Lecture 11

Finishing off Lecture 9

Interactive Proofs for LLM answer verification,

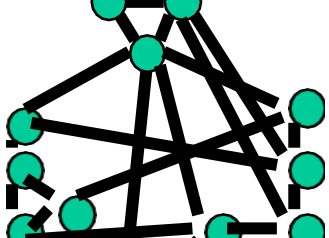
Orr Paradise, EPFL

Zero Knowledge for all of NP

Theorem[GMW87]: If one-way functions exist, then every problem in NP has a computational zero knowledge interactive proofs

- To prove the theorem, enough to show zero knowledge interactive proof for one NP complete problem

3COLOR = all graphs which can be colored with 3 colors s.t for for all edges (u,v) color(u) ≠color(v)

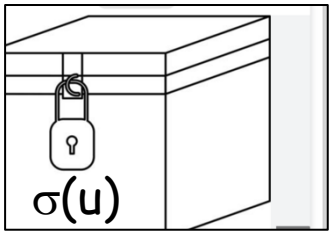


Physical Intuition for Protocol

On common input graph G =(V,E) and

Provers private input coloring $\alpha: V \longrightarrow \{0,1,2\}$

- P picks a random permutation π of the colors {0,1,2}& color the graph with coloring $\sigma=\pi(\alpha)$. It hides the color $\sigma(u)$ of each vertex inside a locked box



- V Select a random edge (u,v)
- P opens boxes corresponding to u and v

- V accepts if and only if $\sigma(u) \neq \sigma(v)$

  [colors are different]

Completeness and Soundness

- Completeness: if prover uses a proper 3-coloring, the verifier will accept.

- Soundness: Let $k = |E|^2$

  If G is not 3-colroable, then for all P*

  Prob[(P*,V)(G) accepts]<1- 1/|E|

  Repeat k times where $k=|E|^2$

  Soundness Prob[(P*,V)(G) accepts]<

  $(1-1/|E|)^{k.} < 1/e^{|E|}$

Commitment Scheme: Digital analogue of locked boxes

- An efficient two-stage protocol between a sender and receiver on security parameter 1^k s.t.:

- commit stage: sender has private input b

  At the end of the commit stage
  - both parties hold output com (called the commitment)
  - sender holds a private output dec (called the de-commitment)

- reveal stage: sender sends the pair (dec' , b') to receiver.

  Receiver accepts or rejects (com, dec, b)

Hiding and Binding Properties

Commitment schemes satisfy two properties:

- **Hiding**: Given comm sent by sender, receiver should not be able to b and b' for $b \neq b'$

- **Binding**: S*ender* can not produce two tuples (comm, dec, b) and (comm, dec, b') that the receiver will accept both with high probability if $b \neq b'$.

Implementation: send computationally secure c=Enc(r,b) to commit to b, and to reveal send r

ZK interactive proof for G3COL

On common input graph $G=(V,E)$ and private prover input coloring $\alpha: V \rightarrow \{0,1,2\}$

- P -> V: Pick a random permutation π of the coloring & color the graph with coloring $\pi(\alpha(v))$. Send commitments $Enc(r_v, \pi(\alpha(v)))$ $\forall$ vertex v.
- V -> P: Select a random edge (u,v) and send it
- P -> V: reveal colors of u and v committed to by releasing r_u and r_v
- If *color u = color v,* V rejects,

  otherwise repeat

  V accepts after k iterations.

Simulation for any Verifier V*

Simulator SIM on input G=(V,E) and verifier

V*: Fix random tape ω for V*

For i = 1 to $|E|^2$:

- Choose random edge (a, b) and permutation π of the colors, generate vector com = Enc(r_v, $\alpha(\pi(v))$) as in honest verifier simulation.
- Run V*(com; ω) to obtain challenge ($a*$, $b*$); if ($a*$, $b*$) = (a,b), then output transcript= (Enc(r_v, $\alpha(\pi(v))$) $\forall v$; (a, b); r_a, r_b)

If all iterations fail, output $\perp$.

Theorem: If Enc is computationally secure with respect to non-uniform adversaries, then

Claim 1: $\forall G$, α (a true coloring) : prob[$\perp$ output]=neg(|E|)

Claim 2 :if $\perp$ is not output, then simulated-view $\approx_c$real-view

Examples of NP-assertions

- Graph G: is 3-colorable

- Graph G: has a traveling salesman tour of cost C,

…

- Tuple (Enc(x), y, program P): y=P(x)

Applications:

Can prove properties about m without ever revealing m, only Enc(m)

Can prove relationships between m_1 and m_2 never revealing either one, only Enc(m_1) and Enc(m_2).

- For example: $L = \{(C_1, C_2)$: *there exists* r_1, r_2, M *s.t.* $C_1=E_1(r_1, M)$ *and* $C_2=E_2(r_2,M)\}$ is in NP

Generally: A tool to enforce honest behavior without forcing to reveal information.

Uses of Zero Knowledge Proofs

2014 Zero Knowledge and Nuclear Disarmament: projects at Princeton and MIT [Barak etal]

2015 Zero Knowledge and Forensics [Naor etal]

2016 Zero Cash, crypto currency which protects the privacy of transactions [BenSasson,Chiesa, Tromer et al]

2017 Proof of "compliance" of FISA with secret laws [Goldwasser etal]

Can you prove more with interactive proofs?

- Is IP greater than NP?
- Theorem[lfkn, shamir]: IP=PSPACE
- Reshaped Complexity Theory

Recall IP = {L s.t. ∃ (P,V) interactive proof system for L with completeness c(x)=2/3 and soundness s(x)=1/3}

Classically: Can Efficiently Verify

		$EQ(x_1,\ldots,x_n)$
NP	✓	∃solution
Co-NP	?	0 solutions
#P	?	2^{100} -13 solutions
PSPACE	?	∀∃∀…∃

Can you prove more via interactive proofs?

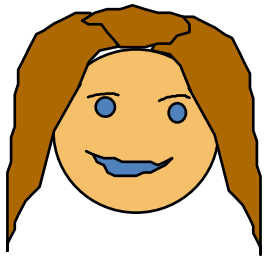
Interactively Provable= IP

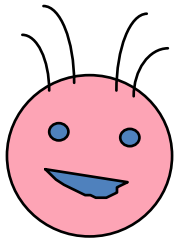
NP ✓

#P ✓

Co-NP ✓

PSPACE =IP ✓





q_1

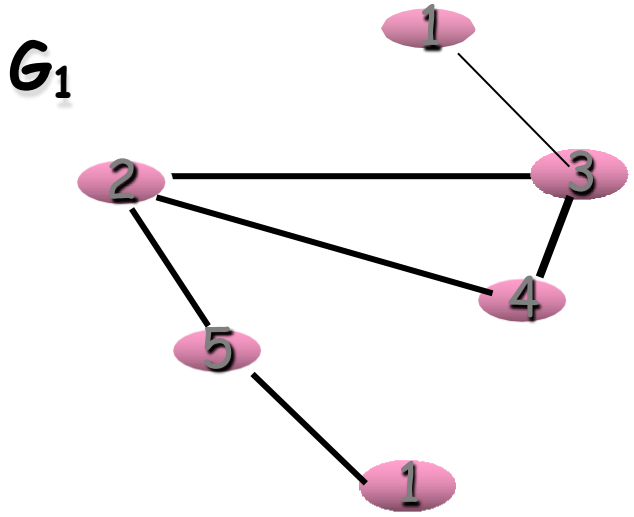
a_1

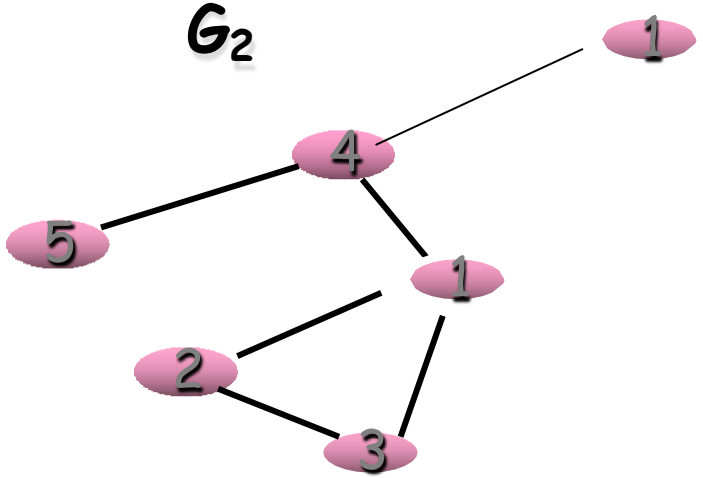
Accept/
Reject

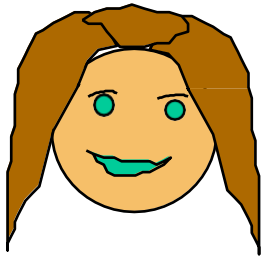
q_2

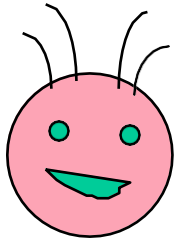
a_2

Example: G_1 is **NOT** isomorphic to G_2





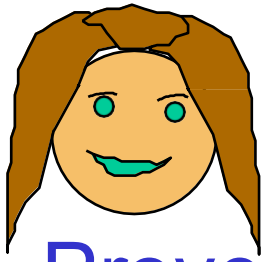




Shortest classical proof:
≈exponential n!

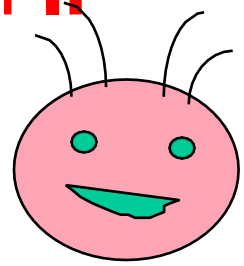
But can convince with an efficient interactive proof

Graph Non-Isomorphism (Non-ISO) in **IP**



Prover

input: (G_0, G_1)



$H = \gamma(G_c)$

Verifier

flip coin $c \in \{0,1\}$;

pick random γ

if H isomorphic to $G_{0,}$ then b = 0,

else b = 1

b

If $b \neq c$, reject

Else repeat

accept after k iterations

Completeness: if $(G_0, G_1) \in Non - ISO$, then

$Prob[(P,V)[(G_0, G_1)]=accept]=1$

Soundness: if $(G_0, G_1) \in ISO$,

$Prob[(P,V)[(G_0, G_1)]=accept] \leqq 1/2^k$

GNI Interactive Proof

- completeness:
  - if G_0 not isomorphic to G_1 then H is isomorphic to exactly one of (G_0, G_1)
  - prover will choose correct b=c
- soundness:

  if G_0 is isomorphic to G_1 then prover sees same distribution on H for c = 0, c = 1

  which provides no information on c $\Rightarrow$

  Prob[prover P* outputs b=c]$\leq$1/2

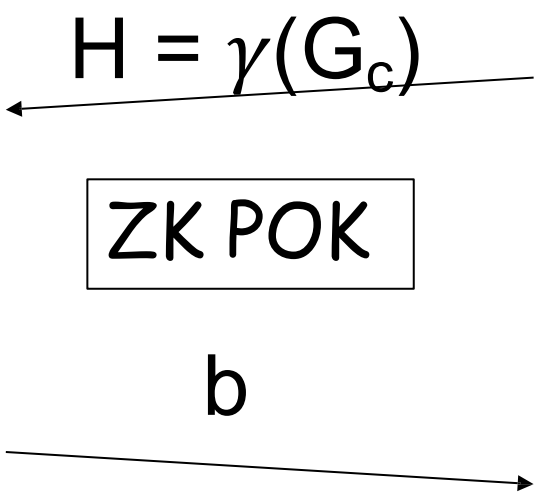
  After k iterations, prob drops to $1/2^k$

Honest Verifier Zero Knowledge

This is obviously honest verifier zero-knowledge (when the graphs are isomorphic):
--All the verifier gets is the coin c he tossed.

But, is it zero-knowledge for all verifiers?
-No. V can use P to find out if H is isomorphic to G_0 or isomorphic to G_1.

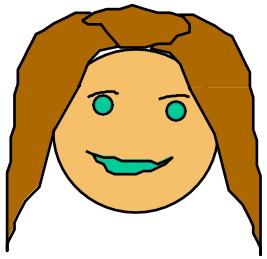
-Instead, the Verifier proves in ZK that he knows y s.t either $H=\gamma(G_0)$ or $H=\gamma(G_1)$

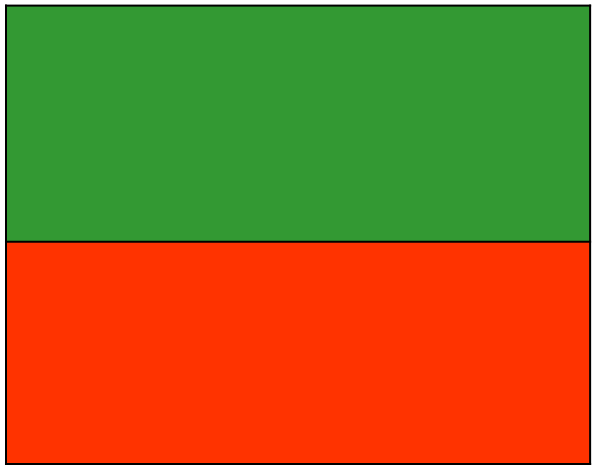


Interactive Proof

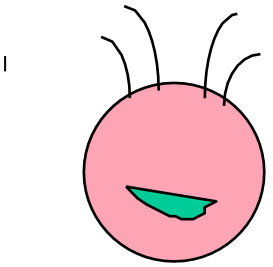
Prove to color blind bob that colors exist

Claim: there are 2 colors on this page, red to





Bob color blind



Alice sends Bob a page with 2 colors

Bob tosses coin 
If Heads do nothing
If Tails flips page
to green on top.

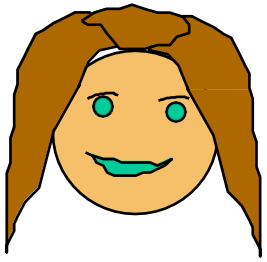
Bob returns altered page

Alice guests Bobs coin'

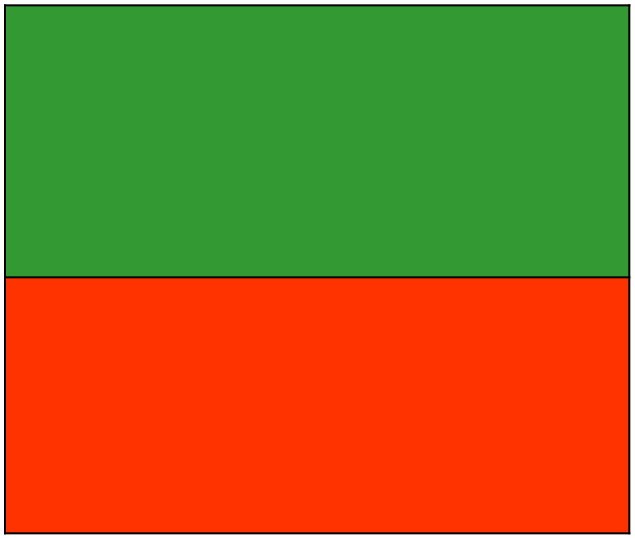
If coin=coin', Bob rejects
Else repeat
After k iteractions, accept

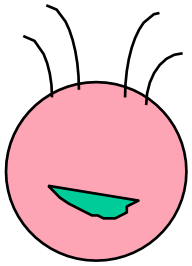
Interactive Proof

Prove to color blind bob that colors exist



Claim: there are 2 colors on this page,





Bob color blind

Completeness: if there are 2 colors, Bob will always accept ,

Soundness: if there is only 1 color, then
Probability[Bob Rejects that after k iterations] $\geq 1-1/2^{100}$