Notes for Lecture 27

Scribed by Madhur Tulsiani, posted May 16, 2009

Summary

In this lecture we begin the construction and analysis of a zero-knowledge protocol for the 3-coloring problem. Via reductions, this extends to a protocol for any problem in NP. We will only be able to establish a weak form of zero knowledge, called "computational zero knowledge" in which the output of the simulator and the interaction in the protocol are computationally indistinguishable (instead of identical). It is considered unlikely that NP-complete problem can have zero-knowledge protocols of the strong type we defined in the previous lectures.

As a first step, we will introduce the notion of a *commitment scheme* and provide a construction based on any one-way permutation.

1 Commitment Scheme

A commitment scheme is a two-phase protocol between a *Sender* and a *Receiver*. The Sender holds a message m and, in the first phase, it picks a random key K and then "encodes" the message using the key and sends the encoding (a *commitment* to m) to the Receiver. In the second phase, the Sender sends the key K to the Receiver can *open* the commitment and find out the content of the message m.

A commitment scheme should satisfy two security properties:

- **Hiding.** Receiving a commitment to a message m should give no information to the Receiver about m;
- **Binding.** The Sender cannot "cheat" in the second phase and send a different key K' that causes the commitment to open to a different message m'.

It is impossible to satisfy both properties against computationally unbounded adversaries. It is possible, however, to have schemes in which the Hiding property holds against computationally unbounded Receivers and the Binding property holds (under appropriate assumptions on the primitive used in the construction) for bounded-complexity Senders; and it is possible to have schemes in which the Hiding property holds (under assumptions) for bounded-complexity Receivers while the Binding property holds against any Sender. We shall describe a protocol of the second type, based on one-way permutations. The following definition applies to one-round implementations of each phase, although a more general definition could be given in which each phase is allowed to involve multiple interactions.

Definition 1 (Computationally Hiding, Perfectly Binding, Commitment Scheme) *A Perfectly Binding and (t,ϵ)-Hiding Commitment Scheme for messages of length ℓ is a pair of algorithms (C,O) such that*

- **Correctness.** *For every message m and key K,*

$$O(K, C(K,m)) = m$$

- **(t,ϵ)-Hiding.** *For every two messages $m, m' \in \{0,1\}^\ell$, the distributions $C(K,m)$ and $C(K,m')$ are (t,ϵ)-indistinguishable, where K is a random key, that is, for every algorithm A of complexity $\leq t$,*

$$|\, \mathbb{P}[A(C(K,m)) = 1] - \mathbb{P}[A(C(K,m')) = 1]| \leq \epsilon$$

- **Perfectly Binding.** *For every message m and every two keys K, K',*

$$O(K', C(K,m)) \in \{m, FAIL\}$$

In the following we shall refer to such a scheme (C,O) as simply a (t,ϵ)-secure commitment scheme.

Given a one-way permutation $f : \{0,1\}^n \to \{0,1\}^n$ and a hard-core predicate P, we consider the following construction of a one-bit commitment scheme:

- $C(K,m) := f(K), m \oplus P(K)$
- $O(K,(c_1,c_2))$ equals $FAIL$ if $f(K) \neq c_1$, and $P(K) \oplus c_2$ otherwise.

Theorem 2 *If P is a (t,ϵ)-secure hard core predicate for f, then the above construction is a $(t - O(1), 2\epsilon)$-secure commitment scheme.*

Proof: The binding property of the commitment scheme is easy to argue as the commitment is a permutation of the key and the message. In particular, given $C(K,m) = (x,y)$, we can find the unique K and m that generate it as

$$K = f^{-1}(x) \quad \text{and} \quad m = y \oplus P(K) = y \oplus P(f^{-1}(x))$$

To prove the hiding property in the contrapositive, we want to take an algorithm which distinguishes the commitments of two messages and convert it to an algorithm which computes the predicate P with probability better than $1/2+\epsilon$. Let A be such an algorithm which distinguishes two different messages (one of which must be 0 and the other must be 1). Then, we have that for A

$$\begin{array}{rl} & |\mathbb{P}[A(C(K,m))=1]-\mathbb{P}[A(C(K,m)=1]| > 2\epsilon \\ \Longrightarrow & |\mathbb{P}[A(f(K),P(K)\oplus 0)=1]-\mathbb{P}[A(f(K),P(K)\oplus 1)=1]| > 2\epsilon \end{array}$$

Assume without loss of generality that the quantity in the absolute value is positive i.e.

$$\mathbb{P}[A(f(K),P(K))=1]-\mathbb{P}[A(f(K),P(K)\oplus 1)=1] > 2\epsilon$$

Hence, A outputs 1 significantly more often when given the correct value of $P(K)$. As seen in previous lectures, we can convert this into an algorithm A' that predicts the value of $P(K)$. Algorithm A' takes $f(K)$ as input and generates a random bit b as a guess for $P(K)$. It then runs $A(f(K),b)$. Since A is correct more often on the correct value of $P(K)$, A' outputs b if $A(f(K),b)=1$ and outputs $b\oplus 1$ otherwise. We can analyze its success probability as below

$$\begin{array}{rcl} \mathbb{P}[A'(f(K))=P(K)] & = & \mathbb{P}[b=P(K)]\cdot\mathbb{P}[A(f(K),P(K))=1] \\ & & +\mathbb{P}[b\neq P(K)]\cdot\mathbb{P}[A(f(K),P(K)\oplus 1)=0] \\ & = & \frac{1}{2}\cdot\mathbb{P}[A(f(K),P(K))=1] \\ & & +\frac{1}{2}\cdot(1-\mathbb{P}[A(f(K),P(K)\oplus 1)=1]) \\ & = & \frac{1}{2}+\frac{1}{2}\cdot(\mathbb{P}[A(f(K),P(K))=1]-\mathbb{P}[A(f(K),P(K)\oplus 1)=1]) \\ & \geq & \frac{1}{2}+\epsilon \end{array}$$

Thus, A' predicts P with probability $1/2+\epsilon$ and has complexity only $O(1)$ more than A (for generating the random bit) which contradicts the fact that P is (t,ϵ)-secure.
□

There is a generic way to turn a one-bit commitment scheme into a commitment scheme for messages of length ℓ (just concatenate the commitments of each bit of the message, using independent keys).

Theorem 3 *Let (O,C) be a (t,ϵ)-secure commitment scheme for messages of length k such that $O(\cdot,\cdot)$ is computable in time r. Then the following scheme $(\overline{C},\overline{O})$ is a $t-O(r\cdot\ell),\epsilon\cdot l)$-secure commitment scheme for message of length $k\cdot\ell$:*

- $\overline{C}(K_1, \ldots, K_\ell, m) := C(K_1, m_1), \ldots, C(K_\ell, m_\ell)$
- $\overline{O}(K_1, \ldots, K_\ell, c_1, \ldots, c_\ell)$ *equals* $FAIL$ *if at least one of* $O(K_i, c_i)$ *outputs* $FAIL$*; otherwise it equals* $O(K_1, c_1), \ldots, O(K_\ell, c_\ell)$*.*

Proof: The commitment to m is easily seen to be binding since the commitments to each bit of m are binding. The soundness can be proven by a hybrid argument.

Suppose there is an A algorithm distinguishing $\overline{C}(K_1, \ldots, K_\ell, m)$ and $C(K_1, \ldots, K_\ell, m)$ with probability more than $\epsilon \cdot \ell$. We then consider "hybrid messages" $m^{(0)}, \ldots, m^{(\ell)}$, where $m^{(i)} = m'_1 \ldots m'_i m_{i+1}, \ldots, m_\ell$. By a hybrid argument, there is some i such that

$$\left|\mathbb{P}[A(K_1, \ldots, K_\ell, m^{(i)}) = 1] - \mathbb{P}[A(K_1, \ldots, K_\ell, m^{(i+1)}) = 1]\right| > \epsilon$$

But since $m^{(i)}$ and $m^{(i+1)}$ differ in only one bit, we can get an algorithm A' that breaks the hiding property of the one bit commitment scheme $C(\cdot, \cdot)$. A', given a commitment c, outputs

$$A'(c) \;=\; A(C(K_1, m_1), \ldots, C(K_i, m_i), c, C(K_{i+2}, m'_{i+2}), \ldots, C(K_\ell, m'_\ell))$$

Hence, A' has complexity at most $t + O(r \cdot l)$ and distinguishes $C(K_{i+1}, m_{i+1})$ from $C(K_{i+1}, m'_{i+1})$. $\Box$

There is also a construction based on one-way permutations that is better in terms of key length.

2 A Protocol for 3-Coloring

We assume we have a (t, ϵ)-secure commitment scheme (C, O) for messages in the set $\{1, 2, 3\}$.

The prover P takes in input a 3-coloring graph $G = ([n], E)$ (we assume that the set of vertices is the set $\{1, \ldots, n\}$ and use the notation $[n] := \{1, \ldots, n\}$) and a proper 3-coloring $\alpha : [n] \to \{1, 2, 3\}$ of G (that is, α is such that for every edge $(u, v) \in E$ we have $\alpha(u) \neq \alpha(v)$). The verifier V takes in input G. The protocol, in which the prover attempts to convince the verifier that the graph is 3-colorable, proceeds as follows:

- The prover picks a random permutation $\pi : \{1, 2, 3\} \to \{1, 2, 3\}$ of the set of colors, and defines the 3-coloring $\beta(v) := \pi(\alpha(v))$. The prover picks n keys $K_1, \ldots, K_n$ for (C, O), constructs the commitments $c_v := C(K_v, \beta(v))$ and sends $(c_1, \ldots, c_n)$ to the verifier;

- The verifier picks an edge $(u, v) \in E$ uniformly at random, and sends (u, v) to the prover;
- The prover sends back the keys K_u, K_v;
- If $O(K_u, c_u)$ and $O(K_v, c_v)$ are the same color, or if at least one of them is equal to $FAIL$, then the verifier rejects, otherwise it accepts

Theorem 4 *The protocol is complete and it has soundness error at most* $(1-1/|E|)$.

Proof: The protocol is easily seen to be complete, since if the prover sends a valid 3-coloring, the colors on endpoints of every edge will be different.

To prove the soundness, we first note that if any commitment sent by the prover opens to an invalid color, then the protocol will fail with probability at least $1/|E|$ when querying an edge adjacent to the corresponding vertex (assuming the graph has no isolated vertices - which can be rivially removed). If all commitments open to valid colos, then the commitments define a 3-coloring of the graph. If the graph is not 3-colorable, then there must be at least one edge e both of whose end points receive the same color. Then the probability of the verifier rejecting is at least the probability of choosing e, which is $1/|E|$. □

Repeating the protocol k times sequentially reduces the soundness error to $(1 - 1/|E|)^k$; after about $27 \cdot |E|$ repetitions the error is at most about 2^{-40}.

3 Simulability

We now describe, for every verifier algorithm V^*, a simulator S^* of the interaction between V^* and the prover algorithm.

The basic simulator is as follows:

Algorithm S^*_{1round}

- Input: graph $G = ([n], E)$
- Pick random coloring $\gamma : [n] \to \{1, 2, 3\}$.
- Pick n random keys $K_1, \ldots, K_n$
- Define the commitments $c_i := C(K_i, \gamma(i))$
- Let (u, v) be the 2nd-round output of V^* given G as input and $c_1, \ldots, c_n$ as first-round message

- If $\gamma(u) = \gamma(v)$, then output FAIL
- Else output $((c_1, \ldots, c_n), (u, v), (K_u, K_v))$

And the procedure $S^*(G)$ simply repeats $S^*_{1round}(G)$ until it provides an output different from $FAIL$.

It is easy to see that the output distribution of $S^*(G)$ is always *different* from the actual distribution of interactions between P and V^*: in the former, the first round is almost always a commitment to an invalid 3-coloring, in the latter, the first round is always a valid 3-coloring.

We shall prove, however, that the output of $S^*(G)$ and the actual interaction of P and V^* have *computationally indistinguishable* distributions provided that the running time of V^* is bounded and that the security of (C, O) is strong enough.

For now, we prove that $S^*(G)$ has efficiency comparable to V^* provided that security of (C, O) is strong enough.

Theorem 5 *Suppose that (C, O) is $(t+O(nr), \epsilon/(n\cdot|E|))$-secure and C is computable in time $\leq r$ and that V^* is a verifier algorithm of complexity $\leq t$.*

Then the algorithm S^*_{1round} as defined above has probability at most $\frac{1}{3}+\epsilon$ of outputting $FAIL$.

The proof of Theorem 5 relies on the following result.

Lemma 6 *Fix a graph G and a verifier algorithm V^* of complexity $\leq t$.*

Define $p(u, v, \alpha)$ to be the probability that V^* asks the edge (u, v) at the second round in an interaction in which the input graph is G and the first round is a commitment to the coloring α.

Suppose that (C, O) is $(t + O(nr), \epsilon/n)$-secure, and C is computable in time $\leq r$.

Then for every two colorings α, β and every edge (u, v) we have

$$|p(u, v, \alpha) - p(u, v, \beta)| \leq \epsilon$$

Proof: If $p(u, v, \alpha)$ and $p(u, v, \beta)$ differ by more than ϵ for any edge (u, v), then we can define an algorithm which distinguishes the n commitments corresponding to α from the n commitments corresponding to β. A simply runs the verifier given commitments for n colors and outputs 1 if the verifier selects the edge (u, v) in the second round.

Then, by assumption, A ϵ-distinguishes the n commitments corresponding to α from the n commitments corresponding to β in time $t + O(nr)$. However, by Theorem 3, this means that (C, O) is not $(t + O(nr), \epsilon/n)$-secure which is a contradiction. □

Given the lemma, we can now easily prove the theorem.

PROOF: (of Theorem 5) The probability that S^*_{1round} outputs $FAIL$ is given by

$$\mathbb{P}\left[S^*_{1round} \text{ outputs } FAIL\right] \;=\; \frac{1}{3^n} \cdot \sum_{c \in \{1,2,3\}^n} \sum_{\substack{(u,v)\in E \\ c(u)\neq c(v)}} p(u,v,c)$$

Let $\mathbf{1}$ denote the coloring which assigns the color 1 to every vertex. Then using Lemma 6 we bound the above as

$$\begin{aligned}
\mathbb{P}\left[S^*_{1round} \text{ outputs } FAIL\right] &\leq \frac{1}{3^n} \cdot \sum_{c \in \{1,2,3\}^n} \sum_{\substack{(u,v)\in E \\ c(u)\neq c(v)}} \left(p(u,v,\mathbf{1}) + \epsilon\right) \\
&= \sum_{(u,v)\in E} p(u,v,\mathbf{1}) \left(\sum_{c:c(u)\neq c(v)} \frac{1}{3^n} \right) \;+\; \epsilon \\
&= \frac{1}{3} \sum_{(u,v)\in E} p(u,v,\mathbf{1}) + \;\epsilon \\
&= \frac{1}{3} + \epsilon
\end{aligned}$$

where in the second step we used the fact that $c(u) \neq c(v)$ for a 1/3 fraction of all the colorings and the last step used that the probability of V^* selecting some edge given the coloring $\mathbf{1}$ is 1. □