A preliminary version of this paper appears in *Advances in Cryptology - ASIACRYPT 2008*, Lecture Notes in Computer Science Vol. 5350 pp. 125–142, J. Pieprzyk ed., Springer-Verlag, 2008. This is the full version.

Hash Functions from Sigma Protocols and Improvements to VSH

Mihir Bellare * Todor Ristov †

October 2008

Abstract

We present a general way to get a provably collision-resistant hash function from any (suitable) Σ-protocol. This enables us to both get new designs and to unify and improve previous work. In the first category, we obtain, via a modified version of the Fiat-Shamir protocol, the fastest known hash function that is provably collision-resistant based on the *standard* factoring assumption. In the second category, we provide a modified version VSH* of VSH which is faster when hashing short messages. (Most Internet packets are short.) We also show that Σ-hash functions are chameleon, thereby obtaining several new and efficient chameleon hash functions with applications to on-line/off-line signing, chameleon signatures and designated-verifier signatures.

*Dept. of Computer Science and Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0404, USA. Email: mihir@cs.ucsd.edu. URL: http://www-cse.ucsd.edu/users/mihir. Supported in part by NSF grants CNS 0524765 and CNS 6627779 and a gift from Intel.

†Dept. of Computer Science and Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0404, USA. Email: tristov@cs.ucsd.edu. URL: http://www-cse.ucsd.edu/users/tristov. Supported in part by above-mentioned grants of first author.

Contents

1 Introduction

The failure of popular hash functions MD5 and SHA-1 [47, 48] lends an impetus to the search for new ones. The contention of our paper is that there will be a "niche" market for proven-secure even if not-so-fast hash functions. Towards this we provide a general paradigm that yields hash functions provably secure under number-theoretic assumptions, and also unifies, clarifies and improves previous constructs. Our hash functions have extra features such as being chameleon [27]. Let us now look at all this in more detail.

THE NEED FOR PROVEN-SECURE HASHING. Suppose an important document has been signed with a typical hash-then-sign scheme much as PKCS#1 [26]. If collisions are found in the underlying hash function the public key needs to be revoked and the signature can no longer be accepted. Yet there are instances in which we want a public key and signatures under it to survive for twenty or more years. This might be the case for a central and highly disseminated certificate or an important contract. Revocation of a widely disseminated public key is simply too costly and error-prone. In such a case, we want to be able to trust that collisions in our hash function will not be found even twenty years down the line.

Given the failure of MD5 and SHA-1, it would be understandable, from this twenty-year perspective, to feel uncertain about any hash function designed by "similar" methods. On the other hand, we may be very willing to pay a (reasonable!) computational price for security because documents or certificates of the ultra-importance we are considering may not need to be signed often. In this case, hash functions with *proven* security are interesting, and the faster they are the better. Our contribution is a general transform that yields a plurality of such hash functions, not only providing new ones but "explaining" or improving old ones.

FROM Σ TO HASH. We show how to construct a collision-resistant hash function from any (suitable) Σ-protocol. Recall that Σ-protocols are a class of popular 3-move identification schemes. Canonical examples are the Schnorr [41], Fiat-Shamir [19] and GQ [22] protocols, but there are many others as well [31, 34, 10, 23, 40, 33, 35]. Briefly, our hash function is defined using the simulator underlying a strong form of the usual honest-verifier zero-knowledge property of Σ-protocols. (We stress that the computation of the hash is deterministic even though the simulator is randomized!) The collision-resistance stems from strong special soundness [9], a well-studied property of Σ-protocols. The advantage of our approach is that there is a rich history in constructing proven-secure Σ-protocols and we can now leverage this to get collision-resistant hash functions. For future reference let us refer to a hash function derived from our approach as a Σ-hash function.

Damgard [18] and Cramer, Damgard and Mckenzie [15] have previously shown that it is possible to design commitment schemes based on Σ-protocols, but prior to our work it has not been observed that one can design collision-resistant hash functions from Σ-protocols. Note that secure commitment is not known to imply collision-resistant hashing and in fact is unlikely to do so because the former can be based on one-way functions [32] and the latter probably not [43]. Perhaps as a consequence, our construction requires slightly stronger properties from the Σ-protocols than do the constructions of [18, 15].

SPECIFIC DESIGNS. The Schnorr [41] and GQ [22] schemes are easily shown to meet our conditions, yielding collision resistant Σ-hash functions $\mathcal{H}$-$\mathcal{Sch}$ and $\mathcal{H}$-$\mathcal{GQ}$ based, respectively, on discrete log and RSA. More interesting is the Fiat-Shamir protocol $\mathcal{FS}$ [19]. It doesn't satisfy strong special soundness but we modify it to a protocol $\mathcal{SFS}$ (strong $\mathcal{FS}$) that we prove does under the factoring assumption,

Pre	$\mathcal{H}$-$\mathcal{D}\mathcal{a}$	$\mathcal{H}$-$\mathcal{ST}$	$\mathcal{H}$-$\mathcal{SFS}$
0	1	0.22	2
2048	1	0.33	4
16384	1	2	8

Figure 1: Performance of factoring-based hash functions. The modulus and output size are 1024 bits and the block size is 512 bits. "Pre" is the amount of pre-computation, in number of group elements stored. The table entry is the rate, defined as the average number of bits of data hashed per modular multiplication.

thereby obtaining a Σ-hash function $\mathcal{H}$-$\mathcal{SFS}$. From a modified version of the Micali-Shamir protocol [31] we obtain a Σ-hash function $\mathcal{H}$-$\mathcal{SMS}$ with security based on the SRPP (Square Roots of Prime Products) assumption of [31]. We also obtain a Σ-hash $\mathcal{H}$-$\mathcal{O}\mathcal{k}\mathcal{a}$ from Okamoto's protocol [34] and a pairing-based Σ-hash $\mathcal{H}$-$\mathcal{HS}$ from an identification protocol of [5] derived from the identity-based signature scheme of Hess [23].

How fast? One question we consider interesting is, how fast can one hash while maintaining a proof of security under the *standard* factoring assumption? Figure 1 compares $\mathcal{H}$-$\mathcal{SFS}$ to the fastest known factoring-based functions and shows that the former emerges as the winner. (VSH is faster than all these, but is based on a non-standard assumption related to the difficulty of extracting modular square roots of products of small primes. We will discuss VSH, and our improvement to it, in a bit.) In Figure 1, $\mathcal{H}$-$\mathcal{D}\mathcal{a}$ is the most efficient factoring-based instantiation known of Damgård's claw free permutation-based hash function [16, 21, 27]. $\mathcal{H}$-$\mathcal{ST}$ is the hash function of Shamir and Tauman [42]. The table entries are the rate, defined as the average number of bits of data hashed per modular multiplication in MD mode with a block size of 512 bits and a modulus and output size of 1024 bits. The figure shows that without pre-computation, $\mathcal{H}$-$\mathcal{SFS}$ is twice as fast as $\mathcal{H}$-$\mathcal{D}\mathcal{a}$ and 9 times as fast as $\mathcal{H}$-$\mathcal{ST}$. But $\mathcal{H}$-$\mathcal{SFS}$ is amenable to pre-computation based speedup and $\mathcal{H}$-$\mathcal{D}\mathcal{a}$ is not, so the gap in their rates increases swiftly with storage. $\mathcal{H}$-$\mathcal{ST}$ is also amenable to pre-computation based speedup but $\mathcal{H}$-$\mathcal{SFS}$ remains a factor 4 faster for any given amount of storage. We also remark that additionally $\mathcal{H}$-$\mathcal{SFS}$ is amenable to parallelization, unlike the other functions. We remark that $\mathcal{H}$-$\mathcal{SMS}$ is faster than $\mathcal{H}$-$\mathcal{SFS}$ but based on a stronger assumption. In Section 4 we recall $\mathcal{H}$-$\mathcal{D}\mathcal{a}$ and $\mathcal{H}$-$\mathcal{ST}$ and justify the numbers in Figure 1. We also discuss implementation results.

Features of Σ-hash functions. Krawczyk and Rabin [27] introduced chameleon hashing. They presented two examples, and others were found by [42, 2, 3]. The analyses, however, are ad hoc. We, in contrast, show a general result (Theorem 3.4), namely, that any Σ-hash is chameleon. As a consequence, we immediately obtain that $\mathcal{H}$-$\mathcal{GQ}$, $\mathcal{H}$-$\mathcal{SFS}$, $\mathcal{H}$-$\mathcal{SMS}$, $\mathcal{H}$-$\mathcal{O}\mathcal{k}\mathcal{a}$, and $\mathcal{H}$-$\mathcal{HS}$ are chameleon.

Chameleon hashing has numerous applications. One of these is Shamir and Tauman's chameleon hash based method for on-line, off-line signing [42]. This means that when one uses a Σ-hash one can completely eliminate the on-line cost of signing. (This cost is shifted entirely to the off-line phase.) This compensates to some extent for the reduced efficiency of Σ-hash functions compared to conventional ones. (MD5 and SHA-1 are not chameleon and do not allow one to use the Shamir-Tauman construction.) Another application is chameleon signatures [27], which provides a recipient with a non-repudiable signature of a message without allowing it to prove to a third party that the signer signed this message. As explained in [27] this is an important tool for privacy-respecting authenticity in the signing of contracts and agreements. Finally, chameleon hash functions are used in designated-verifier signatures to achieve privacy [25, 44]. By adding new and more efficient chameleon hash functions to the pool of existing ones we enable new and more efficient ways to implement all

these applications.

Another attribute of Σ-hash functions is that they are keyed. While one can, of course, simply hardwire into the code a particular key to get an unkeyed function in the style of MD5 or SHA-1, it is advantageous, as explained in [7], to allow each user to choose their own key. The reason is that damage from a collision is now limited to the user whose key is involved, and the attacker must re-invest resources to attack another key. This slows down the rate of attacks and gives users time to get patches in place or revoke keys.

Finally, the reductions underlying the security proofs of Σ-hash functions are tight, so that the proven security guarantees hold with normal values of the security parameters.

Reverse connection. As indicated above, Theorem 3.4 shows that Σ-hash functions are chameleon. Theorem 3.5 shows that the converse is true as well. Namely, all chameleon hash functions are Σ-hash functions. We prove this by associating to any chameleon hash function $\mathcal{H}$ a Σ-protocol $\mathcal{SP}$ such that applying our Σ2H (Σ-to-hash) transform to $\mathcal{SP}$ returns $\mathcal{H}$. We thereby have a characterization of chameleon hash functions as Σ-hash functions, which we consider theoretically interesting. We also obtain numerous new Σ-protocols, and thus identification protocols and, via [15, 18], commitment schemes, from existing chameleon hash functions such as $\mathcal{H}$-$\mathcal{Da}$ [16] and $\mathcal{H}$-$\mathcal{ST}$ [42]. However, we are not aware of any practical benefit of these constructs over known ones.

Unifying previous work. $\mathcal{H}$-$\mathcal{Sch}$ turns out to be exactly the classical hash function of Chaum, Van Heijst and Pfitzmann [12], which was shown to be chameleon by [27]. $\mathcal{H}$-$\mathcal{Oka}$ is an extension thereof [12]. $\mathcal{H}$-$\mathcal{GQ}$ is a special case of a chameleon hash function proposed by Ateniese and de Medeiros [2, 3]. (Our other hash functions $\mathcal{H}$-$\mathcal{SFS}$, $\mathcal{H}$-$\mathcal{SMS}$ and $\mathcal{H}$-$\mathcal{HS}$ are new.) The re-derivation of these hash functions as Σ-hashes sheds new light on the designs and shows how the Σ paradigm explains and unifies previous constructs.

But the most interesting connection in this regard is one we make between VSH [13] and $\mathcal{H}$-$\mathcal{SMS}$, the Σ-hash function emanating from the protocol of Micali and Shamir [31]. The latter is a more efficient version of the Fiat-Shamir protocol in which the public key, rather than consisting of random quadratic residues, consists of small primes. Interestingly $\mathcal{H}$-$\mathcal{SMS}$ turns out to be the VSH compression function [13] modulo some details. We suggest that this provides some intuition for the VSH design. It turns out that we can exploit this connection to get some improvements to VSH.

VSH*. In number-theoretic hashing there is (as elsewhere) a trade-off between speed and assumptions. We saw above that $\mathcal{H}$-$\mathcal{SFS}$ is the fastest known hash function under the standard factoring assumption. We now turn to non-standard factoring-related assumptions. Here the record-holder is VSH [13] with a proof based on the VSSR assumption of [13]. Our contribution is a modification VSH* of VSH that is faster for short messages. (Our implementations show that VSH* is up to 5 times faster than VSH on short messages. On long messages they have the same performance.) This is important because short messages are an important case in practice. (For example, most Internet packets are short.) VSH* remains provably collision-resistant under the same VSSR assumption as VSH.

We provide analogous improvements for the Fast-VSH variant of VSH provided by [13]. Again we can provide Fast-VSH* whose underlying compression function (unlike that of Fast-VSH) is proven collision-resistant, leading to speedups in hashing short messages. However, the speed gains are smaller than in the previous case.

Overall we believe that, even putting performance aside, having a collision resistant compression function underlying a hash function is a plus since it can be used directly and makes the hash function more misuse-resistant.

WHAT Σ-HASH FUNCTIONS AREN'T. Some recent work [14, 6, 1] suggests that general-purpose hash functions should have extra properties like pseudorandomness. Σ-hash functions are merely collision-resistant and chameleon; they do not offer these extra attributes. But as indicated above, Σ-hash functions are not intended to be general purpose. The envisaged applications are chameleon hashing and proven-secure, reasonable cost (purely) collision-resistant hashing.

RELATED WORK. Damgård [16] presents a construction of collision-resistant hash functions from claw-free permutation pairs [21]. As noted above, his factoring-based instantiation, based on [21] and also considered in [27, 42], is slower than our $\mathcal{H}$-$\mathcal{SFS}$.

Ishai, Kushilevitz and Ostrovsky [24] show how to transform homomorphic encryption (or commitment) schemes into collision-resistant hash functions. This is an interesting theoretical connection between the primitives. As far as we can tell, however, the approach is not yet practical. Specifically, their quadratic-residuosity (QR) based instantiation has a rate of 1/40 (that is, 40 modular multiplications per bit) with a 1024 bit modulus. (Their matrix needs 80 rows to get the 80-bit security corresponding to a 1024-bit modulus.) Hence their function is much slower than the constructs of Figure 1 in addition to being based on a stronger assumption (QR as opposed to factoring). Additionally it has a $80 \cdot 1024$ bit output so in a practical sense is not really hashing. Other instantiations of their construction that we know (El Gamal under DDH, Paillier [38] under DCRA) are also both slower than known ones and based on stronger assumptions.

Lyubashevsky, Micciancio, Peikert and Rosen [29] present a fast hash function SWIFFT with an asymptotic security proof based on assumptions about the hardness of lattice problems [39, 28], but the proof would not seem to yield guarantees for the parameter sizes proposed in [29]. In contrast, our reductions are tight and the proofs provide guaranties for standard values of the security parameters.

Bellare and Micciancio's construction [4] (whose goal was to achieve incrementality) uses random oracles, but these can be eliminated by using a small block size, such as one bit. In this case their MuHASH is provably collision-resistant based only on the discrete-log assumption, and runs at 0.33 bits per group operation in MD mode. In comparison, $\mathcal{H}$-$\mathcal{Sch}$ (also discrete log based) is faster, at 0.57 bits per group operation in MD mode.

Charles, Goren and Lauter [11] presented a hash function based on the assumed hardness of some problems related to elliptic curves. However, their construction was shown to not be collision-resistant [46] and in fact to not even be pre-image resistant [37]. Tillich and Zémor [45] present a hash function based on the assumed hardness of some graph problems, whose security properties and efficiency were improved by Petit, Veyrat-Charvillon, and Quisquater [36]. The hash function of [36] is slower than Fast-VSH, and thereby slower than Fast-VSH*, according to the performance results reported in [13] and [36].

2 Definitions

NOTATION AND CONVENTIONS. We denote by $a_1 \| \cdots \| a_n$ a string encoding of $a_1, \ldots, a_n$ from which the constituent objects are uniquely recoverable. We denote the empty string by ε. Unless otherwise indicated, an algorithm may be randomized. If A is a randomized algorithm then $y \leftarrow\!\!\$\, A(x_1, \ldots)$ denotes the operation of running A with fresh coins on inputs $x_1, \ldots$ and letting y denote the output. We denote by $[A(x_1, \ldots)]$ the set of all y that have positive probability of being output by A on input $x_1, \ldots$. If S is a (finite) set then $s \leftarrow\!\!\$\, S$ denotes the operation of picking s uniformly at random from S. If $X = x_1 \| x_2 \| \ldots \| x_n$, then $x_1 \| x_2 \| \ldots x_n \leftarrow X$ denotes the operation of parsing X into its constituents. Similarly, if $X = (x_1, x_2, \ldots, x_n)$ is an n-tuple, then $(x_1, x_2, \ldots, x_n) \leftarrow X$ denotes the

operation of parsing X into its elements. We denote the security parameter by k, and by 1^k its unary encoding. Vectors are denoted in boldface, for example $\mathbf{u}$. If $\mathbf{u}$ is a vector then $|\mathbf{u}|$ is the number of its components and $\mathbf{u}[i]$ is its i-th component. "PT" stands for polynomial time.

Σ-PROTOCOLS. A Σ-protocol is a three-move interactive protocol conducted by a prover and a verifier. Formally, it is a tuple $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet},\ \mathrm{ChSet}, \mathrm{RpSet})$, where K, P are PT algorithms and V is a deterministic boolean algorithm. The key-generation algorithm K takes input 1^k and returns a pair (pk, sk) consisting of a public and secret key for the prover. The latter is initialized with pk, sk while the verifier is initialized with pk. The parties interact as depicted in Figure 2. The prover begins by applying P to pk, sk to yield his first move $\mathrm{Y} \in \mathrm{CmSet}(pk)$, called the commitment, together with state information y, called the ephemeral secret key. The commitment is sent to the verifier, who responds with a challenge c drawn at random from $\mathrm{ChSet}(pk)$. The prover computes its response z by applying P to pk, sk, the challenge and the ephemeral secret key y. (This computation may use fresh coins although in the bulk of protocols it is deterministic.) Upon receiving c the verifier applies V to the public key and transcript $\mathrm{Y}\|\mathrm{c}\|\mathrm{z}$ of the conversation to decide whether to accept or reject. We require *completeness*, which means that an interaction between the honest prover and verifier is always accepting. Formally, for all $k \in \mathbb{N}$ we have $\mathrm{d} = 1$ with probability 1 in the experiment

$$(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k);\ (\mathrm{Y}, \mathrm{y}) \leftarrow\!\!\$\ \mathsf{P}(pk, sk);\ \mathrm{c} \leftarrow\!\!\$\ \mathrm{ChSet}(pk);$$

$$\mathrm{z} \leftarrow\!\!\$\ \mathsf{P}(pk, sk, \mathrm{c}, \mathrm{y});\ \mathrm{d} \leftarrow \mathsf{V}(pk, \mathrm{Y}\|\mathrm{c}\|\mathrm{z}).$$

The verifier given $pk, \mathrm{Y}\|\mathrm{c}\|\mathrm{z}$ should always check that $\mathrm{Y} \in \mathrm{CmSet}(\mathrm{Y})$ and $\mathrm{c} \in \mathrm{ChSet}(pk)$ and $\mathrm{z} \in \mathrm{RpSet}(pk)$ and reject otherwise. We implicitly assume this is done throughout.

SECURITY NOTIONS. We provide formal definitions of strong special soundness (sss) and strong honest verifier zero-knowledge (StHVZK). Strong special soundness of Σ-protocol $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet}, \mathrm{ChSet}, \mathrm{RpSet})$ [9] asks that it be computationally infeasible, given only the public key, to produce a pair of accepting transcripts that are commitment-agreeing but challenge-response-disagreeing. Formally an sss-adversary, on input pk, returns a tuple $(\mathrm{Y}, \mathrm{c}_1, \mathrm{z}_1, \mathrm{c}_2, \mathrm{z}_2)$ such that $\mathrm{Y} \in \mathrm{CmSet}(pk); \mathrm{c}_1, \mathrm{c}_2 \in \mathrm{ChSet}(pk);\ \mathrm{z}_1, \mathrm{z}_2 \in \mathrm{RpSet}(pk)$ and $(\mathrm{c}_1, \mathrm{z}_1) \neq (\mathrm{c}_2, \mathrm{z}_2)$. The advantage $\mathbf{Adv}^{\mathrm{sss}}_{\mathcal{SP},A}(k)$ of such an adversary is defined for all $k \in \mathbb{N}$ as the probability that $\mathsf{V}(pk, \mathrm{Y}\|\mathrm{c}_1\|\mathrm{z}_1) = 1$ and $\mathsf{V}(pk, \mathrm{Y}\|\mathrm{c}_2\|\mathrm{z}_2) = 1$ in the experiment where $\mathsf{K}(1^k)$ is first executed to get (pk, sk) and then $A(pk)$ is executed to get $(\mathrm{Y}, \mathrm{c}_1, \mathrm{z}_1, \mathrm{c}_2, \mathrm{z}_2)$. We say that $\mathcal{SP}$ has strong special soundness if $\mathbf{Adv}^{\mathrm{sss}}_{\mathcal{SP},A}(\cdot)$ is negligible for all PT sss-adversaries A. To define StHVZK, let $\mathcal{Tr}_{\mathcal{SP}}$ be the algorithm that on input pk, sk executes P and V as per Figure 2 and returns the transcript $\mathrm{Y}\|\mathrm{c}\|\mathrm{z}$. Recall that a PT algorithm Sim is a HVZK simulator for $\mathcal{SP}$ if the outputs of the processes

$$(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k);\ \ \mathbf{Return}\ (pk, \mathsf{Sim}(pk))$$

and

$$(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k);\ \ \mathbf{Return}\ (pk, \mathcal{Tr}_{\mathcal{SP}}(pk, sk))$$

are identically distributed. We say that a PT algorithm StSim is a strong HVZK (StHVZK) simulator for $\mathcal{SP}$ if StSim is deterministic and the algorithm Sim defined on input pk by

$$\mathrm{c} \leftarrow\!\!\$\ \mathrm{ChSet}(pk);\ \mathrm{z} \leftarrow\!\!\$\ \mathrm{RpSet}(pk);\ \mathrm{Y} \leftarrow \mathsf{StSim}(pk, \mathrm{c}, \mathrm{z});\ \mathbf{Return}\ \mathrm{Y}||\mathrm{c}||\mathrm{z}$$

is a HVZK simulator for $\mathcal{SP}$. We say that $\mathcal{SP}$ is StHVZK if it has a PT StHVZK simulator. We denote by $\boldsymbol{\Sigma}(\mathrm{sss})$ the set of all Σ-protocols that satisfy strong special soundness and by $\boldsymbol{\Sigma}(\mathrm{StHVZK})$ the set of all Σ-protocols that are strong HVZK.

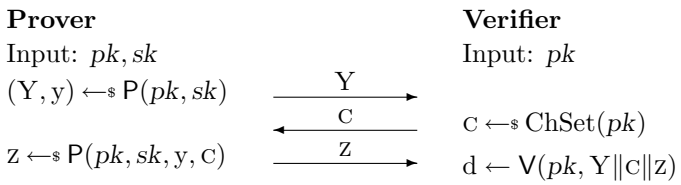

Figure 2: Σ-protocol. Keys pk and sk are produced using key-generation algorithm K.

DISCUSSION. While the basic format of Σ-protocols as 3-move protocols of the type above is agreed upon, when it comes to security properties, there are different choices and variations in the literature. Our formalization of strong special soundness is from [9]. Strong HVZK seems to be new, but is natural since we will find many protocols that posses it.

COLLISION-RESISTANT HASH FUNCTIONS. A family of n-input hash functions (where $n \geq 1$ is a constant) is a tuple $\mathcal{H} = (\mathsf{KG}, \mathsf{H}, \mathrm{D}_1, \ldots, \mathrm{D}_\mathrm{n}, \mathrm{R})$. The key-generation algorithm KG takes input 1^k and returns a key K describing a particular function $\mathsf{H}_K : \mathrm{D}_1(K) \times \ldots \mathrm{D}_\mathrm{n}(K) \to \mathrm{R}(K)$. As this indicates, $\mathrm{D}_1, \ldots, \mathrm{D}_\mathrm{n}, \mathrm{R}$ are functions that given K return sets. A cr-adversary, on input K returns distinct tuples $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ such that $x_i, y_i \in \mathrm{D}_i(K)$ for all $1 \leq i \leq n$. The advantage $\mathbf{Adv}^{\mathrm{cr}}_{\mathcal{H},B}(k)$ of such an adversary B is defined for all $k \in \mathbb{N}$ as the probability that $\mathsf{H}(K, x_1, \ldots, x_n) = \mathsf{H}(K, y_1, \ldots, y_n)$ in the experiment where $\mathsf{KG}(1^k)$ is first executed to get K and then $B(K)$ is executed to get $((x_1, \ldots, x_n), (y_1, \ldots, y_n))$. We say that $\mathcal{H}$ is collision resistant if the cr-advantage of any PT adversary B is negligible.

3 Σ-hash theory

This section covers the theory of Σ-hash functions. We present and justify the Σ2H transform that turns a Σ-protocol $\mathcal{SP} \in \Sigma(\mathrm{sss}) \cap \Sigma(\mathrm{StHVZK})$ into a collision-resistant hash function $\mathcal{H}$-$\mathcal{SP}$. Then we find Σ-protocols which we can prove have the required properties and derive specific Σ-hash functions. Finally we relate Σ and chameleon hash functions. In Section 4 we discuss the practical and performance aspects of our Σ-hash functions.

3.1 The transform

We show how to build a collision-resistant hash function from any Σ-protocol $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet}, \mathrm{ChSet}, \mathrm{RpSet}) \in \Sigma(\mathrm{sss}) \cap \Sigma(\mathrm{StHVZK})$ that satisfies strong special soundness and strong HVZK. Let StSim be a strong HVZK simulator for $\mathcal{SP}$. We define the 2-input family of hash functions $\mathcal{H} = (\mathsf{KG}, \mathsf{H}, \mathrm{ChSet}, \mathrm{CmSet}, \mathrm{RpSet})$ by $\mathsf{KG} = \mathsf{K}^{(1)}$ and $\mathsf{H}_{pk}(\mathrm{c}, \mathrm{z}) = \mathsf{StSim}(pk, \mathrm{c}, \mathrm{z})$, where $\mathsf{K}^{(1)}$ is the algorithm that on input 1^k lets $(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k)$ and returns pk. In other words, the key is the prover's public key. (The secret key is discarded.) The inputs to the hash function are regarded as the challenge and response in the Σ-protocol. The output is the corresponding commitment. The existence of a StHVZK simulator is exploited to *deterministically* compute this output. We refer to a family of functions defined in this way as a Σ-hash. We write $\mathcal{H} = \Sigma\mathsf{2H}(\mathcal{SP})$ to indicate that $\mathcal{H}$ has been derived as above from Σ-protocol $\mathcal{SP}$. The following theorem says that a Σ-hash family is collision-resistant.

Theorem 3.1 Let $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet},\ \mathrm{ChSet}, \mathrm{RpSet}) \in \Sigma(\mathrm{sss}) \cap \Sigma(\mathrm{StHVZK})$ be a Σ-protocol. Let $\mathcal{H} = (\mathsf{KG}, \mathsf{H}, \mathrm{ChSet}, \mathrm{RpSet}, \mathrm{CmSet}) = \Sigma 2\mathsf{H}(\mathcal{SP})$ be the family of hash functions associated to $\mathcal{SP}$ as above. For every cr adversary B against $\mathcal{H}$ there exists an sss-adversary A against $\mathcal{SP}$ such that for all k we have $\mathbf{Adv}^{\mathrm{cr}}_{H,B}(k) \leq \mathbf{Adv}^{\mathrm{sss\text{-}na}}_{\mathcal{SP},A}(k)$, and the running time of B is that of A. ▌

The proof of this theorem, given below, is simple, but we note some subtleties, which is the way it relies on the (strong) HVZK and completeness of the Σ-protocol in addition to the strong special soundness.

Proof of Theorem 3.1: We define adversary A as follows.

Adversary $A(pk)$
$(x_1, x_2) \leftarrow\!\!\$\ F(K)$; $((\mathrm{C}_1, \mathrm{Z}_1), (\mathrm{C}_2, \mathrm{Z}_2)) \leftarrow\!\!\$\ B(pk)$; $\mathrm{Y} \leftarrow \mathsf{H}_{pk}(\mathrm{C}_1, \mathrm{Z}_2)$
Return $(\mathrm{Y}, \mathrm{C}_1, \mathrm{Z}_1, \mathrm{C}_2, \mathrm{Z}_2)$

By definition of a cr-adversary we know that $((\mathrm{C}_1, \mathrm{Z}_1) \neq (\mathrm{C}_2, \mathrm{Z}_2))$. Hence A satisfies the definition of sss adversary. Let $\mathrm{Y}_i = \mathsf{H}_{pk}(\mathrm{C}_i, \mathrm{Z}_i)$ for $i = 1, 2$. The definition of a cr-adversary also implies that $\mathrm{C}_i \in \mathrm{ChSet}(pk)$ and $\mathrm{Z}_i \in \mathrm{RpSet}(pk)$ for $i = 1, 2$. Strong HVZK now implies that the transcripts $\mathrm{Y}_i \| \mathrm{C}_i \| \mathrm{Z}_i$ have positive probability of being produced in the protocol, meaning of being output by $\mathcal{T}r_{\mathrm{SP}}(pk, sk)$. The completeness of the protocol now implies that $\mathsf{V}(pk, \mathrm{Y}_1 \| \mathrm{C}_1 \| \mathrm{Z}_1) = 1$ and $\mathsf{V}(pk, \mathrm{Y}_2 \| \mathrm{C}_2 \| \mathrm{Z}_2) = 1$. Finally, if B succeeds then $\mathrm{Y}_1 = \mathrm{Y}_2$ so A also succeeds. ▌

To construct Σ-hash functions we now seek Σ-protocols which we can show are in $\Sigma(\mathrm{sss}) \cap \Sigma(\mathrm{StHVZK})$.

3.2 Overview of constructions

We begin, as illustrative examples, with the Schnorr [41] and GQ [22] Σ-protocols, which we can easily show to have the desired properties. The hash functions obtained are known [12, 2, 3] and their re-derivation as Σ-hashes sheds new light on their design and also shows how the Σ-hash paradigm unifies and explains existing work. More interesting is the Fiat-Shamir [19] Σ-protocol. It doesn't satisfy strong special soundness, but we modify it to a Σ-protocol $\mathcal{SFS}$ that we prove is in $\Sigma(\mathrm{sss}) \cap \Sigma(\mathrm{StHVZK})$ under the standard factoring assumption. With non-standard factoring-related assumptions (that it is hard to extract modular square roots of products of small primes) we get a faster Σ-hash $\mathcal{H}$-$\mathcal{SMS}$ from a modification of the Micali-Shamir Σ-protocol [31]. We also get another discrete-log based Σ-hash from Okamoto's protocol [34] and a pairing based one from the $\mathcal{HS}$ protocol [23, 5]. Let us now detail all this.

3.3 $\mathcal{Sch}$

We fix a prime-order group generator, by which we mean a PT algorithm $\mathcal{G}$ that on input 1^k returns the description $\langle G \rangle$ of a group G of prime order $p \in \{2^{k-1}, \ldots, 2^k - 1\}$ together with p and a generator g of G. The key-generation process and protocol underlying the $\mathcal{Sch}$ Σ-protocol of [41] are then as shown in Figure 3. The algorithm that on input $pk = (\langle G \rangle, p, g, X)$ picks $\mathrm{C}, \mathrm{Z} \leftarrow\!\!\$\ \mathbb{Z}_p$ and returns $X^{\mathrm{C}} g^{\mathrm{Z}} \| \mathrm{C} \| \mathrm{Z}$ is a HVZK simulator for $\mathcal{Sch}$, so $\mathcal{Sch} \in \Sigma(\mathrm{StHVZK})$ and the derived Σ-hash $\mathcal{H}$-$\mathcal{Sch}$ is as shown in Figure 3. The key observation for strong special soundness is that if $X^{\mathrm{C}_1} g^{\mathrm{Z}_1} = X^{\mathrm{C}_2} g^{\mathrm{Z}_2}$ and $(\mathrm{C}_1, \mathrm{Z}_1) \neq (\mathrm{C}_2, \mathrm{Z}_2)$ then it must be that $\mathrm{C}_1 \neq \mathrm{C}_2$. To sss-adversary A, this leads us to associate the discrete log finder D that on input $\langle G \rangle, p, g, X$ runs A on the same input to get $(\mathrm{Y}, \mathrm{C}_1, \mathrm{Z}_1, \mathrm{C}_2, \mathrm{Z}_2)$ and returns $(\mathrm{Z}_2 - \mathrm{Z}_1)(\mathrm{C}_1 - \mathrm{C}_2)^{-1} \bmod p$. Then for all k we have $\mathbf{Adv}^{\mathrm{sss}}_{\mathcal{Sch},A}(k) \leq \mathbf{Adv}^{\mathrm{dl}}_{\mathcal{G},D}(k)$, where the

Algorithm $\mathsf{K}(1^k)$	Prover		Verifier	$\mathsf{H}_{pk} : \mathbb{Z}_p \times \mathbb{Z}_p \to G$
$(\langle G\rangle, p, g) \leftarrow\!\!\$\ \mathcal{G}(1^k)$	$\mathrm{y} \leftarrow\!\!\$\ \mathbb{Z}_p;\ \mathrm{Y} \leftarrow g^{\mathrm{y}}$	$\xrightarrow{\mathrm{Y}}$		$\mathsf{H}_{pk}(\mathrm{c}, \mathrm{z}) = X^{\mathrm{c}} g^{\mathrm{z}}$
$x \leftarrow\!\!\$\ \mathbb{Z}_p$		$\xleftarrow{\mathrm{c}}$	$\mathrm{c} \leftarrow\!\!\$\ \mathbb{Z}_p$	
$X \leftarrow g^{-x};\ sk \leftarrow x$	$\mathrm{z} \leftarrow \mathrm{y} + x \cdot \mathrm{c} \bmod p$	$\xrightarrow{\mathrm{z}}$	$\mathrm{d} \leftarrow (X^{\mathrm{c}} g^{\mathrm{z}} = \mathrm{Y})$	
$pk \leftarrow (\langle G\rangle, p, g, X)$				
Return (pk, sk)				

Figure 3: $\mathcal{Sch}$ Σ-protocol and the derived Σ-hash family, where $\mathcal{G}$ is a prime-order group generator.

Algorithm $\mathsf{K}(1^k)$	Prover		Verifier
$(N, e) \leftarrow\!\!\$\ \mathcal{G}_{\mathrm{rsa}}(1^k)$	$\mathrm{y} \leftarrow\!\!\$\ \mathbb{Z}_N^*;$		
$x \leftarrow\!\!\$\ \mathbb{Z}_N^*$	$\mathrm{Y} \leftarrow \mathrm{y}^e \bmod N$	$\xrightarrow{\mathrm{Y}}$	
$X \leftarrow x^{-e} \bmod N$		$\xleftarrow{\mathrm{c}}$	$\mathrm{c} \leftarrow\!\!\$\ \{0, \ldots, 2^l - 1\}$
$l \leftarrow L(k); sk \leftarrow x$	$\mathrm{z} \leftarrow x^{\mathrm{c}} \cdot \mathrm{y} \bmod N$	$\xrightarrow{\mathrm{z}}$	$\mathrm{d} \leftarrow (\mathrm{Y} = X^{\mathrm{c}} \cdot \mathrm{z}^e \bmod N)$
$pk \leftarrow (N, e, l, X)$	$\mathsf{H}_{pk} : \{0, \ldots, 2^l - 1\} \times \mathbb{Z}_N^* \to \mathbb{Z}_N^*$		
Return (sk, pk)	$\mathsf{H}_{pk}(\mathrm{c}, \mathrm{z}) = X^{\mathrm{c}} \mathrm{z}^e \bmod N$		

Figure 4: $\mathcal{GQ}$ Σ-protocol and the derived Σ-hash family, where $\mathcal{G}_{\mathrm{rsa}}$ is a prime exponent RSA generator with associated challenge length L.

latter is defined as the probability that $x' = x$ in the experiment where we let $(\langle G\rangle, p, g) \leftarrow\!\!\$\ \mathcal{G}(1^k)$ and $x \leftarrow\!\!\$\ \mathbb{Z}_p$ and then let $x' \leftarrow\!\!\$\ D(\langle G\rangle, p, g, g^x)$. This shows that $\mathcal{Sch}$ has strong special soundness as long as the discrete log problem is hard relative to $\mathcal{G}$. By Theorem 3.1 $\mathcal{H}$-$\mathcal{Sch}$ is collision-resistant under the same assumption.

3.4 $\mathcal{GQ}$

We fix a prime-exponent RSA generator with associated challenge length $L(\cdot)$, by which we mean a PT algorithm $\mathcal{G}_{\mathrm{rsa}}$ that on input 1^k returns an RSA modulus $N \in \{2^{k-1}, \ldots, 2^k - 1\}$ and an RSA encryption exponent $e > 2^{L(k)}$ that is a prime. The key-generation process and protocol underlying Σ-protocol $\mathcal{GQ}$ of [22] are then as shown in Figure 4. The algorithm that on input $pk = (N, e, l, X)$ picks $\mathrm{c} \leftarrow\!\!\$\ \{0,1\}^l; \mathrm{z} \leftarrow\!\!\$\ Z_N^*$ and returns $\mathrm{Y} \| \mathrm{c} \| \mathrm{z}$, where $\mathrm{Y} = \mathrm{z}^{\mathrm{c}} \mathrm{z}^2 \bmod N$, is a HVZK simulator for $\mathcal{GQ}$, so $\mathcal{GQ} \in \Sigma(\mathrm{StHVZK})$ and the derived Σ-hash $\mathcal{H}$-$\mathcal{GQ}$ is as shown in Figure 4. Again observe that if $X^{\mathrm{c}_1} \mathrm{z}_1^e = X^{\mathrm{c}_2} \mathrm{z}_2^e$ and $(\mathrm{c}_1, \mathrm{z}_1) \neq (\mathrm{c}_2, \mathrm{z}_2)$ then $\mathrm{c}_1 \neq \mathrm{c}_2$. To adversary A attacking the strong special soundness, this leads us to associate the inverter I that on input N, e, X runs A on input N, e, l, X where $l = L(\lfloor \log_2(N) \rfloor + 1)$ to get $(\mathrm{Y}, \mathrm{c}_1, \mathrm{z}_1, \mathrm{c}_2, \mathrm{z}_2)$ and returns $(\mathrm{z}_2 \mathrm{z}_1^{-1})^b X^a \bmod N$ where a, b satisfy $ae + b(\mathrm{c}_1 - \mathrm{c}_2) = 1$ and are found via the extended gcd algorithm. (This is where we use the fact that e is prime.) Then for all k we have $\mathbf{Adv}^{\mathrm{sss}}_{\mathcal{GQ},A}(k) \leq \mathbf{Adv}^{\mathrm{rsa}}_{\mathcal{G}_{\mathrm{rsa}},I}(k)$, where the latter is defined as the probability that $x' = x$ in the experiment where we let $(N, e) \leftarrow\!\!\$\ \mathcal{G}_{\mathrm{rsa}}(1^k)$ and $x \leftarrow\!\!\$\ \mathbb{Z}_N^*$ and then let $x' \leftarrow\!\!\$\ I(N, e, x^e \bmod N)$. This shows that $\mathcal{GQ}$ has strong special soundness if RSA is one-way relative to $\mathcal{G}_{\mathrm{rsa}}$. By Theorem 3.1, $\mathcal{H}$-$\mathcal{GQ}$ is collision-resistant under the same assumption.

3.5 $\mathcal{FS}$ and $\mathcal{SFS}$

We fix a modulus generator, namely a PT algorithm $\mathcal{G}_{\mathrm{mod}}$ that on input 1^k returns a modulus $N \in \{2^{k-1}, \ldots, 2^k - 1\}$ and distinct primes p, q such that $N = pq$. We also fix a challenge length $L(\cdot)$. If c

Algorithm $F(N)$
For $i = 1, \ldots, n$ do $\quad \mathbf{s}[i] \leftarrow\!\!\$\ \mathbb{Z}_N^*;\ \mathbf{u}[i] \leftarrow \mathbf{s}[i]^{-2} \bmod N$
$l \leftarrow L(k);\ pk \leftarrow (N, l, \mathbf{u});\ (\mathrm{Y}, \mathrm{C}_1, \mathrm{Z}_1, \mathrm{C}_2, \mathrm{Z}_2) \leftarrow\!\!\$\ A(pk)$
If $\mathrm{C}_1 \neq \mathrm{C}_2$ then
  $g \leftarrow\!\!\$\ \{1, \ldots, l : \mathrm{C}_1[i] \neq \mathrm{C}_2[i]\}$
  $r_1 \leftarrow \left(\frac{\mathrm{Z}_1}{\mathrm{Z}_2} \cdot \prod_{i \neq g} \mathbf{s}[i]^{\mathrm{C}_1[i] - \mathrm{C}_2[i]}\right)^{\mathrm{C}_2[g] - \mathrm{C}_1[g]}$
  $r_2 \leftarrow\!\!\$\ \mathbf{s}[g]$
Else // $\mathrm{C}_1 = \mathrm{C}_2$ and $\mathrm{Z}_1 \neq \mathrm{Z}_2$
  $r_1 \leftarrow \mathrm{Z}_1;\ r_2 \leftarrow \mathrm{Z}_2$
$r \leftarrow \gcd(N, r_1 - r_2)$
Return r

Adversary $B(N, \mathbf{u})$
$pk \leftarrow (N, \mathbf{u})\ ; (\mathrm{Y}, \mathrm{C}_1, \mathrm{Z}_1, \mathrm{C}_2, \mathrm{Z}_2) \leftarrow\!\!\$\ A(pk)$
If $\mathrm{C}_1 \neq \mathrm{C}_2$
  $T \leftarrow \{j : \mathrm{C}_1[j] = 1 \wedge \mathrm{C}_2[j] = 0\}$
  $R \leftarrow \{j : \mathrm{C}_1[j] = 1 \wedge \mathrm{C}_2[j] = 1\}$
  $M \leftarrow \{j : \mathrm{C}_1[j] = 0 \wedge \mathrm{C}_2[j] = 1\}\ ; S \leftarrow M \cup T$
  $x \leftarrow \frac{\mathrm{Z}_1}{\mathrm{Z}_2} \prod_{j \in T} \mathbf{u}[j]$
Else // $\mathrm{C}_1 = \mathrm{C}_2$ and $\mathrm{Z}_1 \neq \mathrm{Z}_2$
  $p \leftarrow \gcd(N, \mathrm{Z}_1 - \mathrm{Z}_2)$ // p is a factor of N
  $x \leftarrow \mathsf{SQR}(\mathbf{u}[1], p, N/p) \bmod N\ ; S \leftarrow \{1\}$
Return (x, S)

Adversary $A(pk)$
$(x_1, x_2) \leftarrow\!\!\$\ F(K)$
$((\mathrm{C}_1, \mathrm{Z}_1), (\mathrm{C}_2, \mathrm{Z}_2)) \leftarrow\!\!\$\ B(pk)$
$\mathrm{Y} \leftarrow \mathsf{H}_{pk}(\mathrm{C}_1, \mathrm{Z}_2)$
Return $(\mathrm{Y}, \mathrm{C}_1, \mathrm{Z}_1, \mathrm{C}_2, \mathrm{Z}_2)$

Figure 5: Factoring algorithm F for proof of Proposition 3.2, adversary A for proof of Theorem 3.1 and spr-adversary B for proof of Proposition 3.3.

is a l-bit string and $\mathbf{u} \in (Z_N^*)^l$ then we let $\mathbf{u}^{\mathrm{C}} = \prod \mathbf{u}[i]^{\mathrm{C}[i]}$ where the product is over $1 \leq i \leq l$ and $\mathrm{C}[i]$ denotes the i-th bit of C. The key-generation algorithm and protocol underlying the $\mathcal{FS}$ Σ-protocol are then as shown in Figure 6. However this protocol does not satisfy strong special soundness because if $\mathrm{Y} \| \mathrm{C} \| \mathrm{Z}$ is an accepting transcript relative to $pk = (N, l, \mathbf{u})$ then so is $\mathrm{Y} \| \mathrm{C} \| \mathrm{Z}'$ where $\mathrm{Z}' = N - \mathrm{Z}$. We now show how to modify $\mathcal{FS}$ so that it has strong special soundness. First, some notation. For $w \in \mathbb{Z}_N$ we let $[w]_N$ equal w if $w \leq N/2$ and $N - w$ otherwise. Let $\mathbb{Z}_N^+ = \mathbb{Z}_N^* \cap \{1, \ldots, N/2\}$. The modified protocol $\mathcal{SFS}$ (Strong $\mathcal{FS}$) is shown in Figure 6. Here CmSet, ChSet are as in $\mathcal{FS}$ but $\mathrm{RpSet}((N, l, \mathbf{u}))$ is now equal to $\mathbb{Z}_N^+$ rather than $\mathbb{Z}_N^*$ as before. Define $\mathbf{Adv}^{\mathrm{fac}}_{\mathcal{G}_{\mathrm{mod}}, B}(k)$, as the probability that $r \in \{p, q\}$ in the experiment where we let $(N, p, q) \leftarrow\!\!\$\ \mathcal{G}_{\mathrm{mod}}(1^k)$ and $r \leftarrow\!\!\$\ B(N)$. The following shows that $\mathcal{SFS}$ has strong special soundness under the *standard* hardness of factoring assumption.

Proposition 3.2 Let $\mathcal{G}_{\mathrm{mod}}$ be a modulus generator and $L(\cdot)$ a challenge length. Let $\mathcal{SFS}$ be the associated Σ-protocol as per Figure 6. If A is a sss-adversary against $\mathcal{SFS}$ then there is a factoring algorithm F against $\mathcal{G}_{\mathrm{mod}}$ such that for all k we have $\mathbf{Adv}^{\mathrm{sss}}_{\mathcal{SFS}, A}(k) \leq 2 \cdot \mathbf{Adv}^{\mathrm{fac}}_{\mathcal{G}_{\mathrm{mod}}, F}(k)$. The running time of F is that of A plus the time for at most $L(\cdot)$ multiplications, one inversion modulo N, and the time for one execution of the gcd algorithm.

Proof: The factoring algorithm F is shown in Figure 5. For the analysis, consider two cases. The first is that $\mathrm{C}_1 \neq \mathrm{C}_2$ and the second is that $\mathrm{C}_1 = \mathrm{C}_2$ and $\mathrm{Z}_1 \neq \mathrm{Z}_2$. In the first case, a simple computation shows that $r_1^2 \equiv r_2^2 \pmod N$. On the other hand, $\mathbf{s}[g] \notin \{r_1, N - r_1\}$ with probability $1/2$ since A gets no information about $\mathbf{s}[g]$ other than its square. So in this case F succeeds with probability $1/2$ times the probability that A succeeds. In the case $\mathrm{C}_1 = \mathrm{C}_2$ but $\mathrm{Z}_1 \neq \mathrm{Z}_2$ we have, modulo N,

$$\mathrm{Y} \equiv \mathrm{Z}_1^2 \cdot \textstyle\prod_{i=1}^{l} \mathbf{s}[i]^{\mathrm{C}_1[i]} \equiv \mathrm{Z}_2^2 \cdot \prod_{i=1}^{l} \mathbf{s}[i]^{\mathrm{C}_2[i]}$$

and $\mathrm{C}_1 = \mathrm{C}_2$ implies $\mathrm{Z}_1^2 \equiv \mathrm{Z}_2^2$. But $\mathrm{Z}_1 \neq \mathrm{Z}_2$ and $\mathrm{Z}_1, \mathrm{Z}_2 \in \mathbb{Z}_N^+$ so it must be that $\mathrm{Z}_1 \notin \{\mathrm{Z}_2, N - \mathrm{Z}_2\}$. So F succeeds with the same probability as A in this case. ∎

Now, the algorithm that on input $pk = (N, l, \mathbf{u})$ lets $\mathrm{C} \leftarrow\!\!\$\ \{0,1\}^l$; $\mathrm{Z} \leftarrow\!\!\$\ Z_N^+$, and $\mathrm{Y} \leftarrow \mathbf{u}^{\mathrm{C}} \cdot \mathrm{Z}^2 \bmod N$ and returns $\mathrm{Y}\|\mathrm{C}\|\mathrm{Z}$ is a HVZK simulator for $\mathcal{SFS}$. Accordingly $\mathcal{SFS} \in \Sigma(\text{StHVZK})$ and we derive from $\mathcal{SFS}$ the Σ-hash family $\mathcal{H}$-$\mathcal{SFS}$ shown in Figure 6. Proposition 3.2 and Theorem 3.1 imply that $\mathcal{H}$-$\mathcal{SFS}$ is collision resistant under the standard factoring assumption.

3.6 $\mathcal{MS}$ and $\mathcal{SMS}$

The Micali-Shamir protocol [31] is a variant of $\mathcal{FS}$ in which verification time is reduced by choosing the coordinates of $\mathbf{u}$ to be small primes. As with $\mathcal{FS}$ it does not satisfy sss, but we can modify it to do so and thereby obtain a collision-resistant hash function $\mathcal{H}$-$\mathcal{SMS}$ that is faster than $\mathcal{H}$-$\mathcal{SFS}$ at the cost of a stronger assumption for security. To detail all this, let $\mathcal{G}_{\mathcal{SP}}$ be a small prime modulus generator with challenge length $L(\cdot)$, by which we mean a PT algorithm that on input 1^k returns a modulus $N \in \{2^{k-1}, \ldots, 2^k - 1\}$, distinct primes p, q such that $N = pq$, and an $L(k)$-vector $\mathbf{u}$ each of whose coordinates is a prime in $\mathrm{QR}(N) = \{x^2 \bmod N : x \in Z_N^*\}$. For efficiency we would choose these primes to be as small as possible. (For example $\mathbf{u}[i]$ is the i-th prime in $\mathrm{QR}(N)$.) An spr-adversary B against $\mathcal{G}_{\mathrm{SP}}, L$ takes input N and $\mathbf{u} \in (\mathbb{Z}_N^*)^{L(k)}$ and returns (x, S) where $x \in \mathbb{Z}_N^*$ and S is a non-empty subset of $\{0,1\}^l$. Its spr-advantage is defined for all k by

$$\mathbf{Adv}^{\mathrm{spr}}_{\mathcal{G}_{\mathrm{SP}},L,B}(k) = \Pr\left[\, x^2 \equiv \textstyle\prod_{i \in S} \mathbf{u}[i] \pmod N : (N, p, q, \mathbf{u}) \leftarrow\!\!\$\ \mathcal{G}_{\mathrm{SP}}(1^k)\ ; (x, S) \leftarrow\!\!\$\ B(N, \mathbf{u}) \,\right].$$

The SRPP (Square Root of Prime Products) assumption [31] says that the spr-advantage of any PT B is negligible. Now, Figure 6 shows our modified version $\mathcal{SMS}$ of the Micali-Shamir protocol. It is in $\Sigma 2\mathsf{H}(\text{StHVZK})$ for the same reason as $\mathcal{SFS}$ and hence the derived hash function is again as shown, where $\mathsf{SQR}(\cdot, p, q)$ takes input $w \in \mathrm{QR}(N)$ and returns at random one of the four square roots of w modulo $N = pq$, computed using the primes p, q. Strong special soundness of $\mathcal{SMS}$ is proven in the following under the SRPP assumption.

Proposition 3.3 Let $\mathcal{G}_{\mathcal{SP}}$ be a small prime modulus generator with associated challenge length L. Let $\mathcal{SMS}$ be the associated Σ-protocol as per Figure 6. If A is a sss-adversary against $\mathcal{SMS}$ then there is a spr-adversary B such that for all k we have $\mathbf{Adv}^{\mathrm{sss}}_{\mathcal{G}_{\mathcal{SP}},L,A}(k) \leq \mathbf{Adv}^{\mathrm{spr}}_{\mathcal{G}_{\mathcal{SP}},L,B}(k)$. The running time of B is that of A plus time $t = \max\{t_1, t_2\}$, where t_1 is the time it takes to execute one inversion modulo N and $L + 1$ multiplications modulo N and t_2 is the time it takes to execute the gcd algorithm and the SQR algorithm.

Proof: Let's explain why the adversary B shown in Figure 5 works. If A succeeds then

$$\mathrm{Z}_1{}^2 \textstyle\prod_{j=1}^n \mathbf{u}[j]^{\mathrm{C}_1[j]} \equiv \mathrm{Z}_2{}^2 \prod_{j=1}^n \mathbf{u}[j]^{\mathrm{C}_2[j]} \pmod N \tag{1}$$

Now, when $\mathrm{C}_1 \neq \mathrm{C}_2$, multiplying both sides of this equation by $\prod_{j \in T} \mathbf{u}[j] / \prod_{j \in R} \mathbf{u}[j]$ gives:

$$\mathrm{Z}_1{}^2 \textstyle\prod_{j \in T} \mathbf{u}[j]^2 \equiv \mathrm{Z}_2{}^2 \prod_{j \in M \cup T} \mathbf{u}[j] \pmod N$$

From $\mathrm{C}_1 \neq \mathrm{C}_2$ it follows that $T \cup M \neq \emptyset$, and therefore by outputting $(\frac{\mathrm{Z}_1}{\mathrm{Z}_2} \prod_{j \in T} \mathbf{u}[j], M \cup T)$ the adversary B succeeds.

In the other case, when $\mathrm{C}_1 = \mathrm{C}_2$, it has to be $\mathrm{Z}_1 \neq \mathrm{Z}_2$ and the equation Equation (1) becomes $\mathrm{Z}_1^2 \equiv \mathrm{Z}_2^2 \bmod N$. By finding the gcd of N and $\mathrm{Z}_1 - \mathrm{Z}_2$, B can factorize N, and so it can compute the square root of $\mathbf{u}[1]$ modulo N. ∎

Proposition 3.3 and Theorem 3.1 imply that $\mathcal{H}$-$\mathcal{SMS}$ is collision-resistant under the SRPP assumption.

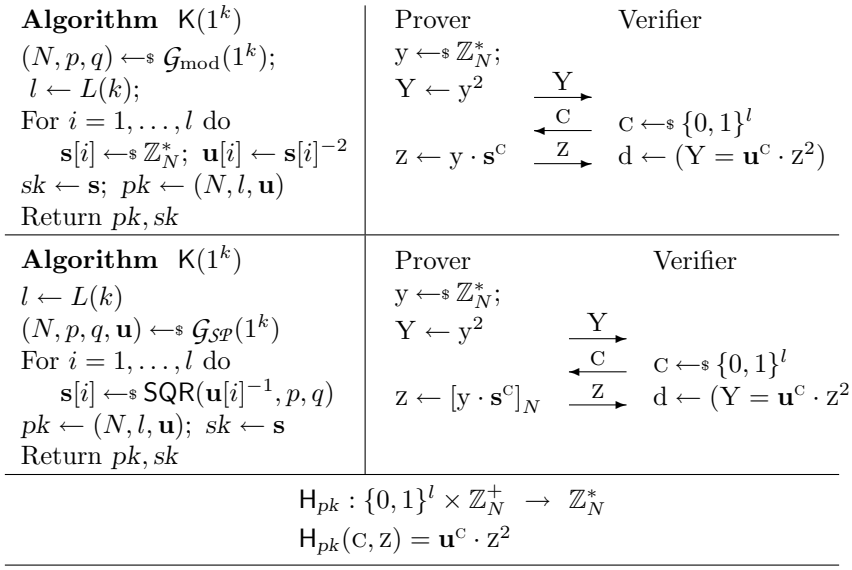

Figure 6: $\mathcal{FS}$, $\mathcal{SFS}$, $\mathcal{MS}$ and $\mathcal{SMS}$ protocols and the Σ-hash derived from $\mathcal{SFS}$, $\mathcal{SMS}$. The upper left key-generation algorithm is that of $\mathcal{FS}$ and $\mathcal{SFS}$, while the lower left one is that of $\mathcal{MS}$ and $\mathcal{SMS}$. The upper protocol is that of $\mathcal{FS}$ and $\mathcal{MS}$ while the lower protocol is that of $\mathcal{SFS}$ and $\mathcal{SMS}$. Here $\mathcal{G}_{\mathrm{mod}}$ is a modulus generator and $\mathcal{G}_{\mathrm{SP}}$ is a small prime modulus generator. The computations are in $\mathbb{Z}_N^*$, meaning modulo N.

3.7 Additional functions

Okamoto's protocol [34] is StHVZK and can be shown to have special soundness if the discrete logarithm problem is hard relative to the underlying prime-order group generator, and hence we obtain a collision-resistant Σ-hash family $\mathcal{H}$-$\mathcal{O}\mathcal{k}\mathcal{a}$. The key has the form $(\langle G\rangle, p, g_1, g_2, X)$, where $g_1, g_2 \in G^*$ and $X \in G$, and $\mathsf{H}_{pk} : \mathbb{Z}_p \times (\mathbb{Z}_p \times \mathbb{Z}_p)$ is defined by $\mathsf{H}_{pk}(\mathrm{C},(\mathrm{z}_1,\mathrm{z}_2)) = X^{\mathrm{C}} g_1^{\mathrm{z}_1} g_2^{\mathrm{z}_2}$. However, this hash function seems to offer no performance advantage over $\mathcal{H}$-$\mathcal{S}\mathcal{c}\mathcal{h}$. A pairing based identification protocol $\mathcal{HS}$, derived from the id-based signature scheme of [23], is noted in [5]. It is shown in [5] to have special soundness under concurrent attack assuming the hardness of the one more computational Diffie-Helman problem relative to an underlying prime-order bilinear group generator. The proof can be easily extended to show strong special soundness while relaxing the assumption to the hardness of the computational Diffie-Helman problem. $\mathcal{HS}$ can also be shown to be StHVZK and hence we obtain a Σ-hash family $\mathcal{H}$-$\mathcal{HS}$. The key has the form $(\langle G_1\rangle, \langle G_2\rangle, q, P, \langle \mathrm{e}\rangle, \alpha)$ where G_1 and G_2 are groups of prime order p; $\mathrm{e} : G_1 \times G_1 \rightarrow G_2$, is non-degenerate bilinear map; $P \in G_1^*$ and $\alpha \in G_2$. The function $\mathsf{H}_{pk} : \mathbb{Z}_p \times G_1 \rightarrow G_2$ is defined by $\mathsf{H}_{pk}(\mathrm{C},\mathrm{z}) = \mathrm{e}(\mathrm{z}, P) \cdot \alpha^{\mathrm{C}}$. Due to the pairing, however, this hash function is slower than $\mathcal{H}$-$\mathcal{S}\mathcal{c}\mathcal{h}$.

3.8 Σ = chameleon

We move from examples of Σ-hash functions to a general property of the class, namely that any Σ-hash function is chameleon and vice-versa.

Definitions. A 2-input hash family $\mathcal{H} = (\mathsf{KG}, \mathsf{H}, \mathrm{D}_1, \mathrm{D}_2, \mathrm{R})$ is said to be trapdoor if there is a PT algorithm K, called the full key-generation algorithm, such that $\mathsf{KG} = \mathsf{K}^{(1)}$ and also there is a deterministic, PT algorithm, I, called the inversion algorithm, such that for all $(K,T) \in \left[\mathsf{K}(1^k)\right]$ and all $\mathrm{c}_1, \mathrm{c}_2 \in \mathrm{D}_1(K)$ and all $\mathrm{Y} \in \mathrm{D}_2(K)$ the map defined by $\mathrm{z} \rightarrow I(K,T,\mathrm{c}_1,\mathrm{c}_2,\mathrm{z})$ is a bijection of

$S_{K,\mathrm{C}_1}(\mathrm{Y})$ to $S_{K,\mathrm{C}_2}(\mathrm{Y})$, where $S_{K,\mathrm{C}}(\mathrm{Y}) = \{\mathrm{Z} \in \mathrm{D}_2(K) : \mathsf{H}_K(\mathrm{C}, \mathrm{Z}) = \mathrm{Y}\}$.[1] We say that $\mathcal{H}$ has the uniformity property if for all K and all $\mathrm{C}_1, \mathrm{C}_2 \in \mathrm{D}_1(K)$ it is the case that $\mathsf{H}_K(\mathrm{C}_1, \cdot)$ and $\mathsf{H}_K(\mathrm{C}_2, \cdot)$ are identically distributed when regarded as random variables over a random choice of their argument Z from $\mathrm{D}_2(K)$. We say that $\mathcal{H}$ is chameleon if it is trapdoor, collision-resistant and has the uniformity property.

The (standard) completeness requirement for a Σ-protocol $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet}, \mathrm{ChSet}, \mathrm{RpSet})$ implies that from a secret key sk and challenge C, one can easily (in PT) compute the response Z, but only if one has the ephemeral secret key y underlying the commitment. To obtain chameleon hash functions from Σ-protocols we need the latter to satisfy a strong form of completeness which says that a response, distributed identically to the response of the real prover P, can be computed even without the ephemeral secret key so long as we have access to some accepting conversation. Formally Σ-protocol $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet}, \mathrm{ChSet}, \mathrm{RpSet})$ satisfies strong completeness if there is a deterministic PT algorithm $\mathcal{P}^*$ such that for all k the output of the following two processes are identically distributed:

$(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k)$	
$\mathrm{Y} \| \mathrm{C}_1 \| \mathrm{Z}_1 \leftarrow\!\!\$\ \mathcal{T}r(pk)$	
$\mathrm{C}_2 \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$	$\mathrm{Y} \| \mathrm{C}_2 \| \mathrm{Z}_2 \leftarrow\!\!\$\ \mathcal{T}r(pk)$
$\mathrm{Z}_2 \leftarrow \mathcal{P}^*(pk, sk, \mathrm{Y}, \mathrm{C}_1, \mathrm{Z}_1, \mathrm{C}_2)$	$\mathrm{Return}(pk, \mathrm{Y} \| \mathrm{C}_2 \| \mathrm{Z}_2)$
$\mathrm{Return}(pk, \mathrm{Y} \| \mathrm{C}_2 \| \mathrm{Z}_2)$	

We let $\boldsymbol{\Sigma}(\mathrm{sc})$ be the class of all Σ-protocols that have the strong completeness property. All the Σ-protocols we have considered are in $\boldsymbol{\Sigma}(\mathrm{sc})$.

The equivalence. The following implies that any Σ-hash is chameleon.

Theorem 3.4 Let $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet}, \mathrm{ChSet}, \mathrm{RpSet}) \in \boldsymbol{\Sigma}(\mathrm{StHVZK}) \cap \boldsymbol{\Sigma}(\mathrm{sss}) \cap \boldsymbol{\Sigma}(\mathrm{sc})$ be a Σ-protocol. Then the Σ-hash family $\mathcal{H}\text{-}\mathcal{SP} = \boldsymbol{\Sigma}2\mathsf{H}(\mathcal{SP}) = (\mathsf{KG}, \mathsf{H}, \mathrm{ChSet}, \mathrm{CmSet}, \mathrm{RpSet})$ is chameleon.

Refer to Appendix B for the proof. As a consequence, we obtain the following new chameleon hash functions: $\mathcal{H}\text{-}\mathcal{GQ}$, $\mathcal{H}\text{-}\mathcal{SFS}$, $\mathcal{H}\text{-}\mathcal{SMS}$, $\mathcal{H}\text{-}\mathcal{Oka}$, $\mathcal{H}\text{-}\mathcal{HS}$. ($\mathcal{H}\text{-}\mathcal{Sch}$ was already known to be chameleon [27].) This yields numerous new and more efficient instantiations of on-line/off-line signatures [42], chameleon signatures [27] and designated-verifier signatures [25, 44].

Even more interestingly, we prove the converse. The following theorem says that any chameleon hash family is a Σ-hash family, meaning the result of applying our $\boldsymbol{\Sigma}2\mathsf{H}$ transform to some Σ-protocol.

Theorem 3.5 Let $\mathcal{H} = (\mathsf{KG}, \mathsf{H}, \mathrm{ChSet}, \mathrm{CmSet}, \mathrm{RpSet})$ be a family of chameleon hash functions. Then there is a Σ-protocol $\mathcal{SP} = (\mathsf{K}, \mathsf{P}, \mathsf{V}, \mathrm{CmSet}, \mathrm{ChSet}, \mathrm{RpSet}) \in \boldsymbol{\Sigma}(\mathrm{StHVZK}) \cap \boldsymbol{\Sigma}(\mathrm{sss}) \cap \boldsymbol{\Sigma}(\mathrm{sc})$ such that $\mathcal{H} = \boldsymbol{\Sigma}2\mathsf{H}(\mathcal{SP})$ is the Σ-hash family corresponding to $\mathcal{SP}$.

The proof is in Appendix B. Applying this to known chameleon-hash functions like $\mathcal{H}\text{-}\mathcal{Da}$ [16, 27] and $\mathcal{H}\text{-}\mathcal{ST}$ [42] yields new Σ-protocols and hence new identification schemes and, via [18, 15], new commitment schemes.

[1] Krawczyk and Rabin [27] only require that $\mathsf{H}_K(\mathrm{C}_2, I(K, T, \mathrm{C}_1, \mathrm{C}_2, \mathrm{Z}_1)) = \mathsf{H}_K(\mathrm{C}_1, \mathrm{Z}_1)$ for all $\mathrm{C}_1, \mathrm{C}_2 \in \mathrm{ChSet}(K)$ and all $\mathrm{Z} \in \mathrm{RpSet}(K)$. Shamir and Tauman require a stronger condition that is essentially a computational version of ours. It seems to us that the non-transferability of chameleon signatures required in [27] requires the hash function to meet one of these stronger conditions. All the Σ-protocols we have considered meet our requirement.

Σ-hash	w	KB/s	space
$\mathcal{H}$-$\mathcal{SFS}$	0	30.85	n/a
$\mathcal{H}$-$\mathcal{SFS}$	4	67.41	2048
$\mathcal{H}$-$\mathcal{SFS}$	8	118.1	16384
$\mathcal{H}$-$\mathcal{SMS}$	0	914.3	n/a

Table 1: Implementation results. Here w is the "width" parameter determining pre-computation and the space is the number of group elements that need to be stored.

---

4 Σ-hash practice and performance

In this section we cover practical issues related to Σ-hash functions, including performance, performance comparison with existing constructions and implementation results.

4.1 Extending the domain

A Σ-hash family $\mathcal{H}$ as defined above is actually a (keyed) compression function since the domain is relatively small. In practice however we need to hash messages of long and variable length. This would not at first appear to be much of a problem since we should be able to do MD iteration [17, 30]. In fact this is essentially true but one has to be careful about a few things. What one would naturally like to do is use the second argument to H_{pk} as the chaining variable. But this requires that outputs of the compression function can be regarded as chaining values, meaning $\mathrm{CmSet}(pk)$ be a subset of $\mathrm{RpSet}(pk)$. Sometimes this is true, as for $\mathcal{H}$-$\mathcal{GQ}$, which in this way lends itself easily and naturally to MD iteration. But in the case of $\mathcal{SFS}$ and $\mathcal{SMS}$ we have $\mathrm{CmSet}((N, l, \mathbf{u})) = \mathbb{Z}_N^* \subsetneq \mathbb{Z}_N^+ = \mathrm{RpSet}((N, l, \mathbf{u}))$. In Appendix A we show how to resolve these problems by appropriate "embeddings" that effectively allow the second input of the compression function to be used as a chaining variable at the cost of 1 bit in throughput and in particular allows us to run any of our Σ-hash functions in MD mode. We won't detail the general transform here, but it is instructive to describe the modified compression function. The public key has the form $(N, l, \mathbf{u}, v)$ where $N, l, \mathbf{u}$ are as before and $v \in \mathrm{QR}(N)$, and $\mathsf{H}_{pk} : \{0,1\}^l \times \mathbb{Z}_N^* \to \mathbb{Z}_N^*$ is defined by

$$\mathsf{H}_{pk}(\mathrm{C}, \mathrm{Z}) = \mathbf{u}^{\mathrm{C}} \cdot \mathrm{Z}^2 \cdot v^{f_N(\mathrm{Z})} \bmod N, \tag{2}$$

where $f_N(\mathrm{Z}) = 0$ if $\mathrm{Z} \in \mathbb{Z}_N^+$ and 1 otherwise. It can be shown that this modified function is also a Σ-hash, meaning the result of applying Σ2H to a suitably modified version of the original Σ-protocol that retains the sss, StHVZK and sc properties of the original. But now $\mathrm{CmSet}((N, l, \mathbf{u}, v)) = \mathbb{Z}_N^* = \mathrm{RpSet}((N, l, \mathbf{u}, v))$ so MD-iteration is possible.

4.2 Metrics

We measure performance of a hash function in terms of rate, which we define as the average number of bits hashed per group operations. (By "average" we mean when the data is random.) In this measure, an exponentiation $a \mapsto A^a$ costs $1.5n$ group operations and a two-fold multi-exponentiation $a, b \mapsto A^a B^b$ costs $1.75n$ group operations where n is the length of a and also of b. We will use these estimates extensively below. We can consider two modes of operation of a given Σ-hash function

$\mathcal{H}$-$\mathcal{SP}$, namely compression and MD. In the first case the data to be hashed by H_{pk} is the full input C, Z, while in the second case it is only C. (The second input is the chaining variable which is not part of the data.) The rate in MD mode is lower than in compression mode for most hash functions. ($\mathcal{SFS}$ is an interesting exception.) Compression mode is relevant when the function is being used as a chameleon hash, since the data can then be compressed with a standard (merely collision-resistant) hash function such as SHA-1 before applying the Σ-hash [27, Lemma 1]. MD mode is relevant when one wants to avoid conventional hash functions and get the full provable guarantees of the Σ-hash by using it alone. Our performance evaluations will consider MD mode.

4.3 Performance of Σ-hash functions

$\mathcal{H}$-$\mathcal{Sch}$ and $\mathcal{H}$-$\mathcal{GQ}$ can be computed with one two-fold multi-exponentiation so that they use 1.75 group operations per bit of data (in MD mode). We now turn to $\mathcal{H}$-$\mathcal{SFS}$. Since we are considering MD mode performance we refer to the MD-compatible version of the function from Equation (2). (But in fact performance is hardly affected by the modification.) On the average about half the bits of C are 1 so $\mathcal{H}$-$\mathcal{SFS}$ comes in at about 0.5 modular multiplications per bit. This explains the claim of Figure 1 in regard to $\mathcal{H}$-$\mathcal{SFS}$ without pre-computation. Now we look at how pre-computation speeds it up, using a block size of $l = 512$ (the same as MD5 and SHA-1) for illustration. The method is obvious. Pick a "width" w that divides l and let $t = l/w$. Letting $pk = (N, l, \mathbf{u}, v)$ denote the public key, pre-compute and store the table T with entries

$$T[i, x] = \textstyle\prod_{j=1}^{w} \mathbf{u}[(i-1)w + j]^x \bmod N \quad (1 \le i \le t,\ x \in \{0, \ldots 2^w - 1\})$$

The size of the table is $t2^w = l2^w/w$ group elements. Now computing $\mathcal{H}$-$\mathcal{SFS}$ takes $t + 2 = 2 + l/w$ multiplications since

$$\mathsf{H}_{pk}(\mathrm{C}, \mathrm{Z}) = \left(\textstyle\prod_{i=1}^{t} T[i, x_i]\right) \cdot \mathrm{Z}^2 \cdot v^{f_N(\mathrm{Z})} \bmod N,$$

where x_i is the integer with binary representation $\mathrm{C}[(i-1)w + 1] \ldots c[iw]$ $(1 \le i \le t)$. The number of group operations per bit is thus $[2 + l/w]/l \approx 1/w$, meaning the rate is w. Figure 1 showed the storage and this rate for $w = 4$ and $w = 8$.

Analytical assessment of the performance of $\mathcal{H}$-$\mathcal{SMS}$ is difficult, but we have implemented both it and (for comparison) $\mathcal{H}$-$\mathcal{SFS}$. The implementation used a 1024 bit modulus and (for MD mode) a 512 bit block size. Table 1 shows that $\mathcal{H}$-$\mathcal{SMS}$ is about 30 times faster than the basic (no pre-computation) version of $\mathcal{H}$-$\mathcal{SFS}$. The gap drops to a factor of 15 and 7.5 when compared with the $w = 4$ and $w = 8$ pre-computation levels of $\mathcal{H}$-$\mathcal{SFS}$, respectively. Note that $\mathcal{H}$-$\mathcal{SMS}$ here is without pre-computation. (The latter does not seem to help it much.) These implementation results are on a Dual Pentium IV, 3.2GHz machine, running Linux kernel 2.6 and using the gmp library [20].

4.4 Comparisons

We now assess performance of previous schemes, justifying claims in Section 1. Damgård [16] shows how to construct collision-resistant hash functions from claw-free permutations [21]. Of various factoring-based instantiations of his construction, the one of [21, 27], which we denote $\mathcal{H}$-$\mathcal{Da}$, seems to be the most efficient. The key is a modulus N product of two primes, one congruent to 3 mod 8 and the other to 7 mod 8, and the hash function $\mathsf{H}_N : \{0,1\}^l \times \mathbb{Z}_N^* \to \mathbb{Z}_N^*$ is defined by $\mathsf{H}_N(m, r) = 4^m \cdot r^s \bmod N$ where $s = 2^l$. Since multiplying by 4 is cheap, we view it as free and the cost is then one multiplication per bit, meaning $\mathcal{H}$-$\mathcal{SFS}$ is twice as fast. But pre-computation does not help $\mathcal{H}$-$\mathcal{Da}$ since r is not fixed, and the gap in rates increases as we allow pre-computation for $\mathcal{H}$-$\mathcal{SFS}$ as shown in Figure 1.

The key of Shamir and Tauman's [42] hash function is a modulus N and an $a \in \mathbb{Z}_N^*$. With a 1024 bit modulus the chaining variable needs to be 1024 bits as well, so that with a 512 bit block size the function would take a $512 + 1024$ bit input, regard it as an integer s, and return $a^s \bmod N$. The computation takes 1.5 multiplications per bit of the full input, which is $1.5 \cdot (1024 + 512)/512 = 4.5$ per bit of data, meaning the rate is $1/4.5 \approx 0.22$ as claimed in Figure 1. Since a is fixed, one can use the standard pre-computation methods for exponentiation. For any v dividing $1024 + 512 = 1536$, the computation takes $1536/v$ multiplications with a table of $2^v \cdot 1536/v$ group elements. Note that per data bit the rate is $512/(1536/v) = v/3$. To compare to $\mathcal{H}$-$\mathcal{SFS}$ we need to choose parameters so that the storage for the two is about the same, meaning $2^w(512/w) \approx 2^v(1536/v)$. This yields $v = 1$ for $w = 4$ and $v = 6$ for $w = 8$. This explains the rates shown in Figure 1.

5 Improvements to VSH

The performance of a hash function on short inputs is important in practice. (For example, a significant fraction of Internet traffic consists of short packets.) We present a variant VSH* of VSH that is up to 5 times faster in this context while remaining proven-secure under the same assumption as VSH. The improvement stems from VSH*, unlike VSH, having a collision-resistant compression function.

BACKGROUND. The key of Contini, Lenstra and Steinfeld's VSH function [13] is a modulus N product of two primes. The VSH compression function $\mathsf{vsh}_N : \{0,1\}^l \times \mathbb{Z}_N^* \to \mathbb{Z}_N^*$ is defined by

$$\mathsf{vsh}_N(\mathrm{C}, \mathrm{Z}) = \mathrm{Z}^2 \cdot \prod_{i=1}^{l} p_i^{\mathrm{C}[i]} \bmod N,$$

where p_i is the i-th prime and $\mathrm{C}[i]$ is the i-th bit of C. The hash function VSH is obtained by MD-iteration of vsh with initial vector 1. A curious feature of VSH is that the compression function is *not* collision-resistant. Indeed, $\mathsf{vsh}_N(c, z) = \mathsf{vsh}_N(c, N - z)$ for any $c \in \{0,1\}^l$ and $z \in \mathbb{Z}_N^*$. Nonetheless, it is shown in [13] that the hash function VSH is collision-resistant based on the VSSR assumption. The latter states that given N, l it is hard to find $x \in \mathbb{Z}_N^*$ and integers $e_1, \ldots, e_l$, not all even, such that $x^2 \equiv p_1^{e_1} \cdot \ldots \cdot p_l^{e_l} \pmod N$. The proof makes crucial use of the fact that the initial vector is set to 1.

VSH*. We alter the compression function of VSH so that it becomes (provably) collision-resistant and then define VSH* by MD iteration with the initial vector being part of the data to be hashed. The first application of the compression function thus consumes much more (1024 bits more for a 1024 bit modulus, for example) of the input, resulting in significantly improved rate for the important practical case of hashing short messages. For example, the implementation results of Table 2 show speed increases of a factor of 5 over VSH when hashing 1024 bit messages. Performance for long messages is the same as for VSH. VSH* and its compression function vsh* are provably collision-resistant under the same VSSR assumption as VSH.

The inspiration comes from $\mathcal{H}$-$\mathcal{SMS}$ which we notice is very similar to vsh but, unlike the latter, is collision-resistant. The difference is that in $\mathcal{H}$-$\mathcal{SMS}$ the primes $\mathbf{u}[1], \ldots, \mathbf{u}[l], v$ —referring to the MD-compatible version of the function from Equation (2)— are quadratic residues. But this turns out to be important for the completeness of the Σ-protocol rather than for collision-resistance. This leads to the compression function $\mathsf{vsh}_N^* : \{0,1\}^l \times \mathbb{Z}_N^* \to \mathbb{Z}_N^*$ defined by

$$\mathsf{vsh}_N^*(\mathrm{C}, \mathrm{Z}) = \left(\prod_{i=1}^{l} p_i^{\mathrm{C}[i]}\right) \cdot p_{l+1}^{f_N(\mathrm{Z})} \cdot \mathrm{Z}^2 \bmod N,$$

where $f_N(\mathrm{Z}) = 0$ if $\mathrm{Z} \in Z_N^+$ and 1 otherwise, p_i is the i-th prime and $\mathrm{C}[i]$ is the i-th bit of C. As a check notice that $\mathsf{vsh}_N^*(\mathrm{C}, \mathrm{Z})$ is unlikely to equal $\mathsf{vsh}_N^*(\mathrm{C}, N - \mathrm{Z})$ because $f_N(\mathrm{Z}) \neq f_N(N - \mathrm{Z})$, meaning the attack showing vsh is not collision-resistant does not apply. Of course this is not the only possible

Hash Function	block size	input size	Iterations	Avg. time
VSH	128	8×128	9	$140\mu s$
VSH*	128	8×128	1	$25\mu s$

Table 2: The size of the modulus used here is 1024. The block and the input size are given in bits.

attack, but the proof of strong special soundness of $\mathcal{SMS}$ Proposition 3.3 can be adapted to show that vsh* is collision-resistant under the VSSR assumption. Finally VSH* is obtained by MD iteration of vsh* but with the initial vector being the first $k-1$ bits of the input. For MD-strengthening, the standard padding method of SHA-1 is used.

The implementation results given in Table 2 were again obtained on a Pentium IV, 3 GHz machine using the gmp library [20]. We set the block size to 128 for both functions and considered hashing a 1024 bit input. In this case (even taking into account the increase in length due to MD strengthening) VSH* needs 1 application of its compression function. On the other hand VSH (with their own form of strengthening) needs 9. The implementation shows that VSH* is 5.6 times faster. We need to add that our implementations (unlike those of [13]) are not optimized, but our goal was more to assess the comparative than the absolute performance of these hash functions, and this is achieved because both are tested on the same platform.

References

[1] E. Andreeva, G. Neven, B. Preneel, and T. Shrimpton. Seven-property-preserving iterated hashing: ROX. In K. Kurosawa, ed., *ASIACRYPT 2007*, vol. 4833 of *LNCS*, 2007. Springer-Verlag.

[2] G. Ateniese and B. de Medeiros. Identity-Based Chameleon Hash and Applications. In A. Juels, ed., FC 2004, vol. 3110 of LNCS, 2004. Springer-Verlag.

[3] G. Ateniese and B. de Medeiros. On the Key Exposure Problem in Chameleon Hashes. In C. Blundo and S. Cimato, eds., SCN 2004, vol. 3352 of LNCS, 2004. Springer-Verlag.

[4] M. Bellare and D. Micciancio. A new paradigm for collision-free hashing: Incrementality at reduced cost. In W. Fumy, ed., *EUROCRYPT'97*, vol. 1233 of *LNCS*, 1997. Springer-Verlag.

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. In C. Cachin and J. Camenisch, ed., *EUROCRYPT 2004*, vol. 3027 of *LNCS*, 2004. Springer-Verlag.

[6] M. Bellare and T. Ristenpart. Multi-property-preserving hash domain extension and the EMD transform. In X. Lai and K. Chen, ed., *ASIACRYPT 2006*, LNCS. Springer-Verlag.

[7] M. Bellare and T. Ristenpart. Hash functions in the dedicated-key setting: Design choices and MPP transforms. In L. Arge, C. Cachin, T. Jurdzinski, and A. Tarlecki, ed., *ICALP 2007*, vol. 4596 of *LNCS*, 2007. Springer-Verlag.

[8] M. Bellare and T. Ristov. Σ-hash Functions. Full Version of this paper. IACR eprint archive, 2008.

[9] M. Bellare and S. Shoup. Two-tier signatures, strongly unforgeable signatures, and Fiat-Shamir without random oracles. In T. Okamoto and X. Wang, ed., *PKC 2007*, vol. 4450 of *LNCS*, 2007. Springer-Verlag.

[10] T. Beth. Efficient zero-knowledge identification scheme for smart cards. In C. G. Günther, ed., *EUROCRYPT'88*, vol. 330 of *LNCS*, 1988. Springer-Verlag.

[11] D. Charles, E. Goren, and K. Lauter. Cryptographic hash functions from expander graphs. Second NIST Hash Function Workshop, 2006.

[12] D. Chaum, E. V. Heijst, and B. Pfitzmann. Cryptographically strong undeniable signatures, unconditionally secure for the signer. In J. Feigenbaum, ed., *CRYPTO'91*, vol. 576 of *LNCS*, 1991. Springer-Verlag.

[13] S. Contini, A. K. Lenstra, and R. Steinfeld. VSH, an efficient and provable collision-resistant hash function. In S. Vaudenay, ed., *EUROCRYPT 2006*, vol. 4004 of *LNCS*, 2006. Springer-Verlag.
[14] J.-S. Coron, Y. Dodis, C. Malinaud, and P. Puniya. Merkle-Damgård revisited: How to construct a hash function. In V. Shoup, ed., *CRYPTO 2005*, vol. 3621 of *LNCS*, 2005. Springer-Verlag.
[15] R. Cramer, I. Damgrd, and P. MacKenzie. Efficient zero-knowledge proofs of knowledge without intractability assumptions. In D. Naccache and P. Paillier, ed., *PKC 2002*, vol. 2274 of *LNCS*, 2002. Springer-Verlag.
[16] I. Damgård. Collision free hash functions and public key signature schemes. In C. G. Günther, ed., *EUROCRYPT'88*, vol. 330 of *LNCS*, 1988. Springer-Verlag.
[17] I. Damgård. A design principle for hash functions. In G. Brassard, ed., *CRYPTO'89*, vol. 435 of *LNCS*, 1990. Springer-Verlag.
[18] I. Damgård. On the existence of bit commitment schemes and zero-knowledge proofs. In G. Brassard, ed., *CRYPTO'89*, vol. 435 of *LNCS*, 1990. Springer-Verlag.
[19] A. Fiat and A. Shamir. How to prove yourself: Practical solutions to identification and signature problems. In A. M. Odlyzko, ed., *CRYPTO'86*, vol. 263 of *LNCS*, August 1987. Springer-Verlag.
[20] The GNU MP bignum library. `http://gmplib.org/`.
[21] S. Goldwasser, S. Micali, and R. Rivest. A Digital Signature Scheme Secure Against Adaptive Chosen-Message Attacks. *SIAM J. on Computing, 17*, 1988.
[22] L. C. Guillou and J.-J. Quisquater. A "paradoxical" indentity-based signature scheme resulting from zero-knowledge. In S. Goldwasser, ed., *CRYPTO'88*, vol. 403 of *LNCS*, 1990. Springer-Verlag.
[23] F. Hess. Efficient identity based signature schemes based on pairings. In K. Nyberg and H. M. Heys, ed., *SAC 2002*, vol. 2595 of *LNCS*, 2003. Springer-Verlag.
[24] Y. Ishai, E. Kushilevitz, and R. Ostrovsky. Sufficient conditions for collision-resistant hashing. In J. Kilian, ed., *TCC 2005*, vol. 3378 of *LNCS*, 2005. Springer-Verlag.
[25] M. Jakobsson, K. Sako, and R. Impagliazzo. Designated verifier proofs and their applications. In U. M. Maurer, ed., *EUROCRYPT'96*, vol. 1070 of *LNCS*, 1996. Springer-Verlag.
[26] J. Jonsson and B. Kaliski. Public-key cryptography standards (PKCS) #1: RSA cryptography, specifications version 2.1., 2003. In Internet RFC 3447.
[27] H. Krawczyk and T. Rabin. Chameleon hashing and signatures. In *ISOC Network and Distributed System Security Symposium (NDSS 2000)*, 2000.
[28] V. Lyubashevsky and D. Micciancio. Generalized compact knapsacks are collision resistant. In M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, ed., *ICALP 2006*, vol. 4052 of *LNCS*, 2006.
[29] V. Lyubashevsky, D. Micciancio, C. Peikert, and A. Rosen. SWIFFT: a modest proposal for FFT hashing. In K. Nyberg, ed., *FSE'2008*, vol. 5086 of *LNCS*, 2008. Springer-Verlag.
[30] R. C. Merkle. A certified digital signature. In G. Brassard, ed., *CRYPTO'89*, vol. 435 of *LNCS*, 1990. Springer-Verlag.
[31] S. Micali and A. Shamir. An improvement of the Fiat-Shamir identification and signature scheme. In S. Goldwasser, ed., *CRYPTO'88*, vol. 403 of *LNCS*, 1990. Springer-Verlag.
[32] M. Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991.
[33] K. Ohta and T. Okamoto. A modification of the Fiat-Shamir scheme. In S. Goldwasser, ed., *CRYPTO'88*, vol. 403 of *LNCS*, 1990. Springer-Verlag.
[34] T. Okamoto. Provably secure and practical identification schemes and corresponding signature schemes. In D. R. Stinson, ed., *CRYPTO'93*, vol. 773 of *LNCS*, 1994. Springer-Verlag.
[35] H. Ong and C.-P. Schnorr. Fast signature generation with a Fiat-Shamir like scheme. In I. Damgård, ed., *EUROCRYPT'90*, vol. 473 of *LNCS*, 1990. Springer-Verlag.
[36] C. Petit, N. Veyrat-Charvillon, and J.-J. Quisquater. Efficiency and Pseudo-Randomness of a Variant of Zémor-Tillich Hash Function. In ICECS 2008. IEEE Proceedings.
[37] C. Petit, K. Lauter and J.-J. Quisquater. Full cryptanalysis of LPS and Morgenstern Hash Functions. In R. Ostrovsky, R. D. Prisco, and I. Visconti, eds., Security and Cryptography for Networks, vol. 5229 of LNCS, 2008. Springer-Verlag.
[38] P. Paillier. Public-key cryptosystems based on composite degree residuosity classes. In J. Stern, ed., *EUROCRYPT'99*, vol. 1592 of *LNCS*, 1999. Springer-Verlag.

[39] C. Peikert and A. Rosen. Efficient Collision-Resistant Hashing from Worst-Case Assumptions on Cyclic Lattices. In S. Halevi and T. Rabin, ed., *TCC 2006*, vol. 3876 of *LNCS*, 2006. Springer-Verlag.
[40] R. Sakai, K. Ohgishi, and M. Kasahara. Cryptosystems based on pairing. In *SCIS 2000*, 2000.
[41] C.-P. Schnorr. Efficient signature generation by smart cards. *Journal of Cryptology*, 4(3):161–174, 1991.
[42] A. Shamir and Y. Tauman. Improved online/offline signature schemes. In J. Kilian, ed., *CRYPTO 2001*, vol. 2139 of *LNCS*, 2001. Springer-Verlag.
[43] D. R. Simon. Finding collisions on a one-way street: Can secure hash functions be based on general assumptions? In K. Nyberg, ed., *EUROCRYPT'98*, vol. 1403 of *LNCS*, 1998. Springer-Verlag.
[44] R. Steinfeld, H. Wang, and J. Pieprzyk. Efficient extension of standard Schnorr/RSA signatures into universal designated-verifier signatures. In F. Bao, R. Deng, and J. Zhou, ed., *PKC 2004*, vol. 2947 of *LNCS* .
[45] J.-P. Tillich and G. Zémor. Hashing with SL_2. In Y. Desmedt, ed., *CRYPTO 1994*, vol. 839 of *LNCS*. Springer-Verlag.
[46] J.-P. Tillich and G. Zémor. Collisions for the LPS expander graph hash function. In N. P. Smart, ed., *EUROCRYPT 2008*, vol. 4965 of *LNCS*. Springer-Verlag.
[47] X. Wang, Y.L. Yin, and H. Yu. Finding collisions in the full SHA-1. In V. Shoup, ed., *CRYPTO 2005*, vol. 3621 of *LNCS*, 2005. Springer-Verlag.
[48] X. Wang, Y.L. Yin, and H. Yu. How to break MD5 and other hash functions. In R. Cramer, ed., *EUROCRYPT 2005*, vol. 3494 of *LNCS*, 2005. Springer-Verlag.

A Extending the domain

Since a Σ-hash has two inputs, there is a natural way to regard it as a compression function and then run it in MD mode to get a full-fledged hash function. Namely, regard the first input of H_{pk} : $\mathrm{ChSet}(pk) \times \mathrm{RpSet}(pk) \to \mathrm{CmSet}(pk)$ as the data and the second as the chaining variable. For this to work however, one must be able to view the output as a chaining value, meaning we need $\mathrm{CmSet}(pk) \subseteq \mathrm{RpSet}(pk)$. Sometimes this is true, as for $\mathcal{H}$-$\mathcal{GQ}$, which in this way lends itself easily and naturally to MD iteration. But in the case of $\mathcal{SFS}$ and $\mathcal{SMS}$ we have $\mathrm{CmSet}((N, l, \mathbf{u})) = \mathbb{Z}_N^* \subsetneqq \mathbb{Z}_N^+ = \mathrm{RpSet}((N, l, \mathbf{u}))$. In the case of $\mathcal{H}$-$\mathcal{Sch}$, we would like to work over on elliptic curve group because then the group size can be smaller (about 2^{160}) and computational costs are reduced. However, when $\mathrm{CmSet}(\langle G\rangle, p, g, X) = G$ is an elliptic curve group, a group element is represented as a pair (x, y) where $x \in \mathbb{Z}_p$ and y is a bit, and this G isn't a subset of $\mathrm{RpSet}((\langle G\rangle, p, g, X)) = \mathbb{Z}_p$. However, we now present a simple and general way to get around these problems and in particular make $\mathcal{H}$-$\mathcal{SFS}$, $\mathcal{H}$-$\mathcal{SMS}$ and $\mathcal{H}$-$\mathcal{Sch}$ amenable to MD iteration. Let $\mathcal{H} = (\mathsf{KG}, \mathsf{H}, \mathrm{ChSet}, \mathrm{CmSet}, \mathrm{RpSet})$ be a Σ-hash family. Let $d(\cdot)$ be an integer valued function called the data length. We now find an *embedding* e. By this we mean that $e_{pk} : \{0,1\}^{d(k)} \times \mathrm{CmSet}(pk) \to \mathrm{ChSet}(pk) \times \mathrm{RpSet}(pk)$ is an injective map for every $pk \in \big[K(1^k)\big]$ and every k. Now define $\mathcal{H}^d = (\mathsf{K}, \mathsf{H}^d, \{0,1\}^{d(\cdot)}, \mathrm{CmSet}, \mathrm{CmSet})$ by

$$\mathsf{H}^d_{pk}(m, w) = \mathsf{H}_{pk}(e^{(1)}_{pk}(m, w), e^{(2)}_{pk}(m, w))$$

where $e^{(i)}_{pk}(m, w)$ is the i-th component of the tuple $e_{pk}(m, w)$ for $i = 1, 2$. Then $\mathcal{H}^d$ is MD-compatible because the range of H^d_{pk} is the domain of its second argument and thus the second argument can be used as a chaining variable. On the other hand it is easy to see that the injectivity of e_{pk} implies that $\mathcal{H}^d$ inherits the collision-resistance of $\mathcal{H}$. So MD-iteration of $\mathcal{H}^d$ yields a full-fledged hash function which is collision-resistant.

Let us now apply this to $\mathcal{H}$-$\mathcal{SFS}$, $\mathcal{H}$-$\mathcal{SMS}$, and $\mathcal{H}$-$\mathcal{Sch}$ by finding suitable embeddings. To maximize data throughput, we should choose d as large as possible.

Suppose $\mathcal{H} = (\mathsf{KG}, \mathsf{H}, \mathrm{ChSet}, \mathrm{CmSet}, \mathrm{RpSet})$ is $\mathcal{H}$-$\mathcal{SFS}$ or $\mathcal{H}$-$\mathcal{SMS}$, so that $\mathrm{ChSet}(pk) = \{0,1\}^l$

and $\mathrm{RpSet}(pk) = \mathbb{Z}_N^+$ and $\mathrm{CmSet}(pk) = \mathbb{Z}_N^*$ where $pk = (N, l, \mathbf{u})$ and $l = L(k)$, with $L(\cdot)$ being the challenge length. Let $d(\cdot) = L(\cdot) - 1$. For every N the map $f_N : \mathbb{Z}_N^* \to \{0,1\} \times \mathbb{Z}_N^+$ defined by $f_N(w) = (0, w)$ if $w \leq N/2$ and $(1, N - w)$ otherwise is a bijection. Let $e_{pk} : \{0,1\}^{l-1} \times \mathbb{Z}_N^* \to \{0,1\}^l \times \mathbb{Z}_N^+$ be defined by $e_{pk}(m, w) = (m \| f_N^{(1)}(w), f_N^{(2)}(w))$ where $f_N^{(i)}(w)$ is the i-th component of the pair $f_N(w)$ for $i = 1, 2$. Then e is an embedding, so from the above $\mathcal{H}^d$ is MD-compatible and collision-resistant as long as $\mathcal{H}$ is collision-resistant. MD iterate $\mathcal{H}^d$ to get a full-fledged hash function.

B Proofs of Theorems 3.4 and 3.5

Proof of Theorem 3.4: Theorem 3.1 implies that $\mathcal{H}$-$\mathcal{SP} = (\mathsf{KG}, \mathsf{H}, \mathrm{ChSet}, \mathrm{CmSet}, \mathrm{RpSet})$ is collision-resistant. We now show that the strong HVZK property of $\mathcal{SP}$ implies uniformity of $\mathcal{H}$-$\mathcal{SP}$. Fix $(pk, sk) \in \left[\mathsf{Kg}(1^k)\right]$ and also fix $\mathrm{C}_1, \mathrm{C}_2 \in \mathrm{ChSet}(pk)$. We want to show that $\mathsf{H}_{pk}(\mathrm{C}_1, \cdot)$ and $\mathsf{H}_{pk}(\mathrm{C}_2, \cdot)$ are identically distributed when their argument is drawn at random from $\mathrm{RpSet}(k)$. We will do this by showing that the statistical distance between these distributions is zero. Consider the games of Figure 7. Let D be any (computationally unbounded) adversary. Then it suffices to show that

$$\Pr\left[\mathrm{S}_1^D \Rightarrow 1\right] = \Pr\left[\mathrm{S}_2^D \Rightarrow 1\right]$$

where "$G^D \Rightarrow 1$" denotes the event that D outputs 1 on input the output of game G, and the probability is over the coins of G and D. But this follows because

$$\Pr\left[\mathrm{S}_1^D \Rightarrow 1\right] = \Pr\left[\mathrm{R}_1{}^D \Rightarrow 1\right] \quad (3)$$

$$\Pr\left[\mathrm{S}_2^D \Rightarrow 1\right] = \Pr\left[\mathrm{R}_2{}^D \Rightarrow 1\right] \quad (4)$$

$$\Pr\left[\mathrm{R}_1^D \Rightarrow 1\right] = \Pr\left[\mathrm{R}_2{}^D \Rightarrow 1\right], \quad (5)$$

Where games $\mathrm{R}_1, \mathrm{R}_2$ are also in Figure 7. Equations (3),(4) are true by the strong HVZK property. (The real and simulated conversation transcripts are equally distributed, and hence continue to be so conditioned on a particular challenge.) Equation (5) is true because Y is generated the same way in both games.

To show that $\mathcal{H}$-$\mathcal{SP}$ is trapdoor we need to exhibit the full key-generation and inversion algorithms. The former is simply the key-generation algorithm K of $\mathcal{SP}$, so that the trapdoor is the secret key of the protocol. On input $pk, sk, \mathrm{C}_1, \mathrm{C}_2, \mathrm{Z}_1$ the inversion algorithm lets $\mathrm{Y} \leftarrow \mathsf{H}_{pk}(\mathrm{C}_1, \mathrm{Z}_1)$ and returns $\mathrm{Z}_2 \leftarrow \mathcal{P}^*(pk, sk, \mathrm{Y}, \mathrm{C}_1, \mathrm{Z}_1, \mathrm{C}_2)$ where $\mathcal{P}^*$ is the strong completeness algorithm. We now claim that this satisfies the condition in the definition, namely that for any $pk, sk, \mathrm{C}_1, \mathrm{C}_2, Y$ the map f defined by $f(z) = \mathcal{P}^*(pk, sk, \mathrm{Y}, \mathrm{C}_1, \mathrm{C}_2, z)$ is a bijection from $S_{pk,\mathrm{C}_1}(\mathrm{Y})$ to $S_{pk,\mathrm{C}_2}(\mathrm{Y})$, where $S_{pk,\mathrm{C}}(\mathrm{Y}) = \{\mathrm{Z} \in \mathrm{RpSet}(pk) : \mathsf{H}_{pk}(\mathrm{C}, \mathrm{Z}) = \mathrm{Y}\}$. But the uniformity property we proved above implies that the sets $S_{pk,\mathrm{C}_1}(\mathrm{Y})$ and $S_{pk,\mathrm{C}_2}(\mathrm{Y})$ have the same size. Now one can infer that f is a bijection from the strong completeness condition. ▌

Proof of Theorem 3.5: Since $\mathcal{H}$ is trapdoor, it has a full key-generation algorithm K and an inversion algorithm I. Let the former be the key-generation algorithm of $\mathcal{SP}$. Now we define the prover algorithm as shown in Figure 8. Define the verifier V on input $pk, \mathrm{Y} \| \mathrm{C} \| \mathrm{Z}$ to output 1 if $\mathsf{H}_{pk}(\mathrm{C}, \mathrm{Z}) = \mathrm{Y}$ and $\mathrm{Y} \in \mathrm{CmSet}(pk)$ and $\mathrm{C} \in \mathrm{ChSet}(pk)$ and $\mathrm{Z} \in \mathrm{RpSet}(pk)$, and 0 otherwise. We now need to show that $\mathcal{SP}$ satisfies strong HVZK, strong special soundness, and strong completeness. (We need to also show that $\mathcal{SP}$ satisfies completeness, but this is implied by strong completeness.)

Game R_1	Game R_2	Game S_1	Game S_2
$(\mathrm{Y}, \mathrm{y}) \leftarrow\!\!\$\ \mathsf{P}(pk, sk)$	$(\mathrm{Y}, \mathrm{y}) \leftarrow\!\!\$\ \mathsf{P}(pk, sk)$	$\mathrm{z} \leftarrow\!\!\$\ \mathrm{RpSet}(pk)$	$\mathrm{z} \leftarrow\!\!\$\ \mathrm{RpSet}(pk)$
$\mathrm{z} \leftarrow\!\!\$\ \mathsf{P}(pk, sk, \mathrm{c}_1)$	$\mathrm{z} \leftarrow\!\!\$\ \mathsf{P}(pk, sk)$	$\mathrm{Y} \leftarrow\!\!\$\ \mathsf{StSim}(pk, \mathrm{c}_1, \mathrm{z}, \mathrm{c}_2)$	$\mathrm{Y} \leftarrow\!\!\$\ \mathsf{StSim}(pk, \mathrm{c}_2, \mathrm{z})$
Return Y	Return Y	Return Y	Return Y

Figure 7: Games for proof of uniformity of $\mathcal{H}$-$\mathcal{SP}$ in proof of Theorem 3.4.

$\mathsf{P}(pk, sk)$	$\mathsf{P}(pk, sk, \mathrm{y})$
$\mathrm{Y} \leftarrow\!\!\$\ \mathrm{CmSet}(pk)$	$(\mathrm{c}_1, \mathrm{z}_1) \leftarrow \mathrm{y}$
$\mathrm{c}_1 \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$	$\mathrm{z}_2 \leftarrow I(pk, sk, \mathrm{c}_1, \mathrm{c}_2, \mathrm{z}_1)$
$\mathrm{z}_1 \leftarrow\!\!\$\ \mathrm{RpSet}(pk)$	Return z_2
$\mathrm{y} \leftarrow (\mathrm{c}_1, \mathrm{z}_1)$	
Return (Y, y)	

Figure 8: Prover algorithm for the proof of Theorem 3.5

Let StSim be defined by $\mathsf{StSim}(pk, \mathrm{c}, \mathrm{z}) = \mathsf{H}_{pk}(\mathrm{c}, \mathrm{z})$. We now show this is a strong HVZK simulator. Fix $(pk, sk) \in \mathsf{K}(1^k)$ and consider the games of Figure 9.

We want to show that

$$\Pr\left[\,\mathrm{R}^D \Rightarrow 1\,\right] = \Pr\left[\,\mathrm{S}^D \Rightarrow 1\,\right] \tag{6}$$

for any (computationally unbounded) adversary D. But the trapdoor property implies that

$$\Pr\left[\,\mathrm{R}^D \Rightarrow 1\,\right] = \Pr\left[\,\mathrm{T}^D \Rightarrow 1\,\right] .$$

On the other hand, by the uniformity property we know that there is a (not necessarily PT) algorithm $\mathcal{Y}$ such that for all $\mathrm{c} \in \mathrm{ChSet}(pk)$ the output of the following processes are identically distributed

$\mathrm{Y} \leftarrow\!\!\$\ \mathcal{Y}(pk)$	$\mathrm{z} \leftarrow\!\!\$\ \mathrm{RpSet}(pk)$
Return Y	Return $\mathsf{H}_{pk}(\mathrm{c}, \mathrm{z})$

This means that

$$\Pr\left[\,\mathrm{T}^D \Rightarrow 1\,\right] = \Pr\left[\,\mathrm{U}^D \Rightarrow 1\,\right] .$$

But,

$$\Pr\left[\,\mathrm{U}^D \Rightarrow 1\,\right] = \Pr\left[\,\mathrm{S}^D \Rightarrow 1\,\right]$$

so we have equation Equation (6).

Game R	Game S	Game T	Game U
$(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k)$	$\mathrm{c} \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$	$(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k)$	$(pk, sk) \leftarrow\!\!\$\ \mathsf{K}(1^k)$
$\mathrm{c}_1 \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$	$\mathrm{z} \leftarrow\!\!\$\ \mathrm{RpSet}(pk)$	$\mathrm{c}_1 \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$	$\mathrm{Y} \leftarrow\!\!\$\ \mathcal{Y}(pk)$
$\mathrm{z}_1 \leftarrow\!\!\$\ \mathrm{RpSet}(pk)$	$\mathrm{Y} \leftarrow \mathsf{H}_{pk}(\mathrm{c}, \mathrm{z})$	$\mathrm{z}_1 \leftarrow\!\!\$\ \mathrm{RpSet}(pk)$	$\mathrm{c}_2 \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$
$\mathrm{Y} \leftarrow \mathsf{H}_{pk}(\mathrm{c}_1, \mathrm{z}_1)$	Return $\mathrm{Y}\|\mathrm{c}\|\mathrm{z}$	$\mathrm{Y} \leftarrow\!\!\$\ \mathsf{H}_{pk}(\mathrm{c}_1, \mathrm{z}_1)$	$\mathrm{z}_2 \leftarrow\!\!\$\ \mathrm{S}_{pk,\mathrm{c}_2}(\mathrm{Y})$
$\mathrm{c}_2 \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$		$\mathrm{c}_2 \leftarrow\!\!\$\ \mathrm{ChSet}(pk)$	Return $(pk, \mathrm{Y}\|\mathrm{c}_2\|\mathrm{z}_2)$
$\mathrm{z}_2 \leftarrow I(pk, sk, \mathrm{c}_1, \mathrm{c}_2, \mathrm{z}_1)$		$\mathrm{z}_2 \leftarrow\!\!\$\ \mathrm{S}_{pk,\mathrm{c}_2}(\mathrm{Y})$	
Return $pk, \mathrm{Y}\|\mathrm{c}_2\|\mathrm{z}_2$		Return $pk, \mathrm{Y}\|\mathrm{c}_2\|\mathrm{z}_2$	

Figure 9: Games for proof of strong HVZK in proof of Theorem 3.5.

Now we claim that the collision-resistance of $\mathcal{H}$ implies strong special soundness of $\mathcal{SP}$. This follows from the fact that StSim as defined above is a stong HVZK simulator.

Finally the trapdoor property of $\mathcal{H}$ implies strong completeness of $\mathcal{SP}$. ∎