COM S 6810 Theory of Computing March 24, 2009

Lecture 17: Arthur-Merlin games, Zero-knowledge proofs

Instructor: Rafael Pass *Scribe: Jean-Baptiste Jeannin*

In the previous lectures, we gave a definition of Arthur-Merlin games, proved that $IP = PSPACE$, then that $MA \subseteq AM$, and that for any fixed integer k, $AM[k] = AM$ where $AM = AM[2]$ by definition.

1 Perfect completeness for MA and AM

First recall that $L \in MA$ if there is a proof system for L that consists of the prover first sending a message, and then the verifier tossing coins and applying a polynomial-time predicate to the input, the prover's message and the coins. If V is the verifier, m is the message sent by the prover and r represent the coins tossed by the verifier, then an MA protocol is such that:

- if $x \in L$, then $\exists m, Pr_r(V(m,r) = 1) \geq \frac{2}{3}$,
- if $x \notin L$, then $\forall m, Pr_r(V(m,r) = 1) \leq \frac{1}{3}$.

We have already seen that by running several proofs in parallel, the following alternate definition of an MA protocol is equivalent: for any n,

- if $x \in L$, then $\exists m, Pr_r(V(m,r) = 1) \geq 1 - \frac{1}{2^n}$,
- if $x \notin L$, then $\forall m, Pr_r(V(m,r) = 1) \leq \frac{1}{2^n}$.

What we would like is to replace this small error in the completeness by a perfect completeness, thus getting the following definition of an MA with perfect completeness:

- if $x \in L$, then $\exists m, Pr_r(V(m,r) = 1) = 1$,
- if $x \notin L$, then $\forall m, Pr_r(V(m,r) = 1) \leq \frac{1}{2}$.

Theorem 1 *From an MA protocol for a language L we can construct an MA protocol with perfect completeness.*

Proof. Let us consider an MA protocol for a language L. Without loss of generality, by running several proofs in parallel, and if $l = |r|$, we can assume that:

- if $x \in L$, then $\exists m, Pr_r(V(m,r)) \geq 1 - \frac{1}{4l}$,
- if $x \notin L$, then $\forall m, Pr_r(V(m,r)) \leq \frac{1}{4l}$.

Now we apply an idea similar to the proof that $BPP \subseteq \Sigma_2$. Given an x and an m, let S_x^m be the set of random coins r such that $V(m,r) = 1$ (i.e., accepts). Furthermore, let $z_1, .., z_l \in \{0;1\}^l$. If S_x^m is small (i.e., of size $\leq \frac{2^l}{4l}$)), then $\bigcup_{i \in [|1;l|]} S_x^m \oplus z_i$ is small (i.e., of size $\leq \frac{2^l}{4}$); this happens when $x \in L$. On the other hand, if S_x^m is big (i.e., of size $\geq 2^l \left(1 - \frac{1}{4l}\right)$), then $\bigcup_{i \in [|1;l|]} S_x^m \oplus z_i = \{0,1\}^l$; this happens when $x \notin L$. A more careful proof was given in the proof that $BPP \subseteq \Sigma_2$.

Now, let us define this new protocol for MA: the verifier (Arthur) sends to the prover $m, z_1, ..., z_l$. Then he wants to check whether $\bigcup_{i \in [|1;l|]} S_x^m \oplus z_i$ is big. To do that, he picks a random $r \in \{0;1\}^l$ and accepts if and only if $r \in \bigcup_{i \in [|1;l|]} S_x^m \oplus z_i$, i.e., if and only if $\bigvee_{i \in [|1;l|]} V(m, r \oplus z_i) = 1$.

■

Now we have proven that from any MA protocol we can get an MA protocol with perfect completeness. What about AM protocols?

Theorem 2 *From an AM protocol for a language L we can construct an AM protocol with perfect completeness.*

Proof. Given an MA protocol consisting of V sending a random string r to P, then P answering with a message m, we know that we can amplify it such that:

- if $x \in L$, for a fraction at least $1 - \frac{1}{4l}$ of r's, there exists an accepting reply m,
- if $x \notin L$, for a fraction at most $\frac{1}{4l}$ of r's, there exists an accepting reply.

Now, let us look at this protocol: the prover P sends $z_1, ..., z_l$ to V; then V sends $r \in \{0;1\}^l$ to P; finally P sends an i and an m to V, and V checks that $V(m, r \oplus z_i)$ accepts. Thus, given a non-perfect AM protocol, we have built an $MAM = AM[3]$ protocol with perfect completeness. But since we know that $MAM = AM$ (i.e., $AM[3] = AM$), we have in fact built an AM protocol with perfect completeness. ■

2 $coSAT$ and AM

Theorem 3 *If $coSAT \subseteq AM$ then $\Sigma_2 \subseteq \Pi_2$ (and the PH collapses).*

Proof. We can first see that $coSAT \in \#P \subseteq PSPACE = IP$, therefore it has an interactive proof with a polynomial number of rounds.

Claim 1 $AM \subseteq \Pi_2$

Given an AM protocol, we now know (from section 1 of this lecture) that we can obtain an AM protocol with perfect completeness. Therefore, if we consider a protocol where V sends r, then P answers with M:

- if $x \in L$, then $\forall r, \exists m_r, Pr_r(V(m_r, r) = 1) = 1$. This can be reformulated as: if $x \in L$, then $\forall r, \exists m_r, V(m_r, r) = 1$;
- if $x \notin L$, then for many r's, $\not\exists m, V(m, r) = 1$. This implies that: if $x \notin L$, then $\exists r, \forall m, V(m, r) \neq 1$.

Put together (and taking the contrapositive of the second), these two assertions mean that $x \in L$ if and only if $\forall r, \exists m_r, V(m_r, r) = 1$, thus proving that $AM \subseteq \Sigma_2$.

Claim 2 *If* $coSAT \subseteq AM$*, then* $\Sigma_2 \subseteq AM$

If $L \in \Sigma_2$, then there is an M such that $x \in L$ if and only if $\exists y, \forall z, M(x, y, z)$. Now, if we define L' as: $(x, y) \in L'$ if and only if $\forall z, M(x, y, z)$, we know that $x \in L$ if and only if $\exists y, (x, y) \in L'$. Clearly, $L' \in coNP$, and because $coSAT \subseteq AM$, $L' \in AM$. Now we get a three-round interactive proof for L: first send y, then run the two-round interactive proof for L'. We know that these three rounds can be collapsed into two rounds, thus proving that $L \in AM$.

From the two claims, we easily get that $\Sigma_2 \subseteq \Pi_2$, and thus that the PH collapses. ■

Corollary 1 *Unless the* PH *collapses, GraphIso is not* NP*-complete.*

Proof. We have already shown a short interactive proof for GraphNonIso (lecture 14, section 2). If GraphIso was NP-complete, GraphNonIso would be $coNP$-complete, thus leading to $coNP \subseteq AM$, making the PH collapse by the previous theorem. ■

More generally, a widely used technique to show that something is not NP-complete is to give a short interactive proof of its complement. Of course, this only holds if the PH does not collapse.

3 Introduction to Zero-knowledge proofs

In Computer Science, we are not interested in philosophical decisions of knowledge. We take a more pragmatic point of view, stating that knowledge and knowing is the ability to perform a task. Therefore something, e.g., a proof, is zero-knowledge if and only if it does not convey any new knowledge.

Let us first take an example: you know that there has been a murder, you are a journalist and want some more information in order to write an article about it. If you call the police, the only thing they will tell you is "there has been a murder", then hang up. This is zero-knowledge, because you did not learn anything from this phone call. You could have thought this phone call in your head without it ever happening. A more complicated example would be to have the police hang up with probability $\frac{1}{2}$, and tell you "there has been a murder" with probability $\frac{1}{2}$. Again, this is zero-knowledge.

What is a good example of a zero-knowledge proof? Let us look at the "Where's Waldo" classical game. In this game you have to find Waldo, whom you have a picture, in a big crowd. How can I prove to you that Waldo is actually in the crowd without revealing where he is? I could have a very big piece of paper with a little hole just the size of Waldo's head; I put the picture behind the piece of paper, putting Waldo behind the hole. Now you are convinced that Waldo is in the picture, but you still do not know where he is.

Another example is the game of the 7 differences: two pictures are very slightly different, I would like to prove to you that they actually are different without revealing the differences. I could give you the two pictures, you can shuffle them, then I have to decide which is which. If we do that say 100 times and I always get it right, then you will be convinced that the pictures are different.