Unprovability of Leakage-Resilient Cryptography Beyond the Information-Theoretic Limit*

Rafael Pass†
Cornell Tech
rafael@cs.cornell.edu

May 11, 2021

Abstract

In recent years, *leakage-resilient* cryptography—the design of cryptographic protocols resilient to bounded leakage of honest players' secrets—has received significant attention. A major limitation of known provably-secure constructions (based on polynomial hardness assumptions) is that they require the secrets to have sufficient *actual* (i.e., information-theoretic), as opposed to *computational*, min-entropy even after the leakage.

In this work, we present barriers to provably-secure constructions beyond the "information-theoretic barrier": Assume the existence of collision-resistant hash functions. Then, no $\mathcal{NP}$ search problem with (2^{n^ϵ})-bounded number of witnesses can be proven (even worst-case) hard in the presence of $O(n^\epsilon)$ bits of computationally-efficient leakage of the witness, using a black-box reduction to any $O(1)$-round assumption. In particular, this implies that $O(n^\epsilon)$-leakage resilient *injective* one-way functions, and more generally, one-way functions with at most 2^{n^ϵ} pre-images, cannot be based on any "standard" complexity assumption using a black-box reduction.

*An extended abstract of this paper appeared in *SCN'20*. This is the full version.

†Supported in part by a JP Morgan Faculty Award, NSF Award SATC-1704788, NSF Award RI-1703846, and AFOSR Award FA9550-18-1-0267. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2019-19-020700006. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

1 Introduction

Modern Cryptography relies on the principle that cryptographic schemes are proven secure based on mathematically precise assumptions; these can be *general*—such as the existence of one-way functions—or *specific*—such as the hardness of factoring products of large primes. The security proof is a *reduction* that transforms any attacker A of the scheme into a machine that breaks the underlying assumption (e.g., inverts an alleged one-way function). This study has been extremely successful, and during the past three decades many cryptographic tasks have been put under rigorous treatment and numerous constructions realizing these tasks have been proposed under a number of well-studied complexity-theoretic hardness assumptions. In this work, we focus on *black-box* (a.k.a. *Turing*) reductions M that only use the presumed attacker A as a black-box.

Leakage-resilient Cryptography In recent years, *leakage-resilient* cryptography [ISW03, MR04, DP08, AGV09]—the design of cryptographic protocols resilient to some forms of leakage of honest players'secrets—has received significat attention. Our focus on one of the most popular and simplest way of formalizing leakage resilience—the *bounded leakage model* [Mau92, AGV09]—where the attacker may receive some bounded amount of leakage on the secret. For concreteness, consider a one-way function f. We say that a one-way function is $\ell(\cdot)$-*leakage resilient* [Mau92, AGV09, KV09, ADW09] if no efficient attacker can invert $y = f(x)$ for a randomly sampled x even if it gets access to an oracle $\mathsf{leak}_x(\cdot)$ that on input a circuit C outputs $C(x)$, as long as the total amount of bits received from the oracle is bounded by $\ell(|x|)$. We emphasize that the attacker may repeatedly access the oracle and may adaptively select the leakage function based on earlier leakage. access In other words, it should be hard to find a pre-image, even if you get to see arbitrary efficiently computable but adaptively-selected "leakage" functions of a pre-image x, as long as the total *length* of the leakage is bounded by $\ell(|x|)$. Clearly, no one-way function f can be leakage resilient to $\ell(\cdot)$-leakage when $\ell(n) = n$, as this enables leaking the whole pre-image. Yet, every one-way function f is $\ell(\cdot)$-leakage resilient for $\ell(n) = O(\log n)$, as an efficient attacker can simply guess the leakage with inverse polynomial probability.

The non-trivial problem is to construct cryptographic schemes that are *polynomially* (i.e., $\ell(n) = n^\epsilon$ for some $\epsilon > 0$) or even *linearly* (i.e., $\ell(n) = O(n)$) leakage resilient. Of course, if we assume that a one-way function f is *subexponentially* (resp. exponentially) secure, then the same guessing argument suffices to conclude that f is polynomially (resp. linearly) leakage resilient. More interestingly, as shown in [KV09, ADW09][1] any sufficiently compressing *second-preimage resistant* hash function (which can be constructed from any (polynomially-secure) one-way function [Rom90]) is also linearly leakage resilient.

It is instructive to recall their argument. Recall that a second-preimage resistant hash function h is a function such that no efficient attacker can, given random $i \leftarrow \{0,1\}^{p(n)}, x \leftarrow \{0,1\}^n$, find an $x' \neq x$ such that $h(i,x) = h(i,x')$. Given a second-preimage resistant hash function h, consider the function $f(i,x) = i, h(i,x)$; we aim to show that f is a leakage resilient one-way function. Towards this, assume for contradiction that there exists an efficient attacker A that, on input $i, y = h(i,x)$ and bounded leakage on x, manages to find a pre-image (i,x'). We can use this attacker to break the second-preimage resistance property of h as follows: Given i, x, we run $A(i, h(i,x))$ and *simulate* the answers to A's leakage queries by computing the leakage function on x until A outputs an inverse (i,x'). Now, if the number of pre-images to $y = h(i,x)$ is sufficiently large—which is guaranteed by the compressing property of h—a bounded number of bits of leakage does not suffice to even

[1] As far as we know, this was first observed by Ramarathan Venkatesan in 2005 (in personal communication).

information-theoretically determine x and thus with high probability, $x' \neq x$ and A must have found a "second" (different) preimage, thus breaking the second-preimage resistance property of h.

Leakage resilience beyond the information-theoretic barrier? Note that a central part of the above argument (as well as arguments to analyze other cryptographic primitives in the presence of leakage, based on the polynomial security of standard assumption), is that even after the leakage queries, there is still sufficient min-entropy in the original secret (i.e., the input x). In essence, the original input/secret x is *information-theoretically* unpredictable given the output of the function and the leakage. A natural question is whether leakage resilience for secrets that are only *computationally* hidden can be based on standard (polynomial) hardness assumptions. For instance:

> *Can we base polynomial leakage resilience of an* injective *one-way function on "standard" complexity assumptions?*

Note that for an injective one-way function f, the output of the function, $y = f(x)$, fully determines the secret x and thus there is no actual entropy in x (even before leakage). More generally, is it possible to just slightly beat the "information-theoretic barrier"? We say that a function has $B(\cdot)$-*bounded number of preimages* if for every $x \in \{0,1\}^n$, there exists at most $B(n)$ values $x' \in \{0,1\}^n$ such that $f(x') = f(x)$. For a function f with $B(n)$-bounded number of preimages, $(\log B(\cdot))$ bits of leakage is required to information-theoretically determine the input, so:

> *Can we base $O(\log B(\cdot))$-leakage resilience of a one-way function with $B(\cdot)$-bounded number of preimages on "standard" complexity assumptions?*

Towards barriers for leakage resilience An elegant work by Wichs from 2013 [Wic13] presents some initial barriers to affirmatively answering question 1. He shows that *certain restricted types of* black-box reductions cannot be used to base polynomially leakage-resilient injective one-way functions on *any* assumption that can be modelled as a security game between a challenger $\mathcal{C}$ and a polynomial-time attacker (such as all standard cryptographic assumptions). The restriction imposed by his result is that the black-box reduction does not get to access the code of the leakage queries issues by the attacker, and can simply "run" the code as a black-box.

The idea behind his result is simple: Consider a black-box reduction M that, given any attacker A that breaks polynomial leakage resilience of some injective one-way function f, breaks some assumption $\mathcal{C}$. Consider an (unbounded) attacker A that given an image y, issues a *random oracle* leakage query H, and next upon getting a response z, inverts f on y to get a pre-image x and returns x if and only if $H(x) = z$, and $\perp$ otherwise. The point is that this attacker A is essentially never useful to the reduction:

- If M does not query the random oracle on the actual (and unique by the injective property of f) pre-image x, then with overwhelming probability, the answer z to the leakage query will not satisfy the condition that $H(x) = z$ (we here rely on the fact that H is a random oracle) and thus A will answer $\perp$;
- and if M queries H on x, it already knows the pre-image itself, and thus can perfectly emulate the answer of A. So M never needs to use A and can break $\mathcal{C}$ on its own!

Let us point out, however, that the restriction to only allowing M to access the leakage query as a black-box (which is what enables the ideal attacker A to issue random oracle leakage queries) is

quite severe: it prevents us from distinguishing between computationally unbounded leakage (which clearly is unrealistic) and *computationally restricted* classes of leakage, which more realistically model "real-life" leakage. In fact, as demontrated in a beautiful work by Barak et al [BHHI10] (see also [MPS16]), in the closely related context of *key-dependent message (KDM) security*, there are non-trivial black-box reductions that (inherently) treat leakage queries in a non black-box way to establish feasibility results for a-priori bounded polynomial-sized leakage. Thus, in light of the above issue, even just question 1 remains largely open.

1.1 Our Results

In this work, assuming the existence of collision-resistant hash functions, we present strong barriers to providing an affirmative answer to question 2 (and thus also question 1) from any "standard" complexity assumption w.r.t. *any* black-box (i.e., Turing) reduction. Our impossibility rules out not just leakage-resilient one-way functions but even (worst-case) leakage-resilient hardness of $\mathcal{NP}$-*search problems* with subexponentially bounded number of witnesses. More precisely, we say that an $\mathcal{NP}$-relation R_L for a language L is (ℓ, s)-*leakage resilient* if there does not exists an efficient attacker A such that for *every* $x \in L$, $y \in R_L(x)$, A given x and $\ell(|x|)$ bits of leakage on y that is computable by $s(|x|)$ size circuits, recovers a witness $y' \in R_L(x)$ with probability 1. We remark that leakage-resilience of search problems was first considered by Aggarwal and Maurer [AM11], where they relate the leakage-resilience limit ℓ of a search problem to various other computational task related to the search problem; most notably, to the success probability of the best PPT algorithm for solving the problem (without leakage). We note that their results do not present any limitations on leakage-resilience for search problems but rather emphasize the importance of studying leakage-resilience of search problems.

Our main result presents a barrier to leakage-resilience of $\mathcal{NP}$-search problems beyond the information-theoretic limit with respect to polynomial-size computable leakage.

Theorem 1 (informally stated)**.** *Assume the existence of collision-resistant hash functions and let R_L be an $\mathcal{NP}$-relation where statements of length n have at most 2^{n^ϵ} witnesses. Then there exists some polynomial s such that if there exists a black-box reduction M for basing $(O(n^\epsilon \cdot r(n), s(n))$-leakage resilience of R_L on some $r(n)$-round assumption $\mathcal{C}$, then $\mathcal{C}$ can be broken in polynomial-time.*

By an $r(\cdot)$-round assumption $\mathcal{C}$, we refer to a security game between a challenger $\mathcal{C}$ and a polynomial-time attacker $\mathcal{A}$ that proceed in $r(n)$ rounds (given security parameter n), and the goal of the attacker is to make the challenger output 1. We note that all "standard" complexity assumptions used in Cryptography (e.g., hardness of factoring, discrete logarithms, lattice-problems etc) can be modeled as 2-round assumptions. $r(\cdot)$-round assumptions for $r(n) > 2$ capture an even larger class of assumption (e.g., the assumption that an $r(n)$-round protocol hides some secret).

We emphasize that every leakage-resilient one-way function f with (2^{n^ϵ})-bounded number of preimages directly yields an leakage-resilient $\mathcal{NP}$-relation with (2^{n^ϵ})-bounded number of witnesses and as such Theorem 1 shows barriers to basing (weak forms of) leakage-resilient one-wayness of any function with 2^{n^ϵ}-bounded number of pre-images on "standard" complexity assumptions using a black-box reduction.

We also highlight that we do not impose any restrictions on the black-box reduction. In particular, we allow the reduction to access the code of the leakage circuit. As a consequence, (in contrast to [Wic13]), our impossibility only applies to assumptions with an *a-priori bounded*, $r(n)$, number of rounds. This is *inherent*, as otherwise the assumption that an efficiently computable function f is an $(\ell(\cdot), s(\cdot))$-leakage resilient one-way function can itself be modeled as an $\ell(\cdot)$-round assumption.

(We remark, however, that our impossibility result relies on the existence of collision-resistant hash functions, whereas Wichs' result is unconditional.)

On non-black-box reductions We emphasize that in our leakage model, the attacker gets "interactive" leakage on the secret x—i.e., it gets to repeatedly and *adaptively* select the leakage functions as a result of answers to earlier queries. A more restricted notion of leakage resilience provides the attacker with just a single *non-adaptive* leakage query. In general, the two leakage models are equivalent (see e.g., [AM11]) as we can view the whole leakage-selection process by the attacker as a single leakage query; this transformation, however, uses the attacker in a non-black-box way. This observation shows that *non-black-box* reductions can be used to overcome our barrier in its most general form: the assumption that a particular function f is (ℓ, s)-leakage resilient can be based on a 2-round assumption where the attacker first submits an "attack circuit" (of unbounded polynomial size) and the challenger checks whether the circuit indeed breaks (ℓ, s)-leakage resilience of f. This "non-black-box feasibility" is not specific to the particular tasks we consider—rather, it shows that any (multi-round) falsifiable assumption (i.e., assumption where the challenger runs in polynomial time) can be based on a 2-round falsifiable assumption using a non-black-box reduction. Given this "trivial" non-black-box reduction, the best way to interpret our results is as a barrier to basing leakage-resilience on "adversary-independent" assumptions where the communication complexity of the security game is some fixed polynomial (independent of the attacker). This restriction on the assumption (which indeed is satisfied by all "standard" complexity assumption) prevents relying on the trivial non-black-box reduction (which requires communicating the code of the attacker).

1.2 Impossibility v.s. Unprovability

We briefly mention a related work by Ilan Komargodski [Kom18] that shows *impossibility* of leakage-resilient one-way functions in the presence of "one-way leakage". More precisely, Komargodksi considers a model where the attacker can receive leakage of *unbounded* length, but the leakage function is restricted to be *one way* (so that the input cannot be trivially leaked). He shows, using various obfuscation-type assumptions, that for every one-way function f, there exists some one-way leakage function h such that given $h(x)$ alone, x is (computationally) hard to recover, yet given both $f(x)$ and $h(x)$, x can be recovered. His result motivates why restricting to *bounded-length* leakage is crucial.

We note that the key difference between our results and his is that he considers a notion of leakage that makes leakage resilience impossible. In contrast (as we restrict the leakage to being short), the notion of leakage-resilience we consider is weak and it is generally believed that the primitives we consider exists. Indeed, as mentioned above, any *subexponentially-secure* one-way permutation is polynomially leakage resilient. Rather, our results are *unprovability* results: we show that leakage-resilient primitives, beyond the information-theoretic barrier, cannot be based on standard (bounded-round) assumptions using *polynomial-time* black-box reductions.

1.3 Proof Overview

Assume there exists a security reduction M such that M^A breaks the assumption $\mathcal{C}$ whenever A breaks leakage resilience of some witness relation R_L with 2^{n^ϵ} bounded number of witnesses. We want to use M to directly break $\mathcal{C}$ *without the help of* A. So, following the meta-reduction paradigm of [BV98], the goal will be to efficiently emulate A for M—that is, we will construct a meta-reduction that uses the underlying reduction M to break $\mathcal{C}$.

Ruling out "simple" reductions To explain the ideas, let us first assume that M only invokes a *single instance* of A and *does not rewind* A—this is clearly oversimplifying a lot, but we shall see later on how to overcome these restrictions. These types of reductions are sometimes referred to as "simple" reductions. Additionally, let us (for now) restrict our attention to assumptions $\mathcal{C}$ that only have two rounds (i.e., the challenger sends a single message, and the attacker provides his response).

Consider a particular unbounded attacker A that on input a statement x picks a 2-universal hash function $h : \{0,1\}^* \to \{0,1\}^{3n^\epsilon}$ and issues h as a leakage query—we highlight that h is not a random oracle but rather a concrete hash function sampled from a 2-universal family of functions. Looking forward, the idea for this leakage query is to ensure that, conditionned on this leakage, the witness is uniquely defined (with high probability).

After receiving back an answer z (supposedly $z = h(y)$ where $y \in R_L(x)$), A next "requests" to hear *a succint interactive argument of knowledge* [Kil92] of the statement that "there exists some y' such that $h(y') = z$ and $y' \in R_L(x)$". Such argument systems exist based on the existence of collision-resistant hash functions [Kil92] and relying on the PCP theorem [BFLS91, FGL+96]. Additionally, the prover in such an argument can be deterministic. As such, and due to the *succinct* nature of the argument—in particular, due to the fact that the prover's messages are "short"—the prover's next-messages function can be viewed as bounded polynomial-time computable leakage query to the witness. Finally, if A gets convinced by the argument of knowledge (that it requests through its leakage queries), it recovers (in exponential time) a witness $y'' \in R_L(x)$ such that $h(y'') = z$, and returns it.

We remark that the idea of using a succint arguments of knowledge to present impossibility results for leakage resilience originated in [NVZ14] and was subsequently used in [OPV15]; this idea goes back to even earlier work in a related context [HMR08]. While we use this idea here for a very similar purpose, we note that these earlier works used them to rule out leakage resilience to significantly more complex primitives (secure computations [Yao86, GMW87] and black-box zero-knowledge proofs [GMR89, GK96]). As such, we will have to work a bit harder to present an impossibility result for just leakage-resilience of a search problem.

We would now like to argue that M can simulate the answer to queries to *this particular* A on its own (and thus break $\mathcal{C}$ in polynomial time). The point is that for A to "say anything useful" (i.e., to provide a witness y'), M first needs to provide a convincing argument of knowledge of the statement "there exists some y' such that $H(y') = z$ and $y' \in R_L(x)$" to A, and from this argument *of knowledge*—and since M is only invoking a single instance of A and does not rewind A—a witness y'' such that $y' \in R_L(x)$" and $H(y'') = z$ can be extracted in (expected) polynomial time. M can thus simulate A's response by sending y'' to itself. Note that we here also need to rely on the fact that $\mathcal{C}$ only has 2 rounds, so that when we extract a witness from A, this can be done without having to rewind $\mathcal{C}$.

It remains, however, to argue that y'' is distributed in the same way as the witness y' that A would have returned. This follows from the fact that, with overwhelming probability over the choice of h, by the 2-universal property of h, there exists a *unique* y' such that $h(y') = z$ and $y' \in R_L(x)$—we here rely on the fact that the leakage is "beyond the information-theoretic limit" so that the output of h can uniquely determine y'.

Dealing with general (i.e., rewinding) reductions The above proof sketch, however, relies on the fact that M does not rewind A; if M can *rewind* A and reset A, and start many *concurrent* sessions with A, then it is no longer clear how to extract a witness from the proof it is providing. In particular, if the proof is not "resettably-sound" [BGGL01], then a rewinding reduction may

convince A of even false statements!

Additionally, as mentionned, it also relies on $\mathcal{C}$ only having 2 rounds, or else, we may rewind $\mathcal{C}$ when extracting the witness. Luckily, both of these problems have arisen in earlier black-box separations for interactive proofs. Consider some assumption $\mathcal{C}$ with $r(n)$ communication rounds. In [Pas11], a technique was developed for extracting witnesses from reductions that arbitrarily rewind its oracle—even if they start many interleaved, concurrent, interactions with the oracle—as long as the reduction needs to provide sufficiently many *sequential* interactive arguments of knowledge of a certain "special-sound" [CDS94] type. More precisely, $O(r(n) \cdot n^{\epsilon})$ sequential proofs of this special-sound type need to be provided to ensure an appropriate level of "resettable-soundness", while avoiding rewindings communication with $\mathcal{C}$. We here show how to adopt the above proof idea to fit into the framework of [Pas11], which we need to generalize as it only considers languages with *unique witnesses*. More precisely, we proceed in two steps:

- We show how to modify Kilian's protocol to satisfy an appropriate special-soundness property needed for the meta reduction in [Pas11]—doing this, however, requires making the communication complexity of the protocol greater than n, and thus the protocol no longer seems useful! The point is that, even if the communication complexity is large, the prover of the protocol is *laconic*—in other words, the length of the prover messages is still small and thus we can still view the prover messages as "short" leakage.

- Rather than reproving the result from [Pas11], we rely on the meta-reduction from [Pas11] in a black-box way, and instead present a meta-reduction that turns our attacker into an attacker for a primitive (namely sequentially witness hiding public-coin arguments for unique-witness relations) that is covered by the impossibility of [Pas11]. In more details, we present a meta-reduction showing that any reduction for leakage-resilience can be turned into a different reduction that satsifies the "unique witness" requirement needed for the result in [Pas11] to kick in. We can then apply the meta-reduction of [Pas11] on top of our meta-reduction. As far as we know, this techniques of "nested" meta-reductions (or composition of meta-reductions) is new and we hope that is may be useful elsewhere.

2 General Preliminaries

In this section we recall some standard definitions and preliminaries.

2.1 Notation

Integer, Strings and Vectors. We denote by N the set of natural numbers: 0, 1, 2, ... Unless otherwise specified, a natural number is presented in its binary expansion (with no *leading* 0s) whenever given as an input to an algorithm. If $n \in N$, we denote by 1^n the unary expansion of n (i.e., the concatenation of n 1's). Given a string x, we let $x|_i$ denote the i'th bit of x. We denote by $\vec{x}$ a finite sequence of elements $x_1, x_2, \ldots x_n$, and we let $|\vec{x}|$ denote the number of elements in the sequence.

Algorithms. We employ the following notation for algorithms.

Probabilistic algorithms. By a probabilistic algorithm we mean a Turing machine that receives an auxiliary random tape as input. If M is a probabilistic algorithm, then for any input x, the notation "$M_r(x)$" denotes the output of the M on input x when receiving r as random tape.

We let the notation "$M(x)$" denote the probability distribution over the outputs of M on input x where each bit of the random tape r is selected at random and independently, and then outputting $M_r(x)$.

Interactive Algorithms. We assume familiarity with the basic notions of an *Interactive Turing Machine* [GMR89] (ITM for brevity) and a *protocol*. (Briefly, a protocol is pair of ITMs computing in turns. In each turn, called a round, only one ITM is active. A round ends with the active machine either halting—in which case the protocol halts—or by sending a message m to the other machine, which becomes active with m as a special input. By an interactive algorithm we mean a (probabilistic) interactive Turing Machine.

Given a pair of interactive algorithms (A, B), we let $\langle A(a), B(b)\rangle(x)$ denote the probability distribution over the outputs of $B(b)$ after interacting with $A(a)$ on the common input x.

Oracle algorithms. An oracle algorithm is a machine that gets oracle access to another machine. We will restrict our attention to oracle algorithms that get access to *deterministic* interactive algorithms. Given a probabilistic oracle algorithm M, and a *deterministic* interactive algorithm A, we let $M^A(1^n)$ denote the probability distribution over the outputs of the algorithm M on input 1^n, when given oracle access to a function that on input a *partial transcript* $T = (q_1, r_1, .., q_l)$ outputs the next message r_i sent by A on input 1^n and receiving the messages $(q_1, .., q_l)$ if $r_1, \ldots r_{k-1}$ are the correct next-messages of A on all the earlier prefixes of T and $\perp$ otherwise. Note that this is well defined since we only consider deterministic oracles A.

Negligible and overwhelming functions. The term "negligible" is used for denoting functions that are asymptotically smaller than the inverse of any fixed polynomial. More precisely, a function $\nu(\cdot)$ from non-negative integers to reals is called *negligible* if for every constant $c > 0$ and all sufficiently large n, it holds that $\nu(n) < n^{-c}$. A function $\mu(\cdot)$ is *overwhelming* if there exists some negligible function $\nu(\cdot)$ such that $\mu(n) \geq 1 - \nu(n)$ for all n.

2.2 Witness Relations

We recall the definition of a witness relation for an $\mathcal{NP}$ language [Gol01].

Definition 1 (Witness relation)**.** *A* witness relation *for a language $L \in \mathcal{NP}$ is a binary relation R_L that is polynomially bounded, polynomial time recognizable and characterizes L by $L = \{x : \exists w \, s.t. \, (x, w) \in R_L\}$.*

We say that w is a witness for the membership $x \in L$ if $(x, w) \in R_L$. We will also let $R_L(x)$ denote the set of witnesses for the membership $x \in L$, i.e., $R_L(x) = \{w : (x, w) \in L\}$. If for each $x \in L$, there exists a single $w \in R_L(x)$, we say that R_L is a *unique witness relation*. If for each $x \in L$, there exists at most $k(\cdot)$ different witnesses $w \in R_L(x)$, we say that R_L is a $k(\cdot)$-witness relation. If there exists some constant ϵ such that R_L is a 2^{n^ϵ}-witness relation, we refer to R_L as a *witness relation with subexponentially bounded number of witnesses*.

2.3 Interactive Proofs and Arguments

We recall the standard definitions of interactive proofs and arguments.

Definition 2 (Interactive Proofs and Arguments [GMR89, BCC88])**.** *A pair of probabilistic interactive algorithms (P, V) is called an* interactive proof system *for a language L with witness relation R_L if V is polynomial-time and the following two conditions hold.*

- Completeness: *For every $x \in L$, and every $y \in R_L(x)$,*

$$\Pr\Big[\langle P(y), V\rangle(x) = 1\Big] = 1$$

- Soundness: *For every interactive algorithm P^*, there exists a negligible function ν such that for every $x \notin L$, every $z \in \{0,1\}^*$,*

$$\Pr\Big[(\langle P^*(z), V\rangle(x) = 0\Big] \geq 1 - \nu(|x|)$$

In case that the soundness condition holds only with respect to a provers P^* whose running-time is polynomially bounded in the common input, the pair (P, V) is called an interactive argument system*. If P is probabilistic polynomial-time, (P, V) is an* efficient prover *interactive proof/argument system.*

2.4 Intractability Assumptions and Black-box Reductions

Our definition of intractability assumptions and black-box reductions closely follows [Pas11] (the text below is taken almost verbatim from there). Following Naor [Nao03] (see also [DOP05, HH09, RV10]), we model an intractability assumption as an interactive game between a probabilistic machine $\mathcal{C}$—called the challenger—and an attacker A. Both parties get as input 1^n where n is the security parameter. For any $t(n) \in [0,1]$ and any "adversary" A, if $\Pr[\langle A, \mathcal{C}\rangle(1^n) = 1] \geq t(n)+p(n)$, then we say that *$A$ breaks $\mathcal{C}$ with advantage $p(n)$* over the "threshold" $t(n)$. When this happens, we might also say that *A breaks $(\mathcal{C}, t)$ with advantage $p(n)$*. Any pair $(\mathcal{C}, t)$ intuitively corresponds to the following assumption:

> *For every polynomial-time adversary A, there exists a negligible function $\nu(\cdot)$ such that for every $n \in \mathbb{N}$, A breaks $\mathcal{C}$ with advantage at most $\nu(n)$ over the threshold $t(n)$.*

We refer to $(\mathcal{C}, t)$ as an $r(\cdot)$-round assumption if $\mathcal{C}$ on input 1^n communicates with the attacker A in at most $r(n)$ communication rounds.

Black-box Reductions. We consider probabilistic polynomial-time Turing reductions—i.e., *black-box reductions*. A black-box reduction refers to a probabilistic polynomial-time oracle algorithm. Roughly speaking, a black-box reduction for basing the security of a primitive P on the hardness of an assumption $(\mathcal{C}, t)$, is a probabilistic polynomial-time oracle machine M such that whenever the oracle O "breaks" P with respect to the security parameter n, then M^O "breaks" $(\mathcal{C}, t)$ with respect to a polynomially related security parameter n' such that n' can be efficiently computed given n; see Figure 1.

We restrict ourselves to the case where $n' = n$, since without loss of generality we can always redefine the challenger $\mathcal{C}$ so that it acts as if its input was actually n' (since n' can be efficiently computed given n). To formalize this notion, we thus restrict ourselves to oracle machines M that on input 1^n always query the oracle on inputs of the form $(1^n, \cdot)$.

Definition 3. *We say that M is a* fixed-parameter black-box reduction *if M is an oracle machine such that $M(1^n)$ only queries its oracle with inputs of the form $(1^n, x)$, where $x \in \{0,1\}^*$.*

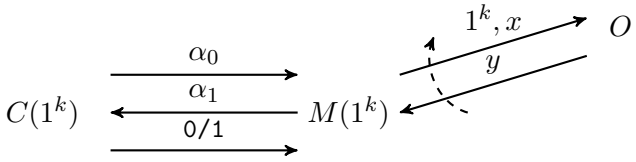

Figure 1: A black-box reduction M.

A more liberal notion of a black-box reduction allows the reduction M to (on input 1^n) query its oracle on multiple security parameters (that are all polynomially related to n). In our eyes, such a liberal notion is less justified from a practical point of view (and as far as we are aware, cryptographic reductions typically do not rely on such liberal reductions); nevertheless, our proofs directly applies also for such a notion of black-box reductions.

Remark 1 (On expected polynomial-time reductions)**.** *It is often helpful to use a more general notion of a probabilistic* expected polynomial-time *black-box reduction M, which is identically defined except that the running time of the reduction needs only be polynomially bounded in expected over the randomness of the reduction and the randomness of the challenger $\mathcal{C}$ (but independent of the attacker A it is communicating with). We note, however, that any reduction M with expected time $T(n)$ and advantage $\frac{1}{p(n)}$ can be turned into a* strict *polynomial-time reduction with running time $2p(n)T(n)$ and avantage $\frac{1}{2p(n)}$ by truncating the execution of M after $2p(n)T(n)$ steps and relying on the Markov bound. Thus, although expected polynomial-time reductions often are more natural to construct, it is without loss of generality to restrict our attention to simply strict polynomial-time reductions.*

3 Preliminaries from [Pas11]

We recall some notions and results from [Pas11] which will be useful to us. (Some of the text is taken almost verbatim from there.)

Special Soundness We start by recalling a strong notion of a *proof of knowledge* [GMR89, FFS87] that will be instrumental to us. Recall that a three-round public-coin interactive proof is said to be *special-sound* [CDS94], if a valid witness to the statement x can be efficiently computed from any two accepting proof-transcripts of x which have the same first message but different second messages. [Pas11] considers a relaxation of this notion—referred to as *computational special-soundness*—where (a) the number of communication rounds is any constant (instead of just three), (b) the extractor may need a polynomial number of accepting transcripts (instead of just two), and (c) extraction need only succeed if the transcripts are generated by communicating with a computationally-bounded prover.

Definition 4 (Computational Special-Soundness)**.** *Let (P, V) be a k-round (where k is a constant) public-coin interactive argument for the language $L \in \mathcal{NP}$ with witness relation R_L. (P, V) is said to be* computationally special-sound *if there exists a constant $i < k$, some polynomial $m(\cdot)$, and a polynomial-time extractor machine X, such that for every polynomial-time deterministic machine P^*, and every polynomial $p(\cdot)$, there exists a negligible function $\mu(\cdot)$ such that the following holds for every $x \in L$ and every auxiliary input z for P^*. Let $\vec{T} = (T_1, T_2, \ldots T_{p(|x|)})$ denote transcripts in $p(|x|)$ random executions between $P^*(x, z)$ and $V(x)$ where V uses the same randomness for the*

first $k-i-1$ messages (thus, the first $k-i-1$ messages are the same in all transcripts). Then, the probability (over the randomness used to generate $\vec{T}$) that:

1. *$\vec{T}$ contains a set of $m(|x|)$ accepting transcripts with different round $k-i$ messages; and*
2. *$X(\vec{T})$ does not output a witness $w \in R_L(x)$*

is smaller than $\mu(|x|)$. We say that a computationally special-sound protocol has a large challenge space *if the length of the verifier challenge is $\omega(\log n)$ on common inputs of length n.*

In this work, we introdue a relaxation of computational special soundness where above extraction property only needs to hold on instances x that have a unique witness (i.e., only for $x \in L$ such that there exits a single $y \in R_L(x)$); we refer to such a notion as *unique-witness computational special-soundness*.

Witness Hiding A desirable property of interactive proofs is that they "hide" the witness used by the prover. We will consider a very weak notion of *worst-case* sequential witness hiding: roughly speaking, a protocol is said to be *worst-case sequential witness hiding* if no polynomial time attacker can *always* recover the witness for any statement x that it hears $\ell(|x|)$ sequential proofs of.

Definition 5 (Worst-case Witness Hiding)**.** *Let (P,V) be an argument for the language L with witness relation R_L. We say that (a potentially unbounded) A* breaks worst-case $\ell(\cdot)$-sequential witness hiding of (P,V) *with respect to R_L if for every $n \in \mathbb{N}$, $x \in L \cap \{0,1\}^n$, and $w \in R_L(x)$, A wins in the following experiment with probability 1: Let $A(x)$ sequentially communicate with $P(x,w)$ $\ell(n)$ times; A is said to win if it outputs a witness w' such that $w' \in R_L(x)$. (P,V) is called* worst-case $\ell(\cdot)$-sequentially witness hiding *w.r.t R_L if no polynomial time algorithm A breaks worst-case $\ell(\cdot)$-sequential witness hiding of (P,V) w.r.t R_L.*

Definition 6 (Black-box Reductions for Worst-case Sequential Witness Hiding)**.** *We say that M is a* black-box reduction for basing worst-case $\ell(\cdot)$-sequential witness hiding of (P,V) *w.r.t R_L on the hardness of $(\mathcal{C},t)$ if M is a probabilistic polynomial-time oracle machine, such that for every deterministic machine A that breaks worst-case $\ell(\cdot)$-sequential witness hiding of (P,V) with respect to R_L, there exists a polynomial $p(\cdot)$ such that for infinitely many $n \in \mathbb{N}$, S^A breaks the assumption $(\mathcal{C},t)$ with advantage $\frac{1}{p(n)}$ on input 1^n.*

The Result of [Pas11] We now state (a simplified form) of the main result of [Pas11].

Theorem 2 (Main Result of [Pas11])**.** *Let (P,V) be a computationally-special-sound argument with large challenge space for the language L with a unique witness relation R_L, and let $(\mathcal{C},t)$ be an $r(\cdot)$-round assumption where $r(\cdot)$ is a polynomial. Let $\ell(n) = r(n)n^\epsilon$ for some constant $\epsilon > 0$. If there exists a fixed-parameter black-box reduction M for basing worst-case $\ell(\cdot)$-sequential witness hiding of (P,V) w.r.t R_L on the hardness of $(\mathcal{C},t)$, then there exists a polynomial $p(\cdot)$ and an efficient algorithm B such that $B(1^n)$ breaks $(\mathcal{C},t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$.*

For convenience of the reader (and because we will need to slightly generalize this result), let us provide a very high-level overview of the proof of this theorem. Assume there exists a security reduction M such that M^A breaks the assumption $\mathcal{C}$ whenever A breaks worst-case sequential witness hiding of a (computationally) special-sound argument (P,V) for a language with unique

witnesses. We want to use M to directly break $\mathcal{C}$ without the help of A—that is, we will construct a "meta-reduction" [BV98] B that uses the underlying reduction M to break $\mathcal{C}$.

Towards this, we consider a particular computationally unbounded oracle A that after hearing an appropriate number of proofs using (P, V) (acting as a verifier) simply outputs a witness to the statement proved. The meta-reduction B next needs to *efficiently* emulate A for M and thereby can efficiently break $\mathcal{C}$ (without the help of A). To enable such an emulation of A, the idea is to "extract" out the witness that A would have provide to M *from* M *itself* by "rewinding" M—since (P, V) is computationally special-sound, M, intuitively, must know a witness for all statements x that it proves to A. See Figure 2 for an illustration of the mechanics of this paradigm.

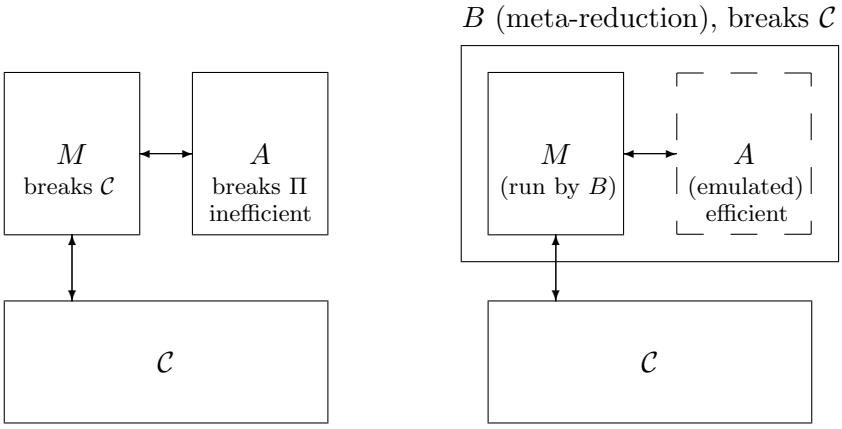

Figure 2: The meta-reduction paradigm. *Left:* M breaks the assumption $\mathcal{C}$ by using the "ideal" *inefficient* adversary A as an oracle. *Right:* the meta-reduction B runs M (forwarding its communication with $\mathcal{C}$) and efficiently emulates A to break $\mathcal{C}$.

There are several obstacles in formalizing this approach. The main one is that the reduction M is not a "stand-alone" prover—it might *rewind and reset* the oracle A—so it is no longer clear that it needs to "know" a witness for x in order to convince A of x. It is here that the proof of [Pas11] relies on the fact that there are *multiple* proofs being provided by S; this gives the meta-reduction more opportunities to rewind M, which enables extraction even if M "nests" its queries to A in an arbitrary way. Let us point out that the unique witness requirement is needed to guarantee that the witness extracted out by the meta-reduction is identically the same as the witness used by the unbounded attacker A.

We remark that the proof in [Pas11] directly also extends to languages *without unique witnesses* as long as the reduction rarely "hits" instances that do not have unique witnesses; additionally, we only require the special soundness condition to hold whenever the statement proved has a unique witness. (As can be seen in the high-level description above, this this suffices to ensure that the witnesses extracted by the meta-reduction is identically the same as the witness used by the unbounded attacker A *with overwhelming probability.*)

Theorem 3 (Slight generalization of [Pas11])**.** *Let (P, V) be a unique-witness computationally-special-sound argument with large challenge space for the language L with witness relation R_L, and let $(\mathcal{C}, t)$ be an $r(\cdot)$-round assumption where $r(\cdot)$ is a polynomial. Let $\ell(n) = r(n)n^{\epsilon}$ for some constant $\epsilon > 0$. Let M be a fixed-parameter black-box reduction for basing worst-case $\ell(\cdot)$-sequential witness hiding of (P, V) w.r.t R_L on the hardness of $(\mathcal{C}, t)$. Assume further that with overwhelming*

probability, $M(1^n)$ only queries its oracle on statements x that have a unique witness $w \in R_L(x)$. Then, there exists a polynomial $p(\cdot)$ and an efficient algorithm B such that $B(1^n)$ breaks $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$.

4 Leakage-Resilient Witness Relations and The Main Theorem

To define leakage-resilient witness relations, we consider an attacker A that receives as input a statement x and may adaptively ask for leakage of a witness y; A succeeds if it recovers any witness to x, while having seen less than $\ell(|x|)$ bits of leakage of y. More formally, let $\mathsf{leak}_y(\cdot)$ be a function that on input a circuit C outputs $C(y)$.

Definition 7 (Leakage-resilient Relations)**.** *Let R_L be a witness relation for the language L. We say that (a potentially unbounded) A* breaks *$(\ell(\cdot), s(\cdot))$*-leakage resilience of *R_L if for every $n \in \mathbb{N}, x \in L \cap \{0,1\}^n, y \in R_L(x)$, with probability 1, it holds that (a) $A_y^{\mathsf{leak}}(1^n, x) \in R_L(x)$ and (b) A received at most $\ell(|x|)$ bits from its oracle and (c) A only queries its oracle with circuits of size at most $s(|x|)$, where the probability is over the randomness of A. R_L is said to be a $(\ell(\cdot), s(\cdot))$-*leakage resilient *if there does not exists an efficient attacker A that breaks $(\ell(\cdot), s(\cdot)$-leakage resilience of R_L.*

Let us make some remarks about this notion of leakage resilience:

- $(0, s)$-leakage resilience of an $\mathcal{NP}$-relation is equivalent to stating that the $\mathcal{NP}$ search problem associated with it is hard for probabilistic polynomial time; thus, $(0, s)$-leakage resilient witness relations exists assuming $\mathcal{NP} \not\subseteq \mathcal{BPP}$. $(O(\log n), s)$-leakage resilience of a witness relation is equivalent to $(0, s)$-leakage resilence since we can simply enumerate all possible answers to the leakage queries.
- By enumerating the answers to all leakage queries, (n^ϵ, s)-leakage resilience of a witness relation R_L can be based on the assumption that the $\mathcal{NP}$ search problem associated with it cannot be solved in time $\text{poly}(2^{n^\epsilon})$.

Our main result will present barriers to basing leakage-resilient relations on polynomial-time hardness assumption. Towards this, we turn to defining what it means to base leakage resilience of R_L on some assumption using a black-box reduction.

Definition 8 (Black-box Reductions for Leakage-resilient Relations)**.** *We say that M is a* black-box reduction for basing *$(\ell(\cdot), s(\cdot))$*-leakage resilience *on the hardness of $(\mathcal{C}, t)$ if M is a probabilistic polynomial-time oracle machine, such that for every deterministic machine A that breaks $(\ell(\cdot), s(\cdot))$-leakage resilience of R_L, there exists a polynomial $p(\cdot)$ such that for infinitely many $n \in \mathbb{N}$, S^A breaks the assumption $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n.*

The Main Theorem We are now ready to state our main theorem.

Theorem 4. *Assume the existence of families of collision-resistant hash functions. Let R_L be a (2^{n^ϵ})-witness relation for some constant $\epsilon > 0$, let $(\mathcal{C}, t)$ be a $r(\cdot)$-round assumption where $r(\cdot)$ is a polynomial.. Then there exists some polynomial s and some constant $c \geq 1$ such that for $\ell(n) = c \cdot r(n) \cdot n^\epsilon$, if there exists a fixed-parameter black-box reduction M for basing $(\ell(\cdot), s(\cdot))$-leakage resilience of R_L on the hardness of $(\mathcal{C}, t)$, then there exists a polynomial $p(\cdot)$ and an efficient algorithm B such that $B(1^n)$ breaks $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$.*

5 Proof of the Main Theorem

Towards proving Theorem 4, we first present a $O(1)$-round unique-witness computationally special-sound argument for $\mathcal{NP}$ with a small prover communication complexity (i.e., with a laconic prover) and with a deterministic, efficient, prover strategy.

Next, as a warm up (and stepping-stone) to the complete proof of Theorem 4, we focus on leakage resilience of *unique witness relations* and observe that any leakage resilient *unique witness relations* with a black-box security proof together with our laconic prover argument, yields a $O(1)$-round computationally special-sound worst-case $O(n^\epsilon)$-sequentially witness hiding argument for a unique witness language: we can simply view the prover messages (which are deterministic and efficiently computable) as leakage—which is small as the length of the prover messages, as well as the number of sequential repetitions, are small—and as such leakage resilience of the relation implies worst-case witness hiding. The special case of Theorem 4 for unique witness relations can next be concluded by appealing to Theorem 2 (which rules out black-box reductions for basing worst-case $O(n^\epsilon)$-sequential witness hiding for unique witness languages on bounded-round assumptions) . In fact, if we rely on the (more general) Theorem 3, we can conclude an even stronger version that applies to any (potentially non-unique) witness relations, as long as we restrict to reductions that (with overwhelming probability) only query its oracle on instances with unique witnesses.

To conclude the full proof of Theorem 4, we next show that for any witness relation R_L with subexponentially bounded number of witnesses, there exists a different witness relation $R_{L'}$ such that any reduction for basing leakage resilience of R_L on some assumption $(\mathcal{C}, t)$ can be turned into a reduction $\tilde{M}$ for basing leakage resilience of $R_{L'}$ on $(\mathcal{C}, t)$ such that, with overwhelming probability, $\tilde{M}$ only queries its oracle on instances that have a unique witness (and as such, by the result above, is ruled out).

This is done by letting $R_{L'}$ be the set of pairs $((x, h, z), y)$ such that $y \in R_L(x)$ and $h(y) = z$ and having $\tilde{M}$ internally emulating M but for every interaction with the oracle that M initiates on input an instance x, $\tilde{M}$ first samples a 2-universal hash function h and internally issues h as a leakage query to M. Upon receiving an answer z (which is supposed to be $h(y)$ such that $y \in R_L(x)$), $\tilde{M}$ next queries the outside oracle on the instance (x, h, z) and subsequentially forwards all communication between M and the leakage oracle. By the 2-universal property of the hash function (and a union bound), it follows that if the output of the hash function is sufficiently long, then with overwhelming probability, there exists a unique pre-image for every image z and thus $\tilde{M}$ only queries its oracle on instances with unique witnesses. The point is that if the number of witnesses of x is bounded by 2^{n^ϵ}, it suffices to make the length of the output of the hash function $3n^\epsilon$ to ensure uniqueness (while at the same time ensuring that the length of the leakage is small).

5.1 A Laconic-Prover Computationally Special-sound Argument for $\mathcal{NP}$

We here present a *laconic-prover* (namely with prover communication-complexity $O(n^\epsilon)$ for any ϵ) unique-witness computationally special-sound argument for every language in $\mathcal{NP}$ with a deterministic prover, based on the existence of collision-resistant hash functions. This construction will rely on Kilian's [Kil92] succinct argument for $\mathcal{NP}$, which in turn relies on the PCP theorem [BFLS91, FGL+96].

Theorem 5. *Assume the existence of families of collision-resistant hash functions. Then, for every $\epsilon > 0$ and every language L with witness relation R_L, there exists a constant c and an efficient-prover unique-witness computationally special-sound interactive argument $\Pi = (P, V)$ for L, R_L*

with large challenge space, where (a) the prover is deterministic, and (b) the prover communication complexity is bounded by $c \cdot n^\epsilon$.

Proof. Recall that by the result of Kilian [Kil92], assuming the existence of families of collision-resistant hash function, for every $\epsilon > 0$ and every language L with witness relation R_L, there exists an efficient-prover 4-round interactive argument for L, R_L where the prover strategy is deterministic, and the prover communication complexity is $O(n^\epsilon)$. The protocol (P, V) proceeds as follows on common input a statement $x \in L$, and a witness $y \in R_L(x)$ as private input to P. Let $p(n)$ be a polynomial upper-bound on the length of witnesses for statements of length n.

- V picks a uniformly random string $r \leftarrow \{0,1\}^{p(|x|)}$ and sends it to P.
- P returns $b = \sum_{i=1}^{p(|x|)} y_i r_i \mod 2$.
- P and V next invokes Kilian's (4-round) interactive argument, letting P prove that $b = \sum_{i=1}^{p(|x|)} y_i r_i \mod 2$ and $y \in R_L(x)$.

It follows directly by Gaussian elimination and the soundness of the efficient argument that (P, V) is a unique-witness computationally special-sound interactive argument. (Note that extraction using Gaussian eliminaiton is only guaranteed to work when the witness is unique or else the malicious prover may potentially mix and match witnesses.) By definition, P is deterministic and for each $\epsilon > 0$, we can ensure that the communication complexity of P is bounded by $c \cdot n^\epsilon$ for some sufficiently large constant c. Additionally, by the efficient prover property of Kilian's argument, P can also be implemented in polynomial time. □

5.2 Unprovability of Leakage-Resilient Unique-Witness Relations

As a stepping stone (and warm-up), we start by showing barriers to showing that *unqiue* witness relations are $(O(n^\epsilon), poly)$-leakage resilient, for any constant $\epsilon > 0$. In fact, we directly show black-box unprovability of $(O(n^\epsilon), poly)$-leakage resilience of (potentially non-unique) witness relations using any reduction that with overwhelming probability only queries its oracle on instances with unique witnesses.

Lemma 1. *Assume the existence of families of collision-resistant hash functions. Let R_L be a witness relation, let $(\mathcal{C}, t)$ be an $r(\cdot)$-round assumption where $r(\cdot)$ is a polynomial, and let $\epsilon > 0$. Then, there exists some polynomial $s(\cdot)$ such that for $\ell(n) = c \cdot r(n) \cdot n^\epsilon$, if (1) there exists a fixed-parameter black-box reduction M for basing (ℓ, s)-leakage resilience of R_L on the hardness of $(\mathcal{C}, t)$ and (2) with overwhelming probability, M only queries its oracle on instances with unique witnesses, then there exists a polynomial $p(\cdot)$ and an efficient algorithm B such that $B(1^n)$ breaks $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$.*

Proof. Consider $R_L, \epsilon, r(\cdot)$ and $(\mathcal{C}, t)$ as in the statement of the lemma and assume the existence of families of collision-resistant hash functions. Pick a constant ϵ' such that $0 < \epsilon' < \epsilon$. By Theorem 5, there exists some polynomial $s(\cdot)$ and a unique-witness computationally special-sound argument (P, V) for R_L with communication complexity $c \cdot n^{\epsilon'}$ and with a deterministic prover with computational complexity $s(n)$. Let $m(n) = r(n)n^{\epsilon - \epsilon'}$ and let $\ell(n) = c \cdot r(n)n^\epsilon$ and consider some fixed-parameter black-box reduction M for basing (ℓ, s)-leakage resilience of R_L on the hardness of $(\mathcal{C}, t)$, such that with overwhelming probability M only queries its oracle on instances with unique witnesses.

We will show how to construct a reduction $\tilde{M}$ for basing worst-case $m(n)$-sequential witness hiding of the protocol (P, V) w.r.t. L, R_L on $(\mathcal{C}, t)$ such that with overwhelming probability $\tilde{M}$ only queries its oracle on instances with unique witnesses. By Theorem 3, this implies that there exists a polynomial $p(\cdot)$ and an efficient algorithm B such that $B(1^n)$ breaks $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$, and thus concludes the proof of the lemma.

Let $\mathsf{code}_P(x, q_1, \ldots, q_k)$ denote the circuit $C(\cdot)$ that on input y computes P's $(k+1)$'st message given the input (x, y) and receiving messages $q_1, \ldots, q_k$. $\tilde{M}(1^n)$ internally emulates $M(1^n)$ and proceeds as follows:

- Whenever M wants to send an oracle query $(x, p_1, q_1, \ldots, p_k)$, $\tilde{M}'$ externally forwards it to its oracle. Upon receiving back a reponse q_{k+1}, if $k = 2m(n)$ (i.e., the response is the last message from the oracle for this interaction, that is a witness), or if $q_{k+1} = \perp$, $\tilde{M}$ simply returns q_{k+1} to M. Otherwise, it returns $\mathsf{code}_P(q_1, \ldots, q_{k+1})$ to M.
- All messages that M wants to send to $\mathcal{C}$ are forwarded externally without modification, and messages that $\tilde{M}$ receives from $\mathcal{C}$ are directly forwarded to M.

Consider some attacker A that breaks worst-case $m(n)$-sequential witness hiding of (P, V) with probability 1. Let $\tilde{A}$ be a "wrapped" version of A that post-processes responses from A in exactly the same way as $\tilde{M}$ does. It directly follows from the definition of $\tilde{A}$ and the fact that (P, V) has prover communication complexity $c \cdot n^{\epsilon'}$ and computation complexity $s(n)$ that $\tilde{A}$ breaks leakage resilience of R_L with probability 1 using

$$c \cdot n^{\epsilon'} \cdot m(n) = c \cdot n^{\epsilon'} \cdot r(n) \cdot n^{\epsilon - \epsilon'} = c \cdot r(n) \cdot n^{\epsilon} = \ell(n)$$

bits of leakage. Thus, $\tilde{A}$ breaks (ℓ, s)-leakage resilience of R_L with probability 1, and consequently, $M^{\tilde{A}}$ breaks $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$ for some polynomial $p(\cdot)$. It follows that the same holds with respect to $\tilde{M}^A = M^{\tilde{A}}$, and thus $\tilde{M}$ is a black-box reduction for basing worst-case witness hiding of (P, V) w.r.t. R_L on $(\mathcal{C}, t)$. Additionally, since with overwhelming probability, M only queries its oracle on instances x that have unique witnesses, the same holds for $\tilde{M}$. □

5.3 Unprovability of Leakage-Resilent Bounded-Witness Relations

We proceed to prove Theorem 4 in its full generality. Recall that we will do this by arguing that it is essentially without loss of generality to consider reductions that with overwhelming probability only query its oracle on instances with unique witnesses. To do this, we will rely on 2-universal hash functions.

Definition 9 (2-Universal Hash Functions [CW79])**.** *A family of hash functions $\mathcal{H} = \{h : S \to T\}$ is* 2-universal *if for every $x_1 \neq x_2 \in S$, it holds that*

$$\Pr\left[h(x_1) = h(x_2)\right] \leq \frac{1}{|T|}$$

where the probability is over $h \leftarrow \mathcal{H}$.

Recall that for every prime p and every $N \in \mathbb{N}$, $\mathcal{H} = \{h_{a,b}(x) = (ax + b \mod p) \mod N\}_{a,b \in \mathbb{Z}_p}$ is a family of 2-universal hash functions [CW79]. Thus for every $m \geq n \in \mathbb{N}$, by picking a prime p between 2^n and 2^{n+1}, letting $N = 2^n$, and using truncation, we have that exists a family of 2-universal hash functions over $\mathcal{H} = \{h_i = \{0,1\}^m \to \{0,1\}^n\}$ where each h_i can be computed using

a polynomial-size circuit. Furthemore, by Chebychev's theorem on the concentration of primes, we can also efficiently sample such functions in expected polynomial time, or in strict polynomial-time with overwhelming probability.

The following simple lemma shows how we can use a 2-universal hash function to turn a subexponentially-bounded witness relation into an "almost" unique witness relation. Given a witness relation R_L, let $R_{L'}$ be the set of pairs $((x, h, z), y)$ such that $y \in R_L(x)$ and $h(y) = z$.

Lemma 2. *Let R_L be a 2^{n^ϵ}-witness relation; let $p(n)$ be a polynomial upper-bound on the length of a witness for statements of length n, $T_n = \{0,1\}^{3n^\epsilon}$ and $\mathcal{H}_n = \{h : \{0,1\}^{p(n)} \to T_n\}$ be a family of 2-universal hash functions. Then, for every n, every $x \in \{0,1\}^n$, with probability $1 - 2^{-n^\epsilon}$ over $h \leftarrow \mathcal{H}_n$, it holds that for every $z \in T_n$, there exists at most one y such that $((x, h, z), y) \in R_{L'}$.*

Proof. Consider some $x \in \{0,1\}^n$. By definition, there exists at most 2^{n^ϵ} witnesses y such that $(x, y) \in R_L$; let S_x be the set of these witnesses. By the 2-universal property of the hash function, for every two $y_1 \neq y_2 \in S_x$,

$$\Pr\left[h(y_1) = h(y_2)\right] \leq \frac{1}{|T_n|} = \frac{1}{2^{3n^\epsilon}}$$

where the probability is over the choice of h. By a union bound over $y_1, y_2 \in S_x$, it follows that the probability (over h) that there exists $y_1 \neq y_2 \in S$ such that $h(y_1) = h(y_2)$ is bounded by

$$\frac{2^{2n^\epsilon}}{2^{3n^\epsilon}} = 2^{-n^\epsilon}.$$

□

We now use this lemma to show how to turn any reduction for basing leakage resilience of a witness relation with subexponentially bounded number of witnesses on some assumption $(\mathcal{C}, t)$ into a reduction $\tilde{M}$ for basing leakage resilience of $R_{L'}$ on $(\mathcal{C}, t)$ such that, with overwhelming probability, $\tilde{M}$ only queries its oracle on instances that have a unique witness.

Lemma 3. *Let $\epsilon > 0$, let $s(\cdot), r(\cdot)$ be polynomials, let R_L be a (2^{n^ϵ})-witness relation, let $(\mathcal{C}, t)$ be an $r(\cdot)$-round assumption and let $\ell(n) \geq n^\epsilon$ be a polynomial. There exist some polynomial $s'(\cdot)$ such that, if there exists a fixed-parameter black-box reduction M for basing $(4\ell(\cdot), s'(\cdot) + s(\cdot))$-leakage resilience of R_L on the hardness of $(\mathcal{C}, t)$, then there exists a witness relation $R_{L'}$ and a fixed-parameter black-box reduction $\tilde{M}$ for basing $(\ell(\cdot), s(\cdot))$-leakage resilience of $R_{L'}$ on the hardness of $(\mathcal{C}, t)$. Furthermore, with overwhelming probability $\tilde{M}$ only queries its oracle on instances with unique witnesses.*

Proof. Consider $R_L, (\mathcal{C}, t), M, r(\cdot), \ell, s, \epsilon$ be as in statement of the lemma. Let $s'(n)$ be a polynomial bounding the circuit size of every $h \in \mathcal{H}_n$. Consider the relation $R_{L'}$ defined above, and consider a black-box reduction $\tilde{M}$ that internally emulates M but for every interaction with the oracle that M initiates on input an (new) instance x, $\tilde{M}$ first samples a $h \leftarrow \mathcal{H}_n$ and returns h as a "leakage query" to M. Recall that sampling $h \leftarrow \mathcal{H}_n$ requires sampling a random poly(n)-bit prime, which can be done with overwhelming probability in strict polynomial time. If the sampling fails, M simply lets h be a dummy circuit that outputs 0. Upon receiving an answer z, $\tilde{M}$ next queries the outside oracle on the instance (x, h, z) and subsequently forwards messages back and forth between the oracle and M for that interaction.

Consider some attacker A that breaks (ℓ, s)-leakage resilience of $R_{L'}$ with probability 1. Let $g(n)$ be a polynomial that bounds the amount of randomness used to sample h (as done by $\tilde{M}$). Given some function $f : \{0,1\}^n \to \{0,1\}^{g(n)}$ (later we will instantiate f with a truly random function), let

$\tilde{A}_f$ be a "wrapped" version of A that on input x picks a hash function $h \in \mathcal{H}_n$ using randomness $f(x)$ (in the same way as $\tilde{M}$ does), and next responds with h; upon receiving z as a response, it feeds the statement (x, h, z) to A and subsequently simply forwards all external messages back and forth to A. Note that since A breaks (ℓ, s)-leakage resilience of $R_{L'}$ with probability 1, $\tilde{A}$ breaks $(\ell(n) + 3n^\epsilon \leq 4\ell(n), s'(n) + s(n))$-leakage resilience of R_L with probability 1. Thus, for *every* function f, $M^{\tilde{A}_f}$ breaks $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$ for some polynomial $p(\cdot)$; it follows that the same holds with respect to $\tilde{M}^A = M^{\tilde{A}^{\mathbf{RO}}}$ where RO is a randomly chosen function over $\{0,1\}^n \to \{0,1\}^{g(n)}$. Additionally, note that by Lemma 2, we have that if the sampling of h succeeds, $\tilde{M}^A$ only queries A on instances x that have unique witnesses, except with probability 2^{-n^ϵ}. Finally, recall that the sampling of h succeeds with overwhelming probibility, thus by a union bound we have that with overwhelming probability, $\tilde{M}$ only queries A on instances with unique witnesses. □

Concluding the proof of Theorem 4. Let R_L be a (2^{n^ϵ})-witness relation for some constant $\epsilon > 0$ and let $(\mathcal{C}, t)$ be a $r(\cdot)$-round assumption where $r(\cdot)$ is a polynomial. Let c, s be, respectively, the constant and polynomial guaranteed to exist due to Lemma 1, and let $\ell(n) = c \cdot r(n) \cdot n^\epsilon$. By Lemma 3, there exists some polynomial s' such that the existence of a fixed-parameter black-box reduction M for basing $(4\ell, s + s')$-leakage resilience of R_L on the hardness of $(\mathcal{C}, t)$ implies the existence of some witness relation $R_{L'}$ and a fixed-parameter black-box reduction $\tilde{M}$ for basing (ℓ, s)-leakage resilience of R_L on the hardness of $(\mathcal{C}, t)$, such that with overwhelming probability M only queries its oracle on instances with unique witnesses. By Lemma 1, this implies the existence of a polynomial $p(\cdot)$ and an efficient algorithm B such that $B(1^n)$ breaks $(\mathcal{C}, t)$ with advantage $\frac{1}{p(n)}$ on input 1^n for infinitely many $n \in N$.

6 Acknowledgments

We are very grateful to the SCN anonymous reviewers for their helpful comments.

References

[ADW09] Joël Alwen, Yevgeniy Dodis, and Daniel Wichs. Survey: Leakage resilience and the bounded retrieval model. In *Information Theoretic Security, 4th International Conference, ICITS 2009, Shizuoka, Japan, December 3-6, 2009. Revised Selected Papers*, pages 1–18, 2009. 1

[AGV09] Adi Akavia, Shafi Goldwasser, and Vinod Vaikuntanathan. Simultaneous hardcore bits and cryptography against memory attacks. In *Theory of Cryptography, 6th Theory of Cryptography Conference, TCC 2009, San Francisco, CA, USA, March 15-17, 2009. Proceedings*, pages 474–495, 2009. 1

[AM11] Divesh Aggarwal and Ueli Maurer. The leakage-resilience limit of a computational problem is equal to its unpredictability entropy. In *Advances in Cryptology - ASIACRYPT 2011 - 17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, December 4-8, 2011. Proceedings*, pages 686–701, 2011. 3, 4

[BCC88] Gilles Brassard, David Chaum, and Claude Crépeau. Minimum disclosure proofs of knowledge. *J. Comput. Syst. Sci.*, 37(2):156–189, 1988. 8

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 21–31. ACM, 1991. 5, 13

[BGGL01] Boaz Barak, Oded Goldreich, Shafi Goldwasser, and Yehuda Lindell. Resettably-sound zero-knowledge and its applications. In *FOCS '02*, pages 116–125, 2001. 5

[BHHI10] Boaz Barak, Iftach Haitner, Dennis Hofheinz, and Yuval Ishai. Bounded key-dependent message security. In *Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Monaco / French Riviera, May 30 - June 3, 2010. Proceedings*, pages 423–444, 2010. 3

[BV98] Boneh and Venkatesan. Breaking RSA may not be equivalent to factoring. In *EUROCRYPT: Advances in Cryptology: Proceedings of EUROCRYPT*, 1998. 4, 11

[CDS94] Ronald Cramer, Ivan Damgård, and Berry Schoenmakers. Proofs of partial knowledge and simplified design of witness hiding protocols. In *CRYPTO*, pages 174–187, 1994. 6, 9

[CW79] Larry Carter and Mark N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18(2):143–154, 1979. 15

[DOP05] Yevgeniy Dodis, Roberto Oliveira, and Krzysztof Pietrzak. On the generic insecurity of the full domain hash. In *CRYPTO*, pages 449–466, 2005. 8

[DP08] Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 293–302, 2008. 1

[FFS87] Uriel Feige, Amos Fiat, and Adi Shamir. Zero knowledge proofs of identity. In *STOC*, pages 210–217, 1987. 9

[FGL+96] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996. 5, 13

[GK96] Oded Goldreich and Hugo Krawczyk. On the composition of zero-knowledge proof systems. *SIAM Journal on Computing*, 25(1):169–192, 1996. 5

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989. 5, 7, 8, 9

[GMW87] Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play any mental game. In *STOC '87: Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 218–229, New York, NY, USA, 1987. ACM. 5

[Gol01] Oded Goldreich. *Foundations of Cryptography — Basic Tools*. Cambridge University Press, 2001. 7

[HH09] Iftach Haitner and Thomas Holenstein. On the (im)possibility of key dependent encryption. In *TCC*, pages 202–219, 2009. 8

[HMR08] Shai Halevi, Steven Myers, and Charles Rackoff. On seed-incompressible functions. In Ran Canetti, editor, *Theory of Cryptography Conference*, pages 19–36, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. 5

[ISW03] Yuval Ishai, Amit Sahai, and David A. Wagner. Private circuits: Securing hardware against probing attacks. In *Advances in Cryptology - CRYPTO 2003, 23rd Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 2003, Proceedings*, pages 463–481, 2003. 1

[Kil92] Joe Kilian. A note on efficient zero-knowledge proofs and arguments (extended abstract). In *STOC '02*, pages 723–732, 1992. 5, 13, 14

[Kom18] Ilan Komargodski. Leakage resilient one-way functions: The auxiliary-input setting. *Theor. Comput. Sci.*, 746:6–18, 2018. 4

[KV09] Jonathan Katz and Vinod Vaikuntanathan. Signature schemes with bounded leakage resilience. In *Advances in Cryptology - ASIACRYPT 2009, 15th International Conference on the Theory and Application of Cryptology and Information Security, Tokyo, Japan, December 6-10, 2009. Proceedings*, pages 703–720, 2009. 1

[Mau92] Ueli M. Maurer. Factoring with an oracle. In *Advances in Cryptology - EUROCRYPT '92, Workshop on the Theory and Application of of Cryptographic Techniques, Balatonfüred, Hungary, May 24-28, 1992, Proceedings*, pages 429–436, 1992. 1

[MPS16] Antonio Marcedone, Rafael Pass, and Abhi Shelat. Bounded KDM security from io and OWF. In *Security and Cryptography for Networks - 10th International Conference, SCN 2016, Amalfi, Italy, August 31 - September 2, 2016, Proceedings*, pages 571–586, 2016. 3

[MR04] Silvio Micali and Leonid Reyzin. Physically observable cryptography (extended abstract). In *Theory of Cryptography, First Theory of Cryptography Conference, TCC 2004, Cambridge, MA, USA, February 19-21, 2004, Proceedings*, pages 278–296, 2004. 1

[Nao03] Moni Naor. On cryptographic assumptions and challenges. In *CRYPTO*, pages 96–109, 2003. 8

[NVZ14] Jesper Buus Nielsen, Daniele Venturi, and Angela Zottarel. On the connection between leakage tolerance and adaptive security. *IACR Cryptology ePrint Archive*, 2014:517, 2014. 5

[OPV15] Rafail Ostrovsky, Giuseppe Persiano, and Ivan Visconti. Impossibility of black-box simulation against leakage attacks. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part II*, volume 9216 of *Lecture Notes in Computer Science*, pages 130–149. Springer, 2015. 5

[Pas11] Rafael Pass. Limits of provable security from standard assumptions. In *STOC*, pages 109–118, 2011. 6, 8, 9, 10, 11

[Rom90] John Rompel. One-way functions are necessary and sufficient for secure signatures. In *STOC ;90*, pages 387–394, 1990. 1

[RV10] Guy N. Rothblum and Salil P. Vadhan. Are pcps inherent in efficient arguments? *Computational Complexity*, 19(2):265–304, 2010. 8

[Wic13] Daniel Wichs. Barriers in cryptography with weak, correlated and leaky sources. In *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, pages 111–126, 2013. 2, 3

[Yao86] Andrew Chi-Chih Yao. How to generate and exchange secrets (extended abstract). In *FOCS*, pages 162–167, 1986. 5