Symmetric-Key Encryption: Towards Constructions

Lecture 4

Pseudorandomness

Story So Far

- We defined (passive) security of Symmetric Key Encryption (SKE)
  - **SIM-CPA** = **IND-CPA** + **almost perfect correctness**
    - Restricts to PPT entities
    - Allows negligible advantage to the adversary
- Roadmap for constructions:
  - Pseudorandomness from One-Way Permutations
  - Construct one-time SKE from Pseudorandomness
  - Upgrade to Multi-message SKE
- Today: The concept of pseudorandomness

Constructing SKE schemes

- Basic idea: "stretchable" pseudo-random one-time pads (kept compressed in the key)
  - (For multiple message encryption, will also need a mechanism to ensure that the same piece of the one-time pad is not used more than once)
- Approach used in practice today: complex functions which are conjectured to have the requisite pseudo-randomness properties (stream-ciphers, block-ciphers)
- Theoretical Constructions: Security relies on certain computational hardness assumptions related to simple functions

Pseudorandomness Generator (PRG)

- Expand a short random **seed** to a "random-looking" string
- First, PRG with fixed stretch: $G_k: \{0,1\}^k \to \{0,1\}^{n(k)}$, $n(k) > k$
- How does one define random-looking?
  - Next-Bit Unpredictability: PPT adversary can't predict i^{th} bit of a sample from its first (i-1) bits (for every $i \in \{1,\ldots,n\}$)
  - A "more satisfactory" definition:
    - PPT adversary can't distinguish between a sample from $\{G_k(x)\}_{x \leftarrow \{0,1\}^k}$ and one from $\{0,1\}^{n(k)}$
  - **Turns out they are equivalent!**

$|\ Pr_{y \leftarrow PRG}[A(y)=0] - Pr_{y \leftarrow rand}[A(y)=0]\ |$ is negligible for all PPT A

Coming up

Indistinguishability

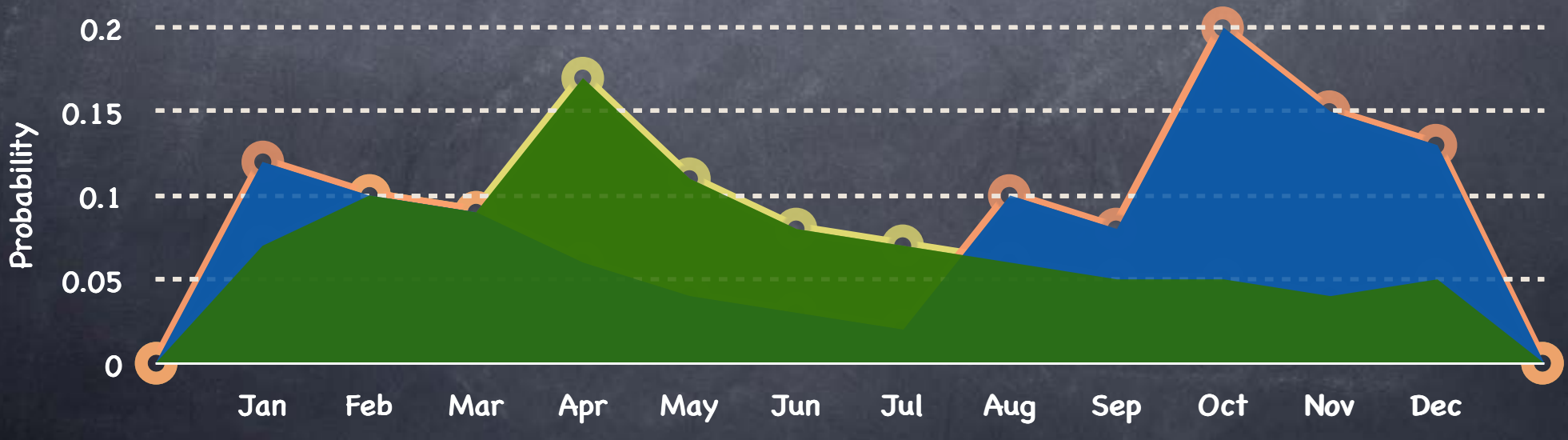
- Security definitions often refer to indistinguishability of two distributions: e.g., REAL vs. IDEAL, or Enc(m_0) vs. Enc(m_1)
  - By a distinguisher who outputs a single bit
- 3 levels of indistinguishability
  - Perfect: the two distributions are identical
  - Computational: for all PPT distinguishers, probability of the output bit being 1 is only negligibly different in the two cases
  - Statistical: the two distributions are "statistically close"
    - Hard to distinguish, irrespective of the computational power of the distinguisher

Statistical Indistinguishability

- Given two distributions A and B over the same sample space, how well can a (computationally unbounded) test T distinguish between them?
  - T is given a single sample drawn from A or B
  - How differently does it behave in the two cases?
- $\Delta(A,B) := \max_T | \Pr_{x \leftarrow A}[T(x)=1] - \Pr_{x \leftarrow B}[T(x)=1] |$

Statistical Difference (Distance) or Total Variation Distance

- Two distribution ensembles $\{A_k\}_k$, $\{B_k\}_k$ are **statistically indistinguishable** from each other if $\Delta(A_k,B_k)$ is negligible in k



Computational Indistinguishability

- Two distribution ensembles $\{X_k\}$ and $\{X'_k\}$ are said to be **computationally indistinguishable** if
  - $\forall$ (non-uniform) PPT distinguisher D, $\exists$ negligible $\nu(k)$ such that $| Pr_{x \leftarrow X_k}[D(x)=1] - Pr_{x \leftarrow X'_k}[D(x)=1] | \leq \nu(k)$

$X_k \approx X'_k$

- cf.: Two distribution ensembles $\{X_k\}$ and $\{X'_k\}$ are said to be **statistically indistinguishable** if $\forall$ functions T, $\exists$ negligible $\nu(k)$ s.t. $| Pr_{x \leftarrow X_k}[T(x)=1] - Pr_{x \leftarrow X'_k}[T(x)=1] | \leq \nu(k)$
  - Equivalently, $\exists$ negligible $\nu(k)$ s.t. $\Delta(X_k,X'_k) \leq \nu(k)$ where $\Delta(X_k,X'_k) := \max_T | Pr_{x \leftarrow X_k}[T(x)=1] - Pr_{x \leftarrow X'_k}[T(x)=1] |$

Pseudorandomness Generator (PRG)

- Takes a short seed and (deterministically) outputs a long string
  - G_k: $\{0,1\}^k \to \{0,1\}^{n(k)}$ where $n(k) > k$
- Security definition: Output distribution induced by random input seed should be "pseudorandom"
  - i.e., Computationally indistinguishable from uniformly random
  - $\{G_k(x)\}_{x \leftarrow \{0,1\}^k} \approx U_{n(k)}$
  - Note: $\{G_k(x)\}_{x \leftarrow \{0,1\}^k}$ cannot be statistically indistinguishable from $U_{n(k)}$ unless $n(k) \leq k$ (Exercise)
    - i.e., no PRG against unbounded adversaries

Equivalent definitions

$|\ Pr_{y \leftarrow PRG}[B(y_1^{i-1}) = y_i\ ] - \frac{1}{2}\ |$ is negligible for all i, all PPT B

$|\ Pr_{y \leftarrow PRG}[A(y)=0] - Pr_{y \leftarrow rand}[A(y)=0]\ |$ is negligible for all PPT A

- Next-Bit Unpredictable ⇔ Pseudorandom
- Pseudorandom ⇒ NBU:
  - Reduction: Given a PPT adversary B (for NBU), will show how to turn it into a PPT adversary A (for Pseudorandomness) with similar advantage. Hence the advantage must be negligible.

    Could be seen as showing the contrapositive: ¬NBU ⇒ ¬Pseudorandom
  - For any PPT B and i, consider PPT A which uses it to predict i^{th} bit and then checks if the prediction was correct
  - Formally, A(y) outputs $B(y_1^{i-1}) \oplus y_i$ (i as specified by B). Then:

    $|\ Pr_{y \leftarrow PRG}[A(y)=0] - Pr_{y \leftarrow rand}[A(y)=0]\ | = |\ Pr_{y \leftarrow PRG}[B(y_1^{i-1}) = y_i\ ] - \frac{1}{2}\ |$

Equivalent definitions

$|\ Pr_{y \leftarrow PRG}[B(y_1^{i-1}) = y_i\ ] - \frac{1}{2}\ |$ is negligible for all i, all PPT B

$|\ Pr_{y \leftarrow PRG}[A(y)=0] - Pr_{y \leftarrow rand}[A(y)=0]\ |$ is negligible for all PPT A

- Next-Bit Unpredictable $\Leftrightarrow$ Pseudorandom
- NBU $\Rightarrow$ Pseudorandom: Using a **Hybrid Argument**
  - Define distributions H_i over n-bit strings: $y \leftarrow$ PRG. Output $y_1^i \,||\, r$ where r is n-i independent uniform bits. H_0 = rand, H_n = PRG.
  - PRG is NBU $\Rightarrow H_i \approx H_{i+1}$: Given a PPT distinguisher A for H_i vs. H_{i+1}, let PPT predictor B be as follows: On input $z \in \{0,1\}^i$, pick $b \leftarrow \{0,1\}$, $r \leftarrow \{0,1\}^{n-i-1}$ and output $A(z \,||\, b \,||\, r) \oplus b$. Then

    $|Pr_{y \leftarrow PRG}[B(y_1^{i-1}) = y_i\ ] - \frac{1}{2}| \;=\; |Pr_{y \leftarrow H_i}[A(y)=0] - Pr_{y \leftarrow H_{i+1}}[A(y)=0]|$
  - Then $H_0 \approx H_n$ (for n(k) that is polynomial)
    - [Exercise]