Entropic Security and the Encryption of High-Entropy Messages

Yevgeniy Dodis, NYU

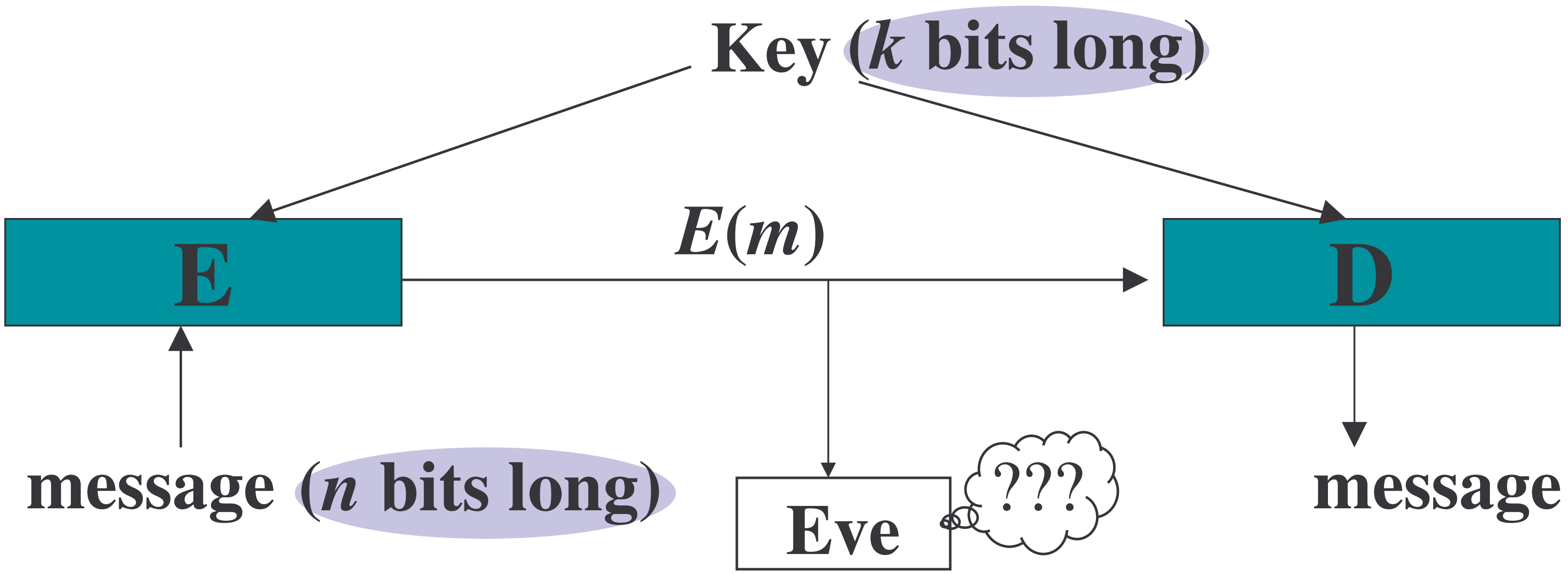
Adam Smith, Weizmann (work done at MIT)

Unconditional Secrecy When Leaking Information is Unavoidable

Yevgeniy Dodis, NYU

Adam Smith, Weizmann (work done at MIT)

Symmetric Encryption



- Shannon: Symmetric Encryption without computational assumptions requires $k \geq n$ (achieved by one-time pad)
- Russell and Wang 2002 [RW02]: What can be said when the message is guaranteed to have high entropy?

“Entropic” Security [CMR98,RW02]

[RW02]: Encryption of high-entropy messages

1. No computational assumptions (statistical secrecy)
2. Assume message distribution has **high entropy**
3. Constructions with short key (not possible without #2)

[CMR98]: Hash functions which hide “partial information”

1. Given H(m) and m', one can check if $m' = m$
2. Assume high entropy
3. H(m) leaks no predicate of m

This Paper

Motivation:

- Systematic study, simplification of entropic security
- Understand “high-entropy secrets” in simple setting
- Develop tools for settings other than encryption

This talk:

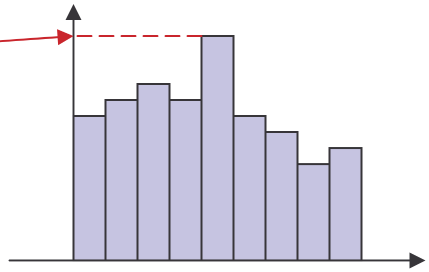
- Definitions
- Equivalent characterizations (extraction)
- Encryption: analysis, constructions, bounds
- Ideas for Other Settings

Entropic Security — Intuition

If Eve is uncertain about M, then $E(M)$ does not reveal any predicate of M.

Min-Entropy of Random Variables

- There are various ways to measure entropy…
- X a random variable on $\{0,1\}^n$
- Probability of predicting X = $\max_x \Pr[X = x]$



- **Min-entropy**: $H_\infty(X) = -\log\left(\max_x \Pr[X=x]\right)$

- " Message has min-entropy t " means that adversary's probability of guessing the message is 2^{-t}

Entropic Security [RW02]

Definition: (E,D) is (λ,ε)-entropically secure if

$\forall$ distributions M on $\{0,1\}^n$ with $H_\infty(M) \geq n\text{-}\lambda$

$\forall$ predicates $g:\{0,1\}^n \to \{0,1\}$

$\forall$ (adversaries) $A:\{0,1\}^* \to \{0,1\}$

$\exists$ random variable A' (independent of M)

$$\left|\ \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)]\ \right| \leq \varepsilon$$

- Statistical version of semantic security à la [GM] but only for high-entropy distributions

Entropic Security [RW02]

Definition: (E,D) is (λ,ε)-entropically secure if

$\forall$ distributions M on $\{0,1\}^n$ with $H_\infty(M) \geq n-\lambda$

$\forall$ predicates $g:\{0,1\}^n \to \{0,1\}$

$\forall$ (adversaries) $A:\{0,1\}^* \to \{0,1\}$

$\exists$ random variable A' (independent of M)

$$\left|\ \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)]\ \right| \leq \varepsilon$$

Caveats:

- Assumes that message has high entropy!
  What if the adversary knows more than you think he knows?
- Composition / computational “issues”: what happens when such a scheme gets plugged into more complex situations?

Entropic Security [RW02]

Definition: (E,D) is (λ,ε)-entropically secure if

$\forall$ distributions M on $\{0,1\}^n$ with $H_\infty(M) \geq n\text{-}\lambda$

$\forall$ predicates $g:\{0,1\}^n \to \{0,1\}$

$\forall$ (adversaries) $A:\{0,1\}^* \to \{0,1\}$

$\exists$ random variable A' (independent of M)

$$\left|\ \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)]\ \right| \leq \varepsilon$$

[RW02] There exist (λ,ε)-ES schemes with

$$k \approx \lambda + 3\log(1/\varepsilon)$$

(Without high entropy, still need $k \approx n$)

Two constructions: twists on the one-time pad. Complicated analysis.

Results (and Outline)

- Equivalent Definitions:
  - Hiding all functions
  - Indistinguishability
  - Intuition: entropic security $\approx$ randomness extraction
- Two Simple, General Constructions (improve [RW02])
  - Step on expander graph
  - Hashing
- Lower Bounds

Is This the Right Definition?

Def: (λ,ε)-entropically secure if $\forall M$ (entropy $\geq n$-λ),

$\forall$ predicates $g:\{0,1\}^n \rightarrow \{0,1\}$

$\forall$ adversaries A, $\exists A'$,

$$\left|\ \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)]\ \right| \leq \varepsilon$$

Before we commit long-term:

- Can we do better? (This one is better than it looks)
- Can we work with this? (Yes, with effort)

Is This the Right Definition?

Def: (λ,ε)-entropically secure if $\forall M$ (entropy $\geq n-\lambda$),

$\forall$ functions $g:\{0,1\}^n \to$ any domain you like

$\forall$ adversaries A, $\exists A'$,

$$\left| \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)] \right| \leq \varepsilon$$

Q: Why only predicates? What about functions? Relations?

(If cryptography is everything, why sell ourselves short?)

A: Functions are equivalent!

(Relations impossible with short key)

Equivalence of Functions and Predicates

For function f, random variable $\mathbf{M}$:

$$\mathbf{pred}_f(\mathbf{M}) = \text{ most likely value } = \max_z\{ \Pr[f(\mathbf{M}) = z] \}$$

Lemma: **If**

- $\mathbf{M}$ random variable (entropy $\geq 2\log(1/\varepsilon)$)
- $E()$, $A()$ randomized maps, f arbitrary function.
- $\Pr[A(E(\mathbf{M})) = f(\mathbf{M})] \geq \mathbf{pred}_f(\mathbf{M}) + \varepsilon$

Then there exist predicates B and g such that

$$\Pr[B(A(E(\mathbf{M}))) = g(\mathbf{M})] \geq \mathbf{pred}_g(\mathbf{M}) + \varepsilon / 4$$

Indistinguishability for High Entropy

Def: (λ,ε)-entropically secure if $\forall\, M$ (entropy $\geq n\text{-}\lambda$),

$\forall$ functions $g:\{0,1\}^n \to$ any domain you like

$\forall$ adversaries A, $\exists\, A'$,

$$\left|\ \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)]\ \right| \leq \varepsilon$$

Recall: (Ordinary) semantic security $\Rightarrow$

$\forall$ distributions M,M': $E(M) \approx_{PPT} E(M')$

Indistinguishability for High Entropy

Def: (λ,ε)-entropically secure if $\forall$ M (entropy $\geq n$-λ),

$\forall$ functions $g:\{0,1\}^n \rightarrow$ any domain you like

$\forall$ adversaries A, $\exists$ A',

$$\left|\ \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)]\ \right| \leq \varepsilon$$

Definition: (E,D) is (t,ε)-indistinguishable (IND) if

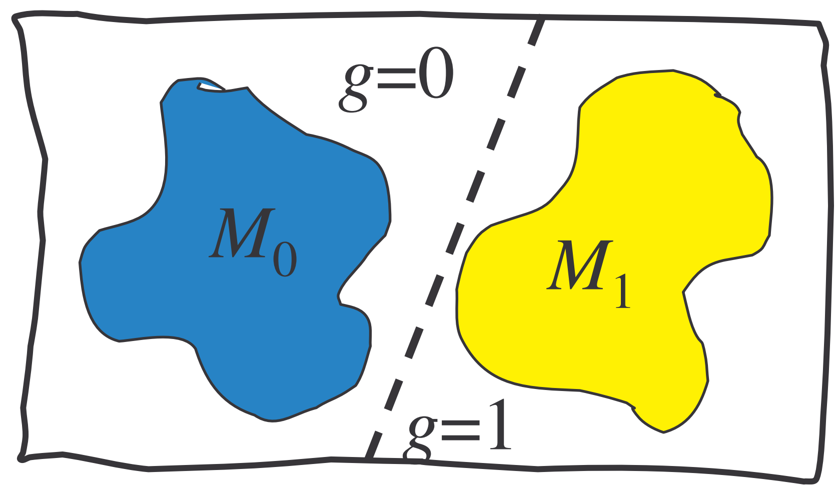
$\forall$ distributions M,M' with $H_\infty(M)$, $H_\infty(M') \geq t$:

$$SD(E(M),E(M')) \leq \varepsilon$$

Proposition: **(λ,ε)-ES** equiv. to **(t , ε')-IND** for $t = n$-λ-1

Proof: (λ,ε)-ES $\Rightarrow$ $(n\text{-}\lambda\text{-}1, 4\varepsilon)$-IND

- Take any M_0, M_1 of min-entropy $\geq t = n\text{-}\lambda\text{-}1$
  (Sufficient to prove lemma for flat distrib's on 2^t points)
- Suppose $M_0 \cap M_1 = \emptyset$
  Use $g(x) = b$ if $x \in \mathrm{supp}(M_b)$
  and $M^* = M_b$ for $b \leftarrow \{0,1\}$



- $H_\infty(M^*) = t+1 = n\text{-}\lambda$
  $\Rightarrow$ No A predicts g better than $\frac{1}{2}+\varepsilon$
  $\Rightarrow$ $SD(\ E(M_0), E(M_1)\ ) \leq 2\varepsilon$
- If M_0, M_1 not disjoint, find M_2 disjoint to both.

Proof: (n-λ-1)-IND ⇒ λ-ES

- Suppose that $\exists\, A, f, M$ such that

  $\Pr[A(E(M)) = f(M)\ ] \geq \mathbf{pred}_f(M) + \varepsilon$

- Define $M_y = M\,|_{\{f(M)=y\}}$
- Group M_y's into bins so that

  $\varepsilon / 2 \leq$ weight of $M_y \leq \mathbf{pred}_f + \varepsilon/2$

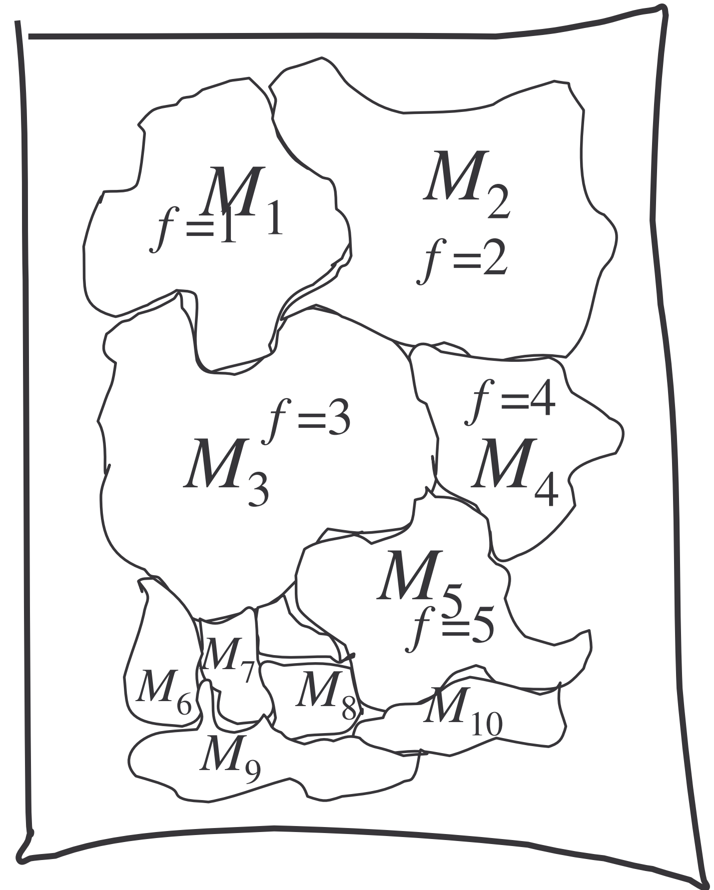
- Advantage still $\geq \varepsilon/2$

  $\Rightarrow$ some pair $E(M_y)$, $E(M_z)$ are far

- Contradicts indistinguishability for

  $\boldsymbol{t = n - \lambda} - \log(1/\varepsilon)$

- Save loss in entropy via balancing & Goldreich-Levin (tricky!)



Recall: Indistinguishability

Def: (λ,ε)-entropically secure if $\forall M$, $H_\infty(M) \geq n-\lambda$, $\forall A\ \forall g$

$\exists A'$: $\left|\ \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)]\ \right| \leq \varepsilon$

Def: (*t*,ε)-indistinguishable (IND) if $\forall M_0, M_1$, $H_\infty(M_b) \geq t$:

$$E(M_0) \approx_\varepsilon E(M_1)$$

Proposition: **(λ,ε)-ES** equiv. to **(*t* , ε')-IND** for $t = n-\lambda-1$

- What does this say?
  - Randomness extractors hide all functions of their source.
- How can we use this?
  - Extractors with "invertibility" give encryption schemes

Outline

- Equivalent Definitions:
  - Hiding all functions
  - Indistinguishability
  - Intuition: entropic security ≈ randomness extraction
- Two Simple, General Constructions (improve [RW02])
  - Step on expander graph
  - Hashing
- Lower Bounds

Expander Graphs

- Important tool in … everything.
- Expander = regular, undirected graph w
  - Let A = adjacency matrix of d-regular (
  - Vector (1,…,1) has eigenvalue d
  - Other eigenvalues $\in [-d,d]$
- G is a β-expander if other
- Random walks converge quickly:

When β is very small, walk converges in 1 step

Fact: If $H_\infty(p) \geq t$, then walk is ε-far from uniform after at most

$$\frac{n - t + 2\log(1/\varepsilon)}{2\log(1/\beta)}$$ steps, where $|G| = 2^n$.

Using Graphs for Encryption

- Encryption of m = random step from m
- Take regular G with $V=\{0,1\}^n$ and $d=2^k$
- Consider $\boldsymbol{E(m,s) = N(m,s)}$

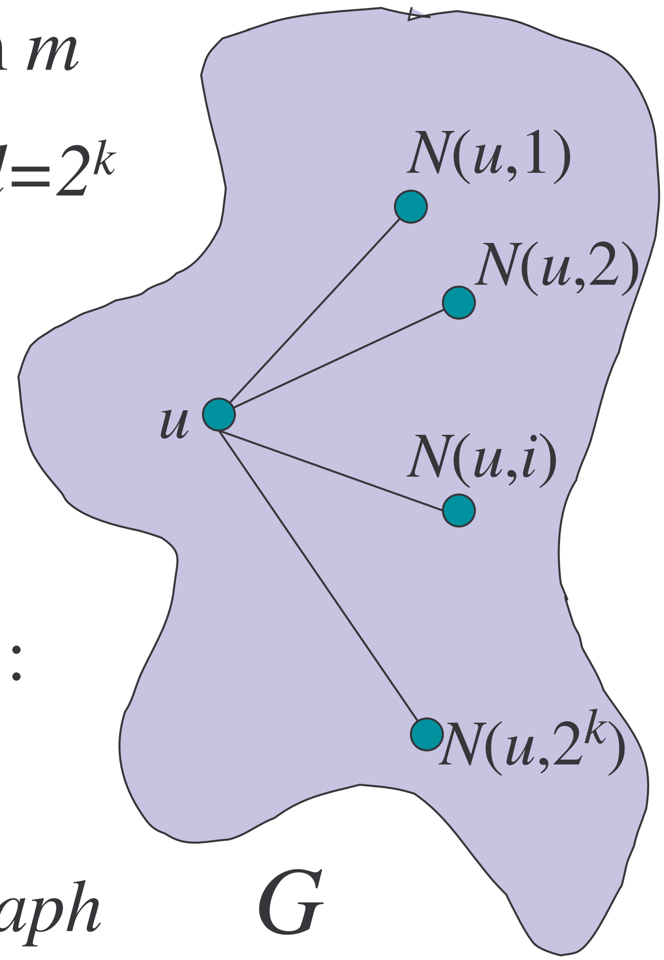
  ($N(u,i) = i^{\text{th}}$ neighbour of node u)

Q: When can you decrypt?

A: Need labeling N with an inverter N':

$$N'(\, N(u,i)\,,\, i) = u$$

Exercise: *Every regular undirected graph has an invertible labeling.*



Using Graphs for Encryption

- Encryption of m = random step from m
- Take regular G with $V=\{0,1\}^n$ and $d=2^k$
- Consider $\mathbf{E(m,s) = N(m,s)}$

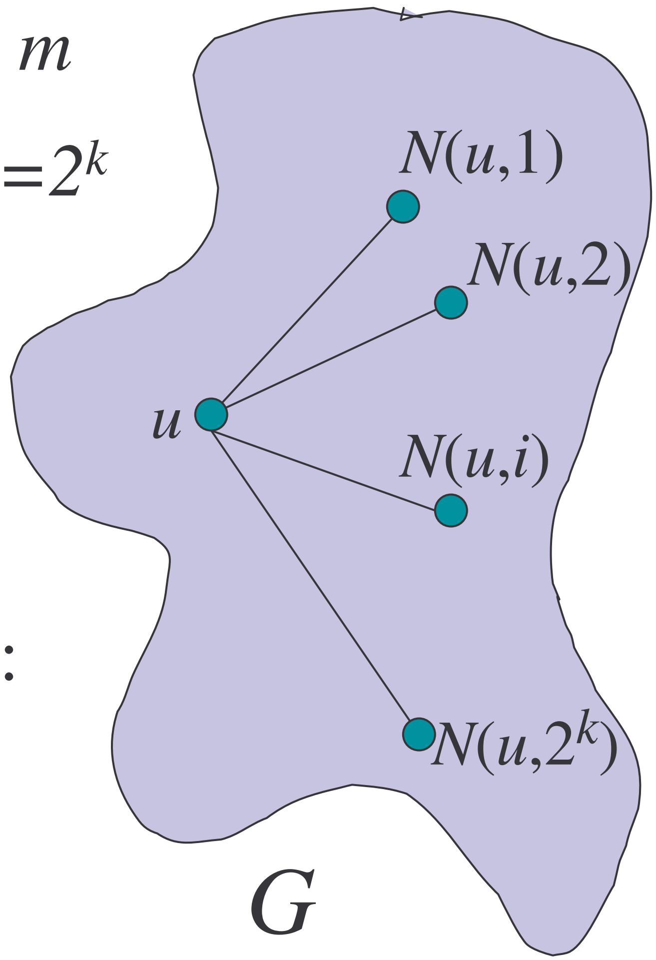
  ($N(u,i) = i^{\text{th}}$ neighbour of node u)

Q: When can you decrypt?

A: Need labeling N with an inverter N':

$$N'(\, N(u,i) \,,\, i) = u$$

Easier exercise: *Cayley graphs are invertible.*



Tangent: Cayley Graphs

- Let (V,*) be a group, $B=\{g_1,\ldots,g_d\}$ a set of generators.

  Cayley graph for $(V,*,B)$ has vertex set V and edges:

  $$E = \{ (u, g*u) \mid u \in V, g \in B \}.$$

- Graph is undirected if B contains its inverses.
  - E.g. hypercube $\{0,1\}^n$ with B ={vectors of weight 1}
- Natural labeling is $N(u,i) = g_i*u$
- Invertible since $N'(w,i) = g_i^{-1}*w$
- Graphs in this talk are Cayley graphs

Using Graphs for Encryption

- Take regular G with $V=\{0,1\}^n$ and $d=2^k$
- Consider ***E(m,s) = N(m,s)***

  ($N(u,i) = i$ th neighbour of node u)

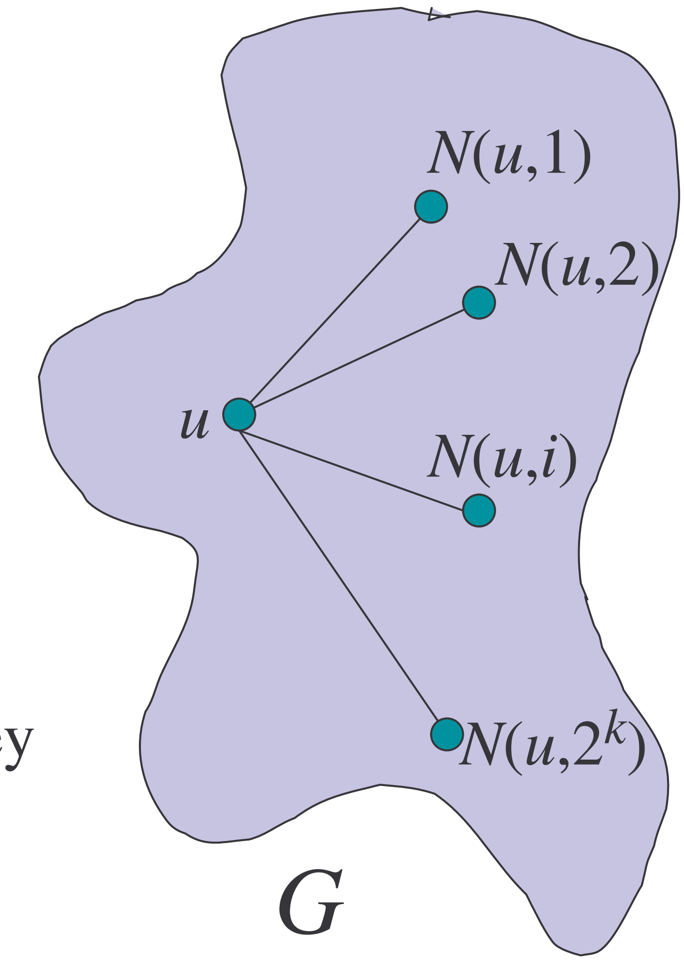
Q: When is E (t,ε)-indistinguishable?

A: When walk converges in 1 step.

Sufficient: G is β-expander with $\beta^2 \leq \varepsilon^2\, 2^{t-n}$

Theorem[LPS]: There exist (explicit) Cayley graphs with $\beta^2 \approx 1/d = 2^{-k}$

Corollary: *There exist (λ,ε)-ES encryption schemes with* $\mathbf{k \approx \lambda + 2 \log(1/\varepsilon)}$



[RW02]: Two constructions

1. $E(m,s) = m \oplus b(s)$, with $b : \{0,1\}^k \rightarrow \{0,1\}^n$.
   - $b(\cdot)$ is carefully chosen: range is "δ-biased set"
   - Fourier-based proof works only for uniform message
   - $k \approx 2 \log n + 3 \log (1/\varepsilon)$ (here $\lambda = 0$)
2. $E(m,s;\ i) = (\phi_i\ ,\ \phi_i(m) + s)$
   - $\{\phi_i\!: \{0,1\}^n \rightarrow \{0,1\}^n\ \}$ are 3-wise independent permutations
   - $k \approx \lambda + 3 \log (1/\varepsilon)$ (works for all λ)
   - $3n$ bits of additional randomness, difficult proof

[RW02]: First construction

1. $E(m,s) = m \oplus b(s)$, with $b : \{0,1\}^k \rightarrow \{0,1\}^n$.
   - $b(\cdot)$ is carefully chosen: range is "δ-biased set"
   - Fourier-based proof works only for uniform message
   - $k \approx 2 \log n + 3 \log (1/\varepsilon)$ (here $\lambda = 0$)

Same scheme, new analysis:

- $G =$ Cayley graph for $\{0,1\}^n$ with generators $\{b(s) \mid s \in \{0,1\}^k\}$
- Observe that G is a δ-expander (degree $= n^2/\delta^2$) (e.g. [BGSW])
- Previous slide $\Rightarrow$ $\boldsymbol{k = \lambda + 2 \log n + 2 \log (1/\varepsilon)}$

  (Same proof works for all λ)

Two General Constructions

#1 : Steps on an expander graph

#2: Random Hashing (not here)

Outline

- Equivalent Definitions:
  - Hiding all functions
  - Indistinguishability
  - Intuition: entropic security ≈ randomness extraction
- Two Simple, General Constructions (improve [RW02])
  - Step on expander graph
  - Hashing
- Lower Bounds

Lower Bounds

- Lower Bound via Shannon Bound:

$$k \geq \lambda$$

- Lower bound via lower bounds on extractors:

$$k \geq \lambda + \log(1/\varepsilon)$$

  – Requires that extra randomness be public, i.e.

$$E(m,s;i) = (i\ ,\ E'(m,s;i)\ )$$

  – All the schemes discussed fit this framework

Simple Lower Bound

Def: (λ,ε)-entropically secure if $\forall M$, $H_\infty(M) \geq n\text{-}\lambda$, $\forall A$ $\forall$ pred. g

$\exists A'$: $\left| \Pr[A(E(M)) = g(M)] - \Pr[A' = g(M)] \right| \leq \varepsilon$

Proof (reduce to bounds on regular encryption):

- $\forall w \in \{0,1\}^\lambda$, define distribution $M_w = w \,||\, U_{n\text{-}\lambda}$
  (i.e.: $M_w = w$ followed by $n\text{-}\lambda$ random bits)
- Indistinguishability $\Rightarrow$ $\forall v, w$: $E(M_v) \approx_\varepsilon E(M_w)$
- This is regular encryption (non-entropic) of w **!**
- Need $k \geq \lambda$

Conclusions

- Systematic study of entropic security [CMR98,RW02]
  - Stronger definition + characterization as **indistinguishability**
  - Extractors hide all functions of their source!
  - Simple constructions, proofs, lower bounds
- Computational question: preserve running time?
- In what other contexts is ES interesting?
  - Password Hashing [CMR98]: similar definition
  - Error Correction (bounded storage, noisy keys) (STOC 05)
  - Database Privacy