# Interactive Proofs and Zero Knowledge: Definitions and a First Example

CS355 Spring 2025

**https://cs355.stanford.edu**

# Recap

$(C, V)$: a **commitment scheme**

Properties: hiding, binding, succinct

$(C, O, V)$: a **vector commitment scheme**

Commit to a vector $v \in W^n$,
later verifiably open some $v[i]$ for $i \in \{0, \ldots, n-1\}$.

Properties: hiding, binding, succinct

# Notation for the rest of the course

- $\mathbb{N} := \{0,1,2,\dots\}$
- $\{0,1\}^* := \bigcup_{n=0}^{\infty}\{0,1\}^n$ (the set of all finite length binary strings)
- For $x \in \{0,1\}^*$ let $|x| := len(x)$

**Def**: $f: \mathbb{N} \rightarrow [0,1]$ is a **negligible function** if

for every polynomial $p: \mathbb{N} \rightarrow \mathbb{N}$,
$\exists N_p$ s.t. $\forall n > N_p$: $f(n) \leq 1/p(n)$

Examples: $f_1(n) = 10^6/2^n$, $f_2(n) = 1/n^{\log n}$

# Algorithms

(modeled at TM)

$A(x, y)$ is **poly-time** if there is a polynomial $p: \mathbb{N} \rightarrow \mathbb{N}$ s.t.

for all $x, y \in \{0,1\}^*$ : $time\big(A(x, y)\big) \leq p(|x| + |y|)$

$A(x, y)$ is **prob. poly-time** (PPT) if there is a poly. $p: \mathbb{N} \rightarrow \mathbb{N}$ s.t.

for all $x, y \in \{0,1\}^*$ , and all $r \in \{0,1\}^{p(|x|+|y|)}$ :

$$time\big(A(x, y; r)\big) \leq p(|x| + |y|)$$

We write $w \leftarrow A(x, y)$ to denote the random variable

w $:=$ $\{ r \leftarrow \{0,1\}^{p(|x|+|y|)}$ , output $A(x, y; r) \}$

# Algorithms

Let $O_1 : X_1 \to Y_1$ , $O_2 : X_2 \to Y_2$ be functions

we write $\boxed{A^{O_1,O_2}(x,y)}$ to denote an **oracle algorithm** that makes queries to $O_1$ and $O_2$ during its execution.

A call to $O_1(w)$ writes the evaluation of $O_1$ at $w$ to memory in one time unit.

# Relations

A **language** $L$ is a subset of $L \subseteq \{0,1\}^*$

examples: $\emptyset$, $\mathrm{PRIMES} \coloneqq \{ \langle p \rangle \mid p \in \mathbb{N} \text{ is a prime} \}$

$\mathrm{3COL} \coloneqq \{ \langle G \rangle \mid G = (V,E) \text{ is 3-colorable} \}$

A **relation** $R$ is a subset $R \subseteq X \times W$

example: $R_{\mathrm{3COL}} \coloneqq \{ (\langle G \rangle, f) \mid G = (V,E),\ f : V \to \{1,2,3\} \}$

is a valid 3-coloring

$R_{hash} \coloneqq \{ (h, m) \mid \mathrm{SHA256}(m) = h \}$

# Relations

**Def**: for a relation $R$:

(1) $L(R) \coloneqq \left\{ x \in X \;\middle|\; \exists w \in W\colon\ (x, w) \in R \right\} \subseteq \{0,1\}^*$

(2) $R$ is an ***NP*-relation** if there is a poly-time alg. A

s.t. $A(x, w) = 1 \iff (x, w) \in R$

example: $L(R_{3\mathrm{COL}}) = 3\mathrm{COL}$ and $R_{3COL}$ is an *NP*-relation

# Distributions

Let $\Omega$ be a finite set.

**Def**: a **distribution** $P$ on $\Omega$ is a function $P\colon \Omega \to [0,1]$ s.t.

$$\sum_{x \in \Omega} P(x) = 1$$

**Def**: for distributions $P, P'$ on $\Omega$ define the **stat. distance** as

$$\boxed{\Delta(P, P') := \frac{1}{2} \sum_{x \in \Omega} |P(x) - P'(x)| \ \in \ [0,1]}$$

We say that $P, P'$ are ***ε*-close** if $\Delta(P, P') \leq \varepsilon$

# Distributions

Example: $m > n$. Define:

$P$ uniform on $\{1,2,\ldots,n\}$ , $P$’ uniform on $\{1,2,\ldots,m\}$

Then: $\Delta(P,P') \coloneqq \frac{1}{2}\left[n \cdot \left(\frac{1}{n} - \frac{1}{m}\right) + (m-n) \cdot \frac{1}{m}\right] = \boxed{\frac{m-n}{m}}$

⇒ if $m$ and $n$ are “close” then $P$ and $P$’ are “close” in stat. distance

# Statistically indistinguishable distributions

**Def**: distribution ensembles $\{P_\lambda \text{ on } \Omega_\lambda\}_{\lambda\in\mathbb{N}}$ and $\{P'_\lambda \text{ on } \Omega_\lambda\}_{\lambda\in\mathbb{N}}$ are **statistically indistinguishable** if

$\varepsilon(\lambda) \coloneqq \Delta(P_\lambda, P'_\lambda)$ is a negligible function

Example: $P_\lambda$ is uniform on $\{1,2,\ldots,2^\lambda\}$ $\Omega_\lambda = \{1,\ldots,2^\lambda\}$

$P'_\lambda$ is uniform on $\{1,2,\ldots,2^\lambda - 1\}$

Then $\Delta(P_\lambda, P'_\lambda) = 1/2^\lambda$ is a negligible function

We write: $\{P_\lambda\}_{\lambda\in\mathbb{N}} \overset{s}{\approx} \{P'_\lambda\}_{\lambda\in\mathbb{N}}$

# Computationally indistinguishable distributions

**Def**: for two distribution ensembles $\{P_\lambda\}_{\lambda\in\mathbb{N}}$ and $\{P'_\lambda\}_{\lambda\in\mathbb{N}}$ and a PPT algorithm A define

$$\mathrm{Adv}_{\mathrm{A}}(\lambda) := \left|\Pr\left[A\left(1^\lambda, x\right) = 1\right] - \Pr\left[A\left(1^\lambda, x'\right) = 1\right]\right|$$

where $x \leftarrow P_\lambda$ and $x' \leftarrow P'_\lambda$

**Def**: ensembles $\{P_\lambda\}_{\lambda\in\mathbb{N}}$ and $\{P'_\lambda\}_{\lambda\in\mathbb{N}}$ are **comp. indistinguishable**

if for all PPT $A$: $\mathrm{Adv}_{\mathrm{A}}(\lambda)$ is a negligible function

⇒ No PPT algorithm can distinguish $P$ from $P'$ . We write $\{P_\lambda\}_{\lambda\in\mathbb{N}} \overset{\mathrm{c}}{\approx} \{P'_\lambda\}_{\lambda\in\mathbb{N}}$ .

# Stat. indist. ⇒ Comp. indist.

**Lemma**: let $\{P_\lambda\}_{\lambda\in\mathbb{N}}$ and $\{P'_\lambda\}_{\lambda\in\mathbb{N}}$ be two distrib. ensembles.

Then for every algorithm $A$:

$$\boxed{\mathrm{Adv}_{\mathrm{A}}(\lambda) \le \Delta(P_\lambda, P'_\lambda)} \quad \text{for all} \quad \lambda \in \mathbb{N}$$

Proof: by an application of the triangular inequality

**Corollary**: if $\{P_\lambda\}_{\lambda\in\mathbb{N}}$ and $\{P'_\lambda\}_{\lambda\in\mathbb{N}}$ are stat. indistinguishable then they are also computationally indistinguishable.

# Interactive Proofs (IP) [Babai, GMR 1985]

A traditional proof: a long text that can be verified in linear time

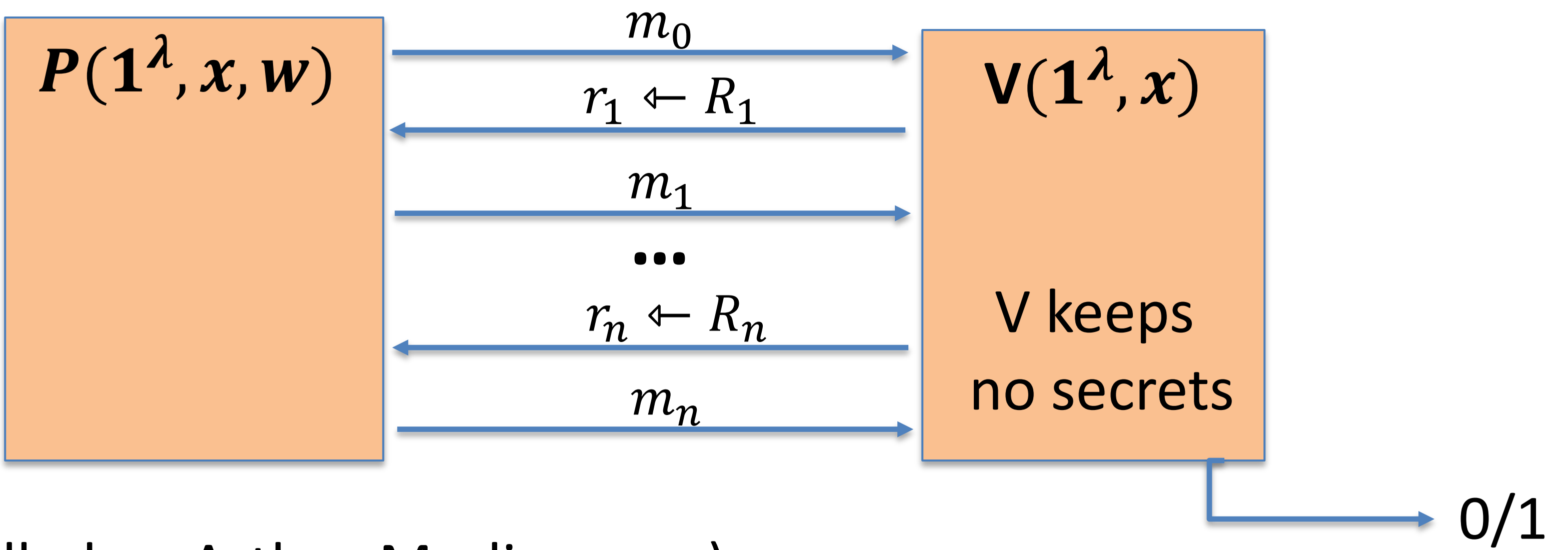
New idea: an **interactive proof** between prover and verifier

**Goal**: for a relation $R \subseteq X \times W$ and $x \in X$

Prover wants to convince Verifier that $x \in L(R)$

# Interactive Proofs (IP) [Babai, GMR 1985]

**Def**: a **(public coin) interactive proof (IP)** for a relation $R \subseteq X \times W$ is a pair of PPT algorithms $(P, V)$ that operate as



(also called an Arthur-Merlin game)

# Interactive Proofs (IP)

**Notation**:

- $\mathrm{out}_\lambda[P,V](x) :=$ output of $V$ at end of interaction with $P$
- $\mathrm{tr}_\lambda[P,V](x) := (x, m_0, r_1, \dots, r_n, m_n)$

called the **transcript** (Verifier's view)

**Def**: $(P,V)$ is **(perfectly) complete** if for all $(x,w) \in R$

$$\Pr[\mathrm{out}_\lambda[P,V](x) = 1] = 1 \qquad \text{for all } \lambda \in \mathbb{N}$$

# Interactive Proofs (IP)

**Notation**:

- $\mathrm{out}_\lambda[P,V](x)$ := output of $V$ at end of interaction with $P$

**Def**: $(P,V)$ is **sound** if for all $x \notin L(R)$ and all $P^*$ :

$$\varepsilon(\lambda) \coloneqq \Pr[\mathrm{out}_\lambda[P^*,V](x) = 1]$$ is a negligible function

$\varepsilon(\lambda)$ is called the **soundness error** of $(P,V)$.

**Def**: $(P,V)$ is **computationally sound** if soundness only holds against PPT provers $P^*$ . (an unbounded prover may fool $V$)

# Interactive Proofs (IP)

If $R$ is an *NP*-relation, then the trivial I.P. for $R$: send $w$ to $V$

To disqualify the trivial I.P. we add two requirements:

(1) $(P, V)$ is **short** if $|\mathrm{transcript}|$ is must less than $|w|$

(2) $(P, V)$ should be **honest verifier zero knowledge** (HVZK)

Each of these requirements, on its own, disqualifies the trivial I.P.

# Honest Verifier Zero-Knowledge (HVZK)

Let $(P, V)$ be a **(public coin) interactive proof (IP)** for a relation $R \subseteq X \times W$

<u>**Goal**</u>: For $x \in X$ Prover wants to convince Verifier that $x \in L(R)$ without revealing "any other information"

How to define this?

- Verifier sees the transcript: $\mathrm{tr} = (x, m_0, r_1, \ldots, r_n, m_n)$
- **Key idea**: $V$ leans nothing from tr if it can generate tr by itself, just given $x$. We say that $V$ can **simulate** the transcript.

# Honest Verifier Zero-Knowledge (HVZK)

**Def**: $(P, V)$ is **honest verifier zero knowledge** (HVZK) if

there exists a PPT simulator $S$ s.t. for all $(x, w) \in R$

(1) **Perfect HVZK:** $\{S(1^\lambda, x)\}_{\lambda \in \mathbb{N}} \equiv \{\mathrm{tr}_\lambda[P, V](x)\}_{\lambda \in \mathbb{N}}$

(2) **Stat. HVZK:** $\{S(1^\lambda, x)\}_{\lambda \in \mathbb{N}} \overset{\mathrm{s}}{\approx} \{\mathrm{tr}_\lambda[P, V](x)\}_{\lambda \in \mathbb{N}}$

(3) **Comp. HVZK:** $\{S(1^\lambda, x)\}_{\lambda \in \mathbb{N}} \overset{\mathrm{c}}{\approx} \{\mathrm{tr}_\lambda[P, V](x)\}_{\lambda \in \mathbb{N}}$

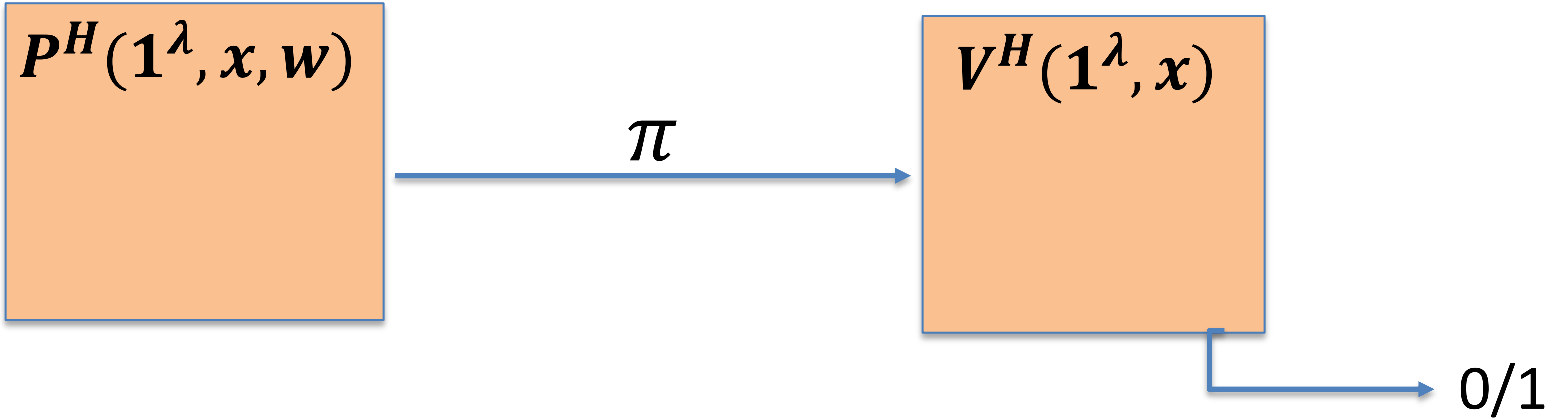
For $(x, w) \in R$, simulator shows that transcript can be generated from $x$ alone

⇒ anything V got from transcript, it could have generated on its own

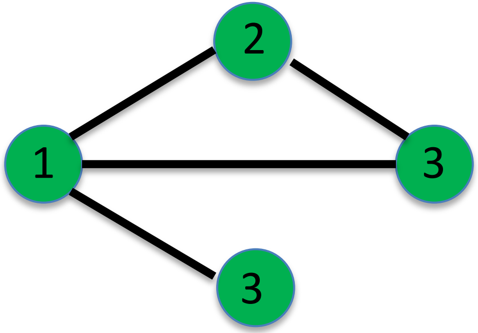
# Is interaction necessary?

We will later see a transformation:

(public-coin) interactive protocol ⇒ a non-interactive protocol



# An HVZK for $R_{3\mathrm{COL}}$



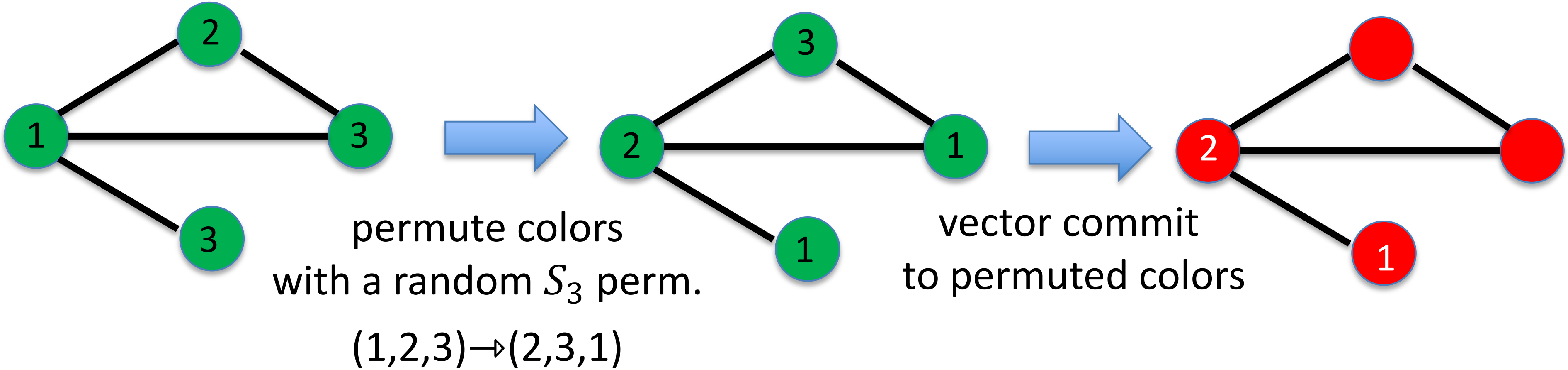
$G = (V, E)$

$f : V \rightarrow \{1,2,3\}$

# An HVZK I.P. for $R_{3\mathrm{COL}}$ [GMW'86]

$\underline{P\left(1^\lambda,\ G=(V,E),\ f: V \to \{1,2,3\}\right)}$

$\underline{V\left(1^\lambda, G=(V,E)\right)}$



permute colors
with a random $S_3$ perm.

(1,2,3)→(2,3,1)

vector commit
to permuted colors

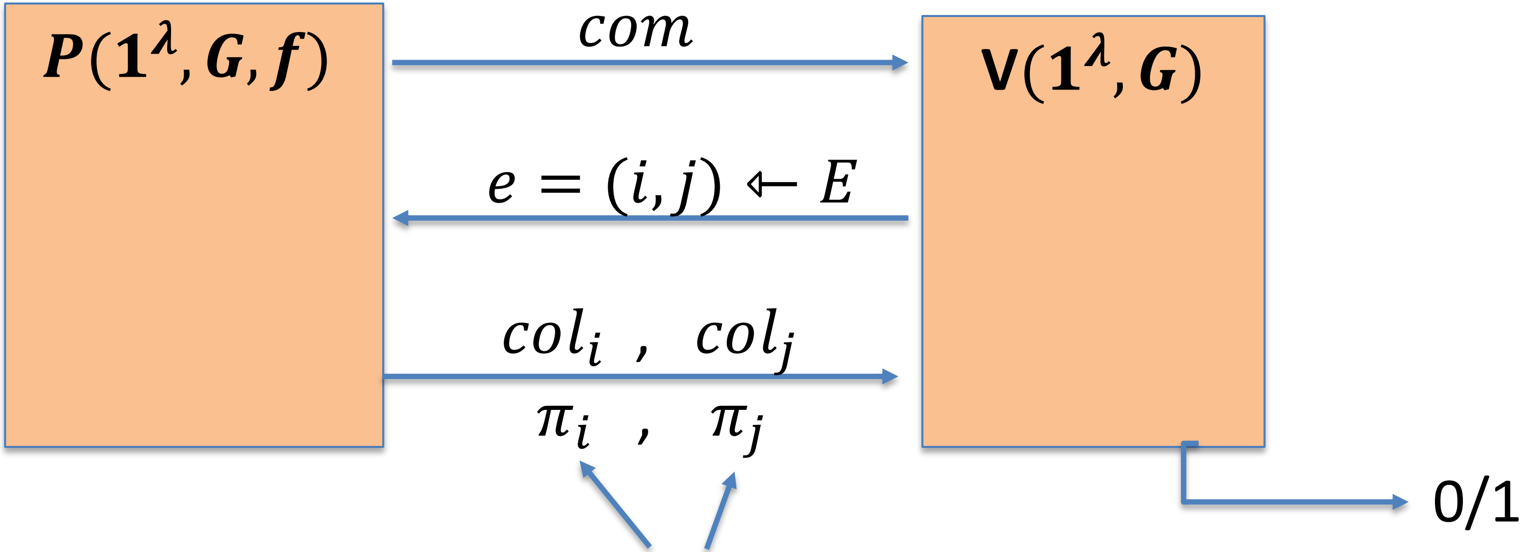
random edge $e=(i,j) \leftarrow E$

open colors of nodes $i$ and $j$

accept if color($i$)≠color($j$)

# An HVZK I.P. for $R_{3\mathrm{COL}}$ [GMW'86]

Protocol sketch:



$\boldsymbol{P(1^\lambda, G, f)}$

$com$

$\mathbf{V}(\mathbf{1}^\lambda, \boldsymbol{G})$

$e = (i, j) \leftarrow E$

$col_i \;,\; col_j$

$\pi_i \;,\; \pi_j$

0/1

opening proofs
for vector commitment

# An HVZK I.P. for $R_{3\mathrm{COL}}$

This is perfectly complete

- Is it computationally sound?

- Is it HVZK?

# Proof of HVZK

**Claim**: (P,V) is a statistical HVZK for $R_{3\text{COL}}$

Proof: Let $(G, f) \in R_{3\text{COL}}$ . We build a simulator $S(1^\lambda, G)$:

- sample $e = (i, j) \leftarrow E$ and $a, a' \leftarrow \{1,2,3\}$ s.t. $a \neq a'$
- set $u' := (1,1, \ldots, \underset{\text{pos. } i}{\underset{\uparrow}{a}}, \ldots, \underset{\text{pos. j}}{\underset{\uparrow}{a'}}, \ldots 1,1) \in \{1,2,3\}^{|V|}$
- $com \leftarrow \text{VectorCommit}(1^\lambda, u', r)$
- Build opening proofs $\pi, \pi'$ for positions $i$ and $j$
- output $\text{tr} := (com, e, a, a', \pi, \pi')$

# Proof of HVZK

**Claim**: (P,V) is a statistical HVZK for $R_{\mathrm{3COL}}$

- set $u' := (1,1, \ldots, a, \ldots, a', \ldots 1,1) \in \{1,2,3\}^{|V|}$
- output $\mathrm{tr} := (com, e, a, a', \pi, \pi')$

(1) The vector commitment is unconditionally hiding $\Rightarrow$

$$\left\{\mathrm{VectorCommit}\left(1^\lambda, \underline{u'}, r\right)\right\}_{\lambda\in\mathbb{N}} \overset{s}{\approx} \left\{\mathrm{VectorCommit}\left(1^\lambda, \underline{\mathrm{real}\ u}, r\right)\right\}_{\lambda\in\mathbb{N}}$$

(2) $e, a, a', \pi, \pi'$ : are distributed exactly as in a real transcript ■

Puzzle: would the protocol be HVZK if $V$ chose $(i,j) \leftarrow |V|^2$ ??

# Computational Soundness

Suppose the vector commitment is unconditionally binding.

**Claim**: if $G \notin L(R_{3\mathrm{COL}})$ then for all PPT $P^*$

$$\Pr[\ \mathrm{out}_\lambda[P,V](G) = 1\ ] \le 1 - \frac{1}{|E|}$$

not negligible!

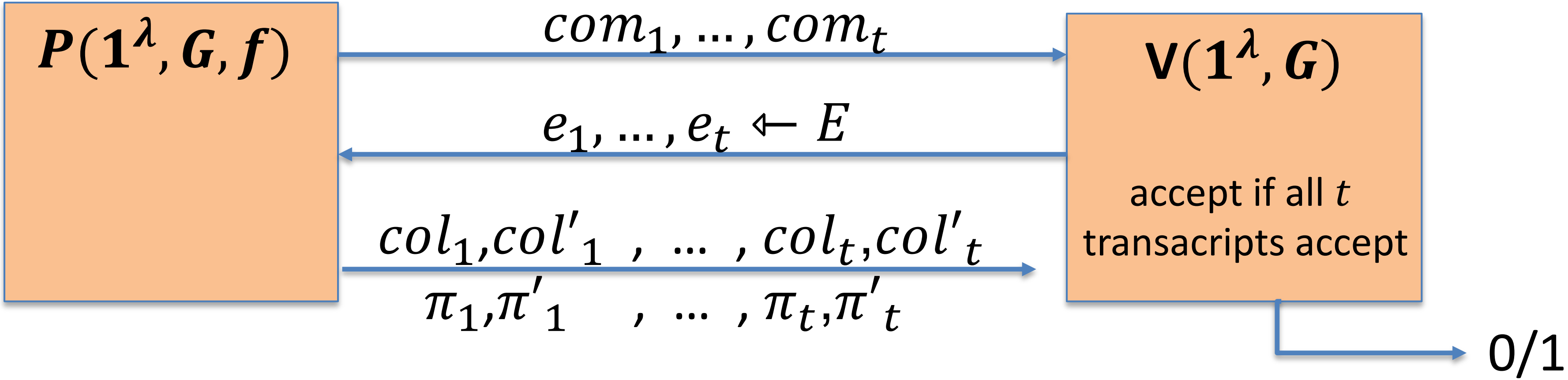
Proof idea: suppose $com$ is a commitment to some $f \in \{1,2,3\}^{|V|}$

Then: $(G,f) \notin R_{3\mathrm{COL}} \ \Rightarrow\ \exists\, e^* = (i,j) \in E$ s.t. $f[i] = f[j]$

$$\Pr[\,V \text{ chooses } e^*] = \frac{1}{|E|} \quad \Rightarrow \quad \Pr[\ \mathrm{out}_\lambda[P,V](G) = 0\ ] \ge \frac{1}{|E|}$$ ■

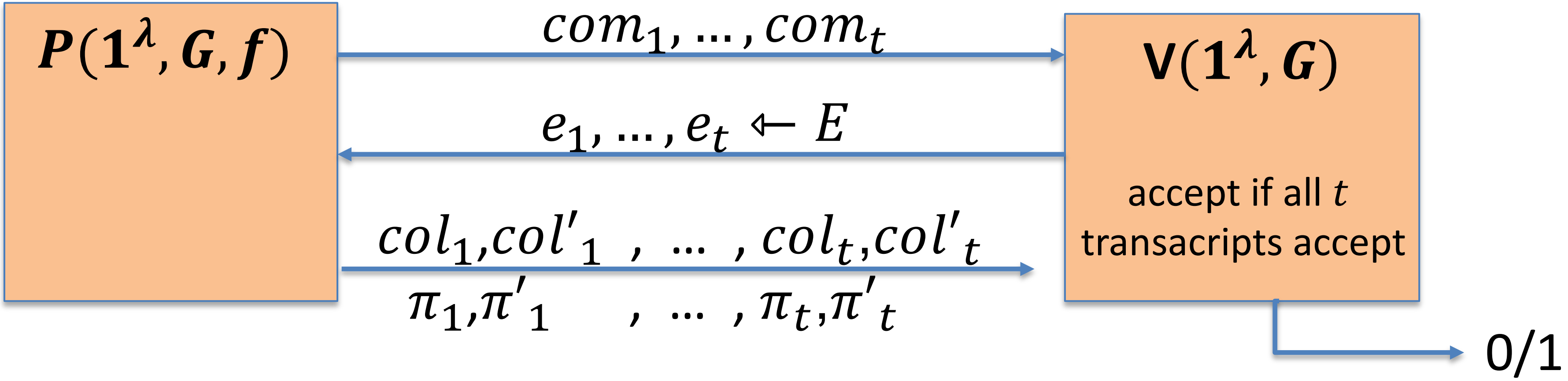
# Amplification by parallel composition

To reduce soundness error to $1/e^{\lambda}$ repeat protocol in parallel $t = \lambda \cdot |E|$ times. Verifier accepts if all iterations accept.


$\boldsymbol{P(1^{\lambda}, G, f)}$
$com_1, \dots, com_t$
$\mathbf{V(1^{\lambda}, G)}$
$e_1, \dots, e_t \leftarrow E$
accept if all $t$ transcripts accept
$col_1, col'_1 \;, \;\dots\; , col_t, col'_t$
$\pi_1, \pi'_1 \;, \;\dots\; , \pi_t, \pi'_t$
0/1

Now: $(G, f) \notin R_{\text{3COL}} \;\Rightarrow\;$ Pr[ $V$ accepts] $\leq \left(1 - \frac{1}{|E|}\right)^{t} \approx 1/e^{\lambda}$

# Amplification by parallel composition



$\boldsymbol{P}(\mathbf{1}^{\lambda}, \boldsymbol{G}, \boldsymbol{f})$

$com_1, \dots, com_t$

$e_1, \dots, e_t \leftarrow E$

$col_1, col'_1 \;,\; \dots \;,\; col_t, col'_t$

$\pi_1, \pi'_1 \;,\; \dots \;,\; \pi_t, \pi'_t$

$\mathbf{V}(\mathbf{1}^{\lambda}, \boldsymbol{G})$

accept if all $t$ transcripts accept

0/1

Note: length of transcript is $O(|E|)$ $\Rightarrow$ not short

**Lemma**: $(P, V)$ is HVZK $\Rightarrow$ $(P^t, V^t)$ is also HVZK

(not true for regular ZK)

Important point: $3COL$ is NP-complete ⇒ all of NP has HVZK I.P.

# END OF LECTURE

Next lecture: a succinct I.P. for every NP-relation