Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 108 (2014)

On Basing Size-Verifiable One-Way Functions on NP-Hardness

Andrej Bogdanov[*]   Christina Brzuska[†]

Abstract

We prove that if the hardness of inverting a size-verifiable one-way function can be based on NP-hardness via a general (adaptive) reduction, then NP $\subseteq$ coAM. This claim was made by Akavia, Goldreich, Goldwasser, and Moshkovitz (STOC 2006), but was later retracted (STOC 2010).

Akavia, Goldreich, Goldwasser, and Moshkovitz [AGGM06] claimed that if there exists an adaptive reduction from an NP-complete problem to inverting an efficient size-verifiable function, then NP $\subseteq$ coAM. They provided a proof for size-verifiable functions that have polynomial pre-image size as well as a proof for general size-verifiable functions, even if the size of the pre-image can only be approximated. The proof for the latter statement was found to be erroneous and has been retracted [AGGM10].[1] In this note we give a proof of their claim. For motivation about the problem, we refer the reader to the work [AGGM06].

Throughout this paper, we consider efficiently computable functions f with $f(\{0,1\}^n) \subseteq \{0,1\}^{m(n)}$, where m is an injective function on integers. We say an oracle I *inverts* f if for every $x \in \{0,1\}^*$, $I(f(x))$ belongs to the set $f^{-1}(f(x))$.

We say that f is *size-verifiable* if the decision problem $N_f = \{(y,s) \colon |f^{-1}(y)| = s\}$ is in AM. We say that f is *approximately size-verifiable* if the following promise problem A_f is in AM:

$$\begin{aligned} \text{YES instances of } A_f\text{:} &\quad (y, s, 1^a) \text{ such that } |f^{-1}(y)| \le s \\ \text{NO instances of } A_f\text{:} &\quad (y, s, 1^a) \text{ such that } |f^{-1}(y)| > (1 + 1/a)s. \end{aligned}$$

A *reduction* from a decision problem L to inverting f is a randomized oracle algorithm $R^?$ such that for every oracle I that inverts f, R^I decides L with probability at least 2/3 over the randomness of $R^?$.

---

[*] This work was partially supported by Hong Kong RGC GRF grant CUHK410113.

[†] Christina Brzuska was supported by the Israel Science Foundation (grant 1076/11 and 1155/11), the Israel Ministry of Science and Technology grant 3-9094), and the German-Israeli Foundation for Scientific Research and Development (grant 1152/2011). Part of this work was done while visiting CUHK.

[1] In the same paper [AGGM06], Akavia et al. also show that the existence of a (randomized) non-adaptive reduction of NP to the task of inverting an arbitrary one-way function implies that NP $\subseteq$ coAM. This result is not affected by the gap found in [AGGM10].

ISSN 1433-8092

Theorem 1. *Let f be an efficiently computable, approximately size-verifiable function. If there exists an efficient reduction from L to inverting f with respect to deterministic inversion oracles, then L is in* $\mathrm{AM} \cap \mathrm{coAM}$.

Corollary 2. *Let f be an efficiently computable, approximately size-verifiable function. There is no efficient reduction from an* NP*-hard language L to inverting f with respect to deterministic inversion oracles, unless* $\mathrm{NP} \subseteq \mathrm{coAM}$.

We first prove a weaker version of the theorem that relies on two simplifying assumptions. Firstly, we assume that the reduction is correct even with respect to *randomized* inversion oracles. These are oracles that have access to an internal source of randomness when answering their queries. Our inversion oracle will simply sample a uniform pre-image amongst all possible pre-images like an inverter for a distributional one-way function [IL89]. Note that a reduction that works for randomized inversion oracles also works with respect to deterministic oracles, as they are a special case of randomized ones. As we prove a negative result, stronger requirements on the reduction weaken our result. We will thus explain later how to remove this additional requirement on the reduction. Secondly, we assume the function to be size-verifiable rather than approximately size-verifiable. We then adapt the proof to the general case.

Randomized inversion oracles Let $R^?$ be a reduction, I a randomized oracle, and z an input. A *valid transcript* of $R^I(z)$ is a string of the form $(r, x_1, \ldots, x_q)$, where r is the randomness of the reduction and $x_1, \ldots, x_q$ are the oracle answers in the order produced by I. We will assume, without loss of generality, that the length of r and the number of queries q depend only on the length of z.

Consider the randomized inversion oracle U that, on query y, returns an x chosen uniformly at random from the set $f^{-1}(y)$, or the special symbol $\perp$ if this set is empty. Let the set C consist of all tuples $(z, r, x_1, \ldots, x_q, p)$, such that $(r, x_1, \ldots, x_q)$ is an accepting valid transcript of $R^U(z)$ and p is an integer between 1 and $\lceil K/(s(y_1)\cdots s(y_q))\rceil$. Here,

- y_i is the i-th query of the reduction,
- $s(y)$ is the size of the set of possible answers on query y:

$$s(y) = \begin{cases} |f^{-1}(y)|, & \text{if } f^{-1}(y) \text{ is non-empty} \\ 1, & \text{otherwise,} \end{cases}$$

- and $K = 2 \cdot 2^{q\ell}$, where ℓ is an upper bound on the length of queries $R^?$ makes on inputs of length $|z|$.

Claim 3. *C is in* AM.

Proof. On input $(z, r, x_1, \ldots, x_q, p)$, the AM verifier for C runs the reduction on input z with randomness r and checks that for for each query y_i that the reduction makes, the answer x_i

is indeed a pre-image of y_i and that the reduction accepts. To see that p is of the right size, we ask the prover to provide $s(y_1), \ldots, s(y_q)$ such that $p \leq K/(s(y_1)\cdots s(y_q))$. We then run the AM verifier for N_f ato check that the numbers $s(y_1), \ldots, s(y_q)$ that the prover provided are correct. $\square$

Let $C(z)$ denote the set of all $(r, x_1, \ldots, x_q, p)$ such that $(z, r, x_1, \ldots, x_q, p)$ is in C.

Claim 4. *$C(z)$ has size at least $\frac{2}{3}2^{|r|}K$ if $z \in L$, and size at most $\frac{1}{2}2^{|r|}K$ if $z \notin L$.*

Proof. Fix the input z. Conditioned on the randomness r, every valid transcript $(r, x_1, \ldots, x_q)$ appears with probability exactly $1/(s(y_1)\ldots s(y_q))$ over the choice of randomness of the inverter. All these probabilities add up to one:

$$\sum_{(x_1,\ldots,x_q)} \frac{1}{s(y_1)\ldots s(y_q)} = 1.$$

If $z \in L$, then at least a $2/3$ fraction of these valid transcripts must be accepting for $R^?(z)$ over the choice of r and so

$$\begin{aligned}
|C(z)| &\geq \tfrac{2}{3}\textstyle\sum_{\text{valid transcript } (r, x_1, \ldots, x_q)} \left\lceil \dfrac{K}{s(y_1)\ldots s(y_q)} \right\rceil \\
&\geq \tfrac{2}{3}\textstyle\sum_r K \sum_{(x_1,\ldots,x_q)} \dfrac{1}{s(y_1)\ldots s(y_q)} \\
&= \tfrac{2}{3}2^{|r|}K.
\end{aligned}$$

If $z \notin L$, then at most a $1/3$ of the valid transcripts are accepting, and

$$\begin{aligned}
|C(z)| &\leq \tfrac{1}{3}\textstyle\sum_{\text{valid transcript } (r, x_1, \ldots, x_q)} \left\lceil \dfrac{K}{s(y_1)\ldots s(y_q)} \right\rceil \\
&\leq \tfrac{1}{3}\textstyle\sum_r (K+1) \sum_{(x_1,\ldots,x_q)} \dfrac{1}{s(y_1)\ldots s(y_q)} \\
&\leq \tfrac{1}{3}\textstyle\sum_r K\Big(\sum_{(x_1,\ldots,x_q)} \dfrac{1}{s(y_1)\ldots s(y_q)} + \sum_{(r,x_1,\ldots,x_q)} 1\Big) \\
&\leq \tfrac{1}{3}2^{|r|}(K + 2^{q\ell}) \\
&\leq \tfrac{1}{2}2^{|r|}K
\end{aligned}$$

by our choice of K. $\square$

Using the set lower bound protocol of Goldwasser and Sipser [GS86], we conclude that L is in AM. Applying the same argument to the reduction $\overline{R}^?$ that outputs the opposite answer of $R^?$, it follows that L is also in coAM.

Deterministic inversion oracles We now prove Theorem 2 for size-verifiable functions and deterministic inversion oracles. Assume $R^?$ is an efficient reduction from L to inverting f with respect to deterministic inversion oracles. Then, for every inversion oracle I for f, R^I decides L with probability at least 2/3. By averaging, it follows that for every distribution $\mathcal{I}$ on inversion oracles I for f, R^I decides L with probability at least 2/3:

$$\Pr_{r, I \sim \mathcal{I}}[R^I(z; r) = L(z)] \geq \tfrac{2}{3} \quad \text{for every } z.$$

If the oracle U could be written as a probability distribution over deterministic inversion oracles for f, then Theorem 2 would follow immediately from Claims 3 and 4. Unfortunately this is not the case: One reason is that a deterministic oracle sampled from any distribution always produces the same answer to the same query, while the oracle U outputs statistically independent answers. We resolve this difficulty by applying a minor modification to the description of U: The modified oracle U' will choose among the answers to a query y using randomness coming from a random function F applied to y. Specifically, if $x_1, \dots, x_{s(y)}$ are the possible inverses of y, then $U'(y) = x_{F(y)}$.

Proof of Theorem 2 for size-verifiable functions. Let $\ell(n)$ and $q(n)$ be polynomial, efficiently computable upper bounds on the query length and query complexity of the reduction on inputs of length n, respectively. Let $\mathcal{F} = \{F_m\}$ be a collection of random functions, where F_m takes as input a string $y \in \{0,1\}^m$ and outputs a number between 1 and $s(y)$. We define the randomized oracle U' as follows:

- **Randomness:** For every query length m, choose a uniformly random F_m, independently of $F_1, \dots, F_{m-1}$.
- **Functionality:** On input y of length m, output $\perp$ if y is not in the range of f, or $U'(y) = x_{F_m(y)}$ if it is, where $x_1, \dots, x_{s(y)}$ are the inverses of y under f.

Observe that U' is determined by a product distribution over $F_1, F_2, \dots$ and any fixing of $F_1, F_2, \dots$ specifies a deterministic inversion oracle for f. Since, for every z, the event $R^{U'}(z; r) = L(z)$ is measurable both over r and over $(F_1, F_2, \dots)$, by averaging

$$\Pr_{r, (F_1, F_2, \dots) \sim \mathcal{F}}[R^{U'}(z; r) = L(z)] \geq \tfrac{2}{3} \quad \text{for every } z.$$

We may now assume, without loss of generality, that $R^{U'}$ never makes the same query twice to the oracle U'. (More formally, we replace $R^?$ by another reduction that memoizes answers to previously made queries, and possibly makes some dummy queries at the end to ensure the number of queries is exactly $q(n)$ on inputs of length n.) We define $C(z)$ as before. Claims 3 and 4 still hold, and so L is in AM $\cap$ coAM. □

Extension to approximately size-verifiable functions Consider the promise problem C', whose YES instances are the same as the YES instances of C, and whose NO instances consist of the $(z, r, x_1, \dots, x_q, p)$ for which either $(r, x_1, \dots, x_q)$ is not an accepting valid transcript of $R^{U'}(z)$ or $p > \lceil \frac{6}{5} K / s(y_1) \dots s(y_q) \rceil$, where $K = \frac{10}{3} 2^{q\ell}$. We now prove the analogues of Claims 3 and 4. We observe that the Goldwasser-Sipser lower bound protocol extends to AM-promise problems and conclude, as before, that L must be in AM $\cap$ coAM.

Claim 5. *C' is in* AM.

Proof. On input $(z, r, x_1, \ldots, x_q, p)$, the AM verifier for C' runs the reduction on input z with randomness r and checks that for for each query y_i that the reduction makes, the answer x_i is indeed a pre-image of y_i and that the reduction accepts. It then asks the prover to provide claims $\hat{s}_i$ for the values $s(y_i), 1 \le i \le q$, runs the AM proof for A_f on input $(y_i, \hat{s}_i, 1^{6q})$, and verifies that $p \le \lceil K/\hat{s}_1 \ldots \hat{s}_q \rceil$. Clearly the verifier accepts YES instances of C'. If $(z, r, x_1, \ldots, x_q, p)$ is a NO instance, then either the transcript is not valid and accepting, or $f(x_i) \ne y_i$ for some i, or $\lceil K/\hat{s}_1 \ldots \hat{s}_q \rceil \ge p > \lceil \frac{6}{5} K / s(y_1) \ldots s(y_q) \rceil$, in which case $s(y_i) > (6/5)^{1/q} \hat{s}_i > (1 + 1/(6q))\hat{s}_i$ for some i and the verifier for A_f rejects. □

Let $C'_{\mathrm{YES}}(z)$ and $C'_{\mathrm{NO}}(z)$ consist of those $(r, x_1, \ldots, x_q, p)$ such that $(z, r, x_1, \ldots, x_q, p)$ are YES and NO instances of C', respectively.

Claim 6. *If $z \in L$, then $C'_{\mathrm{YES}}(z)$ has size at least $\frac{2}{3} 2^{|r|} K$. If $z \notin L$, then $\overline{C'_{\mathrm{NO}}(z)}$ has size at most $\frac{1}{2} 2^{|r|} K$, where $\overline{C'_{\mathrm{NO}}(z)}$ denotes all tuples $(z, r, x_1, ..., x_q, p)$ that are not in $C'_{\mathrm{NO}}(z)$.*

Proof. The proof of the first part is identical to the proof of the first part of Claim 4. For the second part, if $z \notin L$, then by a similar calculation

$$|\overline{C'_{\mathrm{NO}}(z)}| \le \tfrac{1}{3} \sum\nolimits_{\text{valid transcript } (r, x_1, \ldots, x_q)} \left\lceil \frac{6K/5}{s(y_1) \ldots s(y_q)} \right\rceil \le \tfrac{1}{3} 2^{|r|} (\tfrac{6}{5} K + 2^{q\ell}) \le \tfrac{1}{2} 2^{|r|} K. \qquad \square$$

Acknowledgements

We thank Oded Goldreich and Shafi Goldwasser for helpful comments on the presentation.

References

[AGGM06] Adi Akavia, Oded Goldreich, Shafi Goldwasser, and Dana Moshkovitz. On basing one-way functions on NP-hardness. In *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing*, STOC '06, pages 701–710, New York, NY, USA, 2006. ACM.

[AGGM10] Adi Akavia, Oded Goldreich, Shafi Goldwasser, and Dana Moshkovitz. Erratum for: On basing one-way functions on NP-hardness. In *Proceedings of the Forty-second ACM Symposium on Theory of Computing*, STOC '10, pages 795–796, New York, NY, USA, 2010. ACM.

[GS86] Shafi Goldwasser and Michael Sipser. Private coins versus public coins in interactive proof systems. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, STOC '86, pages 59–68, New York, NY, USA, 1986. ACM.

[IL89] Russell Impagliazzo and Michael Luby. One-way functions are essential for complexity based cryptography (extended abstract). In *30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, North Carolina, USA, 30 October - 1 November 1989*, pages 230–235. IEEE Computer Society, 1989.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de