LECTURE 11

LAST TIME

- Interactive proofs
- Multiprover interactive proofs
  Wlog 2 provers
  Can "cross examine" and check consistency
- So provers might as well write down all answers and let verifier check (sort of)
  PCP probabilistically checkable proofs
- $\boxed{NP = PCP(O(\log n), O(1))}$
  For any language L, can write down proof π for x in length poly($|x|$) s.t.
  - Verifier randomly reads constant # bits
  - If $x \in L$ always accepts
  - If $x \notin L$, rejects with prob $\geq 1/2$

CRYPTOGRAPHY

- Perfect secrecy – impossible unless secret key as long as message (then do one-time pad)
- Computational security against poly-time adversaries
- One-way function – easy to compute
  – hard to invert
  Several good candidates
  Can't prove that they are one way — would imply $P \neq NP$
- OWF $\Rightarrow$ can generate pseudorandomness
  Indistinguishable from "real stuff" to any poly-time algorithm

- Small key (random)
  "Stretch" to longer pseudorandom key
  Do "one-time pad" with longer key

ZERO KNOWLEDGE PROOFS

- Interactive proofs where verifier only learns the fact that $x \in L$ from conversation and <u>nothing more</u>
- <u>Formal def</u> Given misbehaving verifier V^*, can construct simulator S^* that produces transcript with the correct distribution (might run V^* as subroutine)
- So the conversation cannot have leaked any info (other than proving to verifier that $x \in L$)
- If one-way functions exist, then any language in NP has a computational ZK proof.

<u>TODAY</u> PROOF COMPLEXITY

"Proof" today:
- <u>efficiently</u> verifiable (as always)
- <u>100% correct</u> (no probabilities involved)
- no interaction

DEF 1 <u>Proof system</u> for language L [Cook & Reckhow '79]
Deterministic algorithm $P(x, \pi)$
- Runs in time poly in $|x| + |\pi|$
- $x \in L \Rightarrow \exists$ <u>proof</u> π s.t. $P(x, \pi) = 1$
- $x \notin L \Rightarrow \forall \pi'$ holds that $P(x, \pi') = 0$

DEF 2 <u>Propositional proof system</u>

Proof system for $\overline{\text{CNFSAT}} = \{ F \mid \text{CNF formula } F \text{ is unsatisfiable} \}$

DEF 3 A proof system $\mathcal{P}$ is <u>polynomially bounded</u> if $\exists$ poly p s.t. $\forall x \in L$ $\exists \pi$ s.t.
$|\pi| \leq p(|x|)$ and $\mathcal{P}(x, \pi) = 1$

<u>THEOREM 4</u> [Cook & Reckhow '79]
NP = coNP iff there exists a polynomially bounded propositional proof system

<u>Proof</u> NP is exactly the set of languages with polynomially bounded proof <u>systems.</u>
Now use coNP-completeness of $\overline{\text{CNFSAT}}$ $\boxed{\text{Q.E.D.}}$

<u>COR 5</u> If there is no p-bounded propositional proof system, then $P \neq NP$.

<u>Proof</u> P is closed under complement. $\boxed{\text{Q.E.D.}}$

1ST REASON FOR PROOF CPLX

<u>Cook's program</u>: Study stronger and stronger proof systems; prove superpoly lower bounds; ultimately hope to get to $P \neq NP$

Get stuck pretty soon ...

At roughly same place as circuit complexity

For roughly similar reasons

2ND REASON FOR PROOF COMPLEXITY

Connections to SAT solving

Used to think of SAT as hard problem.

But there are amazingly good SAT solvers out there that are used routinely to solve formulas with 100 000s or even 1 000 000s of variables

Used in e.g.
- hardware verification
- software testing
- artificial intelligence
- bioinformatics
- cryptography
- ... and the list goes on... (but this is a theory course)

Study corresponding proof systems to understand potential and limitations of such SAT solvers

Best SAT solvers today use technique called conflict-driven clause learning (CDCL)

Can be seen to search for proofs in resolution proof system

Rest of today's lecture:
- define resolution
- Give exposition of lower bound by [Haken '85] early breakthrough result (but we will give different, hopefully simpler, proof)

RESOLUTION PROOF SYSTEM $\quad \pi : F \vdash \bot$

Resolution refutation π of F: sequence of clauses; each clause either

(a) $C \in F$ (axiom clause), or

(b) derived from two previous clauses by **resolution rule**
$$\frac{C \vee x \qquad D \vee \bar{x}}{C \vee D}$$

Refutation ends with empty clause $\bot$ containing no literals

LEMMA 6 F is unsatisfiable iff $\exists$ resolution refutation of F

Proof sketch ($\Leftarrow$) Suppose $\exists$ satisfying assignment α. Satisfies all axiom clauses. Then satisfies all resolvents. But empty clause $\bot$ unsatisfiable. Contradiction.

($\Rightarrow$) Not hard, but requires an argument. $\square$

EXAMPLE 7 $F = (x \vee y) \wedge (x \vee \bar{y} \vee z) \wedge (\bar{x} \vee z) \wedge \bar{z}$

1. $x \vee y$
2. $x \vee \bar{y} \vee z$
3. $\bar{x} \vee z$
4. $\bar{z}$

---

5. $\bar{x}$
6. $x \vee \bar{y}$
7. x
8. $\bot$

Can associate DAG G_π with any refutation π in this way

sources = axioms

sink = $\bot$

Length of refutation = # clauses (here: 8)

Length of refuting F = length of a shortest refutation

Also interested in

- width – size of largest clause in refutation
- space – how many clauses need to remember during verification

Don't have time to talk about this today

PIGEONHOLE PRINCIPLE

$n+1$ pigeons don't fit into n holes if each pigeon should get a separate hole.

Encode (opposite of) this statement as CNF

$x_{i,j}$ $\quad 1 \le i \le m$ pigeons $\quad (m > n)$
$\quad\quad\; 1 \le j \le n$ pigeonholes

we fix $m = n+1$

$x_{i,j}$ true $\Leftrightarrow$ pigeon i goes to hole j

Clauses

$$P^i = \bigvee_{j=1}^{n} x_{i,j} \qquad i \in [m]$$

$$H_j^{i,i'} = \overline{x}_{i,j} \vee \overline{x}_{i',j} \qquad i,i' \in [m],\ i<i',\ j \in [n]$$

P^i = pigeon i gets some hole

$H_j^{i,i'}$ = pigeons i and i' don't both sit in hole j.

$$\boxed{PHP_n^m = \bigwedge_{i=1}^{m} P^i \wedge \bigwedge_{j=1}^{n} \bigwedge_{1\le i<i'\le m} H_j^{i,i'}}$$

Arguably <u>the</u> most studied formula family in all of proof cplx (and still quite a few things we don't know about it)

Fix $m = n+1$

PHP_n^{n+1} has
- $\Theta(n^2)$ variables
- $\Theta(n^3)$ clauses
- size $\Theta(n^3)$

Today we'll prove:

THEOREM 8 [Haken '85]
Resolution refutations of PHP_n^{n+1} require length $\exp(\Omega(n))$.

In terms of formula size $N = \Theta(n^3)$
get lower bound $\exp(\Omega(\sqrt[3]{N}))$.

Also known (for other formulas) truly exponential lower bounds
$\exp(\Omega(N))$ – right up to constant factors in exponent

Prove lower bound via game. Works for any formula – let's focus on PHP

DEFENDANT: claims can fit $n+1$ pigeons into n holes.
PROSECUTOR: Wants to convict defendant of lying

Should be a clear-cut case, but two problems:

(1) JURY TRIAL
And the jury will only be convinced by obviously contradictory answers.

(2) PROSECUTOR OFTEN NEEDS STAND-IN
Has to take care of small kids at home who have cold/fever and can't go to daycare
Stand-in needs super-explicit book with instructions how to cross examine defendant.

<u>Questions</u>: "Does pigeon i go to hole j?"
<u>Answers</u>: Yes / no

Prosecutor stand-in keeps list of information

(i_1, j_1, yes)
(i_2, j_2, no)
(i_3, j_3, yes)
etc

Defendant can see this info

Explicit contradiction (nothing else will convince jury)

(a) (i,j,yes) and (i,j,no)

(b) (i_1,j,yes) and (i_2,j,yes) $i_1 \neq i_2$

(c) $(i,1,\text{no}),(i,2,\text{no}),\ldots,(i,n,\text{no})$

Instructions to prosecutor stand-in

Look up record $(i_1,j_1,\text{yes/no}),(i_2,j_2,\text{yes/no}),\ldots,$
(one per page) $(i_s,j_s,\text{yes/no})$

Instruction one of

(a) ask about pigeon i^* and pigeonhole j^*
(b) forget $(i_q,j_q,\text{yes/no})$

Why forget? To minimize # pages in book with instructions. Not very impressive if prosecutor needs to flip through several times to find out which question to ask...
But if prosecutor asks same question again defendant might answer differently!

LEMMA 9

If $\exists$ resolution refutation $\pi : PHP^{n+1}_n \vdash \bot$ of length L, then $\exists$ prosecutor rule book with $O(L)$ pages.

Proof Look at DAG G_π representing π.
Start at sink node labelled $\bot$
Derived from $\dfrac{x_{ij} \quad \overline{x}_{ij}}{\bot}$ for some x_{ij}

Ask: Does pigeon i go to hole j?
Move to clause falsified by defendant's answer

Invariant Say at clause $C \vee D$ in refutation derived from
$$\frac{C \vee x_{ij} \quad D \vee \bar{x}_{ij}}{C \vee D}$$

Current record: Minimal assignment falsifying current clause $C \vee D$ (for variable x_{ij} resolved over)

Ask "does pigeon i go to hole j?"

Move to clause falsified by answer, say $C \vee x_{ij}$

Forget all record/notes not talking about $Vars(C \vee x_{ij})$

Sooner or later, reach source = clause of PHP_n^{n+1}. By invariant, this clause is falsified. This is an explicit contradiction as defined above. $\boxed{Fin}$

Hence, if we can prove there is no small rule book, then there is no short resolution refutation

LEMMA 10

There is a δ such that any prosecutor stand-in rule book must have $\geq 2^{\delta n}$ pages / rules.

Prove this by exhibiting defendant strategy that forces prosecutor to have many pages. Use randomization.

DEFENDANT STRATEGY

Choose $n/4$ pigeons completely uniformly at random and assign them to $n/4$ uniformly random pigeonholes to get partial matching

[assume for simplicity $4 \mid n$ — no big deal.]

Let α be this $\boxed{\text{partial}}$ matching

Defendant always maintains partial matching

$\beta \supseteq \alpha$

Let's say hole j is <u>prohibited</u> for pigeon i if prosecutor has (i, j, no) on record

<u>How defendant answers question "Does i go to j?":</u>

① If $i \in \mathrm{Dom}(\beta)$, answer yes if $\alpha(i) = j$, else no.

② If $i \notin \mathrm{Dom}(\beta)$, always answer "no".

Then look at # prohibited holes for i

If # prohibited holes $\geq n/2$,

Choose smallest hole j^* that is not prohibited and consistent with matching β and extend β with $(i \mapsto j^*)$

<u>If Prosecutor forgets</u>

If a pigeon $i \in \mathrm{Dom}(\beta) \setminus \mathrm{Dom}(\alpha)$ does <u>not</u> have assigned hole on record and has less than $n/2$ prohibited holes, then <u>remove</u> i from matching β

If defendant cannot do update in ② $\Rightarrow$ gives up

Let's say pigeon i is <u>thoroughly examined</u> on prosecutor record R if R contains

(a) (i, j, yes) for some j, or

(b) (i, j', no) for $\geq n/2$ j's

<u>LEMMA 11</u>

Before the prosecutor gets the defendant convicted, she must create record R with $\geq n/4$ thoroughly examined pigeons.

<u>Proof</u> As long as defendant follows strategy he is consistent. Hence, before conviction, an update of β in (2) must have failed.
How can this happen? Here's how:
Some pigeon i^* reaches $n/2$ prohibited holes, but there is no available hole j^*.
Means $|\mathrm{Dom}(\beta)| \geq n/2$
$\Rightarrow |\mathrm{Dom}(\beta) \setminus \mathrm{Dom}(\alpha)| \geq n/4$

But $i' \in \mathrm{Dom}(\beta) \setminus \mathrm{Dom}(\alpha)$ only if (a) or (b) above holds — QED $\boxed{L_{11}}$

Let's call such a record R <u>informative</u>
Want to prove that prosecutor strategy (i.e., rule book) must contain $\geq$ $2^{\delta n}$ distinct informative records.

We know that with prob 1 for any choice α of Defendant must reach some informative record R.

Prove that for arbitrary fixed informative record R,

$$\Pr[\alpha \text{ and } R \text{ consistent}] \le 2^{-\delta n}$$

Possible to reach R when defendant plays acc to α

Since

$$1 = \Pr[\exists R \text{ consistent with } \alpha]$$

$$\le \sum_{\substack{\text{informative} \\ R}} \Pr[\alpha \text{ and } R \text{ consistent}]$$

$$\le 2^{-\delta n} \cdot (\#\text{ informative } R)$$

we get $2^{\delta n}$ distinct informative records

Let $I_R = \{\text{thoroughly investigated pigeons in } R\}$

$|I_R| \ge n/4$.

Expected size of intersection $I_R \cap \mathrm{Dom}(\alpha)$ is

$$|I_R \cap \mathrm{Dom}(\alpha)| \ge n/16$$

(by linearity of expectation).

By concentration of measure, except with exponentially small probability we have

$$|I_R \cap \mathrm{Dom}(\alpha)| \ge n/32$$

Suppose $|I_R| = n/4$. Then

$$\Pr_\alpha\left[\,|I_R \cap Dom(\alpha)| \leq n/32\right]$$

$$= \frac{\sum_{i=0}^{n/32} \binom{n/4}{i} \binom{n+1-n/4}{n/4-i}}{\binom{n+1}{n/4}}$$

$\leq$ ... calculations ... $\leq 2^{-\delta' n}$

for some $\delta' > 0$

Another way of seeing this:

You (sort of) ~~obtain~~ $|I_R \cap Dom(\alpha)|$

by picking pigeons in α one by one. Every time chance of picking pigeon in I_R is

$\approx 1/4$ and you do this $n/4$ times.

Actual outcome will be sharply concentrated around $n/16$

"coin flips" are not independent

(but this is not a formal argument. The actual calculations are just too boring, though...)

So suppose $|I_R \cap Dom(\alpha)| \geq n/32$

For every $i \in I_R$, it holds that I_R either

(a) specifies one hole j^*

(b) rules out $n/2$ holes.

To be consistent, random α must comply with these restrictions. Choose pigeons in α one by one.

For $(i+1)$ st pigeon

(a) prob $\leq \frac{1}{n-i}$ to hit exactly right hole

(b) prob $\leq \frac{n/2}{n-i}$ to avoid prohibited hole

$i \leq n/32 \Rightarrow$ Probabilities in (a) & (b) $\ll 2/3$

Probability of getting all pigeons consistent $\ll \left(\frac{2}{3}\right)^n < 2^{-\delta'' n}$

$Pr[\alpha$ and R consistent$] \leq$

$Pr[|I_R \cap Dom(\alpha)|$ small$]$ +

$Pr[|I_R \cap Dom(\alpha)|$ large but α consistent with $R]$

$\leq 2^{-\delta' n} + 2^{-\delta'' n} \leq 2^{-\delta n}$ for

some $\delta > 0$, QED $\boxed{L_2}$

This concludes our proof of the lower bound on (refutation length for resolution)

PHP^{n+1}_n.