CS 600.442 – Modern Cryptography 7 October 2015

Lecture 11: Zero-Knowledge Proofs - III

Instructor: Abhishek Jain *Scribe: Navaneeth Krishnan Subramanian*

We finished the last lecture by discussing a physical zero-knowledge protocol for Graph 3-Coloring problem. In this protocol, the prover first generates a permutation $\pi(w)$ of the 3-coloring w for instance graph G, where $\pi(\cdot)$ is a random permutation. He then colors G using $\pi(w)$ and then hides the colored vertices of the graph using inverted cups. When the verifier issues a challenge (i, j), the prover simply removes the inverted cups on the vertices i and j and reveals the colors assigned to them. This process is then repeated several times (where the prover uses a fresh random permutation each time) to achieve a small soundness error.

1 Commitment Schemes

In order to construct an actual zero-knowledge protocol for graph 3-coloring, we will introduce a primitive called commitment schemes that we will use to replace the inverted cups in the physical zero-knowledge protocol discussed above. Intuitively, a commitment scheme must achieve the following two properties of inverted cups:

- Hiding (i.e., the color of a vertex must remain hidden from the verifier when it is covered with an inverted cup)
- Binding (The color of a vertex cannot be changed after it is covered with an inverted cup)

We now proceed to formally define commitment schemes.

Definition 1 (Commitment Scheme) *A PPT algorithm* COM *is called a commitment scheme if there exists some polynomial* $\ell(.)$ *such that the following properties hold:*

- **Perfect binding:** $\forall n \in \mathbb{N}$ *and all* $v_0, v_1 \in \{0,1\}^n$, $r_0, r_1 \in \{0,1\}^{\ell(n)}$ *it holds that:*

$$\mathsf{COM}(v_0, r_0) \neq \mathsf{COM}(v_1, r_1)$$

- **Hiding:** $\forall$ *non-uniform PPT distinguishers* D, *there exists a negligible function* $\mu(\cdot)$ *such that* $\forall n \in \mathbb{N}$, $v_0, v_1 \in \{0,1\}^n$, D *distinguishes the following with probability at most* $\mu(n)$*:*

$$\{r \xleftarrow{\$} \{0,1\}^{\ell(n)} : \mathsf{COM}(v_0, r)\}, \text{ and } \{r \xleftarrow{\$} \{0,1\}^{\ell(n)} : \mathsf{COM}(v_1, r)\}$$

Remark 1 *Commitment schemes also usually come equipped with an algorithm to "open" the commitment. For simplicity, here we will only focus on commitment schemes where to open a commitment* $c = \mathsf{COM}(v, r)$*, we simply reveal the message* v *and randomness* r *used to compute* c*. To verify that the opening is correct, the receiver can recompute* c *and check for correctness.*

We now state a useful property of commitment schemes.

Lemma 1 *Commitment schemes for one-bit message implies commitment schemes for strings.*

To construct a string commitment scheme from a bit commitment scheme, we simply use the bit commitment scheme multiple times to commit each bit of the string separately. The proof of hiding follows easily by a hybrid argument. The proof of binding follows directly from the binding property of the bit commitment scheme.

We now proceed to construct a bit commitment scheme.

Theorem 2 *If one-way permutations exist, then commitment schemes exist.*

Construction. Let f be a one-way permutation and h be its hardcore predicate. We define the commitment algorithm:

$$\mathsf{COM}(b, r) = f(r), b \oplus h(r)$$

The proof of binding and hiding is easy. First note that since f is a permutation, we have that $f(r) \neq f(r')$. Hence $\mathsf{COM}(0, r) \neq \mathsf{COM}(b', r')$. To see why the commitment scheme satisfies hiding property, we observe that $h(r)$ is unpredictable even given $f(r)$. Hence $b \oplus h(r)$ is unpredictable as well.

2 Zero-Knowledge Proof for Graph 3-Coloring

We are now ready to construct a zero-knowledge interactive proof system for graph 3-coloring. Let P and V denote the prover and the verifier, respectively. Let G be an n-vertex graph that is given as the common input (i.e., the instance) to P and V. Let E denote the set of edges in G, where $|E| = m$. Let $w = (C_1, \ldots, C_n)$ denote a witness for 3-coloring of G which is given as private input to P. Here, C_i denotes a color assignment to the ith vertex of G.

The Protocol. The prover P and verifier V proceed as follows:

$P \to V$:

1. Pick a random permutation Π over 3 colors. (Let {R, G, B} denote the set of colors.)
2. $\forall\ i \in [n]$, compute $X_i = \mathsf{COM}(\Pi(C_i), r_i)$.
3. Send $X_1, \ldots, X_n$ to V.

$V \to P$:

1. Choose an edge $(i, j) \in E$ at random.
2. Send (i, j) to P.

$P \to V$:

1. Send $(\Pi(C_i), r_i)$, $(\Pi(C_j), r_j)$ to V.

Let (C'_i, r'_i), (C'_j, r'_j) denote the values that V received from P. V rejects the proof if either of the following conditions is not satisfied:

- $X_i = \mathsf{COM}(C'_i, r'_i)$, and $X_j = \mathsf{COM}(C'_j, r'_j)$

- $C'_i \neq C'_j$

The above process is repeated $n \cdot m$ times sequentially. If V accepts each of the repetitions, then it finally accepts the proof. Otherwise, if it rejects any of the $n \cdot m$ repetitions, then it rejects the proof.

This completes the description of the protocol. The completeness property is easy to verify. We now discuss soundness and then proceed to prove the zero-knowledge property.

Soundness. We need to prove that V rejects the proof if the instance graph G is *not* 3-colorable. Let us first consider a single repetition of the protocol. Note that if G is not 3-colorable, then for any coloring of the graph G, there exists at least one edge $(i^*, j^*) \in E$ such that the vertices i^* and j^* are assigned the same color. If V chooses the challenge edge (i, j) at random, then with probability $\frac{1}{m}$, we will have that $i = i^*$, $j = j^*$. Then, it follows from the perfect binding of COM that in this case, V will output reject.

This means that a cheating prover can still successfully convince an honest verifier of the validity of a false instance with probability $1 - \frac{1}{m}$ in a single repetition of the protocol. However, since we repeat the protocol $n \cdot m$ times, the soundness error decreases to:

$$(1 - \frac{1}{m})^{k.m} \approx e^{-k}$$

which is negligible, as required.

Zero-Knowledge. We will now prove that a single iteration of the protocol satisfies the zero-knowledge property. This suffices for us since it can be shown that the zero-knowledge property composes under sequential repetition.

Simulator. We now construct a simulator S for a single iteration of the protocol. Let V^* denote the (adversarial) verifier. S:

1. Choose an edge $(i', j') \in E$ at random. Choose random colors $(C_{i'}, C_{j'}) \in \{R, G, B\}$ such that $C_{i'} \neq C_{j'}$. Set $C_k = R$ for $k \neq i', j'$
2. $X_{i'} = \mathsf{COM}(C_{i'}, r_{i'})$ where $r_{i'}$ is a random string
3. $(i, j) \leftarrow V^*$
4. If $(i, j) = (i', j')$, open $X_{i'}$, $X_{j'}$. Else, repeat the above steps, at most $n \cdot m$ times
5. If all attempts fail, abort

It follows from description that S runs in polynomial time.

Correctness of Simulation. We first show that the simulator S outputs *fail* with negligible probability. Consider hybrid experiments H_1, H_2 where H_1 corresponds to the simulator S and H_2 corresponds to a modified simulator $\tilde{S}$ that is similar to S, except that it simply sets all the colors $C'_k = R$ (for every vertex k). Note that in H_2, V^* cannot distinguish between any of the edges, hence the probability that it returns $(i, j) = (i', j')$ is $\frac{1}{m}$. Then, it follows from the hiding property

of the commitment scheme COM that $(i,j) = (i',j')$ with probability $\frac{1}{m}$ (minus some negligible amount) in H_1 as well.[1]

It follows then that the probability with which S outputs fail is less than $(1-\frac{1}{m})^{n \cdot m} < e^{-n}$, which is negligible.

Now, we need to show that the transcripts generated by S are computationally indistinguishable from the transcripts generated by $P \leftrightarrow V^*$. Consider a modified version of S, called S' that is given a 3-Coloring of G. In Step 2 of the simulation, S' will choose a random permutation of the colors in $\{R,G,B\}$ for the values of C_k rather than setting all values except $C_{i'}$ and $C_{j'}$ equal to R. This is similar to the way P interacts with V. Note that the transcript generated via $P \leftrightarrow V^*$ is computationally indistinguishable from that generated by S' because S' outputs fail with negligible probability.

We now need to argue that the transcripts generated by S and S' are computationally indistinguishable. Consider two messages m_0 and m_1 of the same length where m_0 consists of $n-2$ instances of commitments to R and two committed colors C_i and C_j (for a random edge $(i,j) \in E$) and m_1 consists of a random 3-coloring of G (with a random $(i',j') \in E$ chosen). Observe that if m_0 is input to V^*, then it corresponds to S and if m_1 is input to V^*, then it corresponds to S'. If we can distinguish between these two cases, then we can distinguish between the sequence of commitments m_0 and m_1, which is a contradiction to the hiding property of COM. This completes the proof.

[1] Here, we use the hiding property of two commitments at once. This follows easily by another hybrid argument where we use the hiding property of one commitment at a time.