

Symmetric Ciphers and PRGs

Symmetric Encryption, Perfect Ciphers, Definitions of PRGs and PRFs

FROM PREVIOUS LECTURE

> Provable security methods:

- Define "security": syntax, model, winning condition
- Propose scheme
- Define assumptions
- Prove security
- Game-based definitions:
 - Algorithms: adversary has necessary information, like in encryption, signature verification, ...
 - Oracles: adversary doesn't have necessary information: decryption, signatures
 - Challenge/response: test to see if adversary wins game

PROBABILITY, ADVANTAGE, REDUCTION

- > Winning probability:
 - Probabilistic schemes: take probability to win over all the randomness
 - Trivial adversary: just guesses the answer
- > Advantage:
 - How much better adversary can do over trivial adversary
 - Distinguishing games: e.g. IND-CCA -- advantage over ¹/₂
 - Guessing games: e.g. EUF-CMA advantage over $1/2^n$

Reduction:

Assume adversary exists against scheme, construct one against assumption

GAME-HOPPING

- > Restricting adversary:
 - Start with original game
 - Restrict adversary's options in future games
- > Each time, prove equivalence of games:
 - Method 1: G_0 , G_1 are the same except output to A (for instance PR output changed to random).
 - Prove that doesn't matter:
 - If A can profit, then construct adversary against assumption
 - Method 2: G_0 , G_1 are same except in G_1 the game aborts if adversary does something (like forge a certificate)
 - Prover that this can't happen
 - If A can profit, then adversary against assumption

ENCRYPTION SCHEMES

- Designed to protect message confidentiality
 - Usually 2 parties, called Alice and Bob; adversary is Eve
 - Plaintext M encrypted by Alice, becoming a ciphertext C
 - Ciphertext C decrypted by Bob to some plaintext M'
 - Necessary: Bob (and maybe Alice) must have a secret k



SECRETS AND NON-SECRETS

- Kerckhoff: Consider the algorithm public
 - If the algorithm is compromised, no problem
 - More eyes to look at the security of a public algorithm
- Symmetric-key encryption (block/stream ciphers)
 - Alice and Bob share secret key *k*



BASIC CIPHERS

Caesar cipher and extensions

- Permutation cipher
- Key is the number of letters we permute by
- Caesar: k = 3
- BLOCKCIPHER becomes EORFNFLSKHU



THE CAESAR CIPHER

- Kerckhoff: algorithm is public
- > We need the key
 - Key space is too small : brute force works in one go with probability $\frac{1}{26}$ and works for sure in 26 attempts
 - Attack works only if message is meaningful

Brute force is base line for attacks against ciphers

ONE-TIME PAD

- > Substitution cipher, C = M + K (e.g. mod 26)
- Key length equal to message length
- ▶ If M = BLOCKCIPHER, and K = PRZANIBQTCS
- Say message is meaningful and key is meaningful
 - Can we do better than brute force?

Yes, look at language statistics

- Say message is meaningful, but key is truly random
 - Key hides message information-theoretically



SECURITY DETAILS

> What if same key used multiple times in N attempts?

- Case 1: Adversary knows it (described in protocol) Passive eavesdropper learns M₁ XOR M₂ Equivalent to using meaningful key
- Case 2: Adversary does not know (accidental collision) Even assuming this is problematic, this happens rarely (w.p. ≤ ^N₂) 2^{-|sk|})

> What does it mean that the key "hides" a message?

 BLOCKCIPHER + "PRZANIBQTCS" = RCNDYLKFAHJ UNIVERSALLY + "XOEHUUSEPWO" = RCNDYLKFAHJ YETIMONSTER + "TYUVLXXNGCS" = RCNDYLKFAHJ

• Message is meaningful: probability bound by dictionary attack

GUARANTEE OF ONE-TIME PAD

> Ingredients:

- Set **§**, which is an alphabet (like A, B, ..., Z)
- Length of messages l
- Subset $\mathcal{M} \in \mathcal{S}^l$ of meaningful messages of length l
- An (Abelian) group operation " + " on \mathcal{S}^{l} , inverse operation "-"

Guarantee:

- The cipher consisting of:
 - Picking K randomly from $\boldsymbol{\delta}^{l}$
 - Encrypting plaintext $M \in \mathcal{M}$ to C := M + K
 - Decrypting plaintext C to M = C K

guarantees that:

Prob[ptext = M | ctext = C] = Prob[ptext = M]

Perfect cipher

PERFECT CIPHERS

Perfect ciphers:

Prob[ptext = M | ctext = C] = Prob[ptext = M]

Ciphertext gives no information on plaintext

> Theorem 1:

- Take a perfect cipher with plaintext alphabet *M* (all messages occuring with non-zero probability) and key space *K*
- Then the size of $\boldsymbol{\mathscr{K}}$ is at least equal to the size of $\boldsymbol{\mathscr{M}}$

> Proof:

• First observation: take plaintexts $M_1 \neq M_2$. Then for all $k \in \mathcal{K}$ it holds that $Enc(k; M_1) \neq Enc(k; M_2)$. Why?

KEY-SIZE OF PERFECT CIPHERS

> Theorem 1:

- Take a perfect cipher with plaintext alphabet *M* (all messages occuring with non-zero probability) and key space *K*
- Then the size of $\boldsymbol{\mathscr{K}}$ is at least equal to the size of $\boldsymbol{\mathscr{M}}$

> Proof:

- Reduction to absurd: Suppose $|\mathcal{K}| \leq |\mathcal{M}| 1$
- Look at mapping $(M, k) \rightarrow C$ (through encryption)
 - Order *M* in some way (lexicographically or just randomly)
 - \circ Take the first message, denote it M_1
 - Pick key k_1 , compute $C = Enc(k_1, M_1)$. If $C = \beth$ (invalid), pick again
 - Continue picking keys $k \neq k_1$ and run *Dec* (*C*, *k*)
- Even if all decryptions give a valid result, Obs 1 tells us there exists at least one *M*^{*} that *C* does not decrypt to.

KEY-SIZE OF PERFECT CIPHERS

> Theorem 1:

- Take a perfect cipher with plaintext alphabet *M* (all messages occuring with non-zero probability) and key space *K*
- Then the size of $\boldsymbol{\mathscr{K}}$ is at least equal to the size of $\boldsymbol{\mathscr{M}}$

> Proof:

- Reduction to absurd: Suppose $|\mathcal{K}| \leq |\mathcal{M}| 1$
- Look at mapping $(M, k) \rightarrow C$ (through encryption)
- Even if all decryptions give a valid result, Obs 1 tells us there exists at least one *M*^{*} that *C* does not decrypt to
- Then for this message it holds that: $Prob[ptext = M^* | ctext = C] = 0 \neq Prob[ptext = M_1 | ctext = C]$
- This is impossible (perfect cipher)
- Hence $|\mathcal{K}| \geq |\mathcal{M}|$

INDISTINGUISHABILITY

Consequence of Theorem 1:

• OTP has optimal key size (and it's long!)

> Another way to phrase perfection property:

• Indistinguishability:

For any messages $M_1 \neq M_2$ and any ciphertext C: Prob[Enc(*, M_1) = C] = Prob[Enc(*, M_2) = C]

> Theorem 2: A cipher is perfect if, and only if, it has the indistinguishability property Proof: in the TDs.

Some Conclusions

> Perfect ciphers:

- Ciphertext reveals nothing about the plaintext
- Equivalently phrased as: each ciphertext could correspond to any plaintext
- ... But they require $|\mathcal{K}| \geq |\mathcal{M}|$
- > One Time Pad (OTP):
 - Is a perfect cipher
 - Requires: changing key at each encryption
 - Key length = message length
 - Unfortunately, this key length is optimal

PART II PTT ADVERSARIES AND GAMES

A RELAXATION OF PERFECTION

- Security of perfect ciphers does not depend on the attacker's computational resources
 - Attacker with 200 years of computation time still learns nothing from ciphertext
- > ... however, we need very large keys
- > We want smaller keys, but sufficient security
 - Idea: bound the adversary's resources
 - Allow some (small) information leakage
 - Adversary can "win" with very small proability

LESS-THAN-PERFECT CIPHERS

- > Now assume that we take $|\mathcal{K}| < |\mathcal{M}|$
- > This introduces some attacks
- Meaningful message, random key:
 - Try to decrypt ciphertext with any possible key
 - This yields a list of "meaningful" possible plaintexts
- Compare to perfect security
 - PS: a ciphertext can hide any meaningful message
 - Imperfect security: ciphertext can "hide" at most |𝔄 messages, with |夫| < |√|
 - Key length determines security

COMPUTATIONAL SECURITY BASICS

- Generic family of ciphers parametrized by "security parameter" n
 - Usually the length of the secret key
- Encryption and Decryption are generic algorithms (no precise description is given)
- Cipher is secure if any adversary A can "break" the encryption scheme with negligible probability
 - Smaller than $\frac{1}{Poly[n]}$ for any polynomial Poly[n]

NEGLIGIBLE PROBABILITIES

- > What is negligible in theory?
 - Our favourite: 2^{-n}
 - Second best: $Poly[n] \cdot 2^{-n}$
 - Another possibility: 2^{-log[n]} is non-negligible, but 2^{-log²[n]} is negligible
- > What is negligible in practice?
 - Say the adversary wins with probability 2^{-n} for a small value of n
 - Trying again and again over a large amount of data, say 1GB, will eventually let *A* succeed
 - In practice, we like a security of at least 2^{-80}

COMPUTATIONAL CIPHER SECURITY

- > Think of it in terms of a game
- The adversary plays this game against our cipher and the parties using it – encryptor, decryptor
- The adversaries can see ciphertexts (possibly very many of them, but polynomial in the size of the key)
- Security notion: indistinguishability (of ciphertexts) from random

PSEUDO-RANDOMNESS

- Intuition:
 - If A can't tell ciphertexts from completely random strings of the same lengths, then:
 - A can't see a plaintext/ciphertext dependence
 - A can't see a key/ciphertext dependence
- Indistinguishability of real cryptographic systems from their idealizations is fundamental to provable security

PSEUDORANDOM GENERATORS (PRG)

Principle: start from a small, random string (called a seed), get a larger string that looks random
 PRG : {0,1}ⁿ → {0,1}^m for m > n

 Security: a "good" PRG outputs strings that are indistinguishable from random (by an adversary)



THE SECURE-PRG GAME

> $s \stackrel{\$}{\leftarrow} \{0,1\}^n$ $b \stackrel{\$}{\leftarrow} \{0,1\}$ $d \leftarrow \mathcal{A}^{Gen_b()}(m,n, PRG)$

 $\boldsymbol{\mathcal{A}}$ wins iff. b = d

 $\frac{Gen_b():}{\text{If } b = 1 \text{ then } x \stackrel{\$}{\leftarrow} U^m}$ Else $x \leftarrow PRG(s)$ Return x

- > Unbounded vs. bounded *A*
 - Unbounded: as many calls to Gen_b as \mathcal{A} wants
 - Bounded: only polynomially many calls, poly-runtime
 k-bounded: only *k* calls, poly-runtime
- (k, ε)-Secure PRG: G is a k-bounded-secure PRG if, and only if, any k-bounded adversary A wins w.p. at most ¹/₂ + ε
 - (asymptotically) k-secure: $\varepsilon \in \text{Negl}[n]$

DISTINGUISHERS/ DISTINGUISHING

- > What is a "random" string?
 - Usually defined as a string for which the probability that any of the bits is 1 is exactly ¹/₂
- > How does the distinguisher distinguish in practice?
 - Fixed bits
 - Fixed relationship between bits
 - Un-fixed, but biased relationship between bits (occurring with prob. p, such that |p 1/2| non-negligible)
 - <u>Theorem</u>: In a random string, the probability that there are less than ^{|m|}/₃ bits equal to 1 is negligible
 Proof in TD

STATISTICAL TESTS

- > Theorem:
 - Consider $\mathcal{J}_{m,k}$ to be the poly-sized set of all statistical tests $T_{m,k}$ which have poly-runtime, which take as input a sample of k bitstrings of length m, for a known, fixed $k \in \text{Poly}[m]$ and which output 0 (if the string sample is not random) and 1 (if the string sample is random)
 - Assume that we have a PRG $G\colon \{0,1\}^n \to \{0,1\}^m$ for m=2n
 - Then: *G* is a secure PRG against a *k*-bounded adversary \mathcal{A} if, and only if, for all $T_{m,k} \in \mathcal{J}_{m,k}$ it holds that for $s \leftarrow \{0,1\}^n$, $T_{m,k}$ run on randomly chosen *k*-sized samples of G(s) returns 0 w.p. at most $\varepsilon \in \text{Negl}[m]$

PROOF BY REDUCTION

> Theorem:

- Assume $T_{m,k} \in \mathcal{J}_{m,k}$ with input a sample of k bitstrings of length m, outputting 0 (if not random) and 1 (if random)
- Assume $G: \{0,1\}^n \to \{0,1\}^m$ for m = 2n
- Then: *G* is *k*-bounded secure iff. for $s \stackrel{\$}{\leftarrow} \{0,1\}^n, \forall T_{m,k} \in \mathcal{J}_{m,k}$ run on the output dist. of *G* returns 0 w.p. $\varepsilon \in \text{Negl}[m]$

\succ Proof : \Rightarrow

- Say *G* is k-bounded secure PRG
- Assume $\exists T_{m,k} \in \mathcal{J}_{m,k}$ which returns 0 w.p. $\delta \notin \text{Negl}[m]$
- Claim: $\delta \notin \text{Negl}[n]$. Why is this true?
- Construct k-bounded A against k-bounded sec. of G s.t. A wins with probability p_A ∉ Negl[n]

PROOF BY REDUCTION

> Theorem:

- Assume $T_{m,k} \in \mathcal{J}_{m,k}$ with input a sample of k bitstrings of length m, outputting 0 (if not random) and 1 (if random)
- Assume $G: \{0,1\}^n \to \{0,1\}^m$ for m = 2n
- Then: *G* is *k*-bounded secure iff. for $s \stackrel{\$}{\leftarrow} \{0,1\}^n, \forall T_{m,k} \in \mathcal{J}_{m,k}$ run on the output dist. of *G* returns 0 w.p. $\varepsilon \in \text{Negl}[m]$

 \succ Proof : \Rightarrow

- \mathcal{A} plays the PRG game. First the game picks: $s \stackrel{*}{\leftarrow} \{0,1\}^n$ bit b
- Query $Gen_b k$ times (ok, \mathcal{A} is k-bounded), get $X = \{x_1, ..., x_k\}$
- Run T_{m,k} on X, get output d ∈ {0,1} (ok, test has poly-runtime)
 o If A does not know which test is good, it can run all of them
- Return output *d* to PRF game
 If *A* tried all tests, return min of all d values

PROOF BY REDUCTION

> Proof :

- A plays the PRG game. First the game picks: $s \stackrel{\$}{\leftarrow} \{0,1\}^n$ bit b
- Query $Gen_b()$ k times (ok, \mathcal{A} is k-bounded), get $X = \{x_1, ..., x_k\}$
- Run $T_{m,k}$ on X, get output $d \in \{0,1\}$ (ok, test has poly-runtime)
- Return output *d* to PRF game

> Analysis:

- Obs 1: $T_{m,k}$ always returns 1 if bit b = 1 ($x_1, ..., x_k$ random)
- Obs 2: if b = 0 then X contains outputs of G. Then $T_{m,k}$ returns 0 w.p. $\delta \notin \text{Negl}[m]$ (by assumption)
- A wins w.p. $\Pr[A \text{ wins } | b = 1] \cdot \Pr[b = 1] + \Pr[A \text{ wins } | b = 0] \cdot \Pr[b = 0] = \frac{1}{2} + \frac{1}{2} \delta$, with $\delta \notin \operatorname{Negl}[n]$
- So *G* not a secure PRG. Contradiction

FOOD FOR THOUGHT

Some significant proof steps:

- $\operatorname{Negl}[m] \cong \operatorname{Negl}[n]$
 - Requiring $m \in Poly[n]$
- J_{m,k} requires a sample of k elements
 Requiring that our A is at least k-bounded!
- $\mathcal{J}_{m,k}$ runs in polynomial time
 - Else, a bounded adversary cannot run this test
- Statement about test holds for randomly chosen seed
 - If it held only for some seeds, we would not be able to transfer winning probability (PRG game first picks seed at rnd.)
 - We could have said it held for ALL keys. But then, it would not be an iff. statement. Let's see why.

NOW THE OTHER WAY

> Theorem:

- Assume $T_{m,k} \in \mathcal{J}_{m,k}$ with input a sample of k bitstrings of length m, outputting 0 (if not random) and 1 (if random)
- Assume $G: \{0,1\}^n \to \{0,1\}^m$ for m = 2n
- Then: *G* is *k*-bounded secure iff. for $s \stackrel{\$}{\leftarrow} \{0,1\}^n, \forall T_{m,k} \in \mathcal{J}_{m,k}$ run on the output dist. of *G* returns 0 w.p. $\varepsilon \in \text{Negl}[m]$

 \succ Proof : \Leftarrow

- Say $\forall T_{m,k} \in \mathcal{J}_{m,k}$ returns 0 w.p. at most $\delta \in \text{Negl}[m]$
- Say $\exists k$ -bounded A winning w.p. $\frac{1}{2} + \varepsilon \notin \text{Negl}[n]$
- Again $\varepsilon \notin \text{Negl}[m]$
- Construct poly-time test T_{m,k} that outputs 0 w.p. p_T ∉ Negl[m]
 o Claim: A is that T_{m,k}

Advantage & Unpredictability

- > In PRG game the adversary's winning probability should not be larger than $1/2 + \varepsilon$
 - We call $\Pr[A \text{ wins}] \frac{1}{2}$ the advantage of \mathcal{A}
- > Unpredictability theorem:
 - If $G: \{0,1\}^n \to \{0,1\}^m$ with m > n is a bounded-secure PRG, then for a randomly chosen $s \leftarrow \{0,1\}^n$, no polyruntime algorithm \mathcal{P} given the first j bits of G(s) can predict the (j + 1)-th bit w.p. $\frac{1}{2} + \varepsilon$ for $\varepsilon \notin \text{Negl}[n]$
 - Proof in TD

PERFECT TO IMPERFECT CIPHER

- > Why would we want that?
 - Well, it's more efficient, since $|\mathcal{A}| < |\mathcal{A}|$

Recall the OTP

- Traditional OTP for $\mathcal{K} = \mathcal{M} = \{0,1\}^m$
 - Choose random $k \stackrel{\$}{\leftarrow} \mathscr{K}$
 - Encrypt message *m* to : $c \coloneqq k \oplus m$
 - Decrypt ciphertext c as: $\widehat{m} \coloneqq c \oplus k$
- Unconditionally secure...
- ... But:
 - Key can only be used one time
 - Key is as long as message

PERFECT TO IMPERFECT OTP USING PRG

Recall the OTP

- Traditional OTP for $\mathcal{K} = \mathcal{M} = \{0,1\}^m$
 - Choose random $k \stackrel{\$}{\leftarrow} \mathscr{R}$
 - Encrypt message *m* to : $c \coloneqq k \oplus m$
 - Decrypt ciphertext c as: $\widehat{m} \coloneqq c \oplus k$

> Now replace random key generation by PRG:

- OTP for $\mathcal{M} = \{0,1\}^m$ with $\mathcal{H} = \{0,1\}^n$ and n < m
- Use a bounded-secure PRG $G: \{0,1\}^n \rightarrow \{0,1\}^m$
 - KeyGen: choose (once) $k \stackrel{\$}{\leftarrow} \boldsymbol{\mathscr{K}}$
 - Encrypt message m as $c \coloneqq G(k) \oplus m$
 - Decrypt message as: $\widehat{m} \coloneqq c \bigoplus G(k)$

PERFECT/IMPERFECT CIPHERS

> Perfect ciphers:

Prob[ptext = M | ctext = C] = Prob[ptext = M]

Alternatively:

For any messages $M_1 \neq M_2$ and any ciphertext *C* :

 $Prob[Enc(*, M_1) = C] = Prob[Enc(*, M_2) = C]$

Semantic security of imperfect ciphers:

• For $k \stackrel{\$}{\leftarrow} K$, $b \stackrel{\$}{\leftarrow} \{0,1\}$, and for any two messages m_0, m_1 , no polynomial-time adversary \mathcal{A} given $\operatorname{Enc}_k(m_b)$ can output d = b with probability $1/2 + \varepsilon$ for $\varepsilon \notin \operatorname{Negl}[|\mathcal{R}]$

OUR IMPERFECT OTP WITH PRG WORKS!

> Theorem:

- The OTP + PRG cipher we considered is semantically secure as long as the PRG is 1-bounded-secure
- Formally: for any adversary *A* against the semantic security of OTP+PRG, there exists a 1-bounded adversary *B* against the PRG-security of *G* such that:
 Pr[*A* wins] ≤ Pr[*B* wins]

If OTP + PRG is insecure, then *G* is insecure

 \simeq

As long as G is secure, OTP + PRG is secure

> Proof:

- Game 0: original semantic security game
- Game 1: replace G(s) by U^m in encryption
- Claim: if there exists a distinguisher *J* between games, then we can construct *B* from *J*



- > Proof:
 - Consider the distinguisher *D*. Depending on a bit b' *D* plays either Game 0 or Game 1
 - We construct \mathcal{B} against the 1-bounded PRG of G.

b B's game starts with sampling
$$s \stackrel{\$}{\leftarrow} \{0,1\}^n$$
 and bit $b'' \stackrel{\$}{\leftarrow} \{0,1\}$

- \mathcal{B} chooses $m_0, m_1 \stackrel{\$}{\leftarrow} \{0,1\}^n$, then queries $Gen_{b''}()$ once to obtain x
- \mathcal{B} draws a random bit d, and sends to \mathcal{D} the value $m_d \oplus x$
- \mathcal{J} returns a guess bit d'. If d = d', then \mathcal{J} returns 0 (Game 0). Else, \mathcal{J} returns 1



- > Proof:
 - Consider the distinguisher *D*. Depending on a bit b' *D* plays either Game 0 or Game 1
 - We construct \mathcal{B} against the 1-bounded PRG of G.
 - \mathcal{B} 's game starts with sampling $s \stackrel{\$}{\leftarrow} \{0,1\}^n$ and bit $b'' \stackrel{\$}{\leftarrow} \{0,1\}$
 - \mathcal{B} chooses $m_0, m_1 \stackrel{\$}{\leftarrow} \{0,1\}^n$, then queries $Gen_{b''}()$ once to obtain x
 - \mathcal{B} draws a random bit d, and sends to \mathcal{D} the value $m_d \oplus x$
 - \mathcal{J} returns a guess bit d', which \mathcal{J} forwards.

> Analysis:

• $\boldsymbol{\mathcal{J}}$ simulates $\boldsymbol{\mathcal{J}}$'s game perfectly and if $\boldsymbol{\mathcal{J}}$ wins w.p. $1/2 + \delta$, for non-negl. δ , then $\boldsymbol{\mathcal{J}}$ wins with same probability

> Proof:

- Game 0: original semantic security game
- Game 1: replace G(s) by U^m in encryption
- Note that

 $\Pr[A \text{ wins } G_0] \leq \Pr[A \text{ wins } G_1] + (\Pr[D \text{ dist. } G_0 \text{ from } G_1] - \frac{1}{2})$

$$= \frac{1}{2} + \left(\Pr[B \text{ wins}] - \frac{1}{2} \right) = \Pr[B \text{ wins}].$$

