Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



Statistical Zero-Knowledge Arguments for **NP** from Any One-Way Function*

4.2 Interactive Hashing

Ostrovsky, Venkatesan and Yung [OVY93] introduced a powerful tool known as *interactive hashing (IH)*, which is a protocol between a sender S_{IH} and receiver R_{IH}. The sender begins with a private input y, and at the end both parties outputs y_0 and y_1 such that $y \in \{y_0, y_1\}$. Informally, the IH protocol has the following properties:

1. *(Hiding)* If the sender's input y is uniformly random, then the receiver does not learn which of y_0 or y_1 equals to y.

2. *(Binding)* The sender can "control" the value of at most one of the two outputs.

Naor, Ostrovsky, Venkatesan and Yung [NOVY98] showed that interactive hashing can be used to construct statistically hiding commitment schemes from one-way permutations.

We extend the notion of interactive hashing to allow multiple outputs (instead of just two output strings). Since we allow the number of outputs to be possibly superpolynomial, we succinctly describe the set of outputs as the image of a polynomial-sized circuit $C\colon \{0,1\}^k \to \{0,1\}^q$, where k and q are polynomially related to the security parameter.

For a relation W, let $W_y = \{z : W(y,z) = 1\}$ and we refer to any $z \in W_y$ as a *valid witness* for y. In the definitions below, we use general relations, and hence do not require that relation W be polynomial-time computable.

Definition 4.2. An *interactive hashing scheme with multiple outputs* is a polynomial-time protocol $(S_{\text{IH}}, R_{\text{IH}})$ where both parties receive common inputs $(1^q, 1^k)$, S_{IH} receives a private input $y \in \{0,1\}^q$, with the common output being a circuit $C\colon \{0,1\}^k \to \{0,1\}^q$, and the private output of S_{IH} being a string $z \in \{0,1\}^k$. We denote q to be the input length and k to be the output length. The protocol $(S_{\text{IH}}, R_{\text{IH}})$ has to satisfy the following security properties:

1. *(Correctness)* For all R^* and all $y \in \{0,1\}^q$, letting $C = (S_{\text{IH}}(y), R^*)(1^q, 1^k)$ and $z = \mathsf{output}_{S_{\text{IH}}}(S_{\text{IH}}(y), R^*)$, we have that $C(z) = y$.

2. *(Perfect hiding)* For all R^*, (V, Z) is distributed identically to (V, U_k), where $V = \mathsf{view}_{R^*}(S_{\text{IH}}(U_q), R^*)$ and $Z = \mathsf{output}_{S_{\text{IH}}}(S_{\text{IH}}(U_q), R^*)$.

3. *("Computational" binding)* There exists an oracle PPT algorithm A such that for every S^* and any relation W, letting circuit $C = (S^*, R_{\text{IH}})(1^q, 1^k)$ and $((x_0, z_0), (x_1, z_1)) = \mathsf{output}_{S^*}(S^*, R_{\text{IH}})$, if it holds that
$$\Pr[x_0 \in W_{C(z_0)} \wedge x_1 \in W_{C(z_1)} \wedge z_0 \neq z_1] > \varepsilon,$$
where the above probability is over the coin tosses of R_{IH} and S^*. Then we have that
$$\Pr_{y \leftarrow \{0,1\}^q}[A^{S^*}(y, 1^q, 1^k, \varepsilon) \in W_y] > 2^{-k} \cdot (\varepsilon/q)^{O(1)}.$$

We make three remarks regarding the above definition.

1. The security requirements should hold for all, even computationally unbounded R^* (for correctness and perfect hiding) and computationally unbounded S^* (even though binding is "computational"). In addition, the relation W need not be polynomial-time computable.

2. To simplify notation, we often write $A^{S^*}(y)$, or even $A(y)$, to denote $A^{S^*}(y, 1^q, 1^k, \varepsilon)$.

3. Although the output of the honest sender S_{IH} is always a string z, the output of the cheating sender S^* is arbitrary; hence, we can assume without loss of generality that S^* breaks binding by producing two pairs of strings (x_0, z_0) and (x_1, z_1).

We think of the string $z \in \{0,1\}^k$ as a k-bit string commitment associated to one of the 2^k outputs strings, namely $y = C(z)$, and a witness $x \in W_y = W_{C(z)}$ as a decommitment to z. Intuitively, the knowledge of x gives the sender the ability to decommit to z. The "computational" binding property, read in its contrapositive, says that if it is hard to find a witness for a uniformly random string y, then it is hard for a sender to successfully decommit to two different values. Notice that this property holds even if the set of "hard" y's is not fixed in advance, but depends on the algorithm trying to find a witness for y, namely an element in W_y.

Protocol 4.5. Interactive Hashing Scheme with Multiple Outputs $(S_{\text{IH}}, R_{\text{IH}})$.

Inputs:

1. Input length 1^q and output length 1^k, both given as common input.
2. String $y \in \{0,1\}^q$, given as private input to sender S_{IH}.

Protocol:

R_{IH}: Select $h_0, h_1, \ldots, h_{q-k-1}$ such that each h_i is a random vector over GF[2] of the form $0^i 1\{0,1\}^{q-i-1}$ (i.e., i number of 0's followed by a 1, and random choice for the last $q-i-1$ positions).

For $j = 0, \ldots, q-k-1$, do the following:

$R_{\text{IH}} \to S_{\text{IH}}$: Send h_j.
$S_{\text{IH}} \to R_{\text{IH}}$: Send $c_j = \langle h_j, y \rangle$.

Output:

- Common output is a circuit $C\colon \{0,1\}^k \to \{0,1\}^q$. computing an affine transformation whose image is $\{y : \langle h_j, y \rangle = c_j \ \ \forall j = 0, \ldots, q-k-1\}$.
- Output of S_{IH} is a string $z \in \{0,1\}^k$ such that $C(z) = y$. (In fact, z can be taken to be the last k bits of y.)

B.1 Hiding Property

Lemma B.1 (perfect hiding)**.** *Protocol 4.5 is perfectly hiding in the sense of the Definition 4.2.*

The proofs presented in this section and the next are very similar in nature to those in [NOVY98], with additional analysis needed to handle interactive hashing for multiple outputs.

Proof. The view of any R^* will be the hash functions $h_0, h_1, \cdots, h_{q-k-1}$ together with S's respond $c_0, c_1, \ldots, c_{q-k-1}$. Given these values, we show that there are 2^{q-k} possible y's that would make $S(y)$ respond to $c_0, c_1, \ldots, c_{q-k-1}$ (given queries $h_0, h_1, \cdots, h_{q-k-1}$ from R^*).

Consider the matrix $H = (h_0, h_1, \cdots, h_{q-k-1})$ whose rows are the h_i's, vector $c = (c_0, c_1, \ldots, c_{q-k-1})$, and the equation $Hy = c$. Since h_i is of the form $0^i 1\{0,1\}^{q-i-1}$, the first $q-k$ columns of the matrix are linearly independent. Hence, any setting of the last k bits of y will fully determine the first $q-k$ bits of it. These are the 2^{q-k} strings y that satisfy $Hy = c$. □

B.2 Binding Property

Lemma B.2 (computational binding)**.** *Protocol 4.5 is computationally binding in the sense of the Definition 4.2.*

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan

We prove Lemma B.2 by providing an algorithm A that finds a valid witness (according to relation W) for a random string $y \leftarrow \{0,1\}^q$ with nonnegligible probability. Before describing A, we provide the following definitions.

Definitions. In the enumerated definitions below, h_i is of the form $0^i 1\{0,1\}^{q-i-1}$, and $h_i(y) = \langle h_i, y \rangle$. Without loss of generality, we can assume that S^* is deterministic because every probabilistic S^* can be converted to a (nonuniform) deterministic one with the same success probability and running time by fixing its random coins to maximize its success probability.

FROM [NOVY98]

We begin by making $\mathcal{S}'$ deterministic which can be done using standard techniques. Suppose that we choose an assignment to the random tape of $\mathcal{S}'$ and count the number of queries of $\mathcal{R}$ (i.e., $h_1, \ldots, h_{n-1}$) on which $\mathcal{S}'$ succeeds in cheating. By assumption, if the assignment is random, then the expected fraction of such queries is at least ε. Let Ω be the set of assignments on which $\mathcal{S}'$ is successful on at least $\varepsilon/2$ of $\mathcal{R}$'s queries. By a simple counting argument we can conclude that Ω consists of at least $\varepsilon/2$ of the possible assignments. Algorithm $\mathcal{A}$ described below requires $\mathcal{S}'$ to be deterministic. Therefore we choose $m = 2n/\varepsilon$ random assignments $\omega_1, \omega_2, \ldots, \omega_m$ and run m times the algorithm $\mathcal{A}$ with the random tape of $\mathcal{S}'$ initialized with $\omega_1, \omega_2, \ldots, \omega_m$. With probability $1 - (1 - \varepsilon/2)^m \geq 1 - e^{-n}$ some $\omega_i \in \Omega$. Therefore from now on we assume that $\mathcal{S}'$ is deterministic and its probability of success over $\mathcal{R}$'s queries is at least $\varepsilon/2$.

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan

Definitions.

1. For $0 \le i < q$, let $\mathcal{H}_i$ denote the set of hash functions of the form $0^i 1\{0,1\}^{q-i-1}$, i.e., $\mathcal{H}_i = \{0^i 1 w : w \in \{0,1\}^{q-i-1}\}$.

2. A *node* N at level i is defined by a series of hash functions $(h_0, h_1, \ldots, h_{i-1})$, where each $h_j \in \mathcal{H}_j$. (Since S^* is deterministic, this determines $c_0, \ldots, c_{i-1}$ where $c_j = S^*(h_0, \ldots, h_j)$.) Let L_i denote the set of nodes at level i.

3. The set of compatible hash functions at node $N \in L_i$ is denoted as

$$\mathrm{Comp}(N, y) = \{h_i \in \mathcal{H}_i : S^*(N, h_i) = h_i(y)\},$$

   where $S^*(N, h_i)$, with $N = (h_0, \ldots, h_{i-1})$, denotes $S^*(h_0, \ldots, h_i)$.

4. A string y is γ-*balanced* at $N \in L_i$ if

$$\frac{1-\gamma}{2} \le \frac{\mathrm{Comp}(N, y)}{|\mathcal{H}_i|} \le \frac{1+\gamma}{2}.$$

   A string y is γ-*fully-balanced* at $N \in L_i$ if it is γ-balanced at all its parental nodes. That is, letting $N = (h_0, \ldots, h_{i-1})$, y is required to be γ-balanced at all $N_0 = (h_0), N_1 = (h_0, h_1), \ldots, N = N_{i-1} = (h_0, \ldots, h_{i-1})$.

5. A string y is said to be *compatible* with a node $N = (h_0, \ldots, h_{i-1})$ if $h_j(y) = S^*(h_1, \ldots, h_j)$ for all $0 \le j < i$. Let $U(N)$ denote the set of compatible y's with node N. Note that for every $N \in L_i$, we have $|U(N)| = 2^{q-i}$.

6. Let $B(N)$ and $F(N)$ denote the set of γ-balanced strings and γ-fully-balanced strings at node N respectively. Moreover, let $G(N) = U(N) \setminus F(N)$ be the set of strings that are not fully-balanced. Note that for every node N, we have $F(N) \subseteq B(N) \subseteq U(N)$.

7. At every node $N \in L_{q-k}$, we can assume WLOG that $S^*(N)$ outputs a pair of strings (x_0, z_0) and (x_1, z_1), but it is not necessarily the case that any of $x_b \in W_{C(z_b)}$.

Description of the witness finding algorithm. Algorithm A: On input $y \in \{0,1\}^q, 1^q, 1^k$ and ε, do the following.

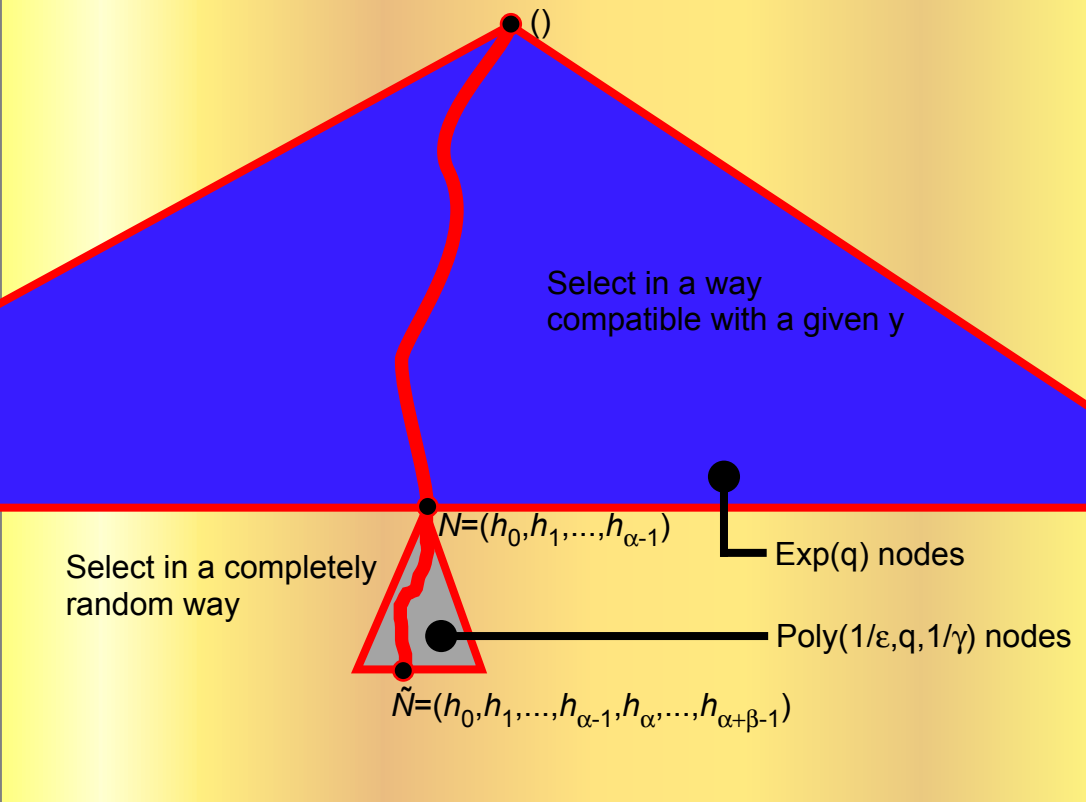
1. Set parameters $\gamma = 1/q$,
$$\beta = \log(1/\varepsilon) + 2\log(q) + 4\log(1/\gamma) + 4, \text{ and}$$
$$\alpha = q - \beta - k.$$
2. Repeat the following for $i = 0, \ldots, \alpha - 1$:

   When A is at node $N \in L_i$, explore along a random $h_i \leftarrow \mathrm{Comp}(N, y)$ to get to a new node $N' = (N, h_i) \in L_{i+1}$. (This can be done efficiently by choosing a random $h_i \leftarrow \mathcal{H}_i$ and querying S^* to make sure that $h_i \in \mathrm{Comp}(N, y)$, and repeat up to $8q$ times if not. If after $8q$ repetitive tries and fail to encounter any $h_i \in \mathrm{Comp}(N, y)$, then output `fail`.)

3. At node $N \in L_\alpha$, choose random $h_\alpha \leftarrow \mathcal{H}_\alpha, \ldots, h_{\alpha+\beta-1} \leftarrow \mathcal{H}_{\alpha+\beta-1}$, to arrive at node $\widetilde{N} = (N, h_\alpha, h_{\alpha+1}, \ldots, h_{\alpha+\beta-1}) \in L_{\alpha+\beta}$. (Note that $q - k = \alpha + \beta$, and hence $\widetilde{N} \in L_{\alpha+\beta} = L_{q-k}$.)
4. Query $S^*(\widetilde{N})$ to get (x_0, z_0) and (x_1, z_1). If either of $C(z_b) = y$, then output x_b. Else, output `fail`.

It is clear that the above algorithm runs in polynomial time (with oracle queries to S^*). All we need to show is that it succeeds with nonnegligible property, and we prove that property in the following claims.

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



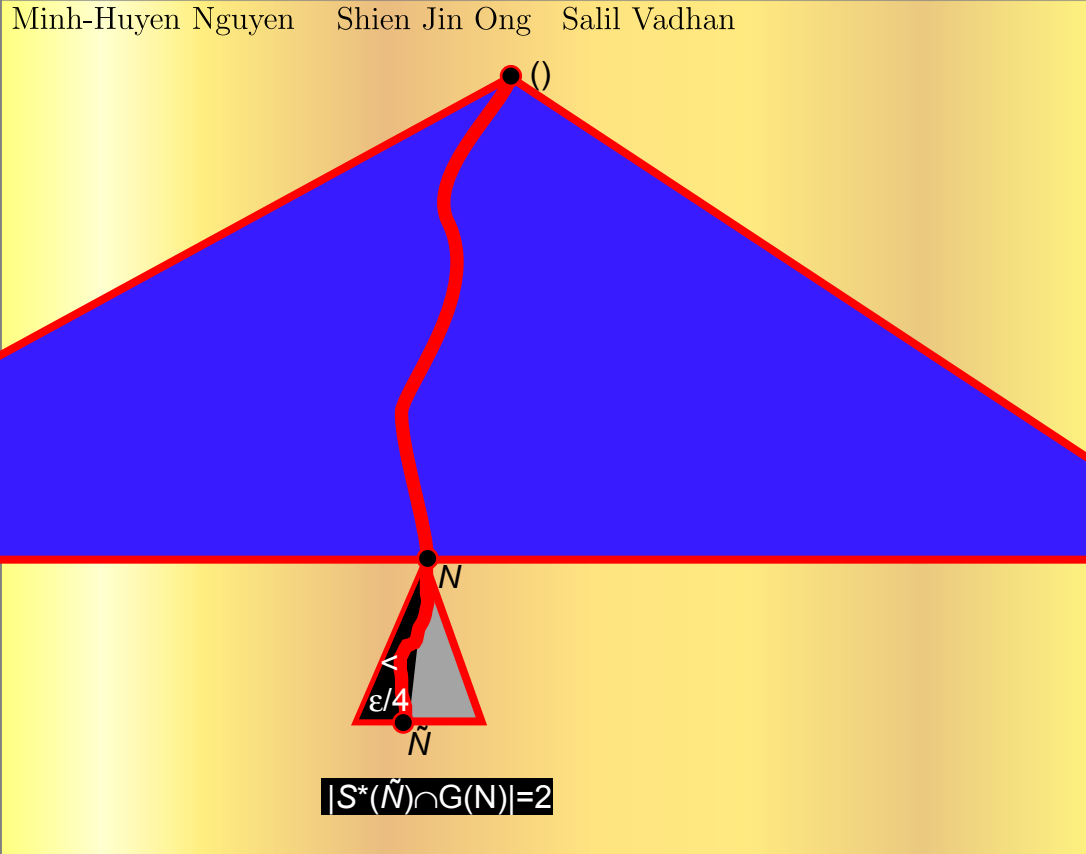
Claim B.3. *For every node $N \in L_i$, the set of unbalanced strings, $U(N) \setminus B(N) \leq 2/\gamma^2$.*

The next claim follows by a union bound on the unbalanced elements.

Claim B.4. *For every node $N \in L_i$, the set of strings that are not fully balanced, $G(N) = U(N) \setminus F(N) \leq 2i/\gamma^2$. In particular, for $\gamma = 1/q$, $|F(N)| \geq |U(N)|/2$ for $i \leq q - 4\log q$.*

Claim B.5. *For every node $N \in L_\alpha$, the fraction of children nodes $N_{\alpha+\beta}$ with greater than one element from $G(N)$ is at most $\varepsilon/4$.*

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



A node $N \in L_{\alpha+\beta} = L_{q-k}$ is *witness revealing* if both of $S^*(N)$'s outputs, namely (x_0, z_0) and (x_1, z_1), satisfy $C(z_b) \in U(N)$ and $x_b \in W_{C(z_b)}$, for $b \in \{0,1\}$. A node $N \in L_\alpha$ is said to be *good* if greater than $\varepsilon/2$ of its children at level $q-k$ are witness revealing.

Claim B.6. *The fraction of good nodes at level α is at least $\varepsilon/2$.*

Claim B.7. *For any fixed $N \in L_\alpha$ and $y' \in F(N)$, we have*

$$\frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1+\gamma)^\alpha} \cdot \frac{1}{|L_\alpha|} \leq \Pr[A \text{ reaches } N \wedge y = y'] \leq \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1-\gamma)^\alpha} \cdot \frac{1}{|L_\alpha|},$$

where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

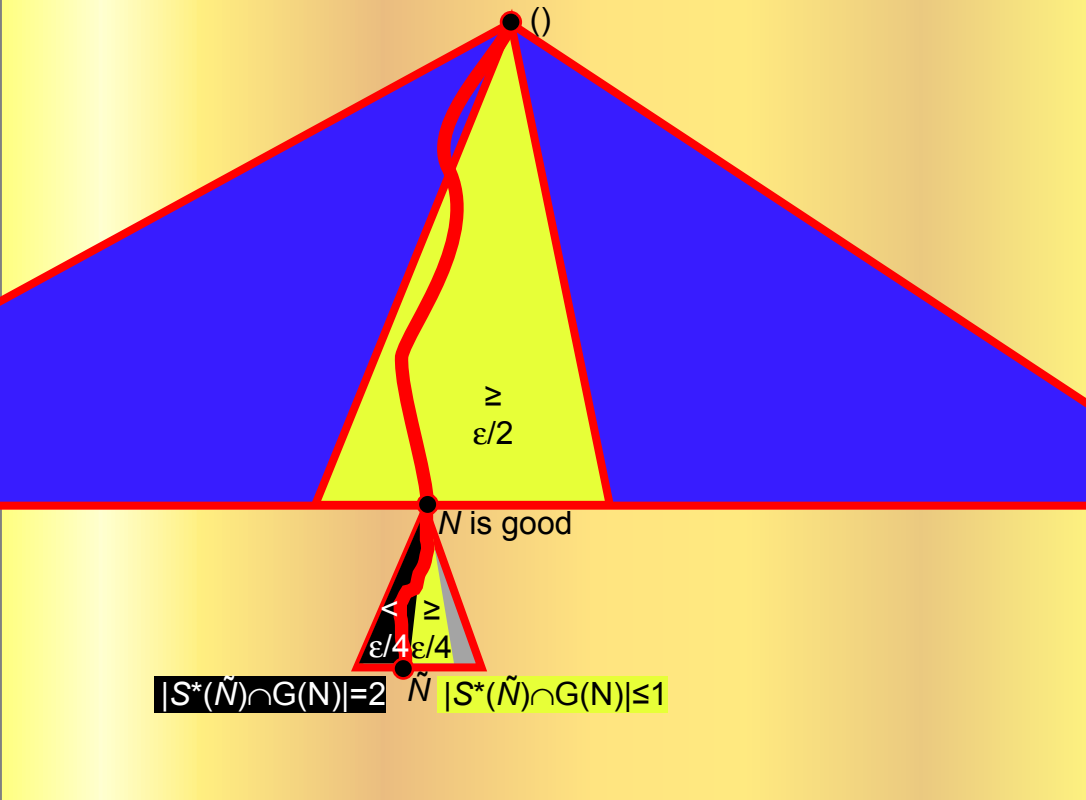
Claim B.8.

$$\Pr[\text{The node } N \text{ reached by } A \text{ is good } \wedge y \in F(N)] \geq \frac{\varepsilon}{4(1+\gamma)^\alpha}.$$

where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

Claim B.9. *In any good node $N \in L_\alpha$, the fraction of nonbinding children of N at level $\alpha+\beta$ that has one or less image in $G(N)$ is at least $\varepsilon/4$.*

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



Claim B.10. *For any fixed $N \in L_\alpha$ and $y' \in F(N)$, we have*

$$\Pr[y = y' | A \text{ reaches } N \wedge y \in F(N)] \geq \frac{1}{|F(N)|} \left(\frac{1-\gamma}{1+\gamma}\right)^\alpha,$$

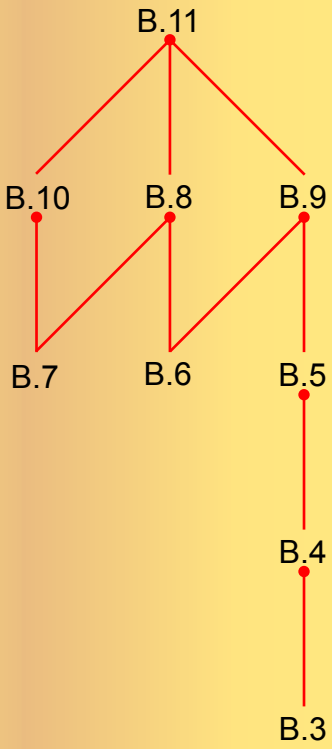
where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

We have now reached our final claim to complete the proof of the binding theorem.

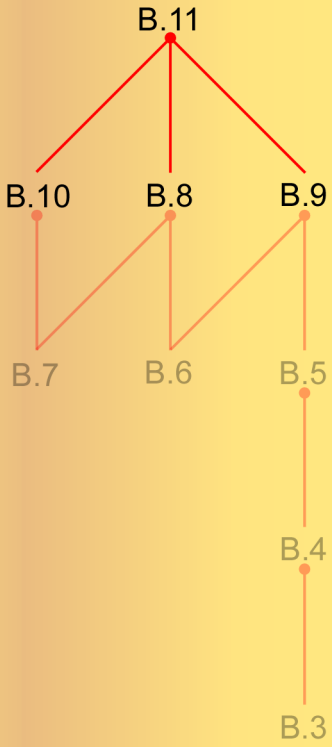
Claim B.11.

$$\Pr_{y \leftarrow \{0,1\}^q}[A(y) \in \tilde{R}_y] > c \cdot (\varepsilon^3 q^{-6} 2^{-k}) - \exp(q), \text{ for some constant } c > 0.$$

W_y


B.11
B.10
B.8
B.9
B.7
B.6
B.5
B.4
B.3

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



Claim B.8.

$$\Pr[\textit{The node } N \textit{ reached by } A \textit{ is good } \wedge y \in F(N)] \geq \frac{\varepsilon}{4(1+\gamma)^{\alpha}}.$$

where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

Claim B.9. *In any good node $N \in L_\alpha$, the fraction of nonbinding children of N at level $\alpha+\beta$ that has one or less image in $G(N)$ is at least $\varepsilon/4$.*

Claim B.10. *For any fixed $N \in L_\alpha$ and $y' \in F(N)$, we have*

$$\Pr[y = y' | A \textit{ reaches } N \wedge y \in F(N)] \geq \frac{1}{|F(N)|}\left(\frac{1-\gamma}{1+\gamma}\right)^{\alpha},$$

where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

We have now reached our final claim to complete the proof of the binding theorem.

Claim B.11.

$$\Pr_{y \leftarrow \{0,1\}^q}[A(y) \in R_y^{W_y}] > c \cdot (\varepsilon^3 q^{-6} 2^{-k}) - \exp(q), \text{ for some constant } c > 0.$$

Proof of claim. Note how A operates. On input y, it follows a random compatible (with y) hash functions h_i out of node $N \in L_i$, for $1 \leq i < \alpha$, and then takes random h_i's (not necessarily compatible with y) when $\alpha \leq i < \alpha + \beta$. (For now, we can ignore failure to obtain compatible hash functions.)

Our algorithm A will find a valid witness for y if the following conditions happen.

1. Algorithm A reaches a good node $N \in L_\alpha$ such that $y \in F(N)$. By Claim B.8, this happens with probability at least $\varepsilon/(4(1+\gamma)^\alpha)$.
2. Algorithm A reaches a witness revealing child with at most one element in $G(N)$. Given that (1) occurs, by Claim B.9, this happens with probability at least $\varepsilon/4$.

   In this case, S^* will output (x_0, z_0) and (x_1, z_1), such that at least one (x_b, z_b) will be such that $x_b \in W_{C(z_b)}$ and $C(z_b) \in U(N) \setminus G(N) = F(N)$. Let $y' = C(z_b)$.
3. The string $y = y' = C(z_b)$. If this happens, then A will output $x_b \in R_y$, a valid witness for y. By Claim B.10, we have that

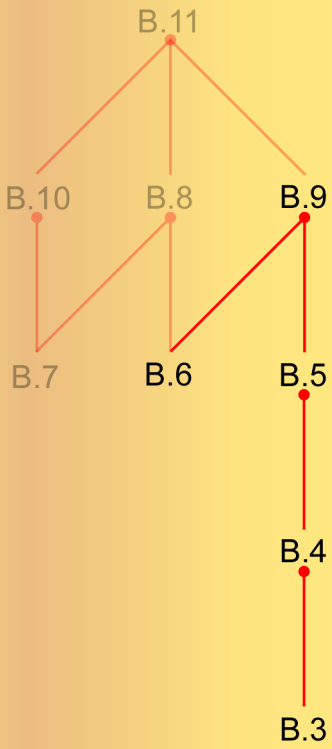
$$\Pr[y = y' | A \text{ reaches } N \wedge y' \in F(N)] \geq \frac{1}{|F(N)|} \left(\frac{1-\gamma}{1+\gamma}\right)^\alpha.$$

Combining all the probabilities, we have

$$\begin{aligned}\Pr_{y \leftarrow \{0,1\}^q}[A(y) \in W_y] &\geq \frac{\varepsilon}{4(1+\gamma)^\alpha} \cdot \frac{\varepsilon}{4} \cdot \frac{1}{|F(N)|} \left(\frac{1-\gamma}{1+\gamma}\right)^\alpha \\ &\geq \frac{1}{2^{\beta+k}} \cdot \frac{\varepsilon^2}{32} \cdot \left(\frac{1-\gamma}{(1+\gamma)^2}\right)^q.\end{aligned}$$

With settings of $\gamma = 1/q$ and $\beta = \log(1/\varepsilon) + 2\log(q) + 4\log(1/\gamma)) + 4$, we have the probability of finding a witness to be greater than $c \cdot (\varepsilon^3 q^{-6} 2^{-k})$, for some constant $c \geq 0$.

Finally, we need to account for the case when we fail to find compatible hash functions h_i out of node $N \in L_i$, for $1 \leq i < \alpha$. However, because our analysis has only focused on fully balanced y, and we repeat $8q$ times to find a compatible hash, the probability of failure is exponentially small. Therefore, the overall success probability is greater than $c \cdot (\varepsilon^3 q^{-6} 2^{-k}) - \exp(q)$. □


B.11
B.10
B.8
B.9
B.7
B.6
B.5
B.4
B.3

Claim B.3. *For every node $N \in L_i$, the set of unbalanced strings, $U(N) \setminus B(N) \leq 2/\gamma^2$.*

Proof of claim. Let $X \subseteq U(N)$ be a set of size 2^d, for some value of d. We also interpret X as a distribution that puts equal weights on each of its 2^d elements.

Let $\mathcal{H}_i$ be the set of hash functions after node N of the form $0^i 1\{0,1\}^{q-i-1}$. Observe that for every $x \neq x'$, $Pr_{h_i \leftarrow \mathcal{H}_i}[h_i(x) = h_i(x')] \leq 1/2$. Also, note that h_i requires exactly $q - i - 1$ bits to describe.

Computing the collision probabilities (using the notation $\mathcal{H}_i$ to denote a random hash function from that family), we get

$$\begin{aligned}
\mathrm{Col}((\mathcal{H}_i, \mathcal{H}_i(X))) &\leq \mathrm{Col}(\mathcal{H}_i)(\mathrm{Col}(X) + \Pr[\mathcal{H}_i(X) = \mathcal{H}_i(X') : X \neq X']) \\
&\leq \mathrm{Col}(\mathcal{H}_i) \cdot (1/2^d + 1/2) \\
&= 2^{-(q-i-1)}(1/2^d + 1/2), \text{ whereas} \\
\mathrm{Col}((\mathcal{H}_i, U_1)) &= \mathrm{Col}(\mathcal{H}_i) \cdot 1/2 \\
&= 2^{-(q-i-1)} \cdot (1/2).
\end{aligned}$$

Definition 2.5: A real-valued function f is a *concave function* on an interval I if

$$f\left(\frac{x+y}{2}\right) \geq \frac{f(x)+f(y)}{2}$$

for all $x, y \in I$. f is a *strictly concave function* on an interval I if

$$f\left(\frac{x+y}{2}\right) > \frac{f(x)+f(y)}{2}$$

for all $x, y \in I$, $x \neq y$.

Here is Jensen's inequality, which we state without proof.

THEOREM 2.5 (Jensen's inequality) *Suppose f is a continuous strictly concave function on the interval I. Suppose further that*

$$\sum_{i=1}^{n} a_i = 1$$

and $a_i > 0$, $1 \leq i \leq n$. Then

$$\sum_{i=1}^{n} a_i f(x_i) \leq f\left(\sum_{i=1}^{n} a_i x_i\right),$$

where $x_i \in I$, $1 \leq i \leq n$. Further, equality occurs if and only if $x_1 = \cdots = x_n$.

Therefore,

$$\begin{aligned}
\Delta((\mathcal{H}_i, \mathcal{H}_i(X)), (\mathcal{H}_i, U_1)) &= 1/2\,|(\mathcal{H}_i, \mathcal{H}_i(X)) - (\mathcal{H}_i, U_1)|_1 \\
&\leq 1/2 \cdot \sqrt{2^{q-i-1}} \sqrt{\mathrm{Col}((\mathcal{H}_i, \mathcal{H}_i(X))) - \mathrm{Col}((\mathcal{H}_i, U_1))} \\
&\leq 1/2 \sqrt{1/2^d} \\
&= 2^{-d/2-1}.
\end{aligned}$$

Setting $d = 2\log(1/\gamma)$, we get that $\Delta((\mathcal{H}_i, \mathcal{H}_i(X)), (\mathcal{H}_i, U_1)) \leq \gamma/2$. Next, assume for sake of contradiction that $U(N) \setminus B(N) > 2^{d+1} = 2/\gamma^2$. Then we will have a set $M \subseteq U(N) \setminus B(N)$ of size greater then 2^d with elements that are unbalanced in one direction (i.e. all $> 1/2 + \gamma$, or all $< 1/2 - \gamma$). But this contradicts the assumption that $\Delta((\mathcal{H}_i, \mathcal{H}_i(T)), (\mathcal{H}_i, U_1)) \leq \gamma/2$ (since $|T| > 2^d$). □

The next claim follows by a union bound on the unbalanced elements.

Claim B.4. *For every node $N \in L_i$, the set of strings that are not fully balanced, $G(N) = U(N) \setminus F(N) \leq 2i/\gamma^2$. In particular, for $\gamma = 1/q$, $|F(N)| \geq |U(N)|/2$ for $i \leq q - 4\log q$.*

Claim B.5. *For every node $N \in L_\alpha$, the fraction of children nodes $N_{\alpha+\beta}$ with greater than one element from $G(N)$ is at most $\varepsilon/4$.*

Proof of claim. Consider any fixed node $N \in L_\alpha$. The number of non-fully-balanced (aka bad) elements in that node is $G(N)$. Hence, the number of pairs of these bad elements is at most $|G(N)|^2$. Since for each $x \neq y \in U(N)$, $\Pr[h_i(x) = h_i(y)] \leq 1/2$ for all $\alpha \leq i < \alpha + \beta$, the fraction of children nodes $N' \in L_{\alpha+\beta}$ with greater than one element from $G(N)$ is at most $|G(N)|^2 / 2^\beta$.

Since $\beta = \log(1/\varepsilon) + 2\log(q) + 4\log(1/\gamma) + 4$, we can bound $|G(N)|^2 / 2^\beta$ as follows:

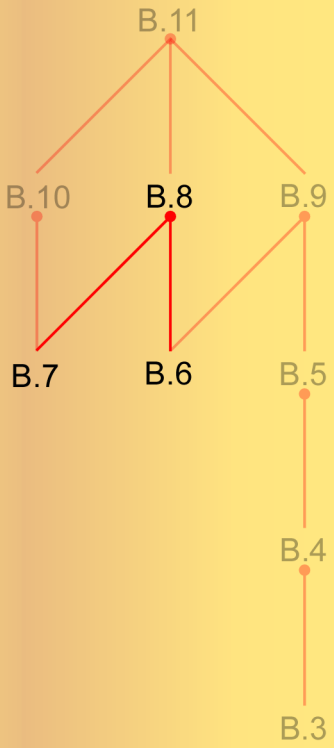
$$\begin{aligned} |G(N)|^2 \cdot 2^{-\beta} &\leq (2\alpha\gamma^{-2})^2 2^{-\beta} \\ &\leq 4q^2\gamma^{-4}2^{-\beta} \\ &< \varepsilon/4. \end{aligned}$$

The result follows. □

Claim B.9. *In any good node $N \in L_\alpha$, the fraction of nonbinding children of N at level $\alpha+\beta$ that has one or less image in $G(N)$ is at least $\varepsilon/4$.*

Proof of claim. The fraction of nonbinding children is greater than $\varepsilon/2$, and by Claim B.5, the fraction of children nodes of N with greater than one element from $G(N)$ is at most $\varepsilon/4$. □

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



A node $N \in L_{\alpha+\beta} = L_{q-k}$ is *witness revealing* if both of $S^*(N)$'s outputs, namely (x_0, z_0) and (x_1, z_1), satisfy $C(z_b) \in U(N)$ and $x_b \in W_{C(z_b)}$, for $b \in \{0,1\}$. A node $N \in L_\alpha$ is said to be *good* if greater than $\varepsilon/2$ of its children at level $q-k$ are witness revealing.

Claim B.6. *The fraction of good nodes at level α is at least $\varepsilon/2$.*

Proof of claim. By the assumption that

$$\Pr[x_0 \in W_{C(z_0)} \wedge x_1 \in W_{C(z_1)} : C = (S^*, R)(1^q, 1^k);\ ((x_0, z_0), (x_1, z_1)) = \mathsf{output}_{S^*}(S^*, R)] > \varepsilon,$$

we know that at least ε fraction of all the nodes at level $q-k$ are nonbinding. And, by a Markov bound, we have that $\varepsilon/2$ fraction of nodes at level α are good. □

Claim B.7. *For any fixed $N \in L_\alpha$ and $y' \in F(N)$, we have*

$$\frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1+\gamma)^\alpha} \cdot \frac{1}{|L_\alpha|} \leq \Pr[A \text{ reaches } N \wedge y = y'] \leq \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1-\gamma)^\alpha} \cdot \frac{1}{|L_\alpha|},$$

where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

Proof of claim. Let $N = (h_0, h_2, \ldots, h_{\alpha-1})$, and for $1 \leq j \leq \alpha$, define $N_j = (h_0, \ldots, h_{j-1})$. To get the upper bound,

$$\begin{aligned}
\Pr[A \text{ reaches } N \wedge y = y'] &= \Pr[y = y'] \cdot \Pr[A \text{ reaches } N] \\
&= 2^{-q} \prod_{j=0}^{\alpha-1} \frac{1}{\mathrm{Comp}(N_j, y)} \\
&\leq 2^{-q} \prod_{j=0}^{\alpha-1} \frac{2}{1-\gamma} \cdot \frac{1}{|\mathcal{H}_j|} \\
&= \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1-\gamma)^\alpha} \cdot \frac{1}{|L_\alpha|}.
\end{aligned}$$

To get the lower bound, we use very similar techniques.

$$\begin{aligned}
\Pr[A \text{ reaches } N \wedge y = y'] &= 2^{-q} \prod_{j=0}^{\alpha-1} \frac{1}{\mathrm{Comp}(N_j, y)} \\
&\geq 2^{-q} \prod_{j=0}^{\alpha-1} \frac{2}{1+\gamma} \cdot \frac{1}{|\mathcal{H}_j|} \\
&= \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1+\gamma)^{\alpha}} \cdot \frac{1}{|L_\alpha|}.
\end{aligned}$$

Our result follows. □

Claim B.8.

$$\Pr[\textit{The node } N \textit{ reached by } A \textit{ is good } \wedge\, y \in F(N)] \geq \frac{\varepsilon}{4(1+\gamma)^{\alpha}}.$$

where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

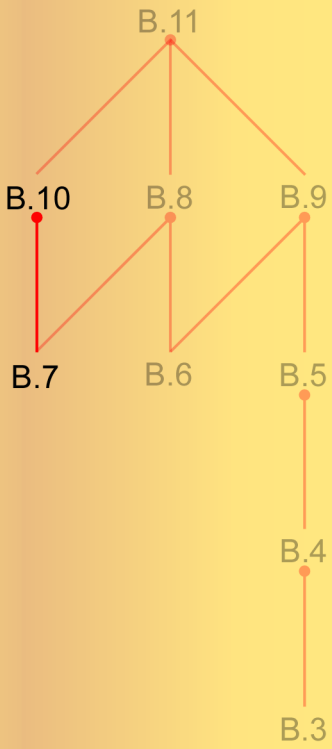
Proof of claim. Let $N \in L_\alpha$ be any good node at level α. Then,

$$\begin{aligned}
\Pr[A \text{ reaches } N \wedge y \in F(N)] &= \sum_{y' \in F(N)} \Pr[A \text{ reaches } N \wedge y = y'] \\
&\geq \sum_{y' \in F(N)} \frac{1}{|L_\alpha|} \cdot \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1+\gamma)^{\alpha}} \\
&= \frac{|F(N)|}{2^{q-\alpha}} \cdot \frac{1}{|L_\alpha|} \cdot \frac{1}{(1+\gamma)^{\alpha}} \\
&= \frac{|F(N)|}{|U(N)|} \cdot \frac{1}{|L_\alpha|} \cdot \frac{1}{(1+\gamma)^{\alpha}} \\
&\geq \frac{1}{2} \cdot \frac{1}{|L_\alpha|} \cdot \frac{1}{(1+\gamma)^{\alpha}},
\end{aligned}$$

with the last inequality following from the fact that $|F(N)| / |U(N)| \geq 1/2$, noting $\alpha \leq q - 3\log q$ (refer to Claim B.4).

There are $|L_\alpha|$ nodes at level α, and at least $\varepsilon/2$ fraction of them are good. Hence, we multiply the above probability by $(\varepsilon/2)\,|L_\alpha|$ to get our stated result. □

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



Claim B.10. *For any fixed $N \in L_\alpha$ and $y' \in F(N)$, we have*

$$\Pr[y = y' | A \text{ reaches } N \wedge y \in F(N)] \geq \frac{1}{|F(N)|} \left(\frac{1-\gamma}{1+\gamma}\right)^\alpha,$$

where the probability is taken over $y \in \{0,1\}^q$ and the random coins of A.

Proof of claim. For any fixed $N \in L_\alpha$ and $y' \in F(N)$,

$$\Pr[y = y' | A \text{ reaches } N \wedge y \in F(N)] = \frac{\Pr[A \text{ reaches } N \wedge y = y']}{\Pr[A \text{ reaches } N \wedge y \in F(N)]}.$$

For the numerator, by Claim B.7,

$$\Pr[A \text{ reaches } N \wedge y = y'] \geq \frac{1}{|L_\alpha|} \cdot \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1+\gamma)^\alpha}.$$

For the denominator, also using Claim B.7,

$$\begin{aligned}
\Pr[A \text{ reaches } N \wedge y \in F(N)] &= \sum_{y' \in F(N)} \Pr[A \text{ reaches } N \wedge y = y'] \\
&\leq \sum_{y' \in F(N)} \frac{1}{|L_\alpha|} \cdot \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1-\gamma)^\alpha} \\
&= |F(N)| \cdot \frac{1}{|L_\alpha|} \cdot \frac{1}{2^{q-\alpha}} \cdot \frac{1}{(1-\gamma)^\alpha}.
\end{aligned}$$

Combining the two, we have our result. □

4.1 Hashing and Randomness Extraction

Entropy. The *entropy* of a random variable X is $H(X) = \mathrm{E}_{x \xleftarrow{R} X}[\log(1/\Pr[X = x])]$, where here and throughout the paper all logarithms are to base 2. Intuitively, $H(X)$ measures the amount of randomness in X *on average* (in bits). The *min-entropy* of X is $H_\infty(X) = \min_x[\log(1/\Pr[X = x])]$; this is a "worst-case" measure of randomness. In general $H_\infty(X) \leq H(X)$, but if X is flat (i.e. uniform on its support), then $H(X) = H_\infty(X) = \log|\mathrm{Supp}(X)|$.

A family of hash functions $\mathcal{H}_{a,b} = \{h : \{0,1\}^a \to \{0,1\}^b\}$ is *pairwise independent* if for any two $x \neq x' \in \{0,1\}^a$ and any two $y, y' \in \{0,1\}^b$, when we randomly choose $h \leftarrow \mathcal{H}_{a,b}$, we have: $\Pr[h(x) = y \wedge h(x') = y'] = \frac{1}{2^{2b}}$. We define $\ell(a,b)$ to be the number of bits required to describe an element of the hash function family $\mathcal{H}_{a,b}$; that is, $\ell(a,b) = \max\{a,b\} + b$. We will use the following strong extractor property of $\mathcal{H}_{a,b}$.

Lemma 4.1 (Leftover Hash Lemma [BBR88, ILL89])**.** *Let $\mathcal{H}_{a,b}$ be a pairwise independent family of hash functions mapping $\{0,1\}^a$ to $\{0,1\}^b$. Let Z be a random variable taking values in $\{0,1\}^a$ such that $\mathcal{H}_\infty(Z) \geq b + 2\log(1/\varepsilon)$. Then the following distribution has statistical difference at most ε from the uniform distribution on $\mathcal{H}_{a,b} \times \{0,1\}^b$: Choose $h \leftarrow \mathcal{H}_{a,b}$ and $x \leftarrow Z$ and output $(h, h(x))$.*

2.2 1-out-of-2-Binding Commitments

Definition 2.4. A *2-phase commitment scheme* (S, R), with security parameter n and message length $k = k(n)$, consists of four interactive protocols: (S_c^1, R_c^1) the first commitment stage, (S_r^1, R_r^1) the first reveal stage, (S_c^2, R_c^2) the second commitment stage, and (S_r^2, R_r^2) the second reveal stage. For us, both reveal phases will always be noninteractive, consisting of a single message from the sender to the receiver. Throughout, both parties receive the security parameter 1^n as input.

1. In the first commitment stage, S_c^1 receives a private input $\sigma^{(1)} \in \{0,1\}^k$ and a sequence of coin tosses r_S. At the end, S_c^1 and R_c^1 receive as common output a commitment $z^{(1)}$. (Without loss of generality, we can assume that $z^{(1)}$ is the transcript of the first commitment stage.)

2. In the first reveal stage, S_r^1 and R_r^1 receive as common input the commitment $z^{(1)}$ and a string $\sigma^{(1)} \in \{0,1\}^k$ and S_r^1 receives as private input r_S. At the end, S_r^1 and R_r^1 receive a common output τ. (Without loss of generality, we can assume that τ is the transcript of the first commitment stage and the first reveal stage and includes R_r^1's decision to accept or reject.)

3. In the second commitment stage, S_c^2 and R_c^2 both receive the common input $\tau \in \{0,1\}^*$, and S_c^2 receives a private input $\sigma^{(2)} \in \{0,1\}^k$ and the coin tosses r_S. S_c^2 and R_c^2 receive as common output a commitment $z^{(2)}$. (Without loss of generality, we can assume that $z^{(2)}$ is the concatenation of τ and the transcript of the second commitment stage.)

4. In the second reveal stage, S_r^2 and R_r^2 receive as common input the commitment $z^{(2)}$ and a string $\sigma^{(2)} \in \{0,1\}^k$, and S_r^2 receives as private input r_S. At the end, R_r^2 accepts or rejects.

- $S = (S^1, S^2) = ((S_c^1, S_r^1), (S_c^2, S_r^2))$ and $R = (R^1, R^2) = ((R_c^1, R_r^1), (R_c^2, R_r^2))$ are computable in probabilistic polynomial time.

- We say that (S, R) is *public-coin* if it is public-coin for R.

Note that instead of providing S with decommitment values as private outputs of the commitment phases, we simply provide it with the same coin tosses throughout (so it can recompute any private state from the transcripts of the previous phases).

As for standard commitment schemes, we define the security of the sender in terms of a hiding property. Loosely speaking, the hiding property for a 2-phase commitment scheme says that *both* commitment phases are hiding. Note that since the phases are run sequentially, the hiding property for the second commitment stage is required to hold even given the receiver's view of the first stage.

Definition 2.5 (hiding)**.** 2-phase commitment scheme (S, R), with security parameter n and message length $k = k(n)$, is *statistically hiding* if for all adversarial receiver R^*,

1. The views of R^* when interacting with the sender in the first phase on any two messages are statistically indistinguishable. That is, for all $\sigma^{(1)}, \widetilde{\sigma}^{(1)} \in \{0,1\}^k$,

$$\left\{\mathsf{view}_{R^*}(S_c^1(\sigma^{(1)}), R^*)(1^n)\right\}_{n\in\mathbb{N}} \approx_s \left\{\mathsf{view}_{R^*}(S_c^1(\widetilde{\sigma}^{(1)}), R^*)(1^n)\right\}_{n\in\mathbb{N}}.$$

2. The views of R^* when interacting with the sender in the second phase are statistically indistinguishable no matter what the sender committed to in the first phase. That is, for all $\sigma^{(1)}, \sigma^{(2)}, \widetilde{\sigma}^{(2)} \in \{0,1\}^k$,

$$\left\{\mathsf{view}_{R^*}(S_c^2(\sigma^{(2)}), R^*)(\Lambda, 1^n)\right\}_{n\in\mathbb{N}} \approx_s \left\{\mathsf{view}_{R^*}(S_c^2(\widetilde{\sigma}^{(2)}), R^*)(\Lambda, 1^n)\right\}_{n\in\mathbb{N}},$$

where $\Lambda = \mathsf{transcript}(S^1(\sigma^{(1)}), R^*)(1^n)$.

We stress that the second condition of the above hiding definition (Definition 2.5) requires that the view of receiver in the second phase be indistinguishable for any two messages even given the transcript of the first phase, $\Lambda = \mathsf{transcript}(S^1(\sigma^{(1)}), R^*)(1^n)$.

Loosely speaking, the binding property says that *at least* one of the two commitment phases is (computationally) binding. In other words, for every polynomial-time sender S^*, there is at most one "bad" phase $j \in \{1, 2\}$ such that given a commitment $z^{(j)}$, S^* can open $z^{(j)}$ successfully both as $\sigma^{(1)}$ and $\widetilde{\sigma}^{(1)} \neq \sigma$ with nonnegligible probability. Actually, we allow this bad phase to be determined dynamically by S^*. Moreover, we require that the second phase be *statistically* binding if the sender breaks the first phase. Our construction achieves this stronger property, and using it simplifies some of our proofs.

Definition 2.6 (1-out-of-2-binding)**.** 2-phase commitment scheme (S, R), with security parameter n and message length $k = k(n)$, is *computationally* $\binom{2}{1}$*-binding* if there exist a set $\mathcal{B}$ of first phase transcripts and a negligible function ε such that:

1. For every (even unbounded) sender S^*, the first-phase transcripts in $\mathcal{B}$ make the second phase statistically binding, i.e. $\forall S^*, \forall \tau \in \mathcal{B}$, with probability at least $1-\varepsilon(n)$ over $z^{(2)} = (S^*, R_c^2)(\tau)$, there is at most one value $\sigma^{(2)} \in \{0,1\}^k$ such that $\texttt{output}(S^*, R_r^2)(z^{(2)}, \sigma^{(2)}) = \texttt{accept}$.

2. $\forall$ nonuniform PPT S^*,[7] S^* succeeds in the following game with probability at most $\varepsilon(n)$ for all sufficiently large n:

   (a) S^* and R_c^1 interact and output a first-phase commitment $z^{(1)}$.

   (b) S^* outputs two full transcripts τ and $\widetilde{\tau}$ of *both* phases with the following three properties:

   - Transcripts τ and $\widetilde{\tau}$ both start with prefix $z^{(1)}$.
   - The transcript τ contains a successful opening of $z^{(1)}$ to the value $\sigma^{(1)} \in \{0,1\}^k$ using a first-phase transcript not in $\mathcal{B}$, and R_r^1 and R_r^2 both accept in τ.
   - The transcript $\widetilde{\tau}$ contains a successful opening of $z^{(1)}$ to the value $\widetilde{\sigma}^{(1)} \in \{0,1\}^k$ using a first-phase transcript not in $\mathcal{B}$, and R_r^1 and R_r^2 both accept in $\widetilde{\tau}$.

   (c) S^* succeeds if all of the above conditions hold and $\sigma^{(1)} \neq \widetilde{\sigma}^{(1)}$.

3 Our Results

Our main theorem, Theorem 1.1, is established via the following theorems.

Theorem 3.1. *If one-way functions exist, then on security parameter n, we can construct in time* $\mathrm{poly}(n)$ *a collection of public-coin 2-phase commitment schemes* $\mathrm{Com}_1, \cdots, \mathrm{Com}_m$ *for* $m = \mathrm{poly}(n)$ *such that:*

- *There exists an index $i \in [m]$ such that scheme* Com_i *is statistically hiding.*
- *For every index $j \in [m]$, scheme* Com_j *is computationally* $\binom{2}{1}$*-binding.*

Theorem 3.2. *Assume that on security parameter n, we can construct in time* $\mathrm{poly}(n)$ *a collection of public-coin 2-phase commitment schemes* $\mathrm{Com}_1, \cdots, \mathrm{Com}_m$ *for* $m = \mathrm{poly}(n)$ *such that:*

- *There exists an index $i \in [m]$ such that scheme* Com_i *is statistically hiding.*
- *For every index $j \in [m]$, scheme* Com_j *is* $\binom{2}{1}$*-computationally binding.*

Then, every language in **NP** *has a public-coin statistical zero-knowledge argument system.*

Protocol 4.6. 2-Phase Commitment Scheme (S, R) based on $f\colon \{0,1\}^n \to \{0,1\}^n$.

Parameters: Integers $t \in \{1,2,\ldots,n\}$, $k \in \{1,2,\ldots,n\}$, $\Delta_1 \in \{0,1,\ldots,t\}$, and $\Delta_2 \in \{0,1,\ldots,n-t\}$.

Sender's private input: String $x \in \{0,1\}^n$. (Note that this is not the value to which the sender is committing, but is rather part of its coin tosses, which will be chosen uniformly at random by S unless otherwise specified.)

First phase commit:

1. S_c^1 sets $y = f(x)$.
2. Let $\mathcal{H}_1 = \{h_1\colon \{0,1\}^n \to \{0,1\}^{t-\Delta_1}\}$ be a family of pairwise independent hash functions. S_c^1 chooses a random hash $h_1 \leftarrow \mathcal{H}_1$, and computes $v = (h_1, h_1(y)) \in \{0,1\}^q$.
3. (S_c^1, R_c^1) run Interactive Hashing Scheme (Protocol 4.5) $(S_{\mathrm{IH}}(v), R_{\mathrm{IH}})(1^q, 1^k)$, with S_c^1 and R_c^1 acting as S_{IH} and R_{IH} respectively.
   Let circuit $C^{(1)}\colon \{0,1\}^k \to \{0,1\}^q$ be the common output and $d^{(1)} \in \{0,1\}^k$ be S_{IH}'s private output in $(S_{\mathrm{IH}}(v), R_{\mathrm{IH}})(1^q, 1^k)$.

First phase sender's private output: String $d^{(1)} \in \{0,1\}^k$.

(to commit to a string $z^{(1)}$, the sender sends $d^{(1)} \oplus z^{(1)}$ to the receiver)

First phase reveal:

S_r^1 sends the tuple $\gamma^{(1)} = (d^{(1)}, y, h_1)$.

Receiver R_r^1 accepts if and only if $C^{(1)}(d^{(1)}) = (h_1, h_1(y))$.

Protocol 4.6. 2-Phase Commitment Scheme (S, R) based on $f\colon \{0,1\}^n \to \{0,1\}^n$.

Second phase commit:

Second phase common input: First-phase transcript $\tau = \mathsf{transcript}(S^1(x), R^1)$, which in particular includes the string y.

1. Let $\mathcal{H}_2 = \{h_2\colon \{0,1\}^n \to \{0,1\}^{n-t-\Delta_2}\}$ be a family of pairwise independent hash functions. S_c^2 chooses a random hash $h_2 \leftarrow \mathcal{H}_2$, and computes $w = (h_2, h_2(x)) \in \{0,1\}^q$.
2. (S_c^2, R_c^2) run Interactive Hashing Scheme (Protocol 4.5) $(S_{\mathrm{IH}}(w), R_{\mathrm{IH}})(1^q, 1^k)$, with S_c^2 and R_c^2 acting as S_{IH} and R_{IH} respectively.
   Let circuit $C^{(2)}\colon \{0,1\}^k \to \{0,1\}^q$ be the common output and $d^{(2)} \in \{0,1\}^k$ be S_{IH}'s private output in $(S_{\mathrm{IH}}(v), R_{\mathrm{IH}})(1^q, 1^k)$.

Second phase sender's private output: String $d^{(2)} \in \{0,1\}^k$.

(to commit to a string $z^{(2)}$, the sender sends $d^{(2)} \oplus z^{(2)}$ to the receiver)

Second phase reveal:

S_r^2 sends the tuple $\gamma^{(2)} = (d^{(2)}, x, h_2)$.

Receiver R_r^2 accepts if and only if $f(x) = y$ and $C^{(2)}(d^{(2)}) = (h_2, h_2(x))$.

Lemma 4.7 (statistical hiding)**.** *If f is a regular function with $H(f(U_n)) \in (t_0 - 1, t_0]$, then Protocol 4.6, with setting of parameters $t = t_0$, $k \leq q(n)$, and $\Delta_1 = \Delta_2 = \omega(\log n)$, is statistically hiding in the sense of Definition 2.5.*

Proof Sketch. For every $y \in \text{Support}(f(U_n))$, we have $p(y) = \Pr[f(U_n) = y] \in [2^{-t_0}; 2^{-t_0+1})$

We denote the distribution $f(U_n)$ by Y. The flat source Y has min-entropy at least $t_0 - 1$. By the Leftover Hash Lemma (Lemma 4.1), the distribution $Z = (H_1, H_1(Y))$ is $2^{-\Omega(\Delta_1)}$-close to the uniform distribution $(H_1, U_{t-\Delta_1})$. By the hiding property of interactive hashing, the first commitment phase is $2^{-\Omega(\Delta_1)}$-statistically hiding.

Let τ be the transcript of the first phase and y the string sent in the first reveal phase. Conditioned on τ, the string x comes from the uniform distribution X over $f^{-1}(y)$ and X is a flat souce with min-entropy at least $n - t_0$. By the Leftover Hash Lemma (Lemma 4.1), the distribution $W = (H_2, H_2(X))$ is $2^{-\Omega(\Delta_2)}$-close to the uniform distribution $(H_2, U_{n-t-\Delta_2})$. By the hiding property of interactive hashing, the second commitment phase is $2^{-\Omega(\Delta_2)}$-statistically hiding. □

Lemma 4.8. *If f is a $s(n)$-secure one-way function (not necessarily regular), then for any value of $t \in \{1, \cdots, n\}$, Protocol 4.6, with setting of parameters $k = O(\log n)$, $\Delta_1 = \Delta_2 \leq (\log(s(n)))/4$, is 1-out-of-2 computationally binding in the sense of Definition 2.6.*

The proposition will be proved in two steps. For every $t \in \{1, \cdots, n\}$, we define the set of "light" strings $L_t = \{y \in \{0,1\}^n : \Pr_{U_n}[f(U_n) = y] \leq 2^{-t-\Delta_3}\}$, for a parameter Δ_3 that we will set at the end of the proof. We define $\mathcal{B}$ to be the set of transcripts where the sender reveals $y \in L_t$. We will first show that if the first commitment transcript is in $\mathcal{B}$, then the second phase will be statistically binding. We will then prove that the first phase is computationally binding, i.e. if there exists an adversary that can break the binding property for the first phase, then there exists an adversary that can invert f with nonnegligible success probability.

Claim 4.9. *For the binding set $\mathcal{B}$ defined above, Condition 1 of Definition 2.6 is satisfied with* $\varepsilon(n) = \mathrm{poly}(n) \cdot 2^{-\Omega(\Delta_3 - \Delta_2)}$.

Proof of Claim 4.9. Let y be the string sent in the first reveal phase. Let $T = \{(h_2, h_2(x)) : h_2 \in \mathcal{H}_{n,n-t-\Delta_2}, x \in f^{-1}(y)\}$ and $\mu(T)$ denote the density of the subset T. Since h_2 maps $\{0,1\}^n$ to $\{0,1\}^{n-t-\Delta_2}$, we have

$$\mu(T) \leq |f^{-1}(y)| \cdot \frac{1}{2^{n-t-\Delta_2}} \leq (2^n \cdot 2^{-t-\Delta_3}) \cdot \frac{1}{2^{n-t-\Delta_2}} = 2^{(\Delta_2 - \Delta_3)}$$

By the binding property of the second execution of the interactive hashing protocol for static sets, we have

$$\Pr\left[(w_0, w_1) = \mathsf{output}(S^{\star}_{\mathrm{IH}}, R_{\mathrm{IH}}) \text{ satisfies } w_0 \in T \wedge w_1 \in T\right] < 2^{-\Omega(\Delta_3 - \Delta_2)} \cdot \mathrm{poly}(q).$$

□

Claim 4.10. *For the binding set $\mathcal{B}$ defined above, if there exists a PPT $S^\star$ that succeeds with nonnegligible success probability ε in the game in Condition 2 of Definition 2.6, then there exists a PPT T that can invert f with success probability at least*

$$\varepsilon^{O(1)} \cdot 1/\mathrm{poly}(n) \cdot 2^{-(k+\Delta_1+\Delta_3)}.$$

Proof of Claim 4.10. We define the relation $\mathcal{R}$:

$$\mathcal{R} = \{((h_1, w), (y, x)) : \ w = h_1(y), \ y = f(x), \ y \notin L_t\}$$

Let $\mathcal{R}_v = \{(y, x) : \mathcal{R}(v, (y, x)) = 1\}$. Suppose we have a PPT $S^\star$ with success probability greater than ε in the game of Definition 2.6. Then we have a PPT S^*_{IH} in the interactive hashing protocol such that

$$\begin{aligned}\Pr[&\mathsf{output}_{S^*_{\mathrm{IH}}}(S^*_{\mathrm{IH}}, R_{\mathrm{IH}}) = ((v_0, v_1), (y, x), (y', x')) \text{ such that} \\ &(v_0, v_1) = \mathsf{output}(S^*_{\mathrm{IH}}, R_{\mathrm{IH}}), (y, x) \in \mathcal{R}_{v_0}, (y', x') \in \mathcal{R}_{v_1}] \ \ \geq \ \ \varepsilon\end{aligned}$$

By the binding property of the interactive hashing protocol, there exists a PPT A such that

$$\Pr_{v \leftarrow H_1 \times U_{t-\Delta_1}}[A(v, 1^{\ell_1}, \varepsilon) \in \mathcal{R}_v] > 2^{-k} \cdot \left(\frac{\varepsilon}{\ell_1}\right)^c$$

Consider the PPT T that on input y picks a hash function h_1 uniformly from $\mathcal{H}_{n,t-\Delta_1}$, runs A on input $v = (h_1, h_1(y))$ and outputs the second component of $A(v)$. Assume without loss of generality that A is deterministic. Then:

$$
\begin{aligned}
&\Pr_{U_n, r_B}[T(f(U_n)) \in f^{-1}(f(U_n))] \\
&= \Pr_{H_1, U_n}[A(H_1, H_1(f(U_n)))_2 \in f^{-1}(f(U_n))] \\
&\geq \sum_{(h_1,w)\in\mathcal{H}_{n,t-\Delta_1}\times\{0,1\}^{t-\Delta_1}} \Pr_{U_n, H_1, r_A}[(H_1, H_1(f(U_n))) = (h_1, w) \wedge A(h_1, w) \in \mathcal{R}_{(h_1,w)}] \\
&= \frac{1}{|\mathcal{H}_{n,t-\Delta_1}|} \sum_{(h_1,w)\ s.t.\ A(h_1,w)\in\mathcal{R}_{(h_1,w)}} \Pr[h_1(f(U_n)) = w] \\
&\geq \frac{1}{|\mathcal{H}_{n,t-\Delta_1}|} \sum_{(h_1,w)\ s.t.\ A(h_1,w)\in\mathcal{R}_{(h_1,w)}} \Pr[f(U_n) = A(h_1, w)_1] \\
&\geq \frac{1}{|\mathcal{H}_{n,t-\Delta_1}|} \cdot \left(|\mathcal{H}_{n,t-\Delta_1}| \cdot 2^{t-\Delta_1} \cdot 2^{-k} \cdot \left(\frac{\varepsilon}{\ell_1}\right)^c\right) \cdot 2^{-t-\Delta_3} \\
&= \left(\frac{\varepsilon}{\ell_1}\right)^c \cdot 2^{-(k+\Delta_1+\Delta_3)}
\end{aligned}
$$

The first inequality comes from considering fixed values of h_1 and w and restricting the success probability of A to the case where $y \notin L_t$. The third inequality comes from considering only values of (h_1, w) such that $w = h_1(y)$ for some $y \notin L_t$. Such strings y have mass at least $2^{-t-\Delta_3}$. $\square$

The lemma follows from the above two claims by setting $\Delta_3 = \Delta_2 + (\log s(n))/4 \leq (\log s(n))/4$. With this, Claim 4.9 shows that Condition 1 in Definition 2.6 is satisfied with $\varepsilon(n) = \text{poly}(n) \cdot 2^{-\Omega(\log s(n))} = \text{neg}(n)$ because $s(n) = n^{\omega(1)}$. Condition 2 of Definition 2.6 is satisfied with negligible probability $\varepsilon(n)$ because otherwise f can be inverted with probability

$$\begin{aligned}\varepsilon^{O(1)} \cdot 1/\text{poly}(n) \cdot 2^{-(k+\Delta_1+\Delta_3)} &\geq \varepsilon^{O(1)} \cdot 1/\text{poly}(n) \cdot 2^{-(O(\log n)+(3/4)\cdot(\log s(n)))} \\ &= \varepsilon^{O(1)} \cdot 1/\text{poly}(n) \cdot s(n)^{-3/4},\end{aligned}$$

which is greater than $1/s(n)$ if ε is nonnegligible.

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan



Statistical Zero-Knowledge Arguments for **NP** from Any One-Way Function*

Minh-Huyen Nguyen Shien Jin Ong Salil Vadhan