Interactive Proofs And Zero-Knowledge

TODAY:
- Proofs
- Interactive Proofs (IP)
- Zero-Knowledge (ZK)
- A ZK Proof (ZKP) for Graph 3-Coloring

What is a proof?

"truth" "can't cheat" "something is correct"

"evidence"

Involves someone convincing someone else that something is true.

"prover" "verifier" "statement"

We formalize statements as membership in a "language".

↑ set of strings over some alphabet

e.g. ~~"n is prime"~~

$n \in \mathcal{L}_{primes}$ ← set of all prime numbers

instance ↑ Language (also called a decision problem)

statement

more examples: "IS is biprime" $\equiv IS \in \{p \cdot q : p, q \in \mathcal{L}_{prime}\}$ ← the language of biprimes

"ϕ is SAT" $\equiv \phi \in \{\text{formula } \Psi : \exists x \text{ s.t. } \Psi(x) = 1\}$

"ϕ is UNSAT" $\equiv \phi \in \{\text{formula } \Psi : \forall x\ \Psi(x) = 0\}$

"MOVE is a winning chess move" $\equiv$ ...

Now that we have statements, how do I convince you it's true?

I need a "proof" that you can "verify".

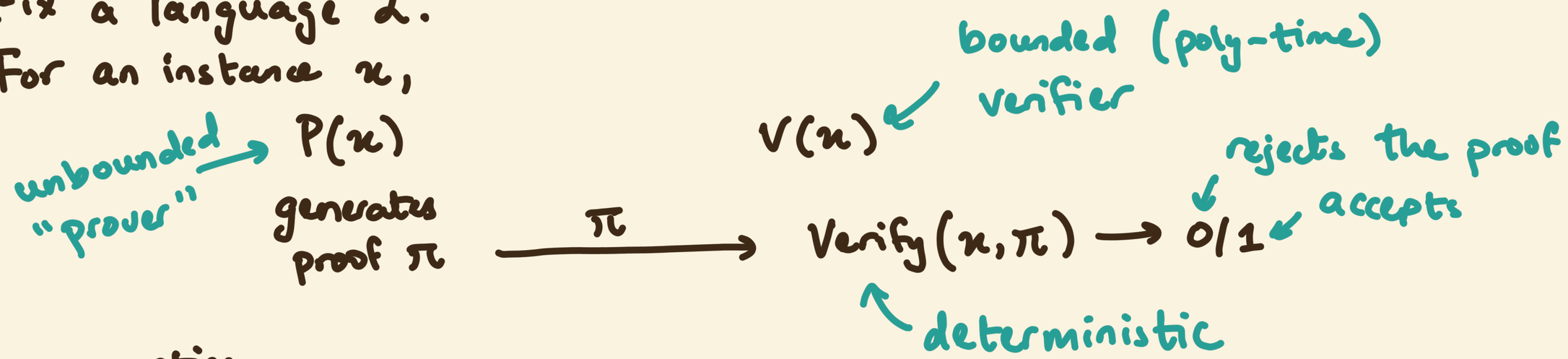
What the type of proof and verification look like is defined by a "proof system": specifies a <u>prover</u> and <u>verifier</u> Turing machine

1 Classical Proofs

Proof is sent and simply read: one-shot or non-interactive

Fix a language $\mathcal{L}$.

For an instance x,



Key properties:

(1) Completeness: $\forall x \in \mathcal{L}$, P outputs π s.t. Verify$(x,\pi) = 1$ $(\exists \pi)$

(2) Soundness: $\forall x \notin \mathcal{L}$, for any π P outputs, Verify$(x,\pi) = 0$ $(\forall \pi)$

In Complexity, a complete and sound system for $\mathcal{L}$

$\Downarrow$

$\mathcal{L} \in NP$ ← complexity class; a set of languages

stands for Non-deterministic Polynomial time

We call π a NP witness or certificate

SAT, 3-COL $\in$ NP

"3-colorable graphs" (we'll see this later)

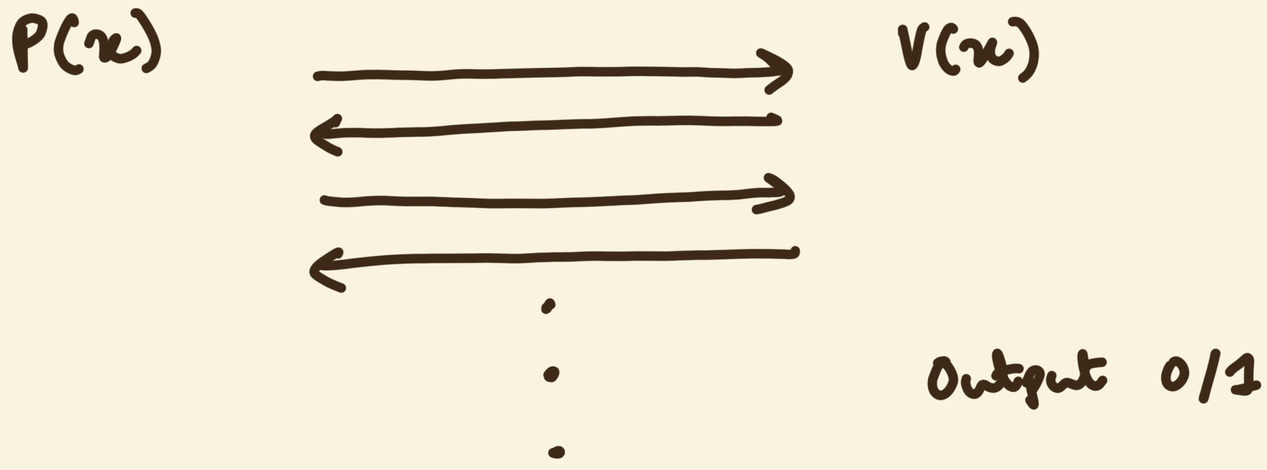
UNSAT $\overset{?}{\in}$ NP

(We don't think so... unless, coNP=NP)

Can we prove more?? E.g. what about $\phi \in$ UNSAT? ← seems hard to come up with a one-shot pf. that ALL assignments are unsatisfiable (but who knows?)

[2] Interactive Proofs [Goldwasser-Micali-Rackoff

Enter interaction, instead of a single message from P to V!



P, V are now randomized, interactive Turing machines

TM with next msg. function:
next(i, msg_i, state_i) → (msg_{i+1}, state_{i+1})
(i: round #)

# of rounds, message length, V complexity are all poly

P is still unbounded.

A language $L \in IP$ "Interactive proofs" if the following properties hold:

- Completeness: $\forall x \in L,\ \Pr_{P,V}[\langle P,V\rangle(x) = 1] \geq 2/3$
- Soundness: $\forall x \notin L,\ \forall P^*\ \Pr_{P^*,V}[\langle P^*,V\rangle(x) = 1] \leq 1/3$

($\langle P,V\rangle(x)$ denotes the output of V when P,V interact on x)

we can amplify these! (to $\geq 1-\text{negl}(\lambda)$ and $\leq \text{negl}(\lambda)$ resp.)

So, how powerful is IP? That is, what languages does it contain?

Thm. IP = PSPACE (!!) [Lund, Fortnow, Karloff, Nisan, Shamir, Shen ~90's]

(!!) believed to be not true! led to many cool subsequent work like MIP = NEXP, PCP theorem, ...

So, very powerful!

Both randomness and interaction are key!

(next HW: IP without randomness = NP)

What about NP + randomness??

Zero-Knowledge

The proof of $x \in L$ may "reveal something" about x

e.g. proving $\phi \in SAT$ may reveal the satisfying instance itself.

Can I prove to you that $x \in L$ and nothing else?

← what does this even mean?

informal examples:
- prove $\phi \in SAT$ without revealing satisfying assignment
- prove $G \in 3\text{-COL}$ without revealing the coloring
- prove you know x s.t. $h = g^x$

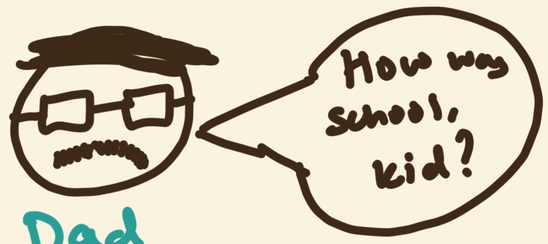
← we'll formalize this in the next class

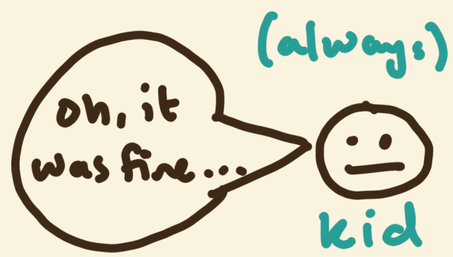
Defining "reveals nothing else" formally

What does it mean to not have learned anything?

Intuitively, you can synthesize what you saw from what you already know

e.g. dad picks up a kid from school





Dad could've *simulated* the conv. himself...

Didn't really learn anything...

In a protocol/interactive proof, the conversation is the "transcript" of messages received and sent, and the input
So, if the Verifier can "fake" the transcript efficiently given what it knows, it's "Zero-Knowledge"

Formally,

Def. (P,V) is honest-verifier zero-knowledge (HVZK) if $\exists$ PPT Sim, s.t. (Sim: fakes a transcript)

$\forall x \in L$

$$\text{view}_V[\langle P,V\rangle(x)] \approx \text{Sim}(x)$$

($\text{view}_V[\langle P,V\rangle(x)]$: notation for V's transcript; $\approx$: computational, statistical or perfect)

What if verifier deviates from the protocol? What can a "malicious" V learn?

Def. (P,V) is zero-knowledge if $\forall$ PPT V^*, $\exists$ Sim s.t.

$\forall x \in L$,

$$\text{view}_{V^*}[\langle P,V^*\rangle(x)] \approx \text{Sim}(x)$$

($\approx$: again, comp./stat./perfect)

If (P,V) satisfies IP-completeness and -soundness,
(P,V) is a (comp., stat. perf.) ZK proof sys

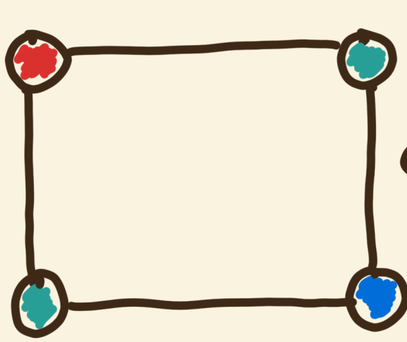
ZK proof system for NP

∃ ZK proof system for any NP language!

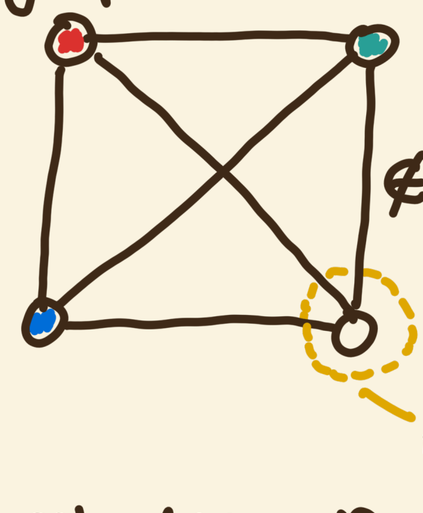
We will show this by constructing ZK proof (ZKP) for the Graph 3-coloring language, which is NP-complete

$$3\text{-COL} := \{G : \text{graph} : G \text{ is } \underline{3\text{-colorable}}\}$$

↖ Can assign labels (or "colors") to vertices s.t. no two adjacent vertices share the same label using at most 3 labels



$\in 3\text{-COL}$



$\notin 3\text{-COL}$

formally, $G = (V, E)$ is 3-colorable if

$$\exists \varphi : V \to \{0,1,2\} \quad \text{s.t.} \quad \forall (u,v) \in E \quad \varphi(u) \neq \varphi(v)$$

φ: coloring; $\{0,1,2\}$: colors; $(u,v) \in E$: adjacent; $\varphi(u) \neq \varphi(v)$: diff. label

The NP-witness for $G \in 3\text{-COL}$ is the coloring, φ.

How do we proceed?

Some observations:

(1) if $G \in 3\text{-COL}$, then by definition $\forall (u,v) \in E$, $\varphi(u) \neq \varphi(v)$

so it's okay to reveal to labels are unequal...

BUT labels themselves may leak info!

So, what if we permute the labels?

Let $\sigma : \{0,1,2\} \to \{0,1,2\}$ be a permutation.

Then, if φ is a valid coloring, so is $\sigma \circ \varphi$.

(2) if $G \notin 3\text{-COL}$, there exists some edge s.t. $\forall \Psi : V \to \{0,1,2\}$, $\Psi(u) = \Psi(v)$

So if I "picked" an edge at random,

prob. of catching "bad" edge is $\geq \frac{1}{|E|}$.

Protocol sketch: P will "commit" to a randomly permuted coloring of G.

V will ask to "open" colors on a random edge.

Since P committed, it can't lie about the coloring it chose.

Recall cryptographic commitment schemes: $Comm : \mathcal{M} \times \mathcal{R} \to \mathcal{C}$ with properties:

- (Comp.) Hiding: $\forall m_0, m_1 \in \mathcal{M}$, $\{Comm(m_0, r) : r \xleftarrow{\$} R\} \approx_c \{Comm(m_1, r) : r \xleftarrow{\$} R\}$
- (Perf.) Binding: $\forall m_0, m_1 \in \mathcal{M}, r_0, r_1 \in R$, if $m_0 \neq m_1$, then $Comm(m_0, r_0) \neq Comm(m_1, r_1)$

A ZKP for 3-COL:

P(G) | **V(G)**

(1) Compute valid 3-coloring, φ.

(2) Sample random permutation,
$\sigma : \{0,1,2\} \to \{0,1,2\}$

(3) Commit permuted coloring to each $v \in V$
$c_v \leftarrow \text{Comm}(\sigma(\varphi(v)), r_v)$

$\{c_v\}_{v \in V}$ →

(4) Sample random edge
$(u,v) \xleftarrow{\$} E$

← (u,v)

(5) Opens commitment to u and v

$(\sigma(\varphi(u)), r_u), (\sigma(\varphi(v)), r_v)$ →

(6) Check
(1) $\sigma(\varphi(u)), \sigma(\varphi(v)) \stackrel{?}{\in} \{0,1,2\}$,
(2) $\sigma(\varphi(u)) \stackrel{?}{\neq} \sigma(\varphi(v))$,
(3) $c_u \stackrel{?}{=} \text{Comm}(\sigma(\varphi(u)), r_u)$,
(4) $c_v \stackrel{?}{=} \text{Comm}(\sigma(\varphi(v)), r_v)$

Output 0 (reject) if any check fails.
Else, output 1.

Completeness. Suppose $G \in 3\text{-COL}$.

Then, P computes a valid 3-coloring φ, and $\sigma(\varphi(u)) \neq \sigma(\varphi(v))$ $\forall$ perm. σ

(Observation (1))

Commitments are det. given randomness

so, all checks pass with prob. 1. ✓

Soundness. suppose $G \notin 3\text{-COL}$. Then $\forall \Psi : V \to \{0,1,2\}$,

$\exists (u,v) \in E$ s.t. $\Psi(u) = \Psi(v)$. (Observation (2))

Let Ψ be the implicit coloring P^* uses in Steps (2) and (3).

(if $Im(\Psi) \subseteq \{0,1,2\}$, V catches with prob. $1/|E|$)

Then. $\exists (u,v) \in E$ s.t. $\Psi(u) = \Psi(v)$

With prob. $1/|E|$, V picks (u,v) in Step (4).

P then must send $(color_u, r_u, color_v, r_v)$.

- if $(color_u = \Psi(u)$ and $color_v = \Psi(v))$ or $(color_u, color_v \notin \{0,1,2\})$, V will reject.
- if $color_u \neq \Psi(u)$, then $c_u \neq Comm(color_u, r_u)$ by perfect binding.
- similarly if $color_v \neq \Psi(v)$, ...

So, V will reject in all cases!

But this only happens when V picks the "bad" edge, which only happens with prob. $\geq 1/|E|$. So, $\Pr[\langle P^*, V\rangle(G) = 1] \leq 1 - \frac{1}{|E|}$ ← we need this to be smaller than 1/3 to be IP...

How do we "amplify" the probability? Repeat!

If V rejects in any iteration, reject!

What is prob. that V accepts in every iteration?

Let t = # of iterations. Then, $\leq \left(1 - \frac{1}{|E|}\right)^t \leq e^{-t/|E|}$

$\Rightarrow$ $(1+x)^n \leq e^x$

So, if we repeat $t \geq |E| \ln(3)$ times,

we get $\Pr[\langle P^*, V\rangle(G) = 1] \leq 1/3$

For crypto, we want $\leq \text{negl}(\lambda)$, so $t \approx \omega(|E|)$, e.g. $t = |E|^2$ works.

HVZK. (computational)

So V behaves honestly, i.e., edges are sampled uniformly at random.

We need to

(1) Construct eff. alg. Sim

(2) Show that $\forall G \in$ 3-COL $\text{view}_V[\langle P, V\rangle(G)] \approx \text{Sim}(G)$

Suppose $G \in$ 3-COL. What is $\text{view}_V[\langle P, V\rangle(G)]$?

It contains (hiding) commitments to all the vertices, a random edge, and an opening of the edge to distinct colors.

How do we simulate?

Easy! Sim will pick a random edge first!

Sim(G):

(a) Sample $(u, v) \xleftarrow{\$} E$

(b) Pick random colors $\text{color}_u \xleftarrow{\$} \{0,1,2\}$, $\text{color}_v \xleftarrow{\$} \{0,1,2\} \setminus \{\text{color}_u\}$

(c) $c_u \leftarrow \text{Comm}(\text{color}_u, r_u)$, $c_v \leftarrow \text{Comm}(\text{color}_v, r_v)$

(d) $\forall w \in V \setminus \{u, v\}$, $c_w \leftarrow \text{Comm}(0, r_w)$.

(e) Output transcript $\{\{c_v\}_{v \in V}, (u, v), ((\text{color}_u, r_u), (\text{color}_v, r_v))\}$

In step (d), we committed to the same color, 0, for all vertices. Most likely not a valid 3-coloring. But V cannot tell since commitment is hiding!

Formally, build hybrids. In each hybrid swap out a simulated commitment with a real commitment. Treat opened edge separately. ← final hybrid is perfectly indist.

If hybrids are distinguishable, can break comp. hiding property of commitment. How?

E.g. $H_0 = \text{view}_V[\langle P,V\rangle(G)] = \{\{c_v\}_{v\in V}, (u,v), ((\sigma(\varphi(u)), r_u), (\sigma(\varphi(v)), r_v))\}$

$H_1 = \{c_{v_1}, \{c_v\}_{v\in V\setminus\{v_1\}}, (u,v), \dots\}$

$v_1 \in V$, a vertex $\neq u, v$ ← $c_{v_1} \leftarrow \text{Comm}(0, r_{v_1})$ as in $\text{Sim}(G)$

$H_0 \approx_c H_1$ because otherwise, can distinguish between $\{\text{Comm}(0,r)\}$ and $\{\text{Comm}(\sigma(\varphi(v_1)),r)\}$ ← Why? Try building the distinguisher!

$\vdots$ $H_{i-1} \approx_c H_i$ $\forall\ i = \{1, \dots, |V|-2\}$

$H_{|V|-2} = \{\{c_w\}_{w\in V\setminus\{u,v\}}, c_u, c_v, (u,v), ((\sigma\circ\varphi(u), r_u), (\sigma\circ\varphi(v), r_v))\}$ ← replaced all vertices $\neq u,v$ with commitment to 0

$H_{|V|-1}$: Sample distinct colors color_u, color_v and set $c_u \leftarrow \text{Comm}(\text{color}_u, r_u)$, $c_v \leftarrow \text{Comm}(\text{color}_v, r_v)$

Replace $c_u \mapsto c_u$, $c_v \mapsto c_v$, $\sigma\circ\varphi(u) \mapsto \text{color}_u$, $\sigma\circ\varphi(v) \mapsto \text{color}_v$

$H_{|V|-2} \equiv H_{|V|-1}$: since σ is random, colors are random in both dist. conditioned on being distinct!

$H_{|V|-1} \equiv \text{Sim}(G)$

Now, what if V is malicious??

Zero-Knowledge. (computational)

1. $\forall\ V^*$, must construct eff. alg. Sim
2. Show Sim satisfies ZK definition:

$\forall G \in 3\text{-Col}$, $\text{view}_{V^*}[\langle P, V^*\rangle(G)] \approx_c \text{Sim}(G)$

Sim(G) can no longer sample edge at random (independently) since V^* may choose its edge based on the commitments it receives (e.g. as a hash of them)

New technique: "rewind" V^* until you get something you can work with...
Good as long as you don't need to rewind too many times!
V^* is just an algorithm (treated as a black-box in this case)

Sim(G): Repeat at most λ times:

(a) Pick random coloring of G, $\varphi: V \to \{0,1,2\}$ ← possibly invalid

(b) Create commitments $c_v \leftarrow \text{Comm}(\varphi(v), r_v) \quad \forall v \in V$

(c) Invoke $(u,v) \leftarrow V^*(G, \{c_v\}_{v \in V})$.

(d) If $\varphi(u) = \varphi(v)$, start again from (a) using fresh randomness

(e) If $\varphi(u) \neq \varphi(v)$, output $(\varphi(u), r_u), (\varphi(v), r_v)$ as the last msg.

(f) If all λ attempts fail, output "$\perp$"

$$\text{Transcript} = \{ \{c_v\}_{v \in V}, (u,v), ((\varphi(u), r_u), (\varphi(v), r_v)) \}$$

↖ from V^*

OR

$\perp$ ← if Sim does not succeed...

First, $\Pr[Sim(G) = \perp]$ = prob. that V^* requests edge with
Since coloring is random, prob of same color = 1/3.
Then, prob. that this happens in all λ attempts
is $(1/3)^\lambda = negl(\lambda)$

So, assume Sim does not output $\perp$. Is $view_{V^*}[\langle P, V^*\rangle(G)] \approx_c Sim(G)$?

Informally, commitments are indist. by commitment hiding,
distribution of edge is identical since drawn from V^*,
final colors opened are random in both $view_V[\langle P, V\rangle(G)]$
and $Sim(G)$ subject to being distinct.

Formally, requires a bit more work. Discuss after class :)