私钥加密

哈尔滨工业大学

张宇

2024春

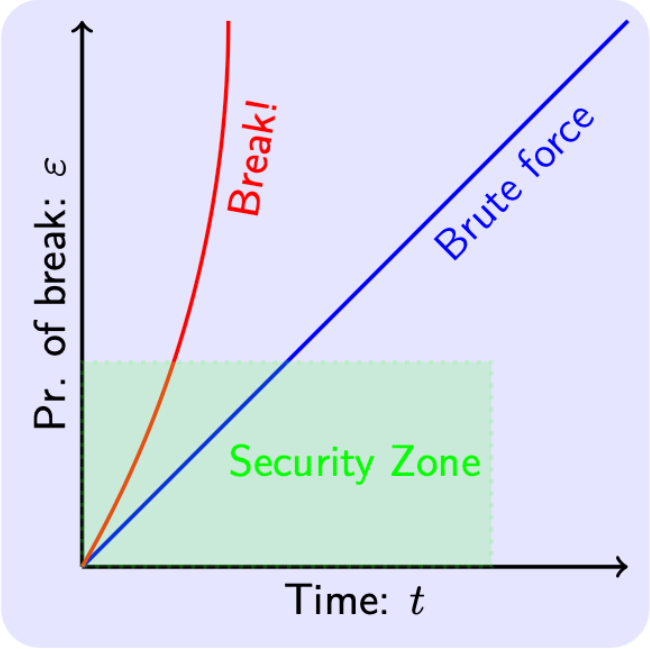
概览

1. 计算安全加密定义
2. 伪随机性（Pseudorandomness）假设
3. 规约证明（Proof by Reduction）
4. 安全加密方案构造与证明

计算安全思想

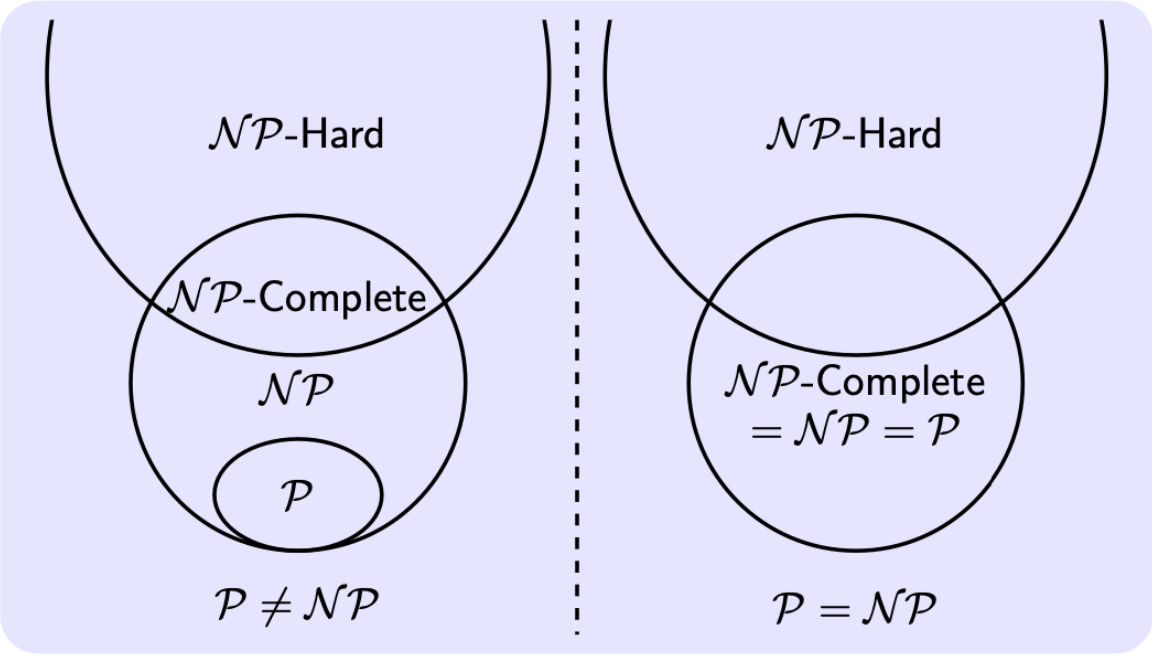
- 完美保密局限性在于密钥需要很长，而且如果密钥不够长，则不能达到完美保密。Kerchhoffs提出另一个原则：一个加密方案如果不是数学上，那必须是实践上不可破解的。
- 不同于在完美保密部的信息论上的安全，计算安全放松了安全条件来追求实践中的安全，使得密钥可以很短。

- 计算安全：
  - 敌手在**可行的时间**内运行
  - 敌手以小到**可忽略**的概率成功



可行的时间与可忽略的概率

- 一个算法是多项式时间的（polynomial time），如果存在一个多项式对于任意输入，算法都在该多项式步骤内结束。
- 一个函数f是可忽略的（negligible），若对于任意多项式$p(\cdot)$，存在一个N使得对于所有整数$n > N$，$f(n) < 1/p(n)$。
- 指数复杂性（2^n）是大到不可行的；$1/2^n$是小到可忽略的



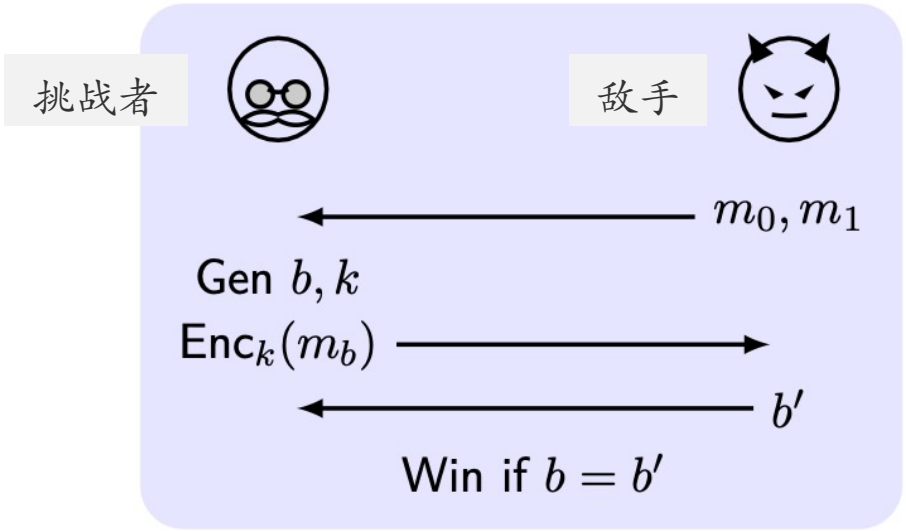
The majority of computer scientists believe $\mathcal{P} \neq \mathcal{NP}$.

窃听不可区分性实验

敌手和挑战者之间进行一个思维实验。敌手根据安全参数产生两个相同长度的不同消息，并发送给挑战者；挑战者根据安全参数生成密钥，并对随机选择的一个消息进行加密，将挑战密文发送给敌手。敌手输出一个比特，来表示对被加密消息的猜测，若猜对，则实验成功；否则，失败。

The eavesdropping indistinguishability experiment $\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi}(n)$:

1. $\mathcal{A}$ is given input 1^n, outputs m_0, m_1 of the same length
2. $k \leftarrow \mathsf{Gen}(1^n)$, a random bit $b \leftarrow \{0,1\}$ is chosen. Then $c \leftarrow \mathsf{Enc}_k(m_b)$ (challenge ciphertext) is given to $\mathcal{A}$
3. $\mathcal{A}$ outputs b'. If $b' = b$, $\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi} = 1$, otherwise 0



窃听安全的私钥加密定义

- 一个加密方案在出现窃听者时是不可区分加密，若对于任意概率多项式时间（PPT）的敌手，使得不可区分实验成功概率与1/2相比（两者间的差异）是可忽略的。
  - 多项式时间和可忽略都是对于“安全参数”的函数。
  - PPT（probabilistic polynomial time）概率多项式时间，概率是指算法具备随机化（“掷硬币”）的能力。

Definition 1

Π has **indistinguishable encryptions in the presence of an eavesdropper** if $\forall$ PPT $\mathcal{A}$, $\exists$ a negligible function negl such that

$$\Pr\left[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi}(n) = 1\right] \leq \frac{1}{2} + \mathsf{negl}(n),$$

where the probability it taken over the random coins used by $\mathcal{A}$.

语义安全（sementic security）

- -在导论部分有一个问题：如何定义不泄漏“meaningful”的信息。下面引入语义安全的概念来解决这个问题。
- 直觉：没有关于明文的任何有意义的信息泄漏
- 关于明文的信息用函数来表示，h(m)表示敌手预先了解的关于明文的信息，f(m)表示敌手希望获取的关于明文的有意义的信息

Definition 2

Π is **semantically secure in the presence of an eavesdropper** if $\forall$ PPT $\mathcal{A}$, $\exists \mathcal{A}'$ such that $\forall$ distribution $X = (X_1, \dots)$ and $\forall f, h$,

$$\left|\Pr[\mathcal{A}(1^n, \mathsf{Enc}_k(m), h(m)) = f(m)] - \Pr[\mathcal{A}'(1^n, h(m)) = f(m)]\right| \leq \mathsf{negl}(n).$$

where m is chosen according to X_n, $h(m)$ is external information.

Theorem 3

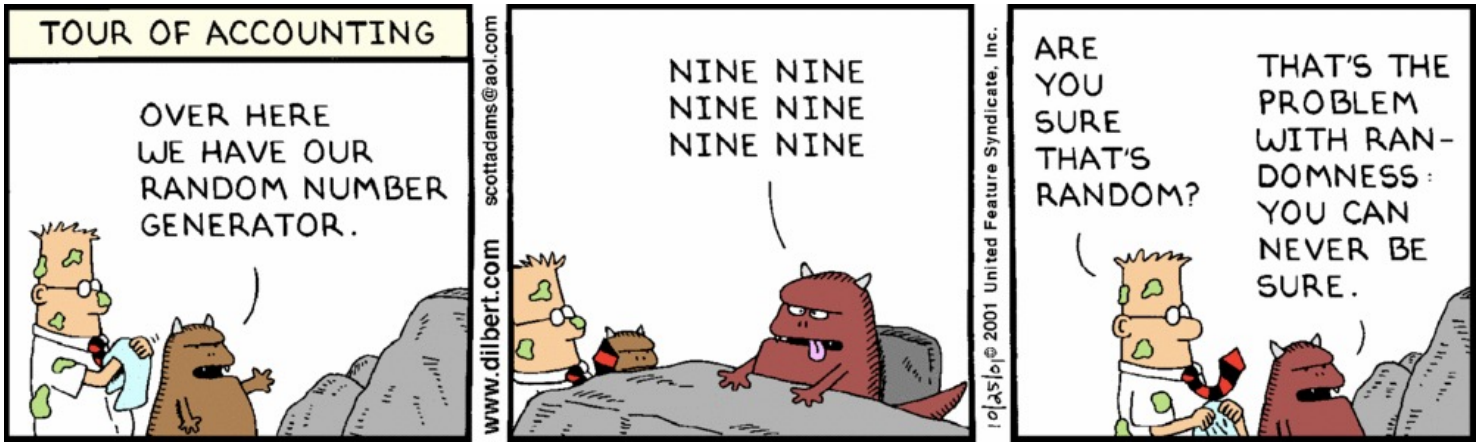
*A private-key encryption scheme has **indistinguishable** encryptions in the presence of an eavesdropper $\iff$ it is **semantically secure** in the presence of an eavesdropper.*

课堂练习（理解安全定义）

- 一次一密方案在出现窃听者时是否是不可区分的？
- 若一个敌手一直在实验中失败，该方案是安全的吗？
- 若从密文中猜测到消息中最低比特的概率是3/4，该方案是安全的吗？
- 相关性：X和Z的分布不可区分，Y和Z的分布不可区分，那么X和Y的分布是不可区分的吗？

伪随机性（pseudorandomness）概念

- 回顾之前完美保密的局限性，密钥长度需要和明文一样长才安全；计算安全中放松了安全的定义，那密钥能不能短一些，或者说能不能放松对随机性的要求，产生足够长但不完全随机的密钥？
- 真随机性不能由一个可描述的机制产生。这里的“机制”不包括“掷骰子”，而是指确定性机制；
- 伪随机对于不知道其机制的观察者来说看起来是真的随机；
- 一个固定的字符串谈不上是否随机/伪随机，随机/伪随机指的是产生字符串的过程；
- 能否证明随机性？不能，我们可能是不知道机制的观察者。



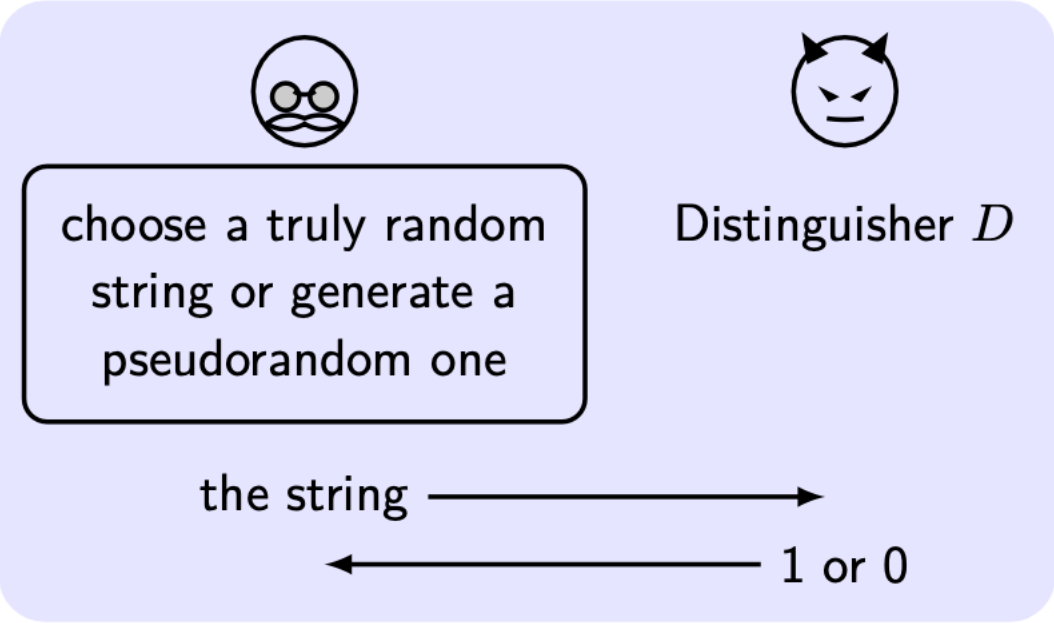
区分器（distinguisher）统计测试

- 一种判断是否随机的务实的方法（区分器）：从一个随机生成器中得到多个随机序列并进行一套统计测试。
  - 例如，序列中0和1的数量之差不应该太大，最大连续0的长度不应该太长等等。
- 伪随机性意味着下一比特不可预测（next-bit unpredictable），通过所有下一比特测试等且仅当通过所有统计测试。（这是姚期智的贡献）
- 问题是难以确定多少测试才足够？

- $D(x) = 0$ if $|\#0(x) - \#1(x)| \leq 10 \cdot \sqrt{n}$
- $D(x) = 0$ if $|\#00(x) - n/4| \leq 10 \cdot \sqrt{n}$
- $D(x) = 0$ if max-run-of-0$(x) \leq 10 \cdot \log n$

定义伪随机性思路

- 直觉：从一个短的真随机种子生成一个长的随机串，这个伪随机串与真随机串是不可区分的。
  - 这是不是和图灵测试类似？
- 区分器输入一个比特串，输出1位比特。*注意：该比特不一定表示输入的串是否是随机的。*



定义伪随机生成器

Definition 4

A deterministic polynomial-time algorithm $G : \{0,1\}^n \to \{0,1\}^{\ell(n)}$ is a **pseudorandom generator (PRG)** if

扩展性

1. (Expansion:) $\forall n, \ell(n) > n$.
2. (Pseudorandomness): $\forall$ PPT distinguishers D,

$$|\Pr[D(r) = 1] - \Pr[D(G(s)) = 1]| \leq \mathsf{negl}(n),$$

where r is chosen *u.a.r* from $\{0,1\}^{\ell(n)}$, the **seed** s is chosen *u.a.r* from $\{0,1\}^n$. $\ell(\cdot)$ is the **expansion factor** of G.

均匀随机 (uniformly at random)

种子

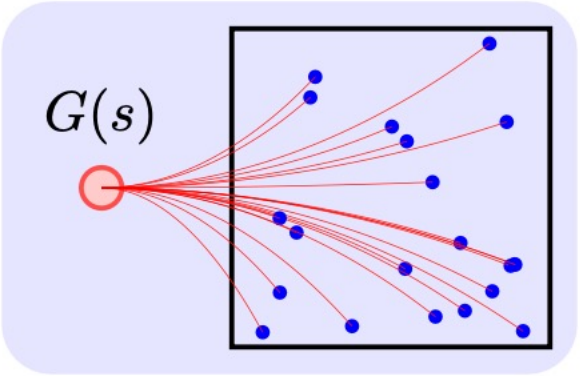
扩展因子

- **Existence**: Under the weak assumption that *one-way functions* exists, or $\mathcal{P} \neq \mathcal{NP}$

单向函数

充分种子空间原则

- 稀疏输出：当扩展因子为2n时，在长度为2n的串中只会产生2^{-n}。
- 蛮力攻击：给定无穷的时间，通过枚举所有种子来产生所有串，能以较高的概率区分出伪随机串。
- 充分种子空间：种子必须长来抵抗蛮力攻击。



真实案例

glibc random()

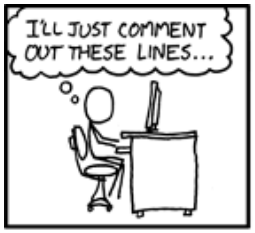
$$r[i] = (r[i-3] + r[i-31])\%2^{32}$$

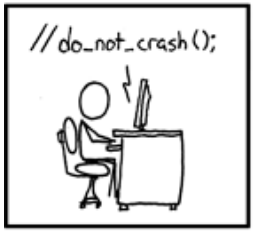
Netscape (by reverse-engineering)

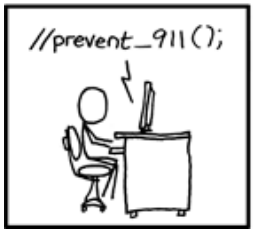
```
global variable seed;
RNG_CreateContext();
    (seconds, microseconds) = time of day;
    pid = process ID; ppid = parent process ID;
    a = mklcpr(microseconds);
    b = mklcpr(pid + seconds + (ppid << 12));
    seed = MD5(a, b);
RNG_GenerateRandomBytes()
    x = MD5(seed);
    seed = seed + 1;
    return x;
```

真实案例

- 2008年，为了避免一个编译警告，Debian的一个发布版本中误删了一行代码，引起OpenSSL中关于随机生成器的漏洞（CVE-2008-0166）。







IN THE RUSH TO CLEAN UP THE DEBIAN-OPENSSL FIASCO, A NUMBER OF OTHER MAJOR SECURITY HOLES HAVE BEEN UNCOVERED:

AFFECTED SYSTEM	SECURITY PROBLEM
FEDORA CORE	VULNERABLE TO CERTAIN DECODER RINGS
XANDROS (EEE PC)	GIVES ROOT ACCESS IF ASKED IN STERN VOICE
GENTOO	VULNERABLE TO FLATTERY
OLPC OS	VULNERABLE TO JEFF GOLDBLUM'S POWERBOOK
SLACKWARE	GIVES ROOT ACCESS IF USER SAYS ELVISH WORD FOR "FRIEND"
UBUNTU	TURNS OUT DISTRO IS ACTUALLY JUST WINDOWS VISTA WITH A FEW CUSTOM THEMES

课堂练习

F is PRG. Is G PRG?

- $G(s)$ is such that $XOR(G(s)) = 1$
- $G(s) = F(s) \| 0$
- $G(s) = F(s \oplus 1^{|s|})$
- $G(s) = F(s) \| F(s)$
- $G(s \| s') = F(s) \| F(s')$
- $G(s) = F(s \| 0)$
- $G : s \leftarrow \{0,1\}^{20}, G(s) = F(s)$

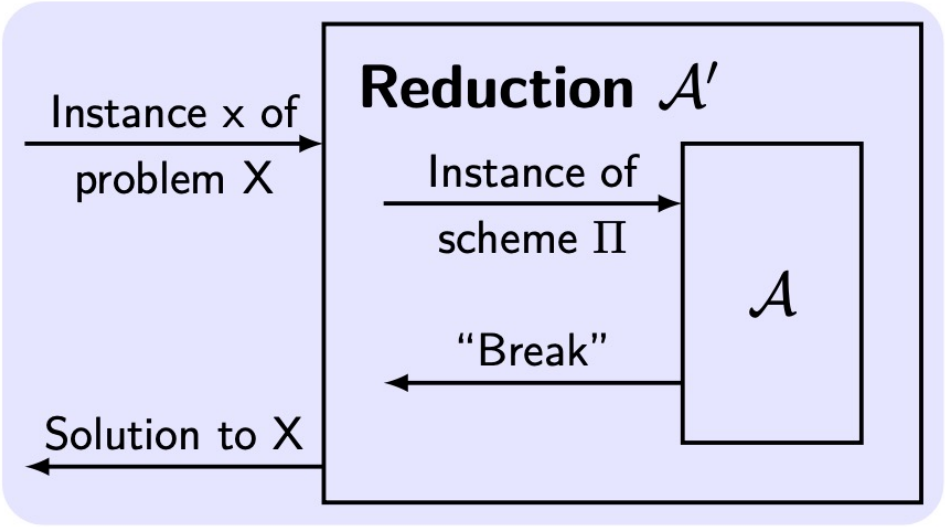
规约法（Reduction）

Reduction $A \leq_m B$ [2]: A is **reducible** to B if solutions to B exist and whenever given the solutions A can be solved.

- 规约法是将一个问题A变换为另一个问题B。变换的意思可以理解为，A可以通过解决B来解决。
- 规约A到B：A可规约为B，如果B的解存在并且给定该解时A可解；可将规约理解为A对B的子函数调用，除了子函数B是黑盒，解决A的步骤都应该是明确的。
- 解决A不能比解决B更难，因为A可通过解决B来得到解决
- 例子
  - 测量矩形面积可规约到测量矩形边长
  - 计算一个数的平方可规约到两个数乘积，相反可以规约吗？

规约证明

- 将解决“假设”的难问题X的算法A’规约到“破解”加密方案的算法A。如果加密方案可以被破解，则假设的难问题也可以解决。这导致矛盾，说明加密方案不可以被破解。



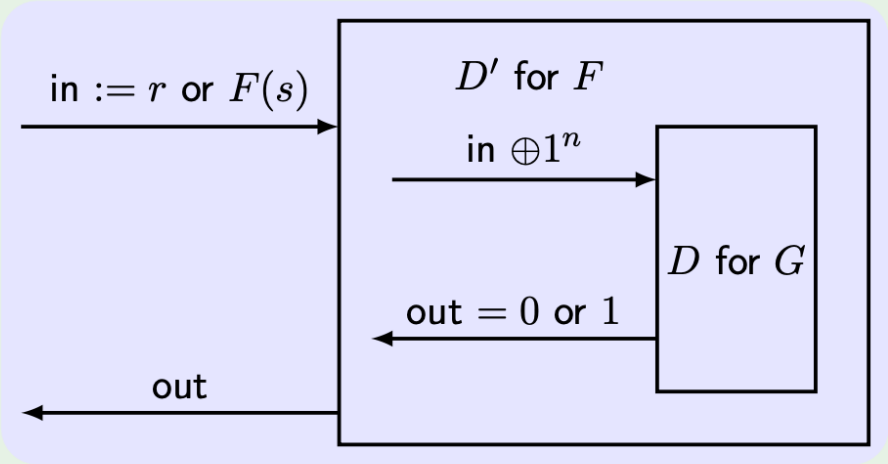
- A PPT $\mathcal{A}$ can break Π with probability $\varepsilon(n)$.
- **Assumption**: Problem X is *hard* to solve.
- **Reduction**: Reduce $\mathcal{A}'$ to $\mathcal{A}$. $\mathcal{A}'$ solves x efficiently with probability $1/p(n)$, running $\mathcal{A}$ as a sub-routine.
- **Contradiction**: If $\varepsilon(n)$ is non-negligible, then $\mathcal{A}'$ solves X efficiently with non-negligible probability $\varepsilon(n)/p(n)$.

规约证明例子

If $F(s)$ is PRG, so is $G(s) = F(s) \oplus 1^{|n|}$?

- Problem A (Assumption): to distinguish $F(s)$ from r
- Problem B (Break the scheme): to distinguish $G(s)$ from r

Idea: Reduce A to B. As $F(s)$ is distinguishable, so is $G(s)$.



$$\Pr[D'(F(s)) = 1] = \Pr[D(G(s) = F(s) \oplus 1^n) = 1]$$

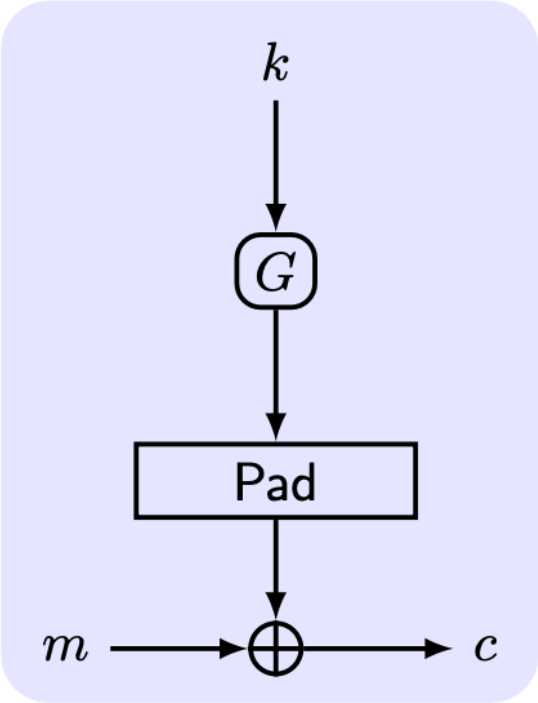
$$\Pr[D'(r) = 1] = \Pr[D(r \oplus 1^n) = 1] = \Pr[D(r) = 1]$$

$$\begin{aligned} \text{negl} &\geq \Pr[D'(F(s)) = 1] - \Pr[D'(r) = 1] \\ &= \Pr[D(G(s)) = 1] - \Pr[D(r) = 1] \end{aligned}$$

According to the definition of PRG, $G(s)$ is a PRG.

安全加密方案构造

- 这个方案和一次一密是类似的，除了密钥更短并且用伪随机生成器生成的比特串来与明文异或。因为伪随机对于任何敌手都可以认为是真随机，所以对于敌手而言，该方案与一次一密是一样的。由此，得到了一个安全加密方案，同时避免一次一密的最大局限性——密钥过长。



Construction 5

- $|G(k)| = \ell(|k|)$, $m \in \{0,1\}^{\ell(n)}$.
- Gen: $k \in \{0,1\}^n$.
- Enc: $c := G(k) \oplus m$.
- Dec: $m := G(k) \oplus c$.

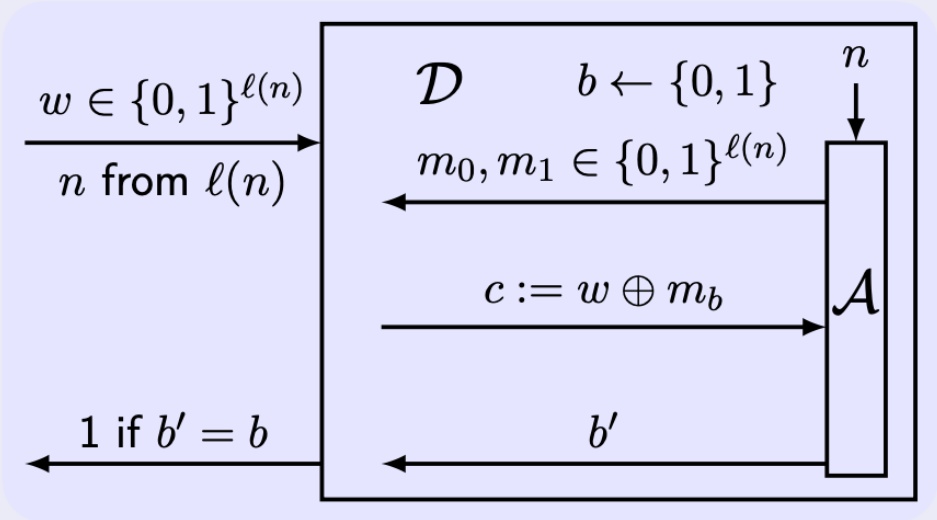
Theorem 6

This fixed-length encryption scheme has indistinguishable encryptions in the presence of an eavesdropper.

不可区分安全证明

- 区分伪随机性为难题假设，破解加密方案为规约的子函数。针对伪随机生成器G的区分器D以$\mathcal{A}$为子函数，使得当$\mathcal{A}$破解了Π则D可以区分出G，与G的伪随机性矛盾。注意这里我们用了符号$\tilde{\Pi}$来表示Π的一个变体，来刻画加密方案中可能使用了真随机串来加密；

Proof.



$$\Pr[D(w) = 1] = \Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\tilde{\Pi}}(n) = 1]$$

□

不可区分安全证明

- 通过规约将A的不可区分实验成功的概率与D的区分器实验输出1的概率建立等式；分析输入真随机串时D输出1的概率（即不可区分实验成功概率）是1/2；根据PRG的定义，输入伪随机串时D输出1的概率（1/2+\varepsilon(n)）与输入真随机串时D输出1的概率（1/2）的差异时可忽略的。

Proof.

To prove $\varepsilon(n) \stackrel{\text{def}}{=} \Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi}(n) = 1] - \frac{1}{2}$ is negligible.

(1) If w is r chosen *u.a.r*, then $\tilde{\Pi}$ is OTP.

$$\Pr[D(r) = 1] = \Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\tilde{\Pi}}(n) = 1] = \frac{1}{2};$$

(2) If w is $G(k)$, then $\tilde{\Pi} = \Pi$.

$$\Pr[D(G(k)) = 1] = \Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi}(n) = 1] = \frac{1}{2} + \varepsilon(n).$$

Use Definition 4:

$$|\Pr[D(r) = 1] - \Pr[D(G(k)) = 1]| = \varepsilon(n) \leq \mathsf{negl}(n).$$

□

作业

Handling Variable-Length Messages (homework)

Definition 7

A **deterministic** polynomial-time algorithm G is a **variable output-length pseudorandom generator** if

1. $G(s, 1^\ell)$ outputs a string of length $\ell > 0$, where s is a string.
2. $G(s, 1^\ell)$ is a prefix of $G(s, 1^{\ell'})$, $\ell' > \ell$.[3]
3. $G_\ell(s) \stackrel{\text{def}}{=} G(s, 1^{\ell(|s|)})$. Then $\forall \ell(\cdot)$, G_ℓ is a PRG with expansion factor ℓ.

Both Construction 5 and Theorem 6 hold here.

[3] for technical reasons to prove security.

本节小结

	Computational	**Info.-theoretical**
Adversary	PPT eavesdropping	no limited eavesdropping
Definition	indistinguishable $\frac{1}{2}$ + negl	indistinguishable $\frac{1}{2}$
Assumption	pseudorandom	random
Key	short random str.	long random str.
Construction	XOR pad	XOR pad
Prove	reduction	prob. theory