Lecture 19: Zero-Knowledge Proofs II

Instructor: Omkant Pandey

Spring 2017 (CSE 594)

Recall: Zero Knowledge

Definition (Zero Knowledge)

An interactive proof (P, V) for a language L with witness relation R is said to be *zero knowledge* if for every non-uniform PPT adversary V^*, there exists a PPT simulator S s.t. for every non-uniform PPT distinguisher D, there exists a negligible function $\nu(\cdot)$ s.t. for every $x \in L$, $w \in R(x)$, $z \in \{0,1\}^*$, D distinguishes between the following distributions with probability at most $\nu(|x|)$:

- $\left\{\mathsf{View}_V^*[P(x,w) \leftrightarrow V^*(x,z)]\right\}$
- $\left\{S(1^n, x, z)\right\}$

- If the distributions are statistically close, then we call it *statistical zero knowledge*
- If the distributions are identical, then we call it *perfect zero knowledge*

Recall: Interactive Proof for Graph Isomorphism

Common Input: $x = (G_0, G_1)$

P**'s witness:** π **s.t.** $G_1 = \pi(G_0)$

Protocol (P, V)**:** Repeat the following procedure n times using fresh randomness

$P \to V$: Prover chooses a random permutation $\sigma \in \Pi_n$, computes $H = \sigma(G_0)$ and sends H

$V \to P$: V chooses a random bit $b \in \{0, 1\}$ and sends it to P

$P \to V$: If $b = 0$, P sends σ. Otherwise, it sends $\phi = \sigma \cdot \pi^{-1}$

$V(x, b, \phi)$: V outputs 1 iff $H = \phi(G_b)$

(P, V) is Perfect Zero Knowledge: Strategy

- Will prove that a single iteration of (P, V) is perfect zero knowledge
- For the full protocol, use the following (read proof online):

Theorem

Sequential repetition of any ZK protocol is also ZK

- To prove that a single iteration of (P, V) is perfect ZK, we need to do the following:
  - Construct a Simulator S for every PPT V^*
  - Prove that expected runtime of S is polynomial
  - Prove that the output distribution of S is correct (i.e., indistinguishable from real execution)

(P, V) is Perfect Zero Knowledge: Simulator

Simulator $S(x, z)$**:**

- Choose random $b' \stackrel{\$}{\leftarrow} \{0, 1\}$, $\sigma \stackrel{\$}{\leftarrow} \Pi_n$
- Compute $H = \sigma(G_{b'})$
- Emulate execution of $V^*(x, z)$ by feeding it H. Let b denote its response
- If $b = b'$, then feed σ to V^* and output its view. Otherwise, restart the above procedure

Correctness of Simulation

Lemma

In the execution of $S(x, z)$,

- *H is identically distributed to $\sigma(G_0)$, and*
- $\Pr[b = b'] = \frac{1}{2}$

Proof:

- Since G_0 is isomorphic to G_1, for a random $\sigma \xleftarrow{\$} \Pi_n$, $\sigma(G_0)$ and $\sigma(G_1)$ are identically distributed
- That is, distribution of H is *independent* of b'
- Therefore, H has the same distribution as $\sigma(G_0)$
- Now, since V^* only takes H as input, its output b' is also independent of b'
- Since b' is chosen at random, $\Pr[b' = b] = \frac{1}{2}$

Correctness of Simulation (contd.)

Runtime of S:

- From Lemma 3: S has probability $\frac{1}{2}$ of succeeding in each trial
- Therefore, in expectation, S stops after 2 trials
- Each trial takes polynomial time, so run time of S is expected polynomial

Indistinguishability of Simulated View:

- From Lemma 3: H has the same distribution as $\sigma(G_0)$
- If we could always output σ, then output distribution of S would be same as in real execution
- S, however, only outputs H and σ if $b' = b$
- But since H is independent of b', this does not change the output distribution

Reflections on Zero Knowledge Proofs

Paradox?

- Protocol execution convinces V of the validity of x
- Yet, V could have generated the protocol transcript on its own

To understand why there is no paradox, consider the following story:

- Alice and Bob run (P, V) on input (G_0, G_1) where Alice acts as P and Bob as V
- Now, Bob goes to Eve: "G_0 and G_1 are isomorphic"
- Eve: "Oh really?"
- Bob: "Yes, you can see this accepting transcript"
- Eve: "Are you kidding me? Anyone can come up with this transcript without knowing the isomorphism!"
- Bob: "But I computed this transcript by talking to Alice who answered my challenge correctly every time!"

Reflections on Zero Knowledge Proofs (contd.)

Moral of the story:

- Bob participated in a "live" conversation with Alice, and was convinced by *how* the transcript was generated
- But to Eve, who did not see the live conversation, there is no way to tell whether the transcript is from real execution or produced by simulator

Zero-Knowledge Proofs for **NP**

Theorem

If one-way permutations exist, then every language in **NP** *has a zero-knowledge interactive proof.*

- The assumption can in fact be relaxed to just one-way functions
- Think: How to prove the theorem?
- Construct ZK proof for every **NP** language?
- Not efficient!

Zero-Knowledge Proofs for **NP** (contd.)

Proof Strategy:

Step 1: Construct a ZK proof for an **NP**-complete language. We will consider *Graph 3-Coloring*: language of all graphs whose vertices can be colored using only three colors s.t. no two connected vertices have the same color

Step 2: To construct ZK proof for any **NP** language L, do the following:

- Given instance x and witness w, P and V reduce x into an instance x' of Graph 3-coloring using Cook's (deterministic) reduction
- P also applies the reduction to witness w to obtain witness w' for x'
- Now, P and V can run the ZK proof from Step 1 on common input x'

Physical ZK Proof for Graph 3-Coloring

- Consider graph $G = (V, E)$. Let C be a 3-coloring of V given to P
- P picks a random permutation π over colors $\{1, 2, 3\}$ and colors G according to $\pi(C)$. It hides each vertex in V inside a locked box
- V picks a random edge (u, v) in E
- P opens the boxes corresponding to u, v. V accepts if u and v have different colors, and rejects otherwise
- The above process is repeated $n|E|$ times
- **Intuition for Soundness:** In each iteration, cheating prover is caught with probability $\frac{1}{|E|}$
- **Intuition for ZK:** In each iteration, V only sees something it knew before – two random (but different) colors

Towards ZK Proof for Graph 3-Coloring

- To "digitze" the above proof, we need to implement locked boxes
- Need two properties from digital locked boxes:
  - **Hiding**: V should not be able to see the content inside a locked box
  - **Binding**: P should not be able to modify the content inside a box once its locked

Commitment Schemes

- Digital analogue of locked boxes
- Two phases:
  Commit phase: Sender locks a value v inside a box
  Open phase: Sender unlocks the box and reveals v
- Can be implemented using interactive protocols, but we will consider non-interactive case. Both commit and reveal phases will consist of single messages

Commitment Schemes: Definition

Definition (Commitment)

A randomized polynomial-time algorithm Com is called a *commitment scheme* for n-bit strings if it satisfies the following properties:

- **Binding:** For all $v_0, v_1 \in \{0,1\}^n$ and $r_0, r_1 \in \{0,1\}^n$, it holds that $\mathsf{Com}(v_0; r_0) \neq \mathsf{Com}(v_1; r_1)$
- **Hiding:** For every non-uniform PPT distinguisher D, there exists a negligible function $\nu(\cdot)$ s.t. for every $v_0, v_1 \in \{0,1\}^n$, D distinguishes between the following distributions with probability at most $\nu(n)$
  - $\left\{ r \xleftarrow{\$} \{0,1\}^n \;:\; \mathsf{Com}(v_0; r) \right\}$
  - $\left\{ r \xleftarrow{\$} \{0,1\}^n \;:\; \mathsf{Com}(v_1; r) \right\}$

Commitment Schemes: Remarks

- The previous definition only guarantees hiding for one commitment
- **Multi-value Hiding:** Just like encryption, we can define multi-value hiding property for commitment schemes
- Using hybrid argument (as for public-key encryption), we can prove that any commitment scheme satisfies multi-value hiding
- **Corollary:** One-bit commitment implies string commitment

Construction of Bit Commitments

Construction: Let f be a OWP, h be the hard core predicate for f

Commit phase: Sender computes $\mathsf{Com}(b; r) = f(r), b \oplus h(r)$. Let C denote the commitment.

Open phase: Sender reveals (b, r). Receiver accepts if $C = (f(r), b \oplus h(r))$, and rejects otherwise

Security:

- Binding follows from construction since f is a permutation
- Hiding follows in the same manner as IND-CPA security of public-key encryption scheme constructed from trapdoor permutations

ZK Proof for Graph 3-Coloring

Common Input: $G = (V, E)$**, where** $|V| = n$

P**'s witness: Colors** $\mathsf{color}_1, \ldots, \mathsf{color}_n \in \{1, 2, 3\}$

Protocol (P, V)**:** Repeat the following procedure $n|E|$ times *using fresh randomness*

$P \to V$: P chooses a random permutation π over $\{1, 2, 3\}$. For every $i \in [n]$, it computes $C_i = \mathsf{Com}(\widetilde{\mathsf{color}}_i)$ where $\widetilde{\mathsf{color}}_i = \pi(\mathsf{color}_i)$. It sends $(C_1, \ldots, C_n)$ to V

$V \to P$: V chooses a random edge $(i, j) \in E$ and sends it to P

$P \to V$: Prover opens C_i and C_j to reveal $(\widetilde{\mathsf{color}}_i, \widetilde{\mathsf{color}}_j)$

V: If the openings of C_i, C_j are valid and $\widetilde{\mathsf{color}}_i \neq \widetilde{\mathsf{color}}_j$, then V accepts the proof. Otherwise, it rejects.

Proof of Soundness

- If G is not 3-colorable, then for any coloring $\mathsf{color}_1, \ldots, \mathsf{color}_n$, there exists at least one edge which has the same colors on both endpoints
- From the binding property of Com, it follows that $C_1, \ldots, C_n$ have unique openings $\widetilde{\mathsf{color}}_1, \ldots, \widetilde{\mathsf{color}}_n$
- Combining the above, let $(i^*, j^*) \in E$ be s.t. $\widetilde{\mathsf{color}}_{i^*} = \widetilde{\mathsf{color}}_{j^*}$
- Then, with probability $\frac{1}{|E|}$, V chooses $i = i^*, j = j^*$ and catches P
- In $n|E|$ independent repetitions, P successfully cheats in all repetitions with probability at most

$$\left(1 - \frac{1}{|E|}\right)^{n|E|} \approx e^{-n}$$

Proving Zero Knowledge

Intuition:

- In each iteration, V only sees two random colors
- Hiding property of Com guarantees that everything else remains hidden from V
- As for Graph Isomorphism, we will only prove zero knowledge for one iteration. For the full protocol, we can prove zero knowledge using Theorem 2

Proving Zero Knowledge: Simulator

Simulator $S(x = G, z)$**:**

- Choose a random edge $(i', j') \xleftarrow{\$} E$ and pick random colors $\mathsf{color}'_{i'}, \mathsf{color}'_{j'} \xleftarrow{\$} \{1, 2, 3\}$ s.t. $\mathsf{color}'_{i'} \neq \mathsf{color}'_{j'}$. For every other $k \in [n] \setminus \{i', j'\}$, set $\mathsf{color}'_k = 1$
- For every $\ell \in [n]$, compute $C_\ell = \mathsf{Com}(\mathsf{color}'_\ell)$
- Emulate execution of $V^*(x, z)$ by feeding it $(C_1, \ldots, C_n)$. Let (i, j) denote its response
- If $(i, j) = (i', j')$, then feed the openings of C_i, C_j to V^* and output its view. Otherwise, restart the above procedure, at most $n|E|$ times
- If simulation has not succeeded after $n|E|$ attempts, then output `fail`

Correctness of Simulation

Hybrid Experiments:

- H_0: Real execution
- H_1: Hybrid simulator S' that acts like the real prover (using witness $\mathsf{color}_1, \ldots, \mathsf{color}_n$), except that it also chooses $(i', j') \xleftarrow{\$} E$ at random and if $(i', j') \neq (i, j)$, then it outputs `fail`
- H_2: Simulator S

Correctness of Simulation (contd.)

- $H_0 \approx H_1$: If S' does not output `fail`, then H_0 and H_1 are identical. Since (i, j) and (i', j') are independently chosen, S' fails with probability at most:

$$\left(1 - \frac{1}{|E|}\right)^{n|E|} \approx e^{-n}$$

  Therefore, H_0 and H_1 are statistically indistinguishable

- $H_1 \approx H_2$: The only difference between H_1 and H_2 is that for all $k \in [n] \setminus \{i', j'\}$, C_k is a commitment to $\pi(\mathsf{color}_k)$ in H_1 and a commitment to 1 in H_2. Then, from the multi-value hiding property of Com, it follows that $H_1 \approx H_2$

Additional Reading

- Zero-knowledge Proofs for Nuclear Disarmament [Glaser-Barak-Goldston'14]
- Non-black-box Simulation [Barak'01]
- Concurrent Composition of Zero-Knowledge Proofs [Dwork-Naor-Sahai'98, Richardson-Kilian'99, Kilian-Petrank'01,Prabhakaran-Rosen-Sahai'02]
- Non-malleable Commitments and ZK Proofs [Dolev-Dwork-Naor'91]
- Non-interactive Zero-knowledge Proofs [Blum-Feldman-Micali'88,Feige-Lapidot-Shamir'90]