RANDOMNESS CONDENSERS FOR EFFICIENTLY SAMPLABLE, SEED-DEPENDENT SOURCES

Joint work with Thomas Ristenpart and Salil Vadhan

Yevgeniy Dodis (New York University)

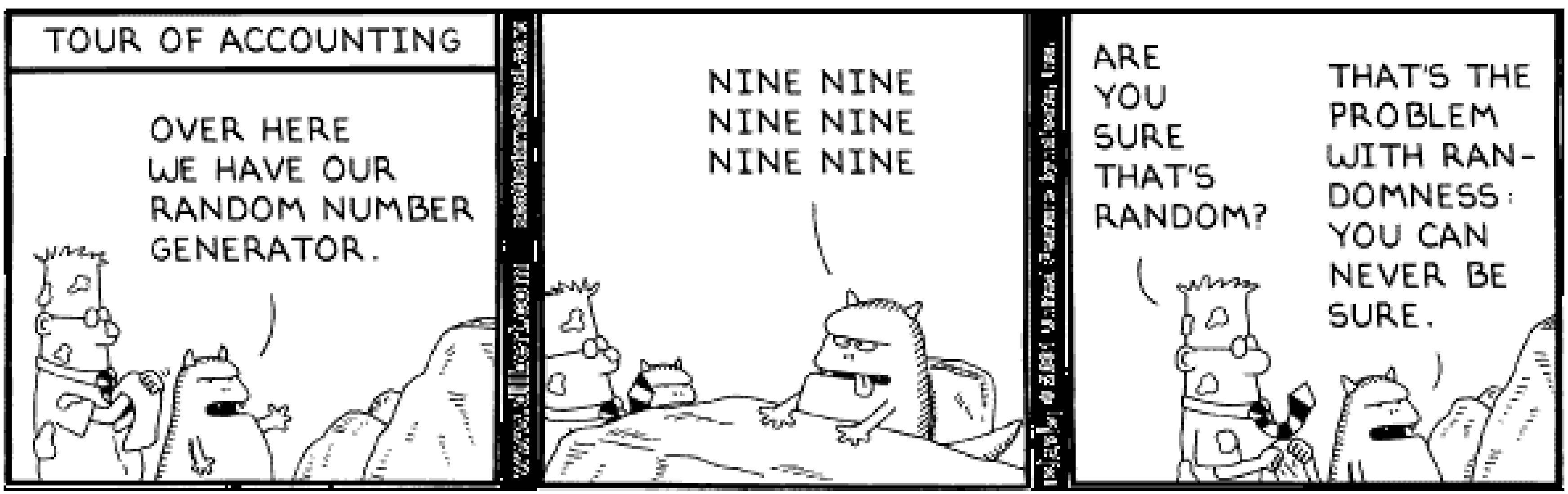
Imperfect Random Sources

- Ideal randomness is crucial in many areas
  - Especially cryptography (i.e., secret keys) [MP91,DOPS04,BD07]
- However, often deal with imperfect randomness
  - physical sources, biometric data, partial knowledge about secrets, extracting from group elements (DH key exchange),…
- Necessary assumption: must have (min-)entropy
  - $\mathbb{H}_\infty(X) \geq k$ if $\Pr[X=x] \leq 2^{-k}$, for all x
- Can we extract (nearly) perfect randomness from such realistic, imperfect sources?

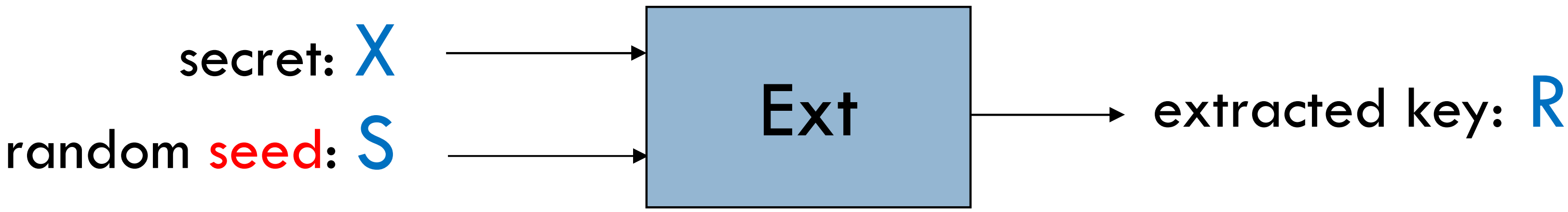
Extractors

secret: X → **Ext** → extracted key: R

- Problem: can't handle general entropy sources
  - Let $X \leftarrow \mathrm{Ext}^{-1}(\mathrm{const})$. High entropy, but $\mathrm{Ext}(X)=\mathrm{const}$



Seeded Extractors [NZ96]



- R is uniformly random even conditioned on the seed S

$$(\mathbf{Ext}(X; S), S) \approx (\text{Uniform}, S)$$

- <u>Advantages</u>:
  - Can extract (almost) all entropy from all k-sources
  - Efficient constructions (leftover hash lemma, no "crypto")
  - Seed can be reused
  - In theory, can make seed very short (often not critical)

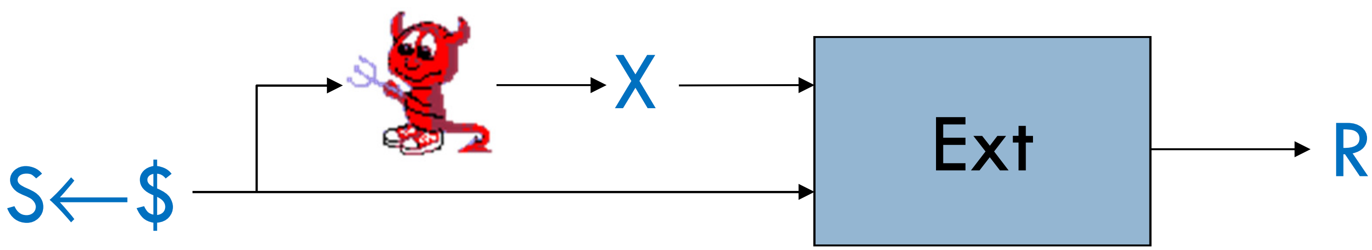
Disadvantages

- Need a (truly random!) seed in the first place
  - Defenses: can arrange in most settings, can be reused
- Must lose some entropy due to extraction
  - Defenses: pretty small, can use PRG to stretch, provably less than we thought for many applications [BDK+11]
- **This work**: seed must be *independent* from the source
  - Main Defense: OK for many applications (e.g., DH exchange)
    - But not all (e.g., RNG computation affects physical source it uses)
    - May find new unexpected applications (stay tuned!)
    - The question is obviously intriguing, let's move on !

Seed-Dependent Extractors

"The tax man says you can't claim your inner child as a dependent."

- (**Ext**(X; S), S) ≈ (Uniform, S), as long as $\boxed{\mathbb{H}_\infty(X \mid S) \geq k}$
- Impossible ☹: same X←Ext^{-1}(const; S) argument
- What if X is efficiently samplable (+ Ext^{-1} is "hard")?



- [TV00]: only possible if complexity of  is (roughly) less than that of the extractor ☹
  -  : keep picking random X until first bit of Ext(X; S)=0

The Attack is Not So Bad !!

"You're paralyzed from the knuckle of your left big toe down – it could have been a lot worse."

- Assume use $R = \mathrm{Ext}(X; S)$ as a secret key
  - If $R=0|$random, then only lost factor of 2 in security!
- <u>Generalization</u>: pick $X \leftarrow \$$ until $\mathrm{Ext}(X; S)$ is "weak"
  - Sampling time $t \approx \frac{1}{\varepsilon}$, where ε = fraction of weak keys
  - Super-polynomial if ε = negligible !
- Is this the best attack?
- Can we formalize a sufficient security notion?

RANDOMNESS CONDENSERS!

- Same syntax as Extractor:

secret: $X \in \{0,1\}^n$ → **Cond** → extracted key: $R \in \{0,1\}^m$

seed: $S \in \{0,1\}^p$ → **Cond**

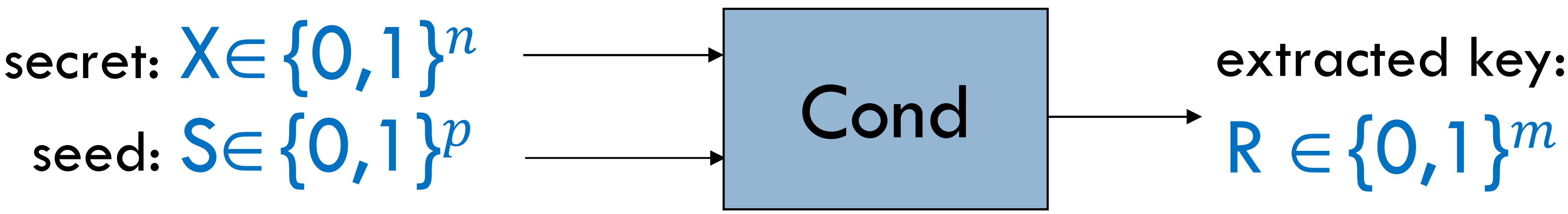
- **Standard Definition**: Cond is $(\frac{k}{n} \rightarrow \frac{v}{m})_\infty$–*condenser* if

$$\mathbb{H}_\infty(X) \geq k \Rightarrow \mathbb{H}_\infty(\mathrm{Cond}(X;\, S) \mid S) \geq v$$

  - Note: no restriction on X being efficiently samplable
  - Non-triviality: want entropy deficiency $m - v \ll n - k$

THE SEED-DEPENDENT CONDENSER

- Same syntax as Extractor:



- **Definition**: Cond is $(\frac{k}{n} \to \frac{v}{m}, t)_\infty$*–seed-dependent (SD) condenser*, if for all A producing $X \leftarrow A(S)$ in time t,

$$\mathbb{H}_\infty(X \mid S) \geq \boldsymbol{k} \Rightarrow \mathbb{H}_\infty(\mathrm{Cond}(X;S) \mid S) \geq \boldsymbol{v}$$

  - As before, want entropy deficiency $d = m - v \ll n - k$
  - Unlike Extractors, Cond can be much faster than A !

Condensers and Key Derivation

- Setting: application P needs a m–bit secret key R
  - Ideal Model: $R \leftarrow U_m$ is uniform
  - Real Model: $R \leftarrow$ Cond(X; S), where $\mathbb{H}_\infty(X \mid S) \geq k$
- Assumption: P is ε–secure in the ideal model
- Desired Conclusion: P is ε'–secure in the real model
- Observation: if Cond is $(\frac{k}{n} \rightarrow \frac{v}{m}, t)_\infty$–SD-*condenser* and $X \leftarrow A(S)$ is sampled in time at most t, then

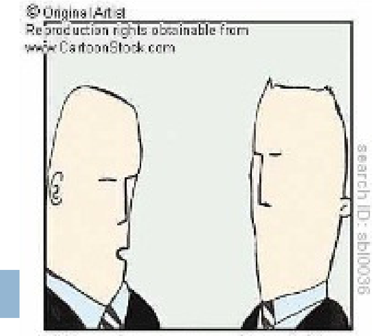
  ε'≤ [security of P with key R s.t. $\mathbb{H}_\infty(R) \geq v$]

  - Reduces key derivation to analysis of P under weak keys!
  - Ahead: *generic* bounds on ε' from ε and $2^{m-v} = 2^d$

Pedantic Viewpoint

- Fix P and any "legal" attacker ***B***
- Let f(r) = [Advantage of ***B*** on key r]
  - Unpredictability apps: f(r)$\in$[0,1]
  - Indistinguishability apps: f(r) $\in$[-½ , ½]
- Ideal adv. of ***B*** $= |\mathbb{E}[\mathrm{f}(U_m)]| = \left|\sum_r \frac{1}{2^m} \cdot f(r)\right|$
- Real adv. of ***B*** $= |\mathbb{E}[\mathrm{f}(R)]| = |\sum_r p(r) \cdot f(r)|$
- <u>Goal</u>: upper bound real advantage of ***B***

Unpredictability Applications



"Massive unpredictablity is absolutely certain, maybe."

- **<u>Lemma1</u>**: If f(r) ≥ 0 and $\mathbb{H}_\infty(R) \geq \boldsymbol{m - d}$ then

$$\boxed{\mathbb{E}[\, f(R)\,] \leq 2^d \cdot \mathbb{E}[\, f(U_m)\,]}$$

- Proof: $\sum p(r)\cdot f(r) \leq 2^m \cdot \max_r (p(r)) \cdot (\sum \frac{1}{2^m} \cdot f(r))$ ∎
- <u>Corollary</u>: any (T, ε)-secure *unpredictability* app. P in the ideal model is also (T, ε')-secure in the $(\boldsymbol{m - d})$-real$_\infty$ model, where $\boxed{\varepsilon' \leq 2^d \cdot \varepsilon}$
  - Exponential loss: OK if negl. ε and polyn. 2^d

Indistinguishability Apps

- **Col**$(R) = \Pr[R_1{=}R_2] = \sum p(r)^2$
  - Renyi: $\mathbb{H}_2(R) = -\log \textbf{Col}(R) \geq \mathbb{H}_\infty(R)$
- **Lemma2**: For all f and $\mathbb{H}_2(R) \geq \boldsymbol{m - d}$,

$$\boxed{|\mathbb{E}[\mathsf{f}(R)]| \leq \sqrt{2^d \cdot \mathbb{E}[\mathsf{f}(U_m)^2]}}$$

- Proof: $|\mathbb{E}[\mathsf{f}(R)]| = |\sum_r p(r) \cdot f(r)|$
- Cauchy-Schwartz:

$$\leq \sqrt{2^m \sum p(r)^2} \cdot \sqrt{\frac{1}{2^m} \sum f(r)^2} \qquad \blacksquare$$

Why is it Nice?

- **Lemma2**: For all f and $\mathbb{H}_2(R) \geq \boldsymbol{m - d}$,

$$\boxed{|\,\mathbb{E}[\mathrm{f}(R)]\,| \leq \sqrt{2^d \cdot \mathbb{E}[\mathrm{f}(U_m)^2]}}$$

  - Works even if f(r) can be negative
  - Renyi entropy $\mathbb{H}_2$ is better than $\mathbb{H}_\infty$
  - Second term is for uniform distribution
- Question: $\mathbb{E}[\mathrm{f}(U_m)] = \varepsilon$, what is $\mathbb{E}[\mathrm{f}(U_m)^2]$?
- Def ([BDK+11]): P is (T, σ)-square secure if for any T-bounded B, $\mathbb{E}[\mathrm{f}_{\mathrm{B}}(U_m)^2] \leq \sigma$

Square Security?

- [BDK+11]: for many natural apps "$\sigma \approx \varepsilon$" (unpredictability, CPA-encryption, weak PRF)
- **Corollary**: Assume P is (T, ε)-secure and "square-friendly". Then P is (T, ε')-secure in the $(m-d)$-real$_2$ model, where $\boxed{\varepsilon' \leq \sqrt{2^d \cdot \varepsilon}}$
  - lost sqrt, but more apps and better $\mathbb{H}_2$ entropy
- In fact, using $\mathbb{H}_\infty$-condensers + Lemma1 got same bounds than $\mathbb{H}_2$-condensers + Lemma2
  - so concentrate on $\mathbb{H}_2$ case

Collisions and Conderensers

- **Theorem**: "Strong enough" collision-resistant hash functions $\{h\}$ are "good" $\mathbb{H}_2$-SD-condensers!
  - Partially explains the use of cryptographic hash for KDF !
- **Formally**: $(2t, \frac{A(t)}{2^m})$-CRHF $\mathcal{H} = \{h:\{0,1\}^n \rightarrow \{0,1\}^m\}$ defines a *seed-dependent* $(\frac{k}{n} \rightarrow \frac{m-d}{m}, t)_2$-condenser $\text{Cond}(x; h) = h(x)$, where $2^d = 2^{m-k} + A(t)$
- $\Pr[h(X_1) = h(X_2)] \leq \Pr[X_1 = X_2] +$

$$\Pr[h(X_1) = h(X_2)\ \&\ X_1 \neq X_2]$$

$$\leq 2^{-k} + \varepsilon_{\text{crhf}}$$

  - Otherwise, find collisions by simply sampling X_1, X_2 !

Collisions and Condensers

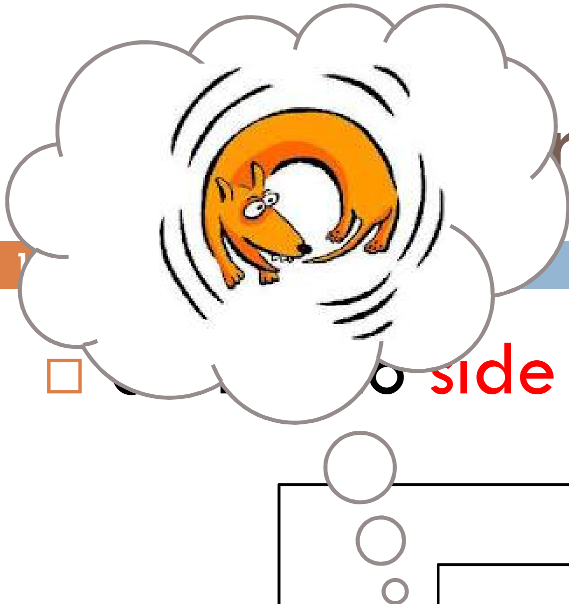
- **Theorem**: "Strong enough" collision-resistant hash functions $\{h\}$ are "good" $\mathbb{H}_2$-SD-condensers!
  - Partially explains the use of cryptographic hash for KDF !
- **Formally**: $(2t, \frac{A(t)}{2^m})$-CRHF $\mathcal{H} = \{h:\{0,1\}^n \rightarrow \{0,1\}^m\}$ defines a *seed-dependent* $(\frac{k}{n} \rightarrow \frac{m-d}{m}, t)_2$-condenser $\mathrm{Cond}(x; h) = h(x)$, where $2^d = 2^{m-k} + A(t)$
  - E.g., if $A(t) = O(t^2)$ and $k \geq m \Rightarrow 2^d = O(t^2)$
- **Corollary**: $\mathcal{H}$ is $(2t, \frac{O(t^2)}{2^m})$-CRHF $\Rightarrow$ $\boxed{\varepsilon' \leq O(t \cdot \sqrt{\varepsilon}\ )}$ for all "square-friendly" ε–secure applications P, against any t-samplable X s.t. $\mathbb{H}_2\,(X \mid h) \geq m$

Collisions and Condensers

- **Theorem**: "Strong enough" collision-resistant hash functions $\{h\}$ are "good" $\mathbb{H}_2$-SD-condensers!
  - Partially explains the use of cryptographic hash for KDF !

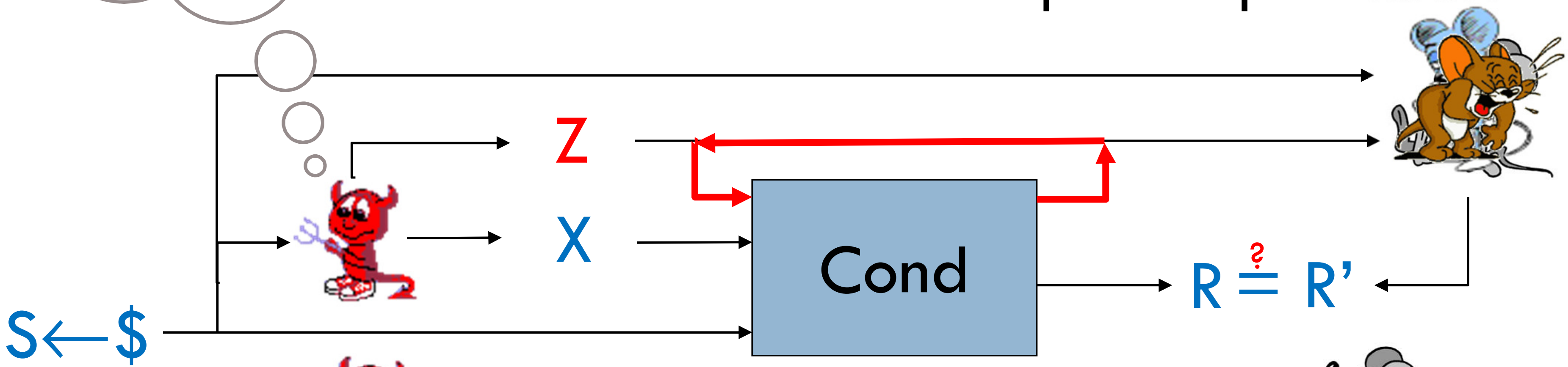
Asymptotic View: negligible ideal security ε
+ polynomial sampling time t $\Rightarrow$
negligible real security ε'

- **Corollary**: $\mathcal{H}$ is $(2t, \frac{O(t^2)}{2^m})$-CRHF $\Rightarrow$ $\boxed{\varepsilon' \leq O(t \cdot \sqrt{\varepsilon}\ )}$ for all "square-friendly" ε–secure applications P, against any t-samplable X s.t. $\mathbb{H}_2\,(X \mid h) \geq m$



…mation

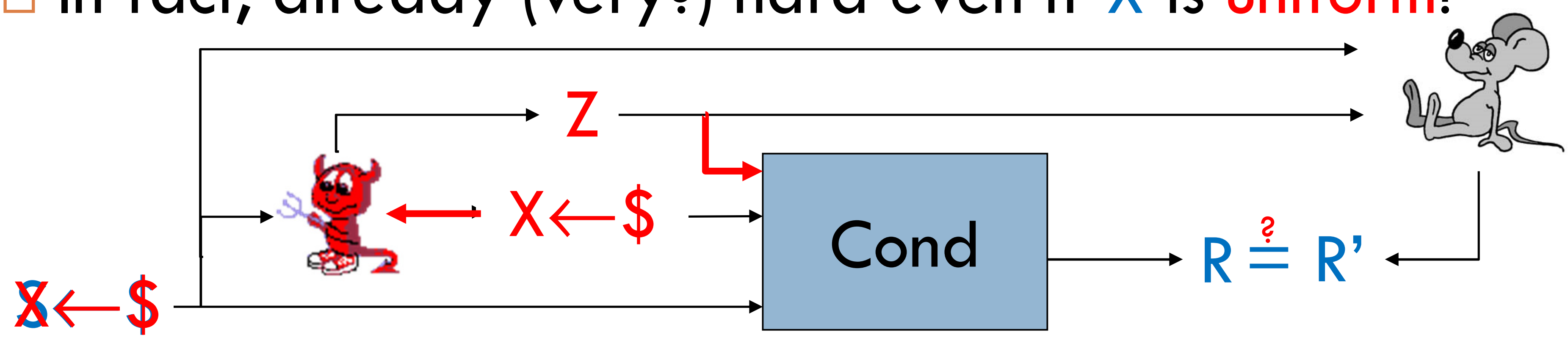
- …o side information from sampler to predictor:



- What if  can pass side information Z to  ?
  - Require $\mathbb{H}_\infty(X \mid S, Z) \geq k$; natural in many settings
- **<u>Problem</u>**:  can now make $Z = \mathrm{Cond}(X; S)$ ☹
- "**<u>Solution</u>**": pass Z to condenser! $R = \mathrm{Cond}((X,Z); S)$
  - Why would  pass Z to Cond? Stay tuned…

Warning: Strong Generalization!

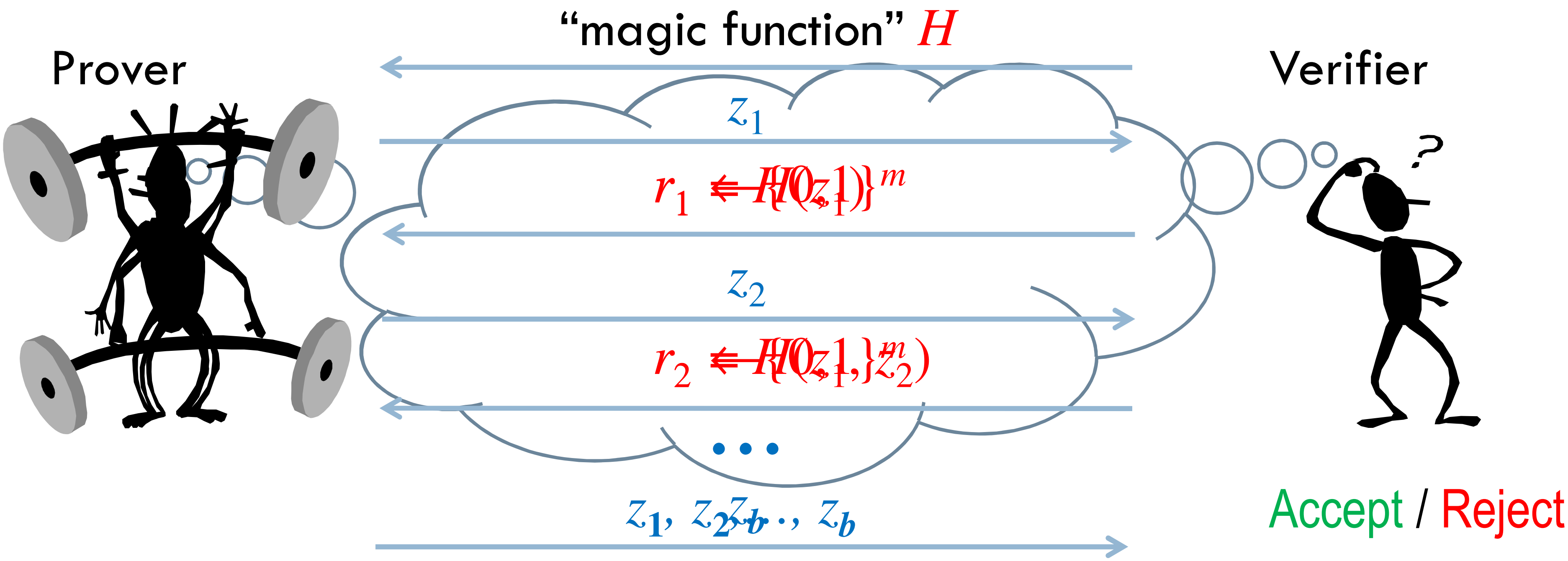
- Conjecture SD-condensers with side info exist, but…
- CRHF scheme no longer works with side information
  - Hard to sample X_1, X_2 conditioned on the same Z
- In fact, already (very?) hard even if X is uniform!



  - Call this important special case Leaky Condenser
- Leaky Condensers enough to instantiate Fiat-Shamir !

Fiat-Shamir

- Public-coin $(2b+1)$-round prot. $\Rightarrow$ public-coin 2-round prot.



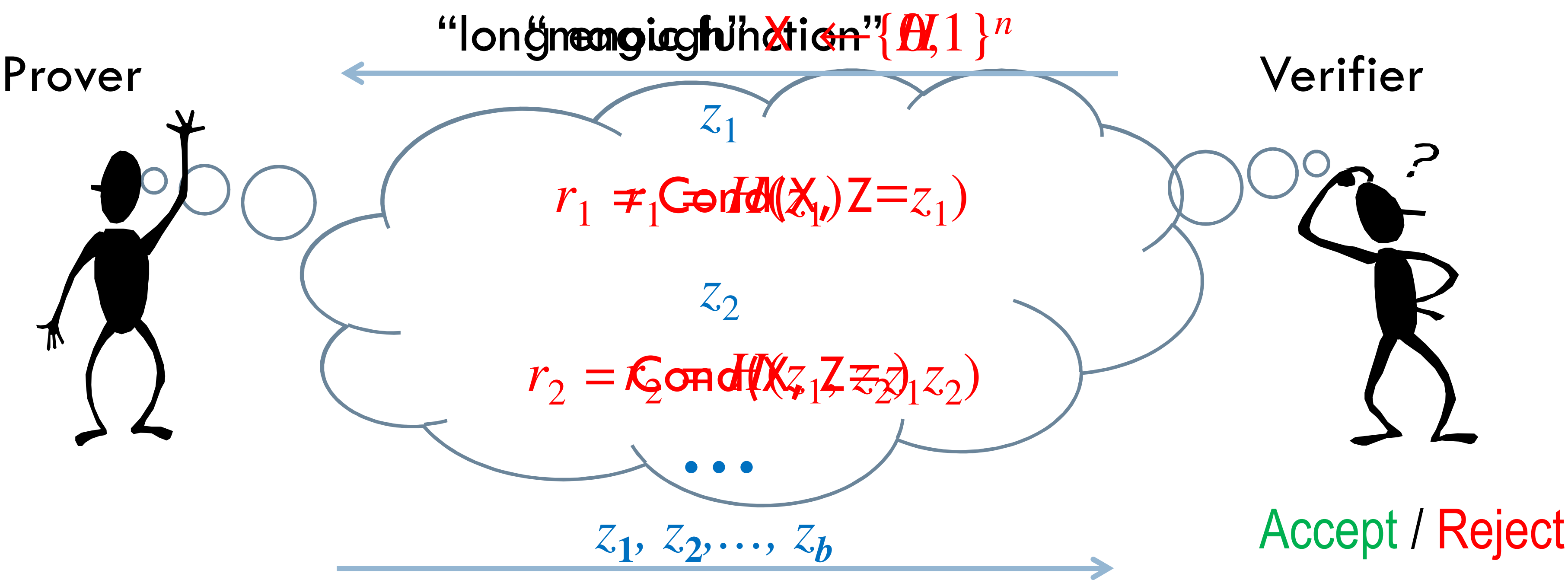
- **<u>Assume</u>**: ε-sound against unbounded Prover (proof)
- **<u>Conclude</u>**: ε’-sound against bounded Prover (argument)

Soundness of Fiat-Shamir?

- True in random oracle model
- Not necessarily true for <u>arguments</u> [Bar01,GK03]
- **<u>Conjecture</u>** [BLV06]: true for constant-round <u>proofs</u>
  - Implies no constant-round, public-coin, ZK proofs outside BPP
- **<u>Our result</u>**: soundness of FS on interactive proofs almost equivalent to existence of non-trivial Leaky Condensers
  - Entropy deficiency d (for $2b+1$ rounds) $\Rightarrow$ $\boxed{\varepsilon' \leq 2^{db} \cdot \varepsilon}$
  - E.g., $2^d = \mathrm{poly}(t)$ and $b = \mathrm{O}(1)$ $\Rightarrow$ $\boxed{\varepsilon' \leq \mathrm{poly}(t) \cdot \varepsilon}$

Leaky Condensers and Fiat-Shamir

- Use Leaky Condenser Cond to implement H



- **Intuition**: view each $z_1 \ldots z_i$ as "short leakage" on X
- Proof + Condenser: soundness increases by $\leq 2^d$ per round

Summary

- Seed-Dependent Condensers against Efficiently Samplable sources
  - Unlike extractors, can be faster than  !
- Application to Key Derivation
  - Importance of Square Advantage
  - Generic bounds on security degradation
- Simple construction from CRHF
- Generalization to Side Information
  - Application: Fiat-Shamir on proofs
  - Open: construction from standard assumptions

Questions?