Towards Non-Interactive Witness Hiding

Benjamin Kuykendall and Mark Zhandry

Princeton University and NTT Research
{brk,mzhandry}@princeton.edu

Abstract. Witness hiding proofs require that the verifier cannot find a witness after seeing a proof. The exact round complexity needed for witness hiding proofs has so far remained an open question. In this work, we provide compelling evidence that witness hiding proofs are achievable *non-interactively* for wide classes of languages. We use non-interactive witness indistinguishable proofs as the basis for all of our protocols. We give four schemes in different settings under different assumptions:

- A *universal* non-interactive proof that is witness hiding as long as any proof system, possibly an inefficient and/or non-uniform scheme, is witness hiding, has a known bound on verifier runtime, and has short proofs of soundness.
- A *non-uniform* non-interactive protocol justified under a worst-case complexity assumption that is witness hiding and efficient, but may not have short proofs of soundness.
- A new security analysis of the *two-message argument* of Pass [Crypto 2003], showing witness hiding for any non-uniformly hard distribution. We propose a heuristic approach to removing the first message, yielding a non-interactive argument.
- A witness hiding non-interactive proof system for languages with *unique witnesses*, assuming the non-existence of a weak form of witness encryption for any language in $\mathsf{NP} \cap \mathsf{coNP}$.

Keywords: Witness hiding · non-interactive proofs

1 Introduction

Zero knowledge proofs [23] prove that an NP statement is true without revealing anything except the truthfulness of the statement. Such proofs, however, must depart from the usual mathematical notion of a proof by allowing multiple rounds of interaction between the prover and verifier. In fact, such proofs require at least three back-and-forth messages [21,5] between the prover and verifier — and likely more if restricted to black-box constructions [29,25] — without an additional resource such as a common reference string or a random oracle.

Weaker security properties. In order to achieve fewer rounds, and in particular to achieve the usual mathematical notion of a non-interactive proof, weaker security guarantees are necessary. Many such notions have been proposed [14,15,33,6,5,7].

Perhaps the most prominent example is witness indistinguishability, which guarantees that the proofs generated using any two witnesses are computationally indistinguishable. Non-interactive witness indistinguishable (NIWI) proofs are known from standard assumptions such as bilinear maps.

However, for general languages, it is unclear what guarantee is provided by witness indistinguishability. If the particular instance has a unique witness, then witness indistinguishability is completely meaningless, and a NIWI proof could simply be the witness itself. Even in settings with multiple witnesses, it is unclear in general what the proof recipient may learn from the proof. For example, perhaps *some* witness can be extracted from such a proof, even if the prover's own witness remains hidden.

For these reasons, NIWI proofs are typically applied to specially crafted languages where witness indistinguishability yields stronger security properties. As a result, NIWIs have been demonstrated to be useful as a building block for higher-level cryptosystems. Yet, they remain of limited use for any given language[1].

This work will focus on a different relaxation of zero knowledge called *witness hiding* [16]. Witness hiding guarantees that the verifier cannot learn any witness for the NP statement, though they may potentially reveal more than just the truthfulness of the statement. Unlike witness indistinguishability, witness hiding provides a clear, intuitive guarantee for arbitrary statements, including the case of unique witnesses.

Though the security guarantees of witness hiding proofs are apparently much weaker than zero knowledge, it has been surprisingly difficult to actually construct witness hiding proofs in fewer than three rounds. In fact, only recently have constructions for *two-message* witness hiding for all of NP been given [28,13,9]. This state of affairs may be at least partially explained by black-box barriers to witness hiding in few rounds [27]. On the other hand, certain restricted settings are known to have *non-interactive* witness hiding proofs, such as NIWI proofs in the special case when instances have two "independent" witnesses [16,6,26], or for particular protocols [12,8].

Given the difficulty of even achieving two-message witness hiding and the limited positive results for the non-interactive setting, the central question in this paper is:

Is non-interactive witness hiding possible,
and if so, what is needed to construct it?

[1] The situation is similar to that of obfuscation, which historically been used to protect intellectual property in software. Here, the ideal notion of Virtual Black Box obfuscation is impossible in general [4], so we consider an indistinguishability notion instead. This weaker notion sees use as a cryptographic building block, but has limited to no meaning for obfuscating general programs, and as such provides no guarantee for the original application to protecting intellectual property.

1.1 Results

In this work, we give a number of positive results for witness hiding in one or two messages. Our protocols work in different settings and rely on different assumptions. Taken together, however, we believe they strongly suggest that non-interactive witness hiding should be possible for all of NP.

In Section 3 we review the two-round proof system of [32] and provide a new proof of soundess. While it was already known that the protocol is witness hiding for quasipolynomially hard distributions, we analyze distributions with standard albeit non-uniform hardness. To achieve this, we weaken the model on interaction, considering the *delayed input* setting where the verifier only gets the instance x after sending its first message.

Theorem 1. *Assume quasipolynomially hard one-way-functions and perfectly sound NIWIs for* NP*. Then for any distribution of instances for which it is hard for efficient non-uniform adversaries to find witnesses, the argument system of [32] is witness hiding argument in the delayed input model.*

In Section 4 we build a non-uniform scheme, meaning that for each distribution and some choice of advice shared by the prover and verifier, the proof system is witness hiding. The result uses super-polynomially secure primitives and relies on a new complexity assumption that can be considered as a quantitative strengthening of $\mathsf{MA} \not\subseteq \mathsf{coNP}$. The choice of parameters is given in the body of the paper.

Theorem 2. *Assume some language in* coNP*, for all but finitely many input lengths, lacks an* MA*-type proof system where the verifier is allowed some specified super-polynomial runtime and witness size. Assume NIWIs for* NP *with some specified super-polynomial security. Then for any distribution of instances for which it is super-polynomially hard for efficient adversaries to find witnesses, there exists a choice of advice such that our construction is witness hiding.*

In Section 5 we build an explicit universal NIWH proof system parameterized by a runtime. If any NIWH scheme exists with a verifier that runs within the time bound and satisfies a *provable soundness* condition we define in the body, then the universal scheme will be witness hiding. Even if the secure scheme has an inefficient prover, the universal scheme will still be efficient. Even if the secure scheme is non-uniform, the universal scheme will still be uniform; although provable soundness must be defined differently in this setting, requiring short proofs of soundness for each input length. We argue that this proof can be extended to arbitrary falsifiable security properties other than witness hiding. In this sense the construction is actually the "best possible" non-interactive proof.

Theorem 3. *Take any distribution D. Assume there exists a non-interactive proof system P, V with an unbounded prover but verifier runtime s. Assume soundness of V is provable in some fixed logical proof system. If P, V is witness hiding for D, then an explicit universal construction (independent of P, V and D but depending on s) is witness hiding as well.*

In Section 6 we present a non-interactive proof system for any language with unique witnesses. As part of the construction, a distribution E over instances of an $\mathsf{NP} \cap \mathsf{coNP}$ problem is used. Security is as follows: a successful adversary against witness hiding yields a weak form of witness encryption where the instances are drawn from E. This alone is slightly hard to interpret as a positive result. But by combining with the best-possible proof system above, we can avoid a concrete choice of E.

Theorem 4. *Assume some language $T \in \mathsf{NP} \cap \mathsf{coNP}$ lacks a witness encryption scheme with average case correctness relative to some ensemble E for all input lengths large enough. Then the best possible proof system above, with a time parameter calculated in the proof, is witness hiding for any distribution over instances with unique witnesses which are hard for efficient adversaries to find.*

1.2 Technical details

To begin, we recall how non-interactive witness hiding proofs are used to construct non-interactive zero knowledge (and hence witness hiding) proofs in the common reference string model. The common reference string will consist of a commitment to 0: $\mathsf{CRS} = \mathsf{Comm}(0; r)$, where r are the random coins. To prove an NP statement x using witness w, compute a NIWI proof π of the statement

$$x' = x \vee (\exists r : \mathsf{CRS} = \mathsf{Comm}(1; r)).$$

Assuming the commitment Comm is perfectly binding, x' is equivalent to x since the second clause is false. Therefore, a proof of x' also proves x. To show zero knowledge, one switches to an experiment where $\mathsf{CRS} = \mathsf{Comm}(1; r)$, which is undetectable by the hiding property of the commitment. At this point, x' can be proven with witness r, and witness indistinguishability guarantees this proof is indistinguishable from the honest one. But the new proof is independent of the witness for x.

Witness hiding arguments in one and two messages (section 3). Building on this idea, we consider the following proof system which eschews the common reference string. Let y be any *false* instance of an NP language. To prove a statement x, compute a NIWI proof of the statement

$$x' = x \vee y$$

As before, this proof is sound because y is false. But what about witness hiding? In the example above using commitments, we switch to a setting where we can generate proofs without knowing a witness for x. In the case now, this would seem to require switching y to be true. But if y is chosen by a single party, this could compromise security. Indeed, a malicious prover could generate y to be true and therefore use the satisfying assignment to generate an invalid proof. Meanwhile, a malicious verifier could ensure that y is always false, preventing us from switching to a true y to prove witness hiding. Addressing these two concerns simultaneously is the goal of each of our constructions.

The work of Pass in [32] presents a solution in two rounds. First, the verifier chooses a true statement y, specifically in the form

$$y_b = \exists r \ : \ f(r) = b$$

where f is a one-way function. Then the prover sends a NIWI[2] of $x \vee y_b$ along with a perfectly binding commitment to a witness.

The proofs of soundness and witness hiding proceed through *complexity leveraging*: the reductions will inefficiently invert the one-way function and commitment. To prove witness hiding, we simulate the proof by brute forcing $r = f^{-1}(b)$ then run the witness hiding adversary on the simulated proof.[3] To prove soundness, we open the commitment by brute force, in turn breaking the one-way function. In order for these attacks to yield contradictions, we need *quasipolynomial* security guarantees: in particular, witness hiding is only guaranteed when finding a witness is hard for quasipolynomial time adversaries.

We present a novel security analysis of the Pass construction that avoids complexity leveraging by using non-uniformity instead. Unfortunately, this analysis only works in the *delayed input model* where the verifier does not learn the instance x until after their message is sent.

Previous works also achieve witness hiding proof in two messages. The work of [28] is also only secure in the delayed input model, so their result is comparable. The proof system of [9] is secure in the usual communication model; however, it require strong primitives such as fully homomorphic encryption and compute-and-compare obfuscation. The protocol presented here also allows the verifier's first message to be reused for an arbitrary number of proofs and for public verification of protocol transcripts, unlike protocols in other works.

We observe that under slightly stronger conditions, we can make the verifier use public coins. Suppose that f is in fact a one-way permutation; then the verifier can sample b directly from the image. This will yield the same distribution of first messages. Since r is not needed for verification, the protocol can still be executed. Witness hiding and soundness follow from the exact same analysis as before. Thus we get a public coin two-message witness hiding protocol under general and plausible assumptions.

We next modify the proof to obtain a heuristic non-interactive protocol. We simply have the public coin verifier's first message be deterministically generated from the security parameter, say setting $b = H(1^\lambda)$ for some hash function H. Witness hiding still follows immediately from the analysis above. By fixing the first message, we also eliminate the delayed-input limitation of the two-round

[2] An appropriate two-message witness hiding proof or "zap" suffices for their work; we stick with the NIWI for simplicity of explication.

[3] This proof actually yields a stronger property called "quasipolynomial simulatability". But we are only interested in witness hiding here.

protocol. Further, computational soundness can be easily justified in the random oracle model for H.

A random oracle model proof of soundness requires some discussion, as it is well known that non-interactive *zero knowledge* exists in the random oracle model. However, we note that such a zero knowledge system inherently requires the simulator to program random oracle outputs. In particular, zero knowledge cannot hold in the standard model without assuming additional resources like a common reference string. As the simulator is needed to prove witness hiding, this means that proving witness hiding of such a protocol requires the full power of programming the random oracle. In contrast, in our scheme witness hiding holds in the standard model without any reliance on the random oracle. Instead, only soundness requires the random oracle. Moreover, soundness requires only for r to be unpredictable. But this follows simply from the hardness of inverting f and the fact that the random oracle output is truly random. Thus, we obtain soundness in a very mild version of the random oracle model. We note that while Lindell [31] constructed NIZKs where zero knowledge similarly does not require the random oracle, the construction also (inherently) requires a common reference string to achieve zero knowledge hence witness hiding. Our non-interactive protocol does not require a common reference string.

Beyond idealized models, the computational soundness of this scheme poses a significant barrier to removing interaction. Any concrete choice of a hash function that outputs true instances yields a non-uniform adversary against soundness. By taking a witness to $y = H(1^\lambda)$ as advice, the adversary can generate a proof of $x \vee y$ for any x. [4] The same barrier applies to the derandomization techniques that remove interaction from ZAPs; for instance, following [6] and taking y as the output of a hitting set generator would require statistical soundness.

Lacking an explicit means to choose a value of y a priori, we turn to the non-uniform setting. We keep the basic scheme the same, but simply let $y \in \textsc{unsat}$ be an advice string for both prover and verifier, guaranteeing soundness. It remains to prove witness hiding.

On non-uniform witness hiding (section 4). We move to the non-uniform setting, where both parties have access to a non-uniform advice string. We will set y to be this advice string. We will also allow adversaries to be non-uniform. We now ask: is there some y such that the protocol above is witness hiding? Suppose to the contrary that the protocol is *not* witness hiding for *any* false y. Then we observe that an adversarial verifier V^* that takes a proof π and extracts a witness for x *itself* serves as a witness to the fact that y is false. Indeed, if y were true, then no such V^* could exist by analogous arguments to above.

So if the protocol fails to be witness hiding for every false y, then we have witnesses for a coNP-complete language. This *suggests* that failure to be witness

[4] The same barrier does not apply to soundness against uniform adversaries. In fact, a closely related construction (described in [3], but analysed in a different setting) can be used to achieve witness hiding proofs sound against uniform adversaries assuming *keyless* collision-resistant hash functions.

hiding for any y implies $\mathsf{coNP} \subseteq \mathsf{NP}$, a widely unexpected outcome. Unfortunately, the verifier sketched above fails to be an NP verifier in three ways:

- It is probabilistic, running the randomized adversary on random inputs.
- It may not succeed for all input sizes, as breaking security only requires successful adversaries for infinitely many input sizes.
- It requires super-polynomial time and witness size, as adversaries can have arbitrary polynomial size and run in arbitrary polynomial time.

Nonetheless, we can strengthen our assumptions to subsume these differences. Define the complexity class $\mathsf{ioMA}(t)$, the analog of MA where the verifier is now allowed to run in time $t(n)$ and is only correct for infinitely many n. We formally describe the verifier sketched above and conclude that either $\mathsf{coNP} \subseteq \mathsf{ioMA}(t)$ for any super-polynomial t — a surprising complexity result— or else for every n there exists *some* false y such that the protocol above is witness hiding.

Unfortunately, we cannot use this protocol in a uniform setting as the y needed to achieve witness hiding may be hard to compute. Furthermore, the choice of y is not universal; it depends on the underlying distribution D from which the statements are drawn. Thus the construction is not a single witness hiding proof system for NP, but rather a family of proof systems, one for each hard distribution. Non-uniform protocols should be viewed as *existential* results: unlike common reference string protocols, the non-uniform model does not require the joint input to be sampleable.

Nevertheless, this result at least suggests a fundamental difficulty of *ruling out* non-interactive witness hiding protocols. Indeed, ruling out such protocols in the non-uniform setting would yield a surprising complexity implication, coming close to showing that the polynomial hierarchy collapses. Given that non-interactive witness hiding cannot be ruled out, we believe our result is also strongly suggestive that it should be possible to actually find a non-interactive witness hiding proof system, under plausible computational assumptions. Finding an explicit procedure for generating appropriate y clearly would suffice to make this scheme uniform; however, it is unclear how to do so.

Best-possible proofs (section 5). As discussed above, our non-uniform construction offers compelling evidence for the existence of non-interactive witness hiding proofs, but gives little indication of how to go about constructing them. Here, we partially close this gap, showing that an inexplicit construction satisfying the right properties is sufficient to build an *explicit* witness hiding protocol.

More concretely, we seek a universal non-interactive proof system, which guarantees witness hiding as long as *some* witness hiding protocol exists. Our inspiration will be the notion of best-possible obfuscation, by Goldwasser and Rothblum [24]. There, they showed that the indistinguishability notion of obfuscation is actually as good of an obfuscator as any other notion of obfuscation, subject to certain minutia regarding program size.

Consider the following first attempt. On input a statement x, a proof will be a NIWI proof of the statement

$$x' := \exists V, \pi' : V(x, \pi')$$

Here, V is a verifier for an arbitrary sound proof system and π' is a proof of x relative to V. The intuition behind witness hiding is that if a witness hiding non-interactive proof system (P, V) exists, then V together with $\pi' = P(x, w)$ is a witness for x'. Such a witness would of course be witness hiding by assumption and can be used to generate the NIWI proof of x'. However, we do not actually need to know (P, V) in order to generate the proof: we can use *any* sound proof system to generate the NIWI proof of x'. For example, take (I, V_L) to be the standard NP proof system: I simply outputs the witness, and V_L checks the NP relation. Of course, this proof system does not hide the witness, but once we use it to generate the NIWI proof of x', witness indistinguishability kicks in and implies that the resulting proof is "as good" as if it had been generated using (P, V). Thus, we obtain witness hiding regardless of the starting proof system.

While the above does indeed demonstrate witness hiding, the protocol is not sound. The problem is that the statement x' does not actually guarantee that V is the verifier of a sound proof system (recall that although soundness is often described as a property of a proof system, it is actually a property of the verifier alone). A cheating prover could simply pick V to accept all inputs; then the proof verifies for any choice of x'.

We need to augment the proof system to check that V is sound. For an arbitrary Turing machine V, there is no way to actually this: the problem is undecidable. Even restricting to circuits, making this determination efficiently would imply a collapse of the polynomial hierarchy. Instead, we require that V is accompanied by a proof attesting to its soundness. In more detail, consider a sound logical system $\mathcal{S}$ that is powerful enough to reason about programs and soundness. A witness (z, V, π) for x' then consists of a witness z for x under V, the code of the verifier V, and an $\mathcal{S}$-proof π that V is sound. In order for this to be an NP relation, we need a polynomial bound on the length of the witness (z, V, π). In particular, we need a bound s on the runtime of V.

Our resulting proof system is sound, assuming the soundness of $\mathcal{S}$. It will also be witness hiding, as long as *some* witness hiding proof system exists whose verifier runs in time at most s and whose soundness can be proving using $\mathcal{S}$. The witness hiding proof system can even have inefficient provers, and our proof system will inherit the witness hiding security and still be efficient. Thus, to demonstrate witness hiding of our protocol, one only has to reason about the *existence* of witness hiding proofs.

The discussion above extends to the case where V is a circuit instead of a Turing machine; this allows us to base witness hiding off of the existence of a non-uniform scheme. But in the non-uniform case, the proof of soundness may be different for each input length. Thus the need for short proofs of soundness becomes a significant obstacle. However, the best-possible proof system remains uniform, even if a non-uniform scheme is used to prove security.

Given this extension to non-uniform schemes, one may hope to combine our best-possible proof system with our non-uniform proof system from the previous section, thereby obtaining a concrete witness hiding proof. Unfortunately, this appears challenging. In order to use our best-possible proof system, we require a proof of soundness in $\mathcal{S}$. But such a proof would demonstrate that the advice string y in the non-uniform proof system is a false statement. Such a proof would be at odds with our justification for the soundness of the protocol. Recall that our soundness proof assumes that a carefully constructed MA-type proof system P_{nu} rejects y. But a proof of soundness implies y has a short proof of satisfiability in $\mathcal{S}$, which in turn defines an NP proof system $P_{\mathcal{S}}$. Thus to find a choice of advice that suffices for the non-uniform protocol and demonstrates provable soundness, we would need to demonstrate a sequence of y that are rejected by P_{nu} but accepted by $P_{\mathcal{S}}$. It is unclear if such instances exist.

Non-interactive witness hiding vs. witness encryption (section 6). To alleviate this difficulty, our next idea is to explicitly choose y with short proofs of unsatisfiability. Concretely, choose y from a distribution over non-instances of some language T in $\mathsf{NP} \cap \mathsf{coNP}$. Making this change means we can no longer rely on the assumption $\mathsf{coNP} \subsetneq \mathsf{MA}$ to prove completeness; but an interesting connection to witness encryption will yield another route to proving security.

We will have the prover sample y from some distribution E over *false* instances and prove the statement $x' = x \vee y$ as before. Since a malicious prover could have chosen a true statement y, this protocol so far is not sound. However, we augment the proof π by also including the witness z for the falseness of y.

Now certainly the protocol is sound, since z means that x' is equivalent to x. But why might this protocol be witness hiding? After all, by including z in the proof, we seem to have again broken any arguments that work by switching y to be true. It appears we are back at square one.

First, we limit ourselves to distributions D over x with *unique* witnesses. [5] Now suppose we actually did *not* include z. Then we can easily prove witness hiding by switching y to be a true instance and using the witness for y to generate the proof. This implies that given π alone it is hard to find the witness for x.

Now suppose that the overall proof is not witness hiding. This means given π alone, the witness w for x is hidden, but given *both* π and z it is possible to recover w. If so, we can turn the protocol plus witness hiding adversary into a type of *witness encryption scheme* for statements $\neg y$. Recall that witness encryption [19] allows for encrypting messages to NP statements; any witness for the statement can recover the message, but messages encrypted to false statements are computationally hidden.

Consider the following first attempt at a witness "key encapsulation" scheme: to encrypt to an instance $\neg y$, sample a random (x, w) from D. Then construct the proof π that $x \vee y$ using the witness w for x. The ciphertext is π and the encapsulated key is w. If $\neg y$ is false (meaning y is true), then we know that π computationally hides w by the NIWI. Thus we get witness encryption security. On the other hand, if $\neg y$ is true and one knows a witness z for $\neg y$, one can run

[5] Recall that for languages with multiple *independent* witnesses, a NIWI proof is already witness hiding, so the unique witness setting covers the other end of the spectrum; we will leave it as an interesting open problem extending our results below to "in between" languages with multiple *dependent* witnesses.

the witness hiding adversary to recover w. Uniqueness of w guarantees that the recovered w is the actual encapsulated key. Of course, w is not pseudorandom as one can verify that w is a valid witness. Instead, we will extract a Goldreich-Levin hardcore bit from w; this hardcore bit will then be used to mask the message bit.

Now, the above scheme fails to satisfy the definition of witness encryption for two reasons:

- The witness hiding adversary might work with only non-negligible probability. This yields a decryption algorithm that succeeds with only non-negligible probability.
- Correctness is only guaranteed with respect to witnesses sampled according to E, not truly arbitrary witnesses.

Nevertheless, such a witness encryption scheme has interesting consequences. If, for distribution D, there is no language T and distribution E that make our protocol witness hiding, then we get such a witness encryption scheme for every language and distribution over instances in $\mathsf{NP} \cap \mathsf{coNP}$. This would be enough to build public key encryption, assuming any hard-on-average problem in $\mathsf{NP} \cap \mathsf{coNP}$. By hard-on-average, we mean that there is a second distribution F over true instances (and valid witnesses) such that y sampled from E or F are computationally indistinguishable. This gives public key encryption from tools that are otherwise not known to imply public key encryption, namely NIWIs and any hard-on-average problem in $\mathsf{NP} \cap \mathsf{coNP}$. It is also enough to build identity-based encryption from any unique signature scheme, following [19].

Another interesting consequence of our connection to witness encryption is the following: general witness encryption is currently known only from very strong and new mathematical tools [19,2]. While many in the community believe witness encryption exists, the case is far from settled. We do not take a position either way, but consider the plausible scenario that there is *some* language in $\mathsf{NP} \cap \mathsf{coNP}$ and some distribution for which no witness encryption scheme exists (in the sense obtained above).[6]

Under this assumed *non*-existence of witness encryption, for appropriate choice of length parameters we immediately see that our best-possible proof system is a non-interactive witness hiding proof system. In fact, it is possible to use any choice of polynomial-length length parameters by re-scaling the security parameters appropriately. Thus we obtain a fully concrete scheme with provable security under plausible assumptions.

[6] It is worth noting that many problems in $\mathsf{NP} \cap \mathsf{coNP}$ do have witness encryption schemes based on their own hardness: for example, the quadratic residuosity problem gives rise to the Goldwasser-Micali pubic key encryption scheme [22], which can be adapted to a witness encryption scheme for the language of quadratic non-residues. However, hardness in $\mathsf{NP} \cap \mathsf{coNP}$ is not known to generically imply witness encryption or even public key encryption. For example, the presumed hard problems of deciding who wins in a stochastic game [11] or determining whether a given knot is the unknot [30] are both in $\mathsf{NP} \cap \mathsf{coNP}$, but neither is known to yield public key encryption.

1.3 Discussion

We observe that our protocols are superficially related to the "proofs of ignorance" approach of Kalai and Deshpande [13]. In their work, they prove x by proving $x \vee y$ and supplying a "proof of ignorance" that the prover does not know a witness for y. For example, the witness w for $\neg y$ in our second construction certainly demonstrates that the prover does not know a witness for y. On the other hand, turning "proofs of ignorance," as defined by [13], into witness hiding, used a very strong KDM security assumption, which was demonstrated to be false [18]. Our justifications for witness hiding proceed by entirely different arguments.

Haitner, Rosen and Shaltiel provide a black-box barrier to witness hiding in few rounds [27]. However, their barrier does not apply to our schemes. Their barrier only applies to specific (but common) approaches to witness hiding by parallel repetition; our schemes do not use parallel repetition. Further, our schemes are certainly non-black-box, using the adversary's code itself to either violate a complexity assumption or build another protocol. Some of our proofs are also not by reduction to the original search problem, again avoiding the barrier.

2 Preliminaries

2.1 Basic building blocks

Let $\mathsf{p.p.t.}$ be the set of probabilistic polynomial time Turing machines. Let negl be the set of negligible functions. We use the standard definition of NP.

Definition 1 (Conventions for NP languages). *Note the verifier characterizes the language: given a two-input machine V that runs in time polynomial in the length of the first input, put $L_V = \{x : \exists y\ V \text{ accepts } (x, w)\}$.*
Define the NP witness relation *for L as $R_L = \{(x, w) : V \text{ accepts } (x, w)\}$.*

Definition 2 (probability ensemble). *A probability ensemble D is a map from $\mathbb{N}$ to distributions over strings. All probability ensembles in this paper will be of polynomial length, meaning there exists a polynomial p such that for all λ and $x \in \mathrm{Sup}(D(\lambda))$ we have $|x| \le p(\lambda)$. They will also be poly-time sampleable. Let $\Delta(S)$ be the set of probability ensembles with support contained in S.*

Definition 3 (search hardness). *Fix $L \in NP$, $D \in \Delta(R_L)$. Say the* search problem over D is hard *when $\forall A \in$ p.p.t.*

$$\Pr_{(x,w)\sim D(\lambda)}[(x, A(x)) \in R_L] = \mathrm{negl}(\lambda).$$

Analogously, say the *search problem over D is hard against non-uniform adversaries* when the same condition holds $\forall A \in \mathsf{p.p.t.}/\,\mathrm{poly}$.

Definition 4 (hard-on-average). *Fix $L \in NP$, $D_0 \in \Delta(\overline{L})$, $D_1 \in \Delta(L)$. Say L is* hard-on-average *when D_0 and D_1 are computationally indistinguishable.*

2.2 Proof systems

The proof systems used in this paper differ in three aspects. First, they use different number of messages. Second, they have different soundness guarantees. Third, they have different guarantees on what information is revealed to the verifier. We do not define a full taxonomy of proof systems but rather only what we will use in the later parts of the paper.

Definition 5 (non-interactive argument system). *Fix an* NP *relation* R_L*. We say a pair of* $\mathsf{p.p.t.}$ *algorithms* P, V *is* a non-interactive argument system for L *when the following two properties hold:*

Completeness: $\forall (x, w) \in R_L$, $\Pr[V(x, P(x, w))] = 1$.
Soundness: $\forall \widetilde{P} \in \mathsf{p.p.t.}$, $\exists \mu \in \text{negl}$, $\forall x \notin L$, $\Pr\Big[V(x, \widetilde{P}(x))\Big] \leq \mu(|x|)$.

Definition 6 (non-interactive proof system). *We say a non-interactive argument system* P, V *is* a non-interactive proof system for L *when the following stronger soundness property holds:*

Soundness: $\exists \mu \in \text{negl}$, $\forall x \notin L$, $\forall \pi$, $\Pr[V(x, \pi)] \leq \mu(|x|)$.

In particular, we say a proof system has *perfect soundness* when the soundness property holds with $\mu = 0$. Analogously, a *non-uniform non-interactive proof system* is a pair of $\mathsf{p.p.t.}/$ poly algorithms for which the same properties hold.

Next we define the delayed input model in the two-message case. We define the steps of the verifier V by two algorithms V_0, V_1. First, V_0 runs on input $|x|$ and outputs the first message m and some internal state q; second, the prover P runs on input x, m and outputs the second message π; third, V_1 runs on input x, q, π and either accepts or rejects. We define completeness and soundness in this model.

Definition 7 (delayed-input two-message argument system). *Fix* $L \in \mathsf{NP}$*. A triple of* $\mathsf{p.p.t.}$ *algorithms* V_0, P, V_1 *is* a two-message argument system for L *when the following two properties hold:*

Completeness: $\forall (x, w) \in R_L$*:*

$$\Pr[V_1(x, q, P(x, w, m)) \mid (q, m) \leftarrow V_0(|x|)] = 1.$$

Soundness: $\forall \widetilde{P} \in \mathsf{p.p.t.}$, $\exists \mu \in \text{negl}$, $\forall x \notin L$

$$\Pr\Big[V_1(x, q, \widetilde{P}(x, m)) \mid (q, m) \leftarrow V_0(|x|)\Big]] \leq \mu(|x|).$$

Definition 8 (witness indistinguishable [10,6]). *Say the prover* P *is* witness indistinguishable *for* $L \in \mathsf{NP}$ *when for any sequence* $I = \{(x, w_1, w_2)\}$ *such that* $(x, w_i) \in R_L$*, the ensembles* Π_1, Π_2 *are computationally indistinguishable, where:*

$$\Pi_i = \{\pi \leftarrow P(x, w_i)\}_{(x, w_1, w_2) \in I}.$$

We assume NIWIs for arbitrary NP relations. Thus when R_L is clear from context, we say simply "a NIWI for x using witness w" to denote $P(x, w)$ and "verify that π is a valid proof of x" to denote $V(x, \pi)$. Implicit is an encoding scheme to write x as an instance of an NP-complete problem and w as the corresponding witness.

Definition 9 (witness hiding). *Fix $L \in \mathsf{NP}$, $D \in \Delta(R_L)$. Say the prover P is* witness hiding *for D when* $\forall A \in \mathsf{p.p.t.}/\operatorname{poly}$

$$\Pr_{(x,w)\sim D(\lambda)}[(x, A(x, P(x, w))) \in R_L] = \operatorname{negl}(\lambda).$$

Though this definition is used in prior work [28,13], it is weaker than the original definition of witness hiding given by Feige and Shamir [17]. Their definition requires an explicit witness extractor M that, by making black-box calls to the adversary A and the sampler for D, achieves $\Pr[(x, A(x, P(x, w))) \in R_L] - \Pr\big[(x, M^{A,D}(x)) \in R_L\big] \leq \operatorname{negl}(\lambda)$. The extractor definition entails explicit black-box security reductions that are not achieved for all of our constructions.

2.3 Fine-grained notions

For Section 4 we define *fine-grained* notions of the above. To make these modifications, we change our notions of "efficient adversaries" and "negligible functions" to concrete measures. Let $\mathrm{SIZE}(S)$ be the class of circuit families of size $S(\lambda)$.

Definition 10 ((S, ε)-hardness of search problem). *Fix $L \in \mathsf{NP}$, $D \in \Delta(R_L)$. Say the* search problem over D is (S, ε)-hard *when $\forall \lambda \in \mathbb{N}$, $\forall A \in$* $\mathrm{SIZE}(S)$

$$\Pr_{(x,w)\sim D(\lambda)}[(x, A(x)) \in R_L] \leq \varepsilon(\lambda).$$

In the standard definitions of proof systems, using the length of the input to quantify the security suffices. But in the finer-grained model, we choose an explicit security parameter λ and provide 1^λ as input to the prover.

Definition 11 ((S, ε)-witness indistinguishable). *Fix $L \in \mathsf{NP}$. We say a proof system (P, V) is* (S, ε)-witness indistinguishable *for L when $\forall \lambda, x, w_1, w_2$ such that $(x, w_i) \in R_L$ and* $\forall A \in \mathrm{SIZE}(S)$

$$\left| \Pr_{\pi \leftarrow P(x, w_1, 1^\lambda)}[A(x, \pi)] - \Pr_{\pi \leftarrow P(x, w_2, 1^\lambda)}[A(x, \pi)] \right| \leq \varepsilon(\lambda).$$

Definition 12 ((S, ε)-witness hiding). *Fix $L \in \mathsf{NP}$, $D \in \Delta(R_L)$. Say a prover P is* (S, ε)-witness hiding *for D when $\forall \lambda \in \mathbb{N}$, $A \in$* $\mathrm{SIZE}(S)$

$$\Pr_{(x,w)\sim D(\lambda)}[(x, A(x, P(x, w, 1^\lambda))) \in R_L] = \varepsilon(\lambda).$$

We also say a proof system is (S, ε)*-witness hiding for all λ large enough* when there exists λ_0 such that for all $\lambda > \lambda_0$ the condition holds.

3 Witness hiding arguments in one and two messages

We review the two-message proof system of Pass from [32]. The referenced work proves a property called quasipolynomial time simulatability. We repeat this analysis, showing how it implies witness hiding for subexponentially hard distributions. The proof system takes the form of a NIWI of $x \vee y$ for a clause in the form $y = \exists r \ : \ f(r) = b$ where f is a one-way function. A perfectly binding commitment to the witness is included for the proof of soundness. The choice of b is made by the verifier. The analysis is completed using complexity leveraging.

We also present a new analysis in the delayed input model. This allows us to replace complexity leveraging with non-uniform choices. This result is incomparable to the original.

The introduction discusses several further properties of the protocol. The protocol allows the verifier's first message to be reused for an arbitrary number of proofs, and allows for public verification of transcripts. Further, if the distribution of b is uniformly random, for example if f is chosen to be a permutation, then the scheme is public coin. Finally, the scheme is amenable to a heuristic implementation by a hash function: simply choose $b = H(1^\lambda)$. In the (non-programmable) random oracle model this is secure.

3.1 Prerequisites

We require surjective one-way functions and commitment schemes with guarantees amenable to complexity leveraging: this entails security against one class of adversaries, but also requires that they can be inverted in some larger runtime.

Definition 13 (one-way function). *Say a one-way function f is* secure against adversaries running in time T_0 *if* $\forall A \in \mathrm{SIZE}(T)$, $\Pr_{x \sim \mathcal{U}}[f(A(f(x))) = f(x)]$. *We say f is* invertible in time T_1 *if there exists a Turing machine B running in time T_1 such that* $\forall x$, $f(B(f(x))) = f(x)$.

Definition 14 (perfectly binding commitment). *Say a commitment scheme* Comm *is* perfectly binding *if* $\forall x_1 \neq x_2, \forall r_1, \forall r_2$, $\mathrm{Comm}(x_1, r_1) \neq \mathrm{Comm}(x_2, r_2)$. *Say* Comm *is* hiding *if commitments to any pairs of sequences of messages, one of each length, are computationally indistinguishable by non-uniform adversaries. We say* Comm *is* extractable in time T_1 *if there exists a Turing machine B running in time T_1 such that* $\forall x, \forall r$, $B(\mathrm{Comm}(x, r)) = x$.

3.2 Construction from Pass 2003

Fix $L \in \mathsf{NP}$ and $D \in \Delta(R_L)$. Consider the following scheme for a two-message witness hiding argument in the delayed input model. Let $f : \{0,1\}^k \to \{0,1\}^\ell$ a surjective one-way function and Comm a perfectly binding commitment scheme.

$\mathsf{TwoMessage.V}_0(|x|)$: sample $r \sim \{0,1\}^k$. Put $b = f(r)$. Save and output b.

$\mathsf{TwoMessage.P}(x, w, b)$: put $c = \mathrm{Comm}((0^\ell, w))$. Compute π a NIWI for

$$S_{b,c} := \exists r', w' \ : \ c = \mathrm{Comm}((r', w')) \wedge (b = f(r') \vee (x, w') \in R_L)$$

using the witness $(0^\ell, w)$. Output $\tau = (c, \pi)$.

$\mathsf{TwoMessage.V}_1(x, b, (c, \pi))$: check that π is a valid proof of $S_{b,c}$.

3.3 Security from complexity leveraging

We sketch the proof of [32] for completeness.

Theorem 5. *Assume the search problem over D is hard against adversaries running in time T_{search}. Assume f is one-way against adversaries running in time T_{owf} but invertible in time T_{invert}. Assume* Comm *is hiding against non-uniform polynomial time adversaries and extractable in time T_{extract}. Assume a perfectly sound NIWI. If the following inequalities hold for all polynomials p:*

$$\begin{aligned} T_{\mathrm{search}} &= O(T_{\mathrm{invert}} + p(n)), \\ T_{\mathrm{owf}} &= O(T_{\mathrm{extract}} + p(n)) \end{aligned}$$

then $\mathsf{TwoMessage}$ is a perfectly complete witness hiding argument system against adversaries running in time $o(T_{\mathrm{search}})$.

In particular, if D, f, and the extraction algorithm for Comm are quasipolynomially hard in their input lengths and security parameters, then input lengths and security parameters can be set to satisfy the above inequalities as in the original paper.

Proof. Completeness follows from the completeness of the underlying NIWI. Witness hiding follows by considering π generated using an alternative witness. Soundness is by reduction to the one-way function game.

Completeness: if $(x, w) \in R_L$ then $(0^n, w)$ is a valid witness for $S_{b,c}$. Thus by the completeness of the NIWI system conclude that the verifier accepts.

Witness hiding: let A be an attacker on witness hiding. We build B simulating the prover in order to break the search problem against D as follows. Sample $x \sim D$ and send it to A, which will reply with some value b. Invert the OWF to find r such that $f(r) = b$. Then compute π' a NIWI of $S_{b,c}$ using witness $(r, 0)$. Let $c = \mathrm{Comm}((r, 0))$. Send (c, b) to A. Interpret the output of A as a possible witness for x.

Argue that $\mathrm{Adv}_B = \mathrm{Adv}_A \pm \mathrm{negl}$ from the witness indistinguishability of the NIWI and the hiding property of the commitment. Standard witness indistinguishability and hiding against non-uniform adversaries suffice (though we omit the proof).

Soundness: fix $x \notin L$ and a possibly malicious prover A outputting (c, π) given honest b. We construct a one-way function adversary B as follows. First, extract the commitment c which yields (r', w') by binding. By the soundness of the NIWI system we know that (r', w') is a witness for $S_{b,c}$. But since $x \notin L$ conclude that $(x, w') \notin R_L$. Thus it must be the case that $f(r') = b$. Output r'.

3.4 Security from non-uniform hardness

The witness hiding proof above can be adjusted to avoid the use of complexity leveraging if we assume the D search problem is hard against non-uniform adversaries and move to the delayed input model.

Theorem 6. *Assume the search problem over D is hard against poly-size circuits. Assume f is one-way against adversaries running in time T_{owf}. Assume* Comm *is hiding against non-uniform polynomial time adversaries and extractable in time T_{extract}. Assume a perfectly sound NIWI. If the following inequality holds for all polynomials p:*

$$T_{\text{owf}} = O(T_{\text{extract}} + p(n))$$

then $\mathsf{TwoMessage}$ is a perfectly complete witness hiding argument system against efficient poly-size circuits.

Proof. Completeness and soundness follow identically. For witness hiding, we build a non-uniform adversary, hard-coding the OWF pre-image.

Witness hiding: let A be an attacker on witness hiding; it plays the role of a malicious verifier and outputs w' a witness for $x \in L$ with probability Adv_A. Recall from the definition that A consists of two algorithms: first, A_0 which takes input $|x|$ and outputs b and some internal state q; second, the honest prover P runs on input x, b yielding c, π; third, A_1 which takes (b, q), (c, π) and x and outputs w.

Let r_0 be the explicit choice of randomness by A_0. Then we can break the experiment that defines Adv_A into two parts:

$$\mathop{\mathrm{E}}_{r_0}[\Pr[(x, A_1(x, q, \tau)) \in R_L \mid (b, q) = A_0(|x|, r_0)]] = \text{Adv}_A,$$

where $\tau = \mathsf{TwoMessage.P}(x, w, b)$ and the inner probability is over choice of $x \sim D$, $z' \sim \{0,1\}^n$, internal randomness of the NIWI prover, and internal randomness of A_1. Now for each input size we can fix some (b, q) such that

$$\Pr[(x, A_1(x, q, \tau)) \in R_L] \geq \text{Adv}_A\,.$$

We define a non-uniform adversary B against the search problem over D. For inputs size λ, the advice is a tuple (b, q, r) with b, q as chosen above and r such that $f(b) = r$. On input x, set $c = \text{Comm}((r, 0))$ and compute a NIWI of $S_{b,c}$ using witness $(r, 0)$. Run A on the new proof. The conclusion that $\text{Adv}_B = \text{Adv}_A \pm \text{negl}$ follows as above.

4 Non-uniform witness hiding

We present a non-interactive *non-uniform* witness hiding proof system. By writing a proof of x as a NIWI of $x \vee y$ for a false statement y fixed non-uniformly, we easily guarantee completeness and soundness. To achieve witness hiding, we give an MA-type verifier relative to which the code of an adversary is itself a witness to the falseness of y.

We begin by quantifying how this proof system differs from the standard complexity class MA. Then we state our construction formally and prove the desired security properties. The construction is unfortunately existential: it is unclear how to instantiate the scheme, even heuristically. However, it provides strong barriers to ruling out non-interactive witness hiding protocols.

4.1 Assumption

Throughout, UNSAT denotes the language of unsatisfiable boolean formulae. An arbitrary coNP-complete language can be used to yield the same result. Now recall the standard definition of MA. (e.g. adapted from Def. 8.10 of [1])

Definition 15 (MA). *We say L has an MA proof system when some $\mathsf{p.p.t.}$ Turing machine V has the following properties for some polynomial q:*

Completeness: $\forall x \in L, |x| = \lambda$, $\exists a \in \{0,1\}^{q(\lambda)}$, $\Pr[V(x,a)] \geq 2/3$,
Soundness: $\forall x \notin L$, $\forall a$, $\Pr[V(x,a)] \leq 1/3$.

From this definition, we get the standard complexity assumption $\mathsf{MA} \not\subseteq \mathsf{coNP}$. Unfortunately, this assumption does not appear sufficient for our NIWH system.

We need to make two changes. First, we allow the verifier to run in some super-polynomial time $T(\lambda)$ and use witnesses of size $R(\lambda)$. Second, while standard assumptions only require that the verifier fail for some input length, we want a proof system that works for all input lengths. Thus we require that any proof system fails for all inputs large enough. Both changes are captured by the following definition.

Definition 16 ($\mathsf{ioMA}(T, R)$). *Take any T, R. Say L has an $\mathsf{ioMA}(T,R)$ proof system when some V running in time $T(\lambda)$ has both of the following properties for infinitely many values of λ:*

Completeness: $\forall x \in L, |x| = \lambda$, $\exists a \in \{0,1\}^{R(\lambda)}$, $\Pr[V(x,a)] \geq 2/3$,
Soundness: $\forall x \notin L, |x| = \lambda$, $\forall a$ $\Pr[V(x,a)] \leq 1/3$.

The complexity assumption $\mathsf{coNP} \not\subseteq \mathsf{ioMA}(T,R)$ is simply a quantitative strengthening of $\mathsf{coNP} \not\subseteq \mathsf{MA}$; we believe it to be justifiable under the same motivation.

4.2 Construction

Fix $L \in \mathsf{NP}$ and $D \in \Delta(R_L)$. We propose the following scheme for a non-interactive witness-hiding proof. The scheme is parameterized by a sequence of circuits $(y_\lambda)_{\lambda \in \mathbb{N}}$ with each $y_\lambda \in \textsc{unsat}$ and $|y_\lambda| = \lambda$. The y_λ serve as advice for the prover and verifier.

$\mathsf{NonUniform.Prove}(x, w, 1^\lambda; y_\lambda)$: output a NIWI for $x \in L \lor y_\lambda \in \textsc{sat}$ using witness w and security parameter λ.

$\mathsf{NonUniform.Verify}(x, \pi; y_\lambda)$: verify π is a valid proof of $x \in L \lor y_\lambda \in \textsc{sat}$.

4.3 Security

Completeness and soundness follow directly from the completeness and soundness of the underlying NIWI, for any choice of $y_\lambda \in \textsc{unsat}$. To prove witness-hiding we use the complexity assumption. The proof and the security of the resulting NIWH proof system are parameterized by the strength of the complexity assumption and the security of the NIWI.

Theorem 7. *Fix a constant $\alpha > 0$. Assume the search problem over D is (S, ε)-hard. Assume a perfectly sound $(S_{\mathrm{NIWI}}, \varepsilon_{\mathrm{NIWI}})$-NIWI. Assume $\mathsf{coNP} \not\subseteq \mathsf{ioMA}(T, R)$. Assume the following inequalities between parameters hold, for some fixed $q(\lambda) = \mathrm{poly}(\lambda)$ and constant β chosen in the proof:*

$$\begin{aligned} &S(\lambda) \geq S_{\mathrm{NIWI}}(\lambda) + q(\lambda), \\ &T(\lambda) \geq \beta((\varepsilon + \varepsilon_{\mathrm{NIWI}})^{-1}\alpha^{-2})(S_{\mathrm{NIWI}}(\lambda) + \mathrm{poly}(\lambda)), \\ &R(\lambda) \geq S_{\mathrm{NIWI}}(\lambda). \end{aligned}$$

Then there exists a sequence of y_λ (depending on D) such that $\mathsf{NonUniform}$ is a perfectly sound proof system with $(S_{\mathrm{NIWI}}, (1 + \alpha)(\varepsilon + \varepsilon_{\mathrm{NIWI}}))$-witness-hiding for all λ large enough.

In particular: take S_{NIWI} slightly super-polynomial in λ and $\varepsilon_{\mathrm{NIWI}} = \mathrm{negl}(\lambda)$. The required T, R, S will be fixed super-polynomial functions as given by the inequalities in the theorem statement. Then under the appropriate assumptions the theorem yields a standard NIWH system; that is, with witness hiding $\mathrm{negl}(\lambda)$ against all $\mathrm{poly}(\lambda)$ adversaries.

Proof. We prove completeness, soundness, and witness hiding.

Completeness: if $(x, w) \in R_L$ then w is a witness for $x \in L \lor y_\lambda \in \textsc{sat}$. By completeness of the NIWI system conclude that $\mathsf{NonUniform.Verify}$ accepts (x, π).

Soundness: consider $x \notin L$. Since $y_\lambda \in \textsc{unsat}$, we know the statement $x \in L \lor y_\lambda \in \textsc{sat}$ is false. Thus by the soundness of the NIWI system we conclude $\mathsf{NonUniform.Verify}$ does not accept (x, π) for any value of π.

Witness-hiding: we prove witness-hiding by constructing an $\mathsf{ioMA}(T, R)$-type protocol for $\textsc{unsat}$. We show the protocol is unconditionally sound and efficient. We show the protocol is complete if and only if there is no choice of $(y_\lambda)_{\lambda\in\mathbb{N}}$ such that $\mathsf{NonUniform}$ is witness-hiding. The verifier is parameterized by a choice of α (in the theorem statement) and k (chosen below).

$\mathsf{UnsatVerifier}(t, A)$: Interpret A as a circuit. If $|A| \geq S_{\mathrm{NIWI}}$ reject. Sample k tuples $(x_i, w_i) \sim D(|t|)$ and compute the sample probability

$$p = \frac{1}{k}\sum_{i\in[k]} \mathbb{1}[(x_i, A(x_i, \mathsf{NonUniform.Prove}(x_i, w_i; t)) \in R_L].$$

Accept if and only $p > (1 + \alpha/2)(\varepsilon + \varepsilon_{\mathrm{NIWI}})$.

Soundness of $\mathsf{UnsatVerifier}$: fix $(t, z) \in R_{\textsc{sat}}$. Let π be a NIWI proof of $x \in L \vee t \in \textsc{sat}$ using witness z. If $|A| < S_{\mathrm{NIWI}}$ then by witness indistinguishability we know

$$\Pr[(x_i, A(x_i, \mathsf{NonUniform.Prove}(x_i, w_i; t))) \in R_L] \leq \Pr[(x_i, A(x_i, \pi)) \in R_L] + \varepsilon_{\mathrm{NIWI}}.$$

Now note that $(x_i, A(x_i, \pi)) \in R_L$ is computed by a circuit of size at most $|A|$, plus the size of the circuit that computes the NIWI, plus the size of the verifier circuit for R_L. Setting q accordingly, then it is bounded in particular by $S \geq S_{\mathrm{NIWI}} + q(\lambda)$. Thus by the hardness of the D-search problem

$$\Pr[(x_i, A(x_i, \pi)) \in R_L] < \varepsilon.$$

Together the two inequalities yield

$$\Pr[(x_i, A(x_i, \mathsf{NonUniform.Prove}(x_i, w_i; t))) \in R_L] \leq \varepsilon + \varepsilon_{\mathrm{NIWI}}.$$

This shows that $\mathrm{E}[p] \leq \varepsilon + \varepsilon_{\mathrm{NIWI}}$. To bound the tail probability use a standard Chernoff bound (e.g. Cor. A.15 in [1]).

$$\Pr\Big[|p - \varepsilon + \varepsilon_{\mathrm{NIWI}}| \geq \frac{\alpha}{2}(\varepsilon + \varepsilon_{\mathrm{NIWI}})\Big] \leq 2\exp\big(-\alpha^2 \Omega(k(\varepsilon + \varepsilon_{\mathrm{NIWI}}))\big).$$

Thus choosing $k = \beta((\varepsilon + \varepsilon_{\mathrm{NIWI}})^{-1}\alpha^{-2})$ for some constant β suffices for the verifier to reject with probability 2/3.

Runtime of $\mathsf{UnsatVerifier}$: from our choice of k and the size of A observe the verifier runs in time

$$k(|A| + \mathrm{poly}(\lambda)) = \beta((\varepsilon + \varepsilon_{\mathrm{NIWI}})^{-1}\alpha^{-2})(S_{\mathrm{NIWI}} + q).$$

Completeness of $\mathsf{UnsatVerifier}$: fix $r \in \textsc{unsat}$, $|r| = \lambda$. Assume $\mathsf{NonUniform}$ with advice r is not sufficiently witness hiding. Then there exists A with $|A| < S_{\mathrm{NIWI}}$ such that

$$\Pr[(x_i, A(x_i, \mathsf{NonUniform.Prove}(x_i, w_i; r))) \in R_L] \geq (1 + \alpha)(\varepsilon + \varepsilon_{\mathrm{NIWI}}).$$

Applying the same Chernoff bound shows that the verifier will accept with overwhelming probability.

Conclude by using the assumption: since we know for all $\lambda > \lambda_0$ that $\mathsf{UnsatVerifier}$ cannot be complete, there must be some $y \in \textsc{unsat}$, $|y| = \lambda$ such that $\mathsf{NonUniform}$ with advice y_λ is witness hiding.

5 Best-possible proofs

We present a construction for a non-interactive witness hiding proof system that is secure as long as any such scheme is secure. In fact, assuming even the existence a proof system with an inefficient prover, the scheme given in this section will be efficient and uniform as long as the original scheme is *provably sound* in a sense made precise later. We discuss how a non-uniform notion of provable soundness can be used to base the same construction off of the existence of a non-uniform witness hiding proof system. We do not know if the scheme given in the previous section meets this requirement.

This construction enjoys security properties beyond witness hiding. In fact, assuming the existence of a provably sound non-interactive proof that achieves any falsifiable security notion, this construction will have the property as well as long as length parameters are picked appropriately. Thus the construction is in fact the "best possible" non-interactive proof that can be achieved, in the same sense as [24].

Unfortunately, this paper does not provide a NIWH for all NP that has provable soundness. Thus, we are left in an odd state of affairs where we know a universal construction but lack the existential proof needed to claim it is secure.

5.1 Prerequisites

Let $\mathcal{S}$ be a proof system for a language powerful enough to encode Turing machines. Assume that $\mathcal{S}$-proofs can be checked in time polynomial in their length. Further assume that $\mathcal{S}$ is sound, meaning that any provable statement in $\mathcal{S}$ is true in the metatheory (or, for the purposes of this work, simply true). A concrete choice of $\mathcal{S}$ would be Peano arithmetic or any standard deductive system for axiomatic set theory.

Fix some verifier V corresponding to the language $L_V \in \mathsf{NP}$. Let D be another polynomial-time verifier. We want a proof that $L_V = L_D$ inside of $\mathcal{S}$. This leads to the following definition.

Definition 17 (soundness for L_V). *We say D is L_V-sound if the following statement holds*

$$\forall x \in \{0,1\}^* \; (\exists y \; D(x,y) \Rightarrow \exists w \; V(x,w)).$$

We require that such a statement can be encoded in the language of $\mathcal{S}$. We also require that the there be a proof that V is L_V-sound. This is a relatively mild assumption, achievable by both concrete choices of proof systems proposed.

5.2 Construction

We begin by constructing V', another verifier for L_V. Fix polynomials q, s, ℓ.

> **$V'(x, w')$:** Interpret $(z, D, \pi) \leftarrow w'$ where
> z is a string of length $q(|x|)$,
> D is a Turing machine description of length $s(|x|)$,
> π is an $\mathcal{S}$-proof of length $\ell(|x|)$.
> Verify that π is a valid proof that D is L_V-sound. If not, reject. Otherwise, simulate D on input (x, z) for s steps and output the result.

Theorem 8. *V' is an NP verifier for L_V for sufficiently large choices of q, s, ℓ.*

Proof. Three things to show:

Polynomial time: the runtime is, as desired,

$$\mathrm{poly}(q(|x|), s(|x|), \ell(|x|)) = \mathrm{poly}(|x|).$$

Completeness: for any $x \in L_V$ let w_x be the shortest witness such that $V(x, w)$ accepts. Since V is an NP verifier, we know that $\max_{x\,:\,|x|=n} |w_x|$ is bounded by some polynomial; choose q larger. Further, the size of V as a Turing machine description is some constant; choose s larger. The runtime of V is bounded by some polynomial; choose s larger. The size of the proof π_V that V is L_V-sound is some constant as well; choose ℓ larger.

Then take any $x \in L$. By construction $V'(x, (w_x, V, \pi_V))$ will accept.

Soundness: assume $V'(x, (z, D, \pi))$ accepts. By the soundness of $\mathcal{S}$, we know that D must be L_V-sound. Since D accepts (x, z), we know by the definition of L_V-soundness that there exists w such that V accepts (x, w). By the soundness of V as an NP verifier, conclude that $x \in L$.

Our final proof will be a NIWI corresponding to V', constructed using the witness given in the completeness proof above.

> $\mathsf{BestPossible.Prove}(x, w)$: output a NIWI that $\exists w'$ such that $V'(x, w')$ accepts using witness (w, V, π_V) with V and π_V as described above.
>
> $\mathsf{BestPossible.Verify}(x, \pi)$: check π is a valid NIWI of the desired statement.

5.3 Security

Theorem 9. *Let $D \in \Delta(R_{L_V})$. Assume the search problem over D is hard for non-uniform adversaries. Assume there exists a non-interactive proof system ($\mathsf{NIWH.Prove}$, $\mathsf{NIWH.Verify}$) with non-uniform witness hiding for D such that for every x the following holds for some polynomials q, s, ℓ and for all $|x|$ large enough:*

the length of $\mathsf{NIWH.Prove}(x, w)$ *is at most* $q(|x|)$,
the size of $\mathsf{NIWH.Verify}$ *written as a Turing machine at most* $s(|x|)$,
the length of an $\mathcal{S}$*-proof of soundness is at most* $\ell(|x|)$.

Assume the NIWI system used in the construction is perfectly sound. Then conclude $\mathsf{BestPossible}$ *with parameters* (q, s, ℓ) *is a perfectly complete and sound NIWH against non-uniform adversaries with an efficient prover.*

Proof. We prove completeness, soundness, and witness hiding.

Completeness: follows from completeness of the NIWI system and the analysis of the witness $(w, V_{|x|}, \pi_{V,|x|})$ from the previous theorem.

Soundness: since the NIWI system is perfectly sound, we know that if $\mathsf{BestPossible.Verify}$ accepts, then $\exists w'$ such that $V'(x, w')$. From the previous theorem, we know V' is an NP-verifier for L_V. Thus conclude $x \in L_V$.

Witness hiding: let A be an attacker (resp. non-uniform attacker) against the witness hiding of $\mathsf{BestPossible}$. We build B an adversary against the witness hiding of NIWH. First, for input size $|x|$, let D be the Turing machine computing $\mathsf{NIWH.Verify}$ and π the $\mathcal{S}$-proof that D is L_V-sound. Then build B as follows: on input (x, π), construct π' a NIWI of $\exists w'$ such that $V'(x, w')$ accepts using witness $(x, (\pi', D, \pi))$. Run A on π' and output the result.

By witness indistinguishability, π' as constructed by B is indistinguishable from $\pi = \mathsf{BestPossible.Prove}(x, w)$. Thus we have

$$\begin{aligned}
\mathrm{Adv}_B &= \Pr_{(x,w)\sim D(\lambda)}[(x, B(x, \pi))] \\
&= \Pr_{(x,w)\sim D(\lambda)}[(x, A(x, \pi'))] \\
&= \Pr_{(x,w)\sim D(\lambda)}[(x, A(x, \pi))] \pm \mathrm{negl}(\lambda) \\
&= \mathrm{Adv}_A \pm \mathrm{negl}(\lambda).
\end{aligned}$$

5.4 Additional properties

Note that non-uniformity is not required in the above proof as long as $\mathsf{NIWH.Verify}$ is uniform and there exists an $\mathcal{S}$-proof that it is sound for all input sizes. In this setting, we get an analogue to the above theorem where the resulting adversary is uniform.

Further note that we never run $\mathsf{NIWH.Prove}$. In fact, the whole proof goes through even if $\mathsf{NIWH.Prove}$ is inefficient. Regardless, $\mathsf{BestPossible}$ will be efficient.

Finally, note that our use of witness hiding was minimal. Observe that witness hiding can be replaced with any falsifiable notion of security. In the sense that this single construction (with appropriate parameters) achieves any desired notion shows that it is the “best possible” non-interactive proof.

5.5 Basing security on non-uniform proofs

In the above construction, consider replacing Turing machines with circuits of size at most s. We replace the soundness condition with the following:

Definition 18 (soundness for inputs of length n). *Say a circuit D is L_V-sound for inputs of length n if the following statement holds*

$$\forall x \in \{0,1\}^n \ (\exists y \ D(x,y) \Rightarrow \exists w \ V(x,w)).$$

Assuming some non-uniform NIWH proof system exists and for each n the $\mathcal{S}$-proof of L_V-soundness for inputs of length n is length at most $\ell(n)$, a slightly modified version of $\mathsf{BestPossible}$ is secure. However, this modification is unnecessary because a non-uniform scheme with this soundness condition actually implies NIWH scheme with an inefficient prover, and per the last section, this suffices for the theorem.

The construction is as follows: let P, V a non-uniform NIWH scheme with advice a_n. Let π_n an $\mathcal{S}$-proof of soundness for inputs of length n. Define V': on input $(x, (\pi', a, \pi))$, check that π is an $\mathcal{S}$-proof that $V(\cdot; a)$ is L_V-sound for inputs of length n. If not, reject. Otherwise run $V(x, \pi'; a)$ and output the result. The inefficient prover P' can simply use brute force to find an acceptable a and π and output $(P(x, w; a), a, \pi)$.

In the construction of Section 4 it is unclear if, for any choice of advice, that soundness for inputs of length n has short proofs. Recall the basic steps of Theorem 7: we showed the protocol was secure as long as the advice $y_\lambda \in \textsc{unsat}$. We constructed an MA-type verifier $\mathsf{UnsatVerifier}$ for the language $\textsc{unsat}$. We concluded that the scheme is witness hiding as long as $\mathsf{UnsatVerifier}$ does not accept y_λ. Thus to achieve probable soundness, we need to prove the existence of y_λ that fulfills the following three conditions:

(a). $y_\lambda \in \textsc{unsat}$,
(b). $\mathsf{UnsatVerifier}$ rejects y_λ,
(c). $\exists$ a poly-size $\mathcal{S}$-proof that $y_\lambda \in \textsc{unsat}$.

By the soundness of $\mathcal{S}$, we know that (c) implies (a). But it is still unclear how to achieve (b) and (c) simultaneously. In general, $\mathsf{NP} \neq \mathsf{coNP}$ establishes that proofs of unsatisfiability are long in the worst case. But this does not rule out short proofs for some appropriate statement.

6 Witness encryption vs. non-interactive witness hiding

Again, we present a non-interactive witness hiding proof system comprised of a NIWI of $x \vee y$; but in this scheme the prover picks y. To maintain soundness, the prover also provides an NP proof that y is false. To prove witness hiding, we restrict to the case where x has a unique witness. Then, an adversary against witness hiding is an algorithm that, from the proof and a witness to $\neg y$, recovers

w. By using w to encode a bit, the adversary serves as the decryptor for a weak *witness encryption* scheme.

We begin by defining witness encryption and a weakened notion of it. We proceed to give the construction and finally prove the desired properties. Recall that witness encryption is currently only known from extremely strong cryptographic tools, namely multilinear maps and obfuscation. Thus it is plausible that our weakened form of witness encryption does not exist for some language in $L \in \mathsf{NP} \cap \mathsf{coNP}$. But then this protocol would indeed be witness hiding. Further, we avoid choosing L concretely by using our best-possible protocol from the previous section.

6.1 Definitions

Consider the usual notion of witness encryption, as introduced by [19]. Fix a language $L \in \mathsf{NP}$ with verifier V.

Definition 19 (witness encryption). *We say* $(\mathsf{Encrypt}, \mathsf{Decrypt})$ *is an* witness encryption scheme for L *when the following two properties hold.*

Correctness: $\forall \lambda \in \mathbb{N}$, $\forall m \in \{0,1\}$, $\forall (x,w) \in R_L$

$$\Pr[\mathsf{Decrypt}(x, w, \mathsf{Encrypt}(x, m)) = m] = 1.$$

Soundness security: $\forall A \in \textsf{p.p.t.}$, $\exists \mu \in \text{negl}$, $\forall x \notin L$,

$$\Pr_{m \sim \{0,1\}}[A(\mathsf{Encrypt}(x, m)) = m] = \frac{1}{2} + \text{negl}(\lambda).$$

Consider a relaxed notion of correctness relative to some distribution T over (x, w) such that $V(x, w)$ is true.

Definition 20 (average case correctness for witness encryption).

Average case correctness: $\exists f \notin \text{negl}$, $\forall \lambda \in \mathbb{N}$, $\forall m \in \{0,1\}$,

$$\Pr_{(x,w) \sim T}[\mathsf{Decrypt}(x, w, \mathsf{Encrypt}(x, m)) = m] = f(\lambda).$$

This definition is weaker in two ways. First, it only guarantees any notion of correctness for infinitely many values of λ (as opposed to for all λ in the original definition). Second, decryption can fail. Since the failure probability is over the choice of instance, it may be the case that some instances always fail. Regardless, even with these limitations, we feel this is a strong cryptographic primitive.

6.2 Construction

Fix $L \in \mathsf{NP}$ and $D \in \Delta(R_L)$. Assume R_L restricted to D has unique witnesses: $\forall (x, w) \in \sup D$ if $(x, w') \in R_L$ accepts then $w' = w$. We propose the following scheme for a non-interactive witness hiding proof. Fix $T \in \mathsf{NP} \cap \mathsf{coNP}$, or equivalently R_T, $R_{\overline{T}}$ two NP relations and a probability ensemble E over $(y, z) \in R_T$.

$\mathsf{VsWE.Prove}(x, w)$: sample $(y, z) \sim E$. Compute π a NIWI for the statement $x \in L \vee y \notin T$ using the witness w. Output $\tau = (y, z, \pi)$.

$\mathsf{VsWE.Verify}(x, \tau)$: parse $(y, z, \pi) \leftarrow \tau$. Accept iff $V_T(y, z)$ accepts and π is a valid proof of the statement $x \in L \vee y \notin T$.

6.3 Security

Completeness and soundness of the scheme follow easily from the properties of the underlying NIWI. Then we argue that if the scheme is not witness hiding, then an adversary yields a witness encryption scheme in the weak sense above.

Theorem 10. *Assume the search problem over D is hard against non-uniform adversaries. Assume a perfectly sound NIWI. Then VsWE is a perfectly sound proof system. Further, any adversary that breaks the witness hiding of VsWE with non-negligible probability yields a witness encryption scheme for the language T with average case correctness with respect to E and the usual sense of soundness security for infinitely many lengths.*

Proof. Completeness and soundness follow from the corresponding properties of the NIWI. Witness hiding is justified by constructing a witness encryption decryptor from a witness hiding adversary.

Completeness: if $(x, w) \in R_L$, then w is a witness for $x \in L \vee y \notin T$. Thus by completeness of the NIWI system conclude that π is valid. Further by construction $V_T(y, z)$ accepts for all $(y, z) \sim E$. Conclude the verifier accepts.

Soundness: fix $x \notin L$. Consider any (y, z, π). Two cases: (1) if $y \in T$ then $x \in L \vee y \notin T$ is false. By NIWI soundness, π fails to verify; (2) if $y \notin T$ then $V_T(y, z)$ cannot accept. In either case, the verifier rejects.

Witness hiding: let A be an attacker that breaks witness hiding, meaning that $p = \Pr[V_L(x, A(x, y, z, \pi))]$ is non-negligible. Then we construct a one-bit witness encryption scheme WE_A for the language T as follows.

$\mathsf{WE}_A.\mathsf{Encrypt}(y, m)$: sample $(x, w) \sim D$. Let π a NIWI for the statement $x \in L \vee y \notin T$ using the witness w. Sample $r \sim \{0, 1\}^{|w|}$. Output $c = (x, \pi, r, \langle w, r \rangle \oplus m)$.

$\mathsf{WE}_A.\mathsf{Decrypt}(y, z, c)$: parse $(x, \pi, r, b) = c$. Let $w' = A(x, y, z, \pi)$. If $V_L(x, w)$ rejects then output $\perp$. Otherwise output $\langle w', r \rangle \oplus b$.

It remains to show that WE_A is correct and secure.

Average case correctness of WE_A: with probability p over choice of y and (x, w), we have w' a valid witness for $x \in L$. Then by the unique witness property we have $w' = w$. By construction this yields $\langle w', r\rangle \oplus b = m$.

Soundness security of WE_A: Fix a $\mathsf{p.p.t.}$ adversary B and sequence of inputs $\{y_\lambda\}_{\lambda\in\mathbb{N}}$, $y_\lambda \notin T$. Consider the following quantity:

$$\mathrm{Adv}_{B,Y}(\lambda) := |\Pr[B(\mathsf{WE}_A.\mathsf{Encrypt}(y_\lambda, 0))] - \Pr[B(\mathsf{WE}_A.\mathsf{Encrypt}(y_\lambda, 1))]|.$$

Let z_λ be a witness for $y_\lambda \notin T$. Such witnesses exist since $T \in \mathsf{coNP}$. Proceed by a series of games, parameterized by λ, for which the challengers are as follows.

$\mathcal{G}_0$**:** Sample $(x, w) \sim D(1^\lambda)$. Let π a NIWI for the statement $x \in L \vee y_\lambda \notin T$ using the witness w. Sample $r \sim \{0,1\}^{|w|}$. Output $c = (x, \pi, r, \langle w, r\rangle \oplus m)$ with $m = 0$.
$\mathcal{G}_1$**:** Same as $\mathcal{G}_0$ but derive π using witness z_λ.
$\mathcal{G}_2$**:** Same as $\mathcal{G}_1$ but output $b \sim \{0,1\}$ instead of $\langle w, r\rangle \oplus m$.
$\mathcal{G}_3$**:** Same as $\mathcal{G}_1$ but with $m = 1$.
$\mathcal{G}_4$**:** Same as $\mathcal{G}_0$ but with $m = 1$.

To show $\mathcal{G}_0 \approx \mathcal{G}_1$. This follows from the witness indistinguishability of the underlying NIWI proof system. Let A_{01} be a $\mathsf{p.p.t.}$ adversary that distinguishes between games 0 and 1. Then we have

$$\mathrm{Adv}_{A_{01}}(\lambda) = \left|\Pr_{c\leftarrow\mathcal{G}_0}[A_{01}(y_\lambda, c)] - \Pr_{c\leftarrow\mathcal{G}_1}[A_{01}(y_\lambda, c)]\right|.$$

Now for each λ choose (x_λ, w_λ) that achieve

$$\mathrm{Adv}_{A_{01}}(\lambda) \leq \left|\Pr_{c\leftarrow\mathcal{G}_0}[A_{01}(y_\lambda, c)|(x,w) = (x_\lambda, w_\lambda)] - \Pr_{c\leftarrow\mathcal{G}_1}[A_{01}(y_\lambda, c)|(x,w) = (x_\lambda, w_\lambda)]\right|.$$

This allows us to define a sequence of NIWI games as follows:

$$\mathcal{I} = \{(x_\lambda \in L \vee y_\lambda \notin T, w_\lambda, z_\lambda)\}_{\lambda\in\mathbb{N}}.$$

Then we give a non-uniform $\mathsf{p.p.t.}$ adversary A_{NIWI} against the NIWI game on sequence $\mathcal{I}$. A_{NIWI} takes $y_\lambda, x_\lambda, w_\lambda$ as an advice string. It receives a proof π as input. It samples $r \sim \{0,1\}^{|w|}$ and outputs $A_{01}(y_\lambda, (x_\lambda, \pi, r, \langle w_\lambda, r\rangle \oplus m))$ with $m = 0$.

Note that A_{NIWI} simulates $\mathcal{G}_0|(x,w) = (x_\lambda, w_\lambda)$ when the challenger uses witness w_λ and $\mathcal{G}_0|(x,w) = (x_\lambda, w_\lambda)$ when it uses witness z_λ. Thus $\mathrm{Adv}_{A_{\mathrm{NIWI}}}(\lambda) \geq \mathrm{Adv}_{A_{01}}(\lambda)$. Conclude by NIWI security that $\mathrm{Adv}_{A_{01}}$ is negligible.

To show $\mathcal{G}_1 \approx \mathcal{G}_2$. This follows from the fact that $\langle w, r\rangle$ is determined by the Goldreich-Levin hardcore predicate associated with $(x, w) \mapsto x$. Let A_{12} be a $\mathsf{p.p.t.}$ adversary that distinguishes between games 1 and 2 with

$$\mathrm{Adv}_{A_{12}}(\lambda) = \left|\Pr_{c\leftarrow\mathcal{G}_1}[A_{12}(y_\lambda, c)] - \Pr_{c\leftarrow\mathcal{G}_2}[A_{12}(y_\lambda, c)]\right|.$$

We give a non-uniform p.p.t. adversary A_{HCP} that guesses the hardcore predicate. A_{HCP} takes non-uniform input y_λ, z_λ. The challenger picks $(x, w) \sim D$, $q \sim \{0,1\}^{|x|}$, $r \sim \{0,1\}^{|w|}$ and A_{HCP} gets input (x, q, r). It derives π from x, y_λ, z_λ. It picks $b \sim \{0,1\}$ and computes $a = A_{12}(y_\lambda, (x, \pi, r, b))$. If $a = 1$ then it outputs $b \oplus \langle x, q\rangle \oplus m$ with $m = 0$. Otherwise it outputs a random bit.

Let $t = \langle w, r\rangle \oplus m$. Considering the designs of $\mathcal{G}_1$ and $\mathcal{G}_2$ we have

$$\begin{aligned}
\Pr_{c \leftarrow \mathcal{G}_1}[A_{12}(y_\lambda, c)] &= \Pr[a = 1 | b = t], \\
\Pr_{c \leftarrow \mathcal{G}_2}[A_{12}(y_\lambda, c)] &= \frac{1}{2}(\Pr[a = 1 | b = t] + \Pr[a = 1 | b \neq t]).
\end{aligned}$$

Now using these values we have

$$\begin{aligned}
\Pr[A_{\mathsf{HCP}}(x, q, r) = \langle (x, w), (r, q)\rangle] &= \Pr[a = 1 | b = t] \Pr[b = t] + \frac{1}{2}\Pr[a = 0] \\
&= \frac{1}{2}\left(1 + \Pr_{c \leftarrow \mathcal{G}_1}[A_{12}(y_\lambda, c)] - \Pr_{c \leftarrow \mathcal{G}_2}[A_{12}(y_\lambda, c)]\right) \\
&= \frac{1}{2} \pm \mathrm{Adv}_{A_{12}}.
\end{aligned}$$

However, since $\langle (x, w), (r, q)\rangle$ is the Goldreich-Levin hardcore predicate [20] of the one-way function $(x, w) \mapsto x$, we know that

$$\Pr[A_{\mathsf{HCP}}(x, q, r) = \langle (x, w), (r, q)\rangle] = \frac{1}{2} + \mathrm{negl}(\lambda).$$

Conclude that $\mathrm{Adv}_{A_{12}}$ is negligible.

Conclude. By repeating the above arguments with $m = 1$, observe that $\mathcal{G}_2 \approx \mathcal{G}_3$ and $\mathcal{G}_3 \approx \mathcal{G}_4$. Then note that $\mathcal{G}_0$ and $\mathcal{G}_4$ are the honest soundness security game with plaintexts $m = 0$ and $m = 1$ respectively. As $\mathcal{G}_0 \approx \mathcal{G}_4$ conclude that $\mathrm{Adv}_{B,Y}$ is negligible.

6.4 Applicability for best-possible proofs

We do not have a candidate for a specific language $T \in \mathsf{NP} \cap \mathsf{coNP}$ for which witness encryption with average case correctness does not exist. However, even lacking such a candidate, we argue that the construction of section 5 is secure for any choice of polynomial length parameters. In particular, it is witness hiding for D with unique witnesses assuming any T exists.

Theorem 11. *Let D as above. Assume some $T \in \mathsf{NP} \cap \mathsf{coNP}$ lacks a witness encryption scheme with average case correctness relative to some ensemble E for all input lengths large enough. Then the* BestPossible *with known length parameters, calculated below, is witness hiding for D.*

Proof. Let $t'(n, m)$ be the runtime of $\mathsf{VsWE.Verify}$ where n is the length of x and m is the length of y encoded as an instance of a fixed NP-complete language; note $t' = \mathrm{poly}(n, m)$. We will run the best-possible proof with witness-length parameter $q = n$, runtime parameter $s = t'(n, n)$, and proof-length parameter n and adjust the length of y appropriately.

Recall that T has an NP verifier V_T. Let $q'(\lambda)$ be the maximum over inputs of length λ of the length of the shortest witness; note $q' = \mathrm{poly}(\lambda)$. Let $s'(\lambda)$ be the size of V_T as a Turing machine; note $s' = O(1)$. Writing y as an instance of our NP-complete language we have $m = \mathrm{poly}(\lambda)$.

Now lift the proof of soundness from Theorem 10 into the deductive system $\mathcal{S}$. The size of this proof depends on the size of V_T, but should still be $\ell'(\lambda) = O(1)$.

So we have s' and ℓ' constant; thus $n > s', \ell'$ for all n large enough. Further we have that q' is polynomial in λ. Choose $\lambda(n)$ so that $q'(\lambda) < n$ and $m(\lambda) < n$. Define the ensemble E' such that $E'(n) = E(\lambda(n))$. Then use E' to instantiate VsWE.

Note that an average case correct witness encryption scheme relative to E' for infinitely many n immediately gives an average case correct witness encryption scheme relative to E for infinitely many λ. Now apply Theorem 10. By construction, s, q, ℓ are large enough for inputs large enough to describe the appropriate witness, verifier, and soundness proof. Conclude $\mathsf{BestPossible}$ is witness hiding.

References

1. Arora, S., Barak, B.: Computational Complexity: A Modern Approach. Cambridge University Press, USA, 1st edn. (2009)
2. Badrinarayanan, S., Miles, E., Sahai, A., Zhandry, M.: Post-zeroizing obfuscation: New mathematical tools, and the case of evasive circuits. In: Fischlin, M., Coron, J.S. (eds.) EUROCRYPT 2016, Part II. LNCS, vol. 9666, pp. 764–791. Springer, Heidelberg (May 2016). https://doi.org/10.1007/978-3-662-49896-5_27
3. Ball, M., Dachman-Soled, D., Kulkarni, M.: New techniques for zero-knowledge: Leveraging inefficient provers to reduce assumptions, interaction, and trust. In: Micciancio, D., Ristenpart, T. (eds.) CRYPTO 2020. LNCS, vol. 12172, pp. 674–703. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-56877-1_24
4. Barak, B., Goldreich, O., Impagliazzo, R., Rudich, S., Sahai, A., Vadhan, S.P., Yang, K.: On the (im)possibility of obfuscating programs. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 1–18. Springer, Heidelberg (Aug 2001). https://doi.org/10.1007/3-540-44647-8_1
5. Barak, B., Lindell, Y., Vadhan, S.P.: Lower bounds for non-black-box zero knowledge. In: 44th FOCS. pp. 384–393. IEEE Computer Society Press (Oct 2003). https://doi.org/10.1109/SFCS.2003.1238212
6. Barak, B., Ong, S.J., Vadhan, S.P.: Derandomization in cryptography. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 299–315. Springer, Heidelberg (Aug 2003). https://doi.org/10.1007/978-3-540-45146-4_18
7. Barak, B., Pass, R.: On the possibility of one-message weak zero-knowledge. In: Naor, M. (ed.) TCC 2004. LNCS, vol. 2951, pp. 121–132. Springer, Heidelberg (Feb 2004). https://doi.org/10.1007/978-3-540-24638-1_7

8. Bellare, M., Palacio, A.: GQ and Schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 162–177. Springer, Heidelberg (Aug 2002). https://doi.org/10.1007/3-540-45708-9_11
9. Bitansky, N., Khurana, D., Paneth, O.: Weak zero-knowledge beyond the black-box barrier. In: Charikar, M., Cohen, E. (eds.) 51st ACM STOC. pp. 1091–1102. ACM Press (Jun 2019). https://doi.org/10.1145/3313276.3316382
10. Bitansky, N., Paneth, O.: ZAPs and non-interactive witness indistinguishability from indistinguishability obfuscation. In: Dodis, Y., Nielsen, J.B. (eds.) TCC 2015, Part II. LNCS, vol. 9015, pp. 401–427. Springer, Heidelberg (Mar 2015). https://doi.org/10.1007/978-3-662-46497-7_16
11. Condon, A.: The complexity of stochastic games. Inf. Comput. **96**(2), 203–224 (Feb 1992). https://doi.org/10.1016/0890-5401(92)90048-K
12. Deng, Y., Song, X., Yu, J., Chen, Y.: On instance compression, schnorr/guillou-quisquater, and the security of classic protocols for unique witness relations. Cryptology ePrint Archive, Report 2017/390 (2017), http://eprint.iacr.org/2017/390
13. Deshpande, A., Kalai, Y.: Proofs of ignorance and applications to 2-message witness hiding. Cryptology ePrint Archive, Report 2018/896 (2018), https://eprint.iacr.org/2018/896
14. Dwork, C., Naor, M.: Zaps and their applications. In: 41st FOCS. pp. 283–293. IEEE Computer Society Press (Nov 2000). https://doi.org/10.1109/SFCS.2000.892117
15. Dwork, C., Stockmeyer, L.J.: 2-round zero knowledge and proof auditors. In: 34th ACM STOC. pp. 322–331. ACM Press (May 2002). https://doi.org/10.1145/509907.509958
16. Feige, U., Shamir, A.: Witness indistinguishable and witness hiding protocols. In: 22nd ACM STOC. pp. 416–426. ACM Press (May 1990). https://doi.org/10.1145/100216.100272
17. Feige, U., Shamir, A.: Zero knowledge proofs of knowledge in two rounds. In: Brassard, G. (ed.) CRYPTO'89. LNCS, vol. 435, pp. 526–544. Springer, Heidelberg (Aug 1990). https://doi.org/10.1007/0-387-34805-0_46
18. Freitag, C., Komargodski, I., Pass, R.: Impossibility of strong KDM security with auxiliary input. Cryptology ePrint Archive, Report 2019/293 (2019), https://eprint.iacr.org/2019/293
19. Garg, S., Gentry, C., Sahai, A., Waters, B.: Witness encryption and its applications. In: Boneh, D., Roughgarden, T., Feigenbaum, J. (eds.) 45th ACM STOC. pp. 467–476. ACM Press (Jun 2013). https://doi.org/10.1145/2488608.2488667
20. Goldreich, O., Levin, L.A.: A hard-core predicate for all one-way functions. In: 21st ACM STOC. pp. 25–32. ACM Press (May 1989). https://doi.org/10.1145/73007.73010
21. Goldreich, O., Oren, Y.: Definitions and properties of zero-knowledge proof systems. Journal of Cryptology **7**(1), 1–32 (Dec 1994). https://doi.org/10.1007/BF00195207
22. Goldwasser, S., Micali, S.: Probabilistic encryption and how to play mental poker keeping secret all partial information. In: 14th ACM STOC. pp. 365–377. ACM Press (May 1982). https://doi.org/10.1145/800070.802212
23. Goldwasser, S., Micali, S., Rackoff, C.: The knowledge complexity of interactive proof systems. SIAM Journal on Computing **18**(1), 186–208 (1989)
24. Goldwasser, S., Rothblum, G.N.: On best-possible obfuscation. In: Vadhan, S.P. (ed.) TCC 2007. LNCS, vol. 4392, pp. 194–213. Springer, Heidelberg (Feb 2007). https://doi.org/10.1007/978-3-540-70936-7_11

25. Gordon, S.D., Wee, H., Xiao, D., Yerukhimovich, A.: On the round complexity of zero-knowledge proofs based on one-way permutations. In: Abdalla, M., Barreto, P.S.L.M. (eds.) LATINCRYPT 2010. LNCS, vol. 6212, pp. 189–204. Springer, Heidelberg (Aug 2010)
26. Groth, J., Ostrovsky, R., Sahai, A.: Non-interactive zaps and new techniques for NIZK. In: Dwork, C. (ed.) CRYPTO 2006. LNCS, vol. 4117, pp. 97–111. Springer, Heidelberg (Aug 2006). https://doi.org/10.1007/11818175_6
27. Haitner, I., Rosen, A., Shaltiel, R.: On the (im)possibility of Arthur-Merlin witness hiding protocols. In: Reingold, O. (ed.) TCC 2009. LNCS, vol. 5444, pp. 220–237. Springer, Heidelberg (Mar 2009). https://doi.org/10.1007/978-3-642-00457-5_14
28. Jain, A., Kalai, Y.T., Khurana, D., Rothblum, R.: Distinguisher-dependent simulation in two rounds and its applications. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017, Part II. LNCS, vol. 10402, pp. 158–189. Springer, Heidelberg (Aug 2017). https://doi.org/10.1007/978-3-319-63715-0_6
29. Katz, J.: Which languages have 4-round zero-knowledge proofs? In: Canetti, R. (ed.) TCC 2008. LNCS, vol. 4948, pp. 73–88. Springer, Heidelberg (Mar 2008). https://doi.org/10.1007/978-3-540-78524-8_5
30. Lackenby, M.: The efficient certification of knottedness and thurston norm (2016), https://arxiv.org/abs/1604.00290
31. Lindell, Y.: An efficient transform from sigma protocols to NIZK with a CRS and non-programmable random oracle. In: Dodis, Y., Nielsen, J.B. (eds.) TCC 2015, Part I. LNCS, vol. 9014, pp. 93–109. Springer, Heidelberg (Mar 2015). https://doi.org/10.1007/978-3-662-46494-6_5
32. Pass, R.: On deniability in the common reference string and random oracle model. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 316–337. Springer, Heidelberg (Aug 2003). https://doi.org/10.1007/978-3-540-45146-4_19
33. Pass, R.: Simulation in quasi-polynomial time, and its application to protocol composition. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 160–176. Springer, Heidelberg (May 2003). https://doi.org/10.1007/3-540-39200-9_10