing.

We show this for the different types of elements in  $\text{out}_j, \text{out}'_j$ :

- $cv = v \cdot g_v + rcv \cdot g_r, cv' = v' \cdot g_v + rcv' \cdot g_r$ : are independent and uniform in  $\mathbb{G}$  because of the independent uniform choices of  $rcv, rcv' \in \mathbb{F}_r$  in **makerandomizedtx**.
- $cm = \mathbf{NC}(g, pk, v, rcm), cm' = \mathbf{NC}(g', pk', v', rcm')$ : are uniform and independent in  $\mathbb{G}$  because of the independent uniform choices of  $rcm, rcm' \in \mathbb{F}_r$  in **makerandomizedout**.
- $epk = esk \cdot g, epk' = esk' \cdot g$  are uniform and independent in  $\mathbb{G}$  because of the independent and uniform choices of  $esk, esk' \in \mathbb{F}_r$  in **makerandomizedout**.
- $\pi, \pi'$  - Assuming the public inputs  $(epk, cm, cv) = (epk', cm', cv')$ , it follows from the witness indistinguishability of the SNARK that  $\pi$  and  $\pi'$  are identically distributed. They are independent for any fixing of the previous values, as given this fixing the value of  $\pi, \pi'$  depends only on the inner randomness of the SNARK prover.
- $enc = \mathbf{ENC}_{\mathbf{KDF}(k)}((g, pk, v)), enc' = \mathbf{ENC}_{\mathbf{KDF}(k')}((g', pk', v'))$  where  $k := (esk \cdot pk, epk)$  and  $k' := (esk' \cdot pk', epk')$ : Assuming  $k \neq k'$ , and moreover  $\{k, k'\}$  are different from all the “key seeds”  $\{k_j, k'_j\}$  used in previous outputs; we have that the encryption keys  $\mathbf{KDF}(k), \mathbf{KDF}(k')$  are uniform and independent of all previous variables. And thus by the theorem’s assumption that  $\mathbf{KDF}$  is a random oracle  $enc, enc'$  are uniform and independent in this case. Thus there are at most  $\ell$  values of the preceding  $esk$  and at most  $\ell$  values of the preceding  $esk'$  that can prevent  $enc$  and  $enc'$  from being uniform and independent; which is a  $\text{negl}(\lambda)$ -fraction of the possible values of the preceding values.

It is now left to deal with the signature elements.  $\sigma_{\text{bind}}, \sigma'_{\text{bind}}, \{\sigma_i, \sigma'_i\}$ .

The distribution of these elements is determined by the public key  $pk = rk_i$  (or  $pk = S$  the sum of value commitments in the case of  $\sigma_{\text{bind}}$ ), the message  $\mathbf{m} = \mathbf{sighash}(\text{raw}_{\text{tx}})$  they are signing, the internal randomness of the signing algorithm and the reply of the random oracle  $\mathcal{R}_{\text{sig}}$  on the query point  $(R, pk, \mathbf{m})$ . Thus, given a fixing of previous variables, the only case where a dependence between  $\sigma_i$  or  $\sigma'_i$  could be created is if there is a collision between the signatures in the choice of  $R$  which happens w.p.  $\text{negl}(\lambda)$ .  $\square$

## 5.1 Balance

The following claim states an adversary should not be able to create “money out of thin air”; or more specifically, extract more money from the shielded pool than was put in it. In Sapling, the value  $v^{\text{bal}} = v^{\text{bal}}(\text{tx})$  in a transaction  $\text{tx}$  corresponds to the alleged difference of spend and output values (see Section 4.12 in the spec) and  $\text{tx}$  is thought of as having ; thus over-extracting from the pool corresponds to a constructing a ledger where the sum of all  $v^{\text{bal}}$  values is strictly positive.

**Claim 5.5.** *The probability that an efficient  $\mathcal{A}$  generates ledger  $L = (\text{tx}_1, \dots, \text{tx}_n)$  such that*

$$\sum_{\text{tx} \in L} v^{\text{bal}}(\text{tx}) > 0$$

*is  $\text{negl}(\lambda)$ .*

*Proof.* Given  $\mathcal{A}$  that produces a ledger as in the claim statement w.p.  $\gamma$ , we construct an efficient  $\mathcal{A}'$  that w.p.  $\gamma/2 - \text{negl}(\lambda)$  produces a collision of **IVK**, **NC**, **treehash** or **VC**. It follows that  $\gamma = \text{negl}(\lambda)$ .

1.  $\mathcal{A}'$  begins by running  $\mathcal{A}$  and aborting if  $\mathcal{A}$  hasn't output a ledger as in the claim.
2. Otherwise, given such a ledger  $L$ ,  $\mathcal{A}'$  can apply an extractor for each SNARK proof in all inputs and outputs in all transactions. For each transaction input  $\text{inp} \in \text{tx} \in L$ ,  $\text{inp} = (\text{cv}, \text{nf}, \text{rt}, \text{rk}, \pi)$ , the extractor except w.p.  $\text{negl}(\lambda)$  outputs an input witness  $\text{inpwit} = (\text{input} = (\text{note}, \text{path}, \text{pos}), \text{pak}, \text{rcv}, \alpha)$ . We denote by  $\text{posnote}$  the *positioned note* corresponding to  $\text{inp}$ ,  $\text{posnote} := (\text{note}, \text{pos})$ . Similarly for every transaction output in some  $\text{tx}$  in  $L$ ,  $\text{out} = (\text{cv}, \text{cm}, \text{epk}, \pi, \text{enc})$ , the extractor outputs  $\text{outwit} = (\text{note}, \text{esk}, \text{rcv})$ . The value  $\text{pos}$  for the output note can be deduced from when it was added to  $L$ , i.e., the location of  $\text{cm}$  in the commitment tree. So again we can define for each  $\text{out}$ , the corresponding positioned note  $\text{posnote} = (\text{note}, \text{pos})$ . For  $i \in [n]$  let us denote respectively by  $\mathcal{I}_i, \mathcal{O}_i$  the positioned input and output notes in  $\text{tx}_i$  with non-zero value<sup>4</sup>.

We also use the extractor from theorem 2.4 to find  $s$  such that  $S = s \cdot \mathbf{g}_r$  where

$$S := \sum_{i=1}^{\ell} \text{cv}_i - \sum_{i=\ell+1}^{\ell+s} \text{cv}_i - \mathbf{v}^{\text{bal}} \cdot \mathbf{g}_v$$

is the public key in the value binding signature  $\sigma_{\text{bind}}$ .

If one of the extractor runs fails  $\mathcal{A}'$  aborts. Note that w.p. at least  $\gamma/2 - \text{negl}(\lambda)$   $\mathcal{A}'$  doesn't abort.

3.  $\mathcal{A}'$  checks if for some  $i \in [n]$  and  $\text{inp} \in \text{tx}_i$ ,  $\text{posnote}(\text{inp}) \notin \mathcal{O}_{< i}$ .

If so, let  $\text{tx} = \text{tx}_i$ . Let  $\text{rt}$  be the root of the tree used in the public input of  $\text{inp}$ ; this is the tree  $T_j$  formed from  $\{\text{tx}_1, \dots, \text{tx}_j\}$  for some  $j < i$ . Let  $\text{posnote} = (\mathbf{g}, \text{pk}, \mathbf{v}, \text{rcm}, \text{pos})$  and  $\text{cm} = \mathbf{NC}(\mathbf{g}, \text{pk}, \mathbf{v}, \text{rcm})$ .  $\text{inpwit}$  contains a path  $\text{path}$  from  $\text{cm}$  to  $\text{rt}$ . If  $\text{pos}$  is an index of a leaf in  $T_j$ , there exists an extended note  $\text{posnote}'$  that was inserted to this position when constructing the ledger and from  $\text{posnote}'$  we can derive a path  $\text{path}'$  from  $\text{cm}' = \mathbf{NC}(\mathbf{g}', \text{pk}', \mathbf{v}', \text{rcm}')$  in position  $\text{pos}$  to  $\text{rt}$ . If  $\text{path} \neq \text{path}'$ , then going down from  $\text{rt}$  to the first difference between  $\text{path}$  and  $\text{path}'$  (ask Sean/Daira : is  $T$  always a full tree with zeroes on other leaves? No you have filler values for the empty subtrees, need to check this are values that are hard to find route to - their impossible to find rout to - have no preimage) this difference gives a collision of *treehash* that  $\mathcal{A}'$  can output.

Otherwise, we have  $\text{cm} = \text{cm}'$ .  $\text{note}$  must be different from  $\text{note}'$  because  $\text{posnote}' = (\text{note}', \text{pos}) \in \mathcal{O}_{< i}$  but  $(\text{note}, \text{pos}) \notin \mathcal{O}_{< i}$ .

Thus  $\text{note}, \text{note}'$  is a collision of  $\mathbf{NC}$ . In this case,  $\mathcal{A}'$  outputs this collision and terminates.

Now suppose  $\text{pos}$  is not a position of a leaf in  $T_j$ . This means there is only a partial path  $\text{path}'$  in  $T_j$  from  $\text{rt}$  to a filler value with no preimage (see spec for details). So, similarly we follow  $\text{path}$  and  $\text{path}'$  to their first difference - a difference that must exist because of the filler value; and this gives us a collision of *treehash* that  $\mathcal{A}'$  outputs.

4. Now  $\mathcal{A}'$  checks if as a multiset  $\mathcal{I} := \mathcal{I}_1 \cup \dots \cup \mathcal{I}_n$  contains a repetition. That is, there exists  $\text{posnote} = (\mathbf{g}, \text{pk}, \mathbf{v}, \text{rcm}, \text{pos})$  such that for two distinct transaction inputs  $\text{inp} = (\text{cv}, \text{nf}, \text{rt}, \text{rk}, \pi)$ ,  $\text{inp}' =$

---

<sup>4</sup>Sapling enables the creation of dummy notes with zero value, for which the spend statement doesn't check Merkle path validity, cf. Section 4.7.2 in the spec).



$(cv', nf', rt', rk', \pi')$  in  $L$ ; if the corresponding extracted witnesses are  $\text{inpwt} = (\text{input} = (\text{note}, \text{path}, \text{pos}), \text{pak}, \text{rcv}, \alpha)$ ,  $\text{inpwt}' = (\text{input}' = (\text{note}', \text{path}', \text{pos}'), \text{pak}', \text{rcv}', \alpha')$ ; then  $(\text{note}, \text{pos}) = (\text{note}', \text{pos}') = \text{posnote}$ .

We show in this case that  $\mathcal{A}'$  can output a collision of **IVK**:

Let  $\text{cm} = \mathbf{NC}(\mathbf{g}, \text{pk}, \mathbf{v}, \text{rcm})$ . Since  $\text{nf} \neq \text{nf}'$ , and  $\text{nf} = \mathbf{NF}(\text{nk}, \text{note}, \text{pos})$ ,  $\text{nf}' = \mathbf{NF}(\text{nk}', \text{note}, \text{pos})$ ; we have  $\text{nk} \neq \text{nk}'$ .

Also  $\text{ivk} = \mathbf{IVK}(\text{ak}, \text{nk})$ ,  $\text{ivk}' = \mathbf{IVK}(\text{ak}', \text{nk}')$ , and  $\text{pk} = \text{ivk} \cdot \mathbf{g} = \text{ivk}' \cdot \mathbf{g}$ . So  $\text{ivk} = \text{ivk}'$ . And thus,  $\mathcal{A}'$  can output  $(\text{ak}, \text{nk}), (\text{ak}', \text{nk}')$  as a collision of **IVK**.

5. Let us denote by  $\text{bal}(\text{tx})$  the (integer) sum of values in inputs of  $\text{tx}$  minus the sum of values in output of  $\text{tx}$  (notes meaning those output by the extractors); and by  $\mathbf{rcv}(\text{tx})$  the sum of values  $\text{rcv}$  in input witnesses of  $\text{tx}$  minus the sum of values  $\text{rcv}$  in output witnesses of  $\text{tx}$ . When reaching this point with no output we know that:

For each  $i \in [n]$ ,  $\mathcal{I}_i \subset \mathcal{O}_1 \cup \dots \cup \mathcal{O}_{i-1} \setminus (\mathcal{I}_1 \cup \dots \cup \mathcal{I}_{i-1})$ .

This implies

$$\sum_{\text{tx} \in L} \text{bal}(\text{tx}) \leq 0.$$

We claim that we must have for some  $\text{tx} \in L$ ,  $\text{bal}(\text{tx}) \neq v^{\text{bal}}(\text{tx})$ : Otherwise, we would have

$$\sum_{\text{tx} \in L} v^{\text{bal}}(\text{tx}) = \sum_{\text{tx} \in L} \text{bal}(\text{tx}) \leq 0,$$

contradicting the fact that  $\mathcal{A}$  has managed to output  $L$  with a positive sum of  $v^{\text{bal}}$  values.

Thus, let  $\text{tx} = \text{tx}_i$  be such that  $\text{bal}(\text{tx}) \neq v^{\text{bal}}(\text{tx})$ . We show in the next step how  $\mathcal{A}'$  uses this to output a collision of **VC**.

6. At this point, we know that  $\text{bal}(\text{tx}) \neq v^{\text{bal}}(\text{tx})$ . As both these values are in the open interval<sup>5</sup>  $(-r/2, r/2)$ , we have also  $\text{bal}(\text{tx}) \neq v^{\text{bal}}(\text{tx}) \pmod{r}$ . Suppose we are in this case with probability  $\gamma$ . We show how to find a collision of **VC** with probability  $\gamma/\text{poly}(\lambda)$ . Since  $\text{tx}$  verifies, we know that  $\text{verifySig}_{\mathbf{g}_r}^{\mathcal{R}}(S, \text{sighash}(\text{raw}_{\text{tx}}), \sigma_{\text{bind}})$  for

$$S = \sum_{i=1}^{\ell} \text{cv}_i - \sum_{i=\ell+1}^{\ell+s} \text{cv}_i - v^{\text{bal}} \cdot \mathbf{g}_v = \left( \sum_{i=1}^{\ell} v_i - \sum_{i=\ell+1}^s v_i \right) \cdot \mathbf{g}_v + \left( \sum_{i=1}^{\ell} \text{rcv}_i - \sum_{i=\ell+1}^s \text{rcv}_i \right) \cdot \mathbf{g}_r - v^{\text{bal}} \cdot \mathbf{g}_v.$$

Using Theorem 2.4, we can with probability  $\gamma/2$  we can use the forking lemma to rewind  $\mathcal{A}$  while altering the response of  $\mathcal{R}$  on the signature challenge in  $\sigma_{\text{bind}}$ , and find  $s$  such that  $s \cdot \mathbf{g}_r = S$ . Thus, we have  $\mathbf{VC}(0, s) = S$ .

Let  $R := \sum_{i=1}^{\ell} \text{rcv}_i - \sum_{i=\ell+1}^s \text{rcv}_i$  and  $v := \text{bal}(\text{tx}) - v^{\text{bal}}(\text{tx})$ . We also have  $\mathbf{VC}(v, R) = S$ . Hence  $\mathcal{A}'$  can output  $(0, s), (v, R)$  as a collision of **VC**.

□

---

<sup>5</sup>See the spec for details:  $v^{\text{bal}}$  and  $\mathbf{v}$  in each transaction input/output are at most  $2^{64}$  in absolute value, so assuming less than, e.g.,  $2^{r-66}$  transaction inputs and outputs in any transaction, this is true.

## 5.2 Spendability

**Valid transaction bases:** A sequence  $x = (\overrightarrow{\text{input}}, \overrightarrow{\text{output}}, v^{\text{bal}})$  is a *valid transaction base* if  $v^{\text{bal}} = \sum v(\text{input}_i) - \sum v(\text{output}_j)$ .

We review note encryption and decryption from the spec in our notation.

**Decrypting notes:**

$\text{dec}(\text{ivk}, \text{out} = (\text{cv}, \text{cm}, \text{epk}, \pi, \text{enc}))$

1. Let  $K := \text{KDF}(\text{epk} \cdot \text{ivk})$
2. Let  $\text{np} = \text{DEC}_K(\text{enc})$ . If  $\text{DEC}()$  fails output *rej*.
3. Suppose  $\text{np} = (\text{d}, \text{v}, \text{rcm}, \text{memo})$ . If  $\text{rcm} \geq r$  output *rej*.
4. Let  $g := \text{GH}(\text{d})$ .
5. Let  $\text{pk} := g \cdot \text{ivk}$ . Let  $\text{note} := (g, \text{pk}, \text{v}, \text{rcm})$ .
6. Check that  $\text{cm} = \text{NC}(\text{note})$ . Output *rej* if not.
7. Output *note*.

We define

$$\begin{aligned} \text{dec}(\text{ivk}, \text{tx}) &:= \cup_{\text{out} \in \text{tx}} \text{dec}(\text{ivk}, \text{out}), \\ \text{dec}(\text{ivk}, \text{L}) &:= \cup_{\text{tx} \in \text{L}} \text{dec}(\text{ivk}, \text{tx}) \end{aligned}$$

And also

$$\text{nf}(\text{tx}) := \cup_{\text{inp} \in \overrightarrow{\text{inp}}(\text{tx})} \text{nf}(\text{inp}), \text{nf}(\text{L}) := \cup_{\text{tx} \in \text{L}} \text{nf}(\text{tx})$$

In the spendability game  $\mathcal{A}$  tries to create a ledger where a note successfully decrypted with  $\text{ivk}$  cannot be spent. Formally, the game proceeds as follows.

1. We choose uniform  $\text{sk} = (\text{ask}, \text{nsk})$ ; and give  $\text{pak} = (\text{ask} \cdot \text{g}_{\text{sig}}, \text{nsk})$  to  $\mathcal{A}$ .
2.  $\mathcal{A}$  outputs a ledger  $\text{L}$ , a positioned note  $(\text{note}, \text{pos})$ , a set of output notes  $\overrightarrow{\text{output}}$ , and a set of incoming viewing keys  $\overrightarrow{\text{ivk}}$ .
3. We choose random  $\overrightarrow{\text{rcv}} \in \mathbb{F}_r^{\ell+s}$  and compute  $\text{tx} = \text{maketx}(\overrightarrow{\text{input}}, \overrightarrow{\text{output}}, v^{\text{bal}}, \text{ask})$ .
4. Let  $\text{ivk} := \text{IVK}(\text{ak}, \text{nk})$ .  $\mathcal{A}$  wins iff
  - (a)  $\text{note} \in \text{dec}(\text{ivk}, \text{L})$ .
  - (b)  $((\text{note}), \overrightarrow{\text{output}}, v^{\text{bal}})$  is a valid transaction base.
  - (c) For each  $i \in [s]$ ,  $\text{output}_i$  belongs to  $\text{ivk}_i$ .
  - (d)  $\text{verify-tx}(\text{L}, \text{tx})$ .
  - (e) For some  $i \in [s]$ ,  $\text{dec}(\text{ivk}_i, \text{out}_i)$  does not return  $\text{output}_i$ .

We wish to show that the success of any efficient  $\mathcal{A}$  in this game is  $\text{negl}(\lambda)$ .

Let  $\text{nk} = \text{nsk} \cdot \text{g}_n$ . Inspection of the protocol shows this exactly corresponds to the nullifier of  $\text{note}$  with nullifier key  $\text{nk}$  already appearing in the ledger. Thus, it suffices to prove the following.

**Claim 5.6.** Fix any efficient  $\mathcal{A}$ . Suppose that  $\mathcal{A}$  is given uniformly chosen  $\text{pak}$ , and let  $\text{ivk} := \text{IVK}(\text{pak})$ . The probability that  $\mathcal{A}$  generates a ledger  $L$  and positioned note  $(\text{note}, \text{pos})$  such that

1.  $(\text{note}, \text{pos}) \in \text{dec}(\text{ivk}, L)$
2.  $\text{NF}(\text{nk}, \text{NC}(\text{note}), \text{pos}) \in \text{nf}(L)$

is  $\text{negl}(\lambda)$ .

*Proof.* Let  $\gamma$  be the probability that  $\mathcal{A}$  outputs  $L, \text{note}$  satisfying the two properties in the claim. We construct an efficient  $\mathcal{A}'$  that receives a forgery challenge  $\text{ak}$  of Schnorr and w.p.  $\gamma - \text{negl}(\lambda)$  does one of the following.

- Output a collision of either **NF**, **NC** or **IVK**.
- Output a forgery w.r.t to randomization of Schnorr for the challenge  $\text{ak}$ .

$\mathcal{A}'$  works as follows.

1.  $\mathcal{A}'$  receives a challenge  $\text{ak}$ ; chooses random  $\text{nsk} \in \mathbb{F}_r$  and sends  $\text{pak} = (\text{ak}, \text{nsk})$  to  $\mathcal{A}$ .
2.  $\mathcal{A}'$  receives the output  $(L, \text{note}, \text{pos})$  of  $\mathcal{A}$ .
3.  $\mathcal{A}'$  checks that  $L, (\text{note}, \text{pos})$  satisfy the two properties in the claim; if not it aborts.
4. Let  $\text{nf} := \text{NF}(\text{nk}, \text{NC}(\text{note}), \text{pos})$ . Fix the  $\text{out}, \text{tx}$  with  $\text{out} \in \text{tx} \in L$  such that  $\text{dec}(\text{ivk}, \text{out}) = (\text{note}, \text{pos})$ .  $\text{out}$  contains a valid SNARK proof for  $\text{SPEND}(\text{rt}, \text{cv}, \text{nf}, \text{rk})$  for some  $\text{cv}, \text{rt}$ . Apply the relevant extractor  $\xi$  relating to the snark proof to obtain e.w.p  $\text{negl}(\lambda)$  a witness  $\text{path}, \text{pos}', g', \text{pk}', v', \text{rcm}', \text{cm}', \text{rcv}', \alpha, \text{ak}', \text{nsk}'$  for the statement.
5. Let  $\text{nk}' := \text{nsk}' \cdot g_{\text{n}}$ . If  $(\text{nk}, \text{cm}, \text{pos}) \neq (\text{nk}', \text{cm}', \text{pos}')$ ,  $\mathcal{A}'$  outputs  $(\text{nk}, \text{cm}, \text{pos}), (\text{nk}', \text{cm}', \text{pos}')$  as a collision of **NF**.
6. Otherwise, let  $\text{note}' = (g', \text{pk}', v', \text{rcm}')$ . We have  $\text{cm} = \text{NC}(\text{note}) = \text{NC}(\text{note}')$ . If  $(g', \text{pk}', v') \neq (g, \text{pk}, v)$ ,  $\mathcal{A}'$  outputs  $(\text{note}, \text{note}')$  as a collision of **NC**.
7. Otherwise, we must have  $\text{ivk}' = \text{ivk}$  (cause  $g \cdot \text{ivk} = g \cdot \text{ivk}' = \text{pk}$ ). Then  $\text{ivk} = \text{IVK}(\text{ak}', \text{nk})$  (by this stage we know  $\text{nk} = \text{nk}'$ ). If  $\text{ak} \neq \text{ak}'$ ,  $\mathcal{A}'$  outputs  $(\text{ak}, \text{nk}), (\text{ak}', \text{nk})$  as a collision of **IVK**.
8. Otherwise  $\text{ak} = \text{ak}'$ , and  $\text{rk} = \text{ak} + \alpha \cdot g$ . Let  $\sigma$  be the signature of  $\text{raw}_{\text{tx}}$  with public key  $\text{rk}$  in  $\text{inp.}$  and  $\mathcal{A}'$  outputs  $(\alpha, \text{raw}_{\text{tx}}, \sigma)$  as a forgery of Schnorr with challenge  $\text{ak}$ .

□

**Remark 5.7.** Note that in the spendability and non-malleability property  $\mathcal{A}$  can choose what value  $\text{nf}$  to work with. It seems likely that in a weaker model where the values  $\text{nf}$  are generated randomly via honest users' notes, a second preimage resistance property of **NF** would suffice (Thanks to Sean Bowe and Zooko Wilcox for mentioning this).

## Acknowledgements

We thank Matthew D. Green for conversations on simulation extractability that in particular inspired the definition and use of invertible group samplers.

## References

- [1] E. Ben-Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza. Zerocash: Decentralized anonymous payments from bitcoin. In *2014 IEEE Symposium on Security and Privacy, SP 2014, Berkeley, CA, USA, May 18-21, 2014*, pages 459–474, 2014.
- [2] N. Fleischhacker, J. Krupp, G. Malavolta, J. Schneider, D. Schröder, and M. Simkin. Efficient unlinkable sanitizable signatures from signatures with re-randomizable keys. *IET Information Security*, 12(3):166–183, 2018.
- [3] D. Hopwood, S. Bowe, T. Hornby, and N. Wilcox. Zcash protocol spec - <https://github.com/zcash/zips/blob/master/protocol/protocol.pdf>.
- [4] D. Pointcheval and J. Stern. Security arguments for digital signatures and blind signatures. *J. Cryptology*, 13(3):361–396, 2000.