基于非完美随机源的密码学原语的安全性研究综述

姚燕青, 李舟军

北京航空航天大学计算机学院 北京 中国 100191

摘要 密码学是信息安全的核心研究内容。传统的密码学原语理想地假设秘密服从均匀随机分布。然而, 在现实世界中往往并非如此。例如, 若秘密源为生物数据、物理源、部分泄漏的秘密等时, 相应的分布并不服从均匀分布。这样的一些分布构成的集合称为“非完美随机源”。因此, 基于非完美随机源的密码学原语是否仍具有安全性? 这已成为当今密码学前沿研究领域的热点和难点课题之一。本文阐述了基于非完美随机源的密码学原语的研究背景、意义及发展历程, 重点介绍了该领域的最新进展, 即 Dodis 和 Yao[CRYPTO 2015]发现的基于一般的非完美随机源的传统隐私(包括位抽取器、加密、承诺、秘密分享方案)和差分隐私的(不)可能性结果。最后, 指出了当前该领域值得探索的问题。

关键词 非完美随机源; 密码学原语; 安全性; 传统隐私; 差分隐私
中图法分类号 TP309.7 DOI 号 10.19363/j.cnki.cn10-1380/tn.2016.02.003

Survey: Security of Cryptographic Primitives with Imperfect Randomness

YAO Yanqing, LI Zhoujun

School of Computer Science and Engineering, Beihang University, Beijing 100191, China

Abstract Cryptography is a core area in information security. Traditional cryptographic primitives take for granted the availability of perfect random sources. However, in many situations it seems unrealistic to expect a source to be perfectly random, and one must deal with various imperfect sources of randomness. Some well known examples are physical sources, biometric data, secrets with partial leakage, etc. Hence, can cryptographic primitives with imperfect randomness be secure? It has become a frontier and hot research topic in Cryptography. This paper reviews the background, significance, and the development history of this topic. It also reviews the latest advances of this topic. Namely, some general impossibility results of traditional privacy (e.g., bit extractor, encryption, commitment, secret sharing scheme) and differential privacy under a general imperfect source proposed by Dodis and Yao in CRYPTO 2015. Finally, the paper analyzes the problems worth exploring in this area.

Key words imperfect random sources; cryptographic primitives; security; traditional privacy; differential privacy

1 引言

密码学是网络信息安全的理论基础和必要保证。密码系统的输出必须在敌手看来是一系列的随机值, 看起来与原始信息毫无关联。在设计密码系统时, 密钥必须为随机的, 否则破解密文将不再困难。密码学中著名的 Kerckhoff 准则为: “一个密码系统的安全性都应该基于密钥的安全性, 而不是基于算法的细节的安全性”。可见, 密钥的随机性在密码系统中起着极其关键的作用。在公钥加密体制中, 为了达到选择明文攻击下的不可区分性(简称为 IND-CPA), 除了要求公钥和明文在敌手看来是随机的, 每次加密中还需引入新鲜独立的随机串。因此, 随机性在某种程度上决定了密码系统的安全性。

在传统的密码学研究中, 假设秘密服从均匀随机分布。密码应用大多适用算法来生成随机数。这些算法是确定的, 所以产生的序列并非统计随机的。但当算法好时, 产生的序列可以经受住随机性检测。

通讯作者 1: 姚燕青, 博士, 讲师, Email: yaoyanqing1984@buaa.edu.cn。**通讯作者 2:** 李舟军, 博士, 教授, Email: lizj@buaa.edu.cn。
本课题得到国家 863 计划(No.2015AA016004), 国家自然科学基金(Nos.61170189, 61370126, 61202239), 北航软件开发环境国家重点实验室探索性自选课题以及校级基本科研业务费项目(No.30486301)资助。
收稿日期: 2015-11-29; 修改日期: 2016-4-19; 定稿日期: 2016-4-28

这样的数称为伪随机数。然而在诸多现实应用中, 秘密连伪随机的要求都达不到。例如: 若秘密源为生物数据[6,21]、物理源[8,11]、部分泄漏的秘密[3,40]、Diffie-Hellman 密钥交换中的群元素[26,31] 时, 则它不是完美随机的。相应的分布不服从均匀分布。一些非均匀分布构成的集合称为“非完美随机源(Imperfect Random Source)”。例如, 多数已有的安全模型假设在加密过程中不会泄漏关于密钥和用于加密的随机值的信息, 而真实环境中敌手可能会通过各种密钥泄漏攻击(例如计算时间、功耗、故障检测、电磁辐射、散热等)来获得关于秘密状态的部分信息。又如, 有些机构利用生物特征来提取密钥, 由于生物特征(例如指纹、脸相、虹膜等)的唯一性和终身不变性, 用这种特征来提取密钥具有不被遗忘或丢失, 随身可用, 不易被伪造等优点, 不过这样的密钥亦不服从均匀分布。

密码学原语是密码系统的基本构件。密码学原语包括: 哈希函数、消息认证码、签名方案、抽取器、加密方案、承诺、秘密分享、差分隐私、密钥生成函数、伪随机数生成器等。当今密码学领域的一个研究热点和难点课题是: 原有的密码学原语在非完美随机源下是否仍然具有安全性? 该领域的研究在秘密取自物理源、生物源、部分泄漏或篡改的秘密源的现实社会中具有广阔的应用前景。

1.1 国内外的相关研究团队

目前, 国际上已有一些著名研究团队研究基于非完美随机源的密码学, 如: 美国麻省理工学院的图灵奖获得者 Shafi Goldwasser 教授[7] 的研究小组、哈佛大学的 Vadhan 和 Sudan 教授[14,17] 的研究小组、美国纽约大学的 Yevgeniy Dodis 教授[5,17,20] 的研究小组、美国 University of Texasat Austin 的 David Zuckerman 教授[32,46] 的研究小组、以色列 Weizmann Institute of Science 的 Moni Naor 教授[4] 的研究小组、澳大利亚伍伦贡大学的 Yi Mu 教授[15] 的研究小组等。基于非完美随机源的密码体制(尤其是泄漏容忍的密码体制)也引起了国内一些机构的研究兴趣, 如中科院信息安全国家重点实验室[36,47]、清华大学[30]、上海交通大学[25,33]、武汉大学[22]、暨南大学[35]、山东大学[29]等。

到目前为止, 学者们已得到关于该课题的一些结果。概括来讲, 认证方案(例如消息认证码、数字签名方案)对于密钥的随机性程度的要求较为宽松[1,20,16,38], 但加密方案以及其他涉及“隐私”的方案却并非如此(例如, 文献[1,5,20,23,37,41])。本文将论述几种典型的非完美随机源, 然后分别阐述基于非完美随机源的认证方案、传统隐私体制、差分隐私体制的安全性研究的发展历程及发展动态, 其中将重点介绍 Dodis 和 Yao 在 CRYPTO 2015 中的最新结果, 接着介绍计算理论意义下基于弱随机密钥的密码学原语的安全性研究历程及存在的问题。最后, 指出该领域目前值得探索的问题。

2 非完美随机源模型

目前, 已出现了一些形式化的模型(例如弱源、Block 源、Santha-Vazirani 源、偏差控制受限源)来抽象非完美随机源[2,10,12,13,19,34,41,42,46]。粗略地说, 它们分为可抽取的和不可抽取的。位抽取问题可概括如下: 该问题有三个参数 m, n, 和 t。用户选取函数 $f:\{0,1\}^m \to \{0,1\}^n$, 敌手选取 m 个输入位中的 t 个位, 并对这 t 个位的取值进行固定, 用户不知道敌手选了哪 t 个位以及这 t 个位的取值是多少。剩余的 n-t 个输入位的取值为独立无偏的抛币游戏的输出。用户用函数 f 作用于这 n 位输入。用户的目标是使得函数的输出服从 $\{0,1\}^n$ 上的均匀分布, 而敌手的目标则是使得函数的输出分布是有偏的。位抽取问题归结为用户的取胜问题。文献[12]给出了使敌手失败的 m 的上下界。概括地讲, 可抽取的源(例如文献[10,12,34,42])允许几乎均匀分布的确定性抽取。尽管使得抽取比率和效率最优化是很有趣的问题, 但从质的角度, 适用于完美随机源的应用中的源都可用可抽取的源来代替。不幸的是, 人们很快意识到现实中的许多源不是可抽取的(例如弱源、Block 源、Santha-Vazirani 源、偏差控制受限源)[13,19,41]。

Santha-Vazirani 源 Santha 和 Vazirani 于 1986 年提出了 Santha-Vazirani 源[41](简记为 SV 源)。该源产生一系列的位 r_1, r_2, ..., 满足 $\Pr[r_i = 0 \mid r_1 r_2 \ldots r_{i-1}] \in [(1-\gamma)/2, (1+\gamma)/2]$, 这里 i 为任意正整数。可高效取样的 SV 源[1] 是逐位产生的, 对于产生的位 r_i, 敌手在获悉已产生的 i-1 个位 r_1, r_2, ..., r_{i-1} 的信息的前提下, 以“在线”的方式高效地对服从均匀分布的 r_i 以独立概率 p 进行篡改。不难看出, 在均匀随机的 n 位二元串上进行上述篡改得到的随机变量服从 $\gamma = p$ 的 SV 分布。事实上, 任何 SV 源既可看成物理源, 又可看成由敌手对均匀随机的 n 位二元串进行上述篡改而得到的源[1]。不管有多少个 SV 位, 均不存在 1 位的确定性抽取器, 能够从 SV 源中抽取出偏差严格小于 γ 的 1 位[41]。

偏差控制受限源 在现实世界中, 流源产生的每一位可能不是均匀随机的, 由于噪音、测量错误及其它一些缺陷, 微小的错误是不可避免的; 由于内

部联系、测量条件限制或设置不当, 一些位非平凡地依赖于前面的位, 其极端情况是有的位由前面的位完全确定。2001年, Dodis [19] 给出了对上述源的模型化表示, 称之为偏差控制受限源(简称为 BCL 源)。该源产生一系列的位 x_1, x_2, ..., x_n: 对 i=1,2, ..., n 来说, x_i 的值按以下两种方式之一依赖于 $x_1, x_2, \ldots, x_{i-1}$:

(A) x_i 由 $x_1, x_2, \ldots, x_{i-1}$ 完全确定, 但这种情况发生的次数至多为某常数 b;

(B) $\Pr[r_i = 0 \mid r_1 r_2 \ldots r_{i-1}] \in [(1-\gamma)/2, (1+\gamma)/2]$

这里 $0 \leqslant \gamma < 1$。特别地, 当 $b = 0$ 时, 此源退化为 Santha 和 Vazirani 引入的 SV 源[41]。与 SV 源相比, BCL 源看起来更切合实际, 尤其是当 b 选择适当的时候。

弱源和 Block 源 最小熵至少为 k 的 n-位弱源(Weak Source) 定义为集合 $\{R \mid H_\infty(R) \geqslant k$, 其中 R 为 $\{0,1\}^n$ 上的分布$\}$ Block 源作为弱源的推广, 允许有 n/m 个块 $R_1, \ldots, R_{n/m}$, 每块在假设已知前面的块的前提下有新鲜的 k 位熵。与 SV 源相比, Block 源不含最大熵的信息, 从这个意义上来说, Block 源的限制性更少, 也比 SV 源更切合实际。

3 基于非完美随机源的概率多项式时间的算法和认证研究

尽管不可抽取的源排除了对于所有应用来说从非完美源到完美源的“黑盒编译”, 但我们仍希望特定的不可抽取的源足以适用于模拟概率算法及密码学中的某些应用。一系列结果[2,13,41,43,46] 表明: 很“弱”的源(包括 SV 源、甚至更弱的切合实际的“弱”源和“Block”源)对于模拟概率多项式时间的算法已经足够; 也就是说, 对于那些在本质上不需要随机源, 不过用随机源时将潜在地提高效率的问题来说, 用很“弱”的源已经足够。另外, 即使在以随机源为基本考虑对象的密码学(例如密钥生成算法)中, 许多不可抽取的源(包括 SV 源等)足以适用于认证应用(例如适当的困难性假设下的消息认证码[16,38] 及签名方案[1,20])。直观地讲, 由于认证应用仅仅要求: 完全猜测(即伪造)某些长布尔串对于攻击者来说是困难的, 因此知道源的最小熵就足以实现成功认证。

4 基于非完美随机源的隐私体制的安全性研究

该部分阐述抽取器、加密体制、承诺、秘密分享、零知识证明、差分隐私等涉及“隐私”的密码体制(简称为“隐私体制”)在非完美随机源下的安全性结果。

4.1 基于非完美随机源的传统隐私体制

当处理传统隐私体制(例如抽取器、加密、承诺、秘密分享、零知识证明等)时, 关于认证的分析不再适用。首先, McInnes 和 Pinkas[37] 得到了结论: 不能基于 SV 源来建立非条件安全的对称加密方案, 即使限制为只加密 1 位。 该结果接着被 Dodis 等人进一步强化, 他们得到: SV 源甚至不能用来构造计算意义上安全的加密方案(即使是加密 1 位), 以及任意其它隐私体制(例如承诺、零知识、秘密分享等)[20]。Austrin 等人[1] 得到了更强的结果: 即使 SV 源可以高效取样(efficient sampling), 这些负面结果仍然成立。 另外, Bosley 和 Dodis[5] 得到了甚至更负面的结果: 若随机源足以用来生成一个能加密 k 位的密钥, 则能确定性地从该随机源中抽取大约 k 位几乎均匀随机位。这就意味着传统隐私需要可抽取的随机源。从积极方面讲, 文献[5]和[23]得到: 可抽取的源对于加密“很少的”几位不是严格必需的。不过, 对于自然的不可抽取的源(例如 SV 源)来说, 文献[1,20,41]已得出结论: 即使仅加密 1 位也是不可能的。

4.2 基于非完美随机源的差分隐私体制

尽管上述一系列负面结果似乎指明了方向: 隐私本质上需要可抽取的随机源, 但是最近 Dodis 等人[17] 却发现 SV 源足以实现隐私中的一个较新的概念(即差分隐私)。差分隐私的研究对象是统计数据库。具有隐私保护作用的统计数据库能保证在不泄漏用户的私密数据的情况下, 让他人获得较宽松的统计事实。差分隐私可保证删除或添加数据库的一条记录不会(大幅)影响体制的输出。通俗地讲, 体制 $M(D, f; r)$ 把随机值 r(即“噪音”)添到真实回答 $f(D)$ 中去, 这里 D 为用户构成的敏感数据库, $f(D)$ 为关于 D 的用户的一些有用的聚合信息。该噪音以某种方式添进去, 满足以下两条似乎相冲突的性质:

(a) ε-差分隐私: 返回值 $z = M(D, f; r)$ 至多以“优势” ε 不泄漏关于单独用户 i 的值 $D(i)$的任意信息;

(b) ρ-效用:在关于 r 的平均意义上, $\|z-q(D)\|$ 以 ρ 为上界, 表示被干扰的回答离真实回答并不远。

加法噪音机制[18,27,28]具有形式 $M(D,f;r)=f(D)+X(r)$,其中 X 为一个适当的“噪音”分布, 被加到真实回答中以保证有差分隐私性。例如, 对于完美随机源来说, 当考虑统计查询时, 适当的分布 X 为拉普拉斯分布[18]。然而, 对于 SV 源来说, 我们不能找到参

数合适的拉普拉斯分布来得到差分隐私机制。事实上, 若存在基于非完美随机源的具有差分隐私性和效用性的机制, 则存在基于该源的随机抽取器, 故由 SV 源的不可抽取性可得:对于 SV 源来说, 不存在具有差分隐私性和效用性的加法噪音机制[17]。从另一个角度讲, 不妨设 T_i (i=1,2)为满足 $M(D_i, f; r)=z$ 的 r 构成的集合, 加法噪音机制必须满足 $T_1 \cap T_2 = \Phi$, 在此基础上, SV 敌手总能成功地扩大比率 $|\Pr[r \in T_1] - \Pr[r \in T_2]|$[24] 或 $\Pr[r \in T_1]/\Pr[r \in T_2]$[17]。

尽管不存在基 SV 源的形如 $M(D,f; r)=wt(D)+X(r)$ 的加法噪音差分隐私机制, 这里 wt 为汉明重量函数, 但 Dodis 等人[17] 于 2012 年构造出了一个结构更好的机制, 该机制关于这种源具有差分隐私性。更具体地, 他们采用“一致采样”(即“consistent sampling”)来建立 SV-鲁棒的机制[17], 并从 SV 源的位到位的性质出发引入了另一条件。这两个条件的组合称为 SV-一致采样。他们还利用截断和算术编码等技术构造了明确的具有差分隐私性和效用性的拉普拉斯机制。这样的机制对于所有的 SV 分布来说都适用, 前提是效用 ρ 的上界被放松为关于 $1/\rho$ 的多项式, 该多项式的度和系数依赖于 γ 而不依赖于数据库 D 的规模。另外, 值 ε 可以为任意小的常数(例如 ε 远远小于 γ)。这与传统隐私中的基于 SV 源的不可能性结果[37,20] 不同, 那里 $\varepsilon=\Omega(\gamma)$(意味着连固定常数安全都不可能实现, 更不用说安全参数为“可忽略的”的值时的情形了)。这些结果表明传统隐私与差分隐私有一定的差距。Dodis 等人[17] 留下公开问题: 差分隐私能否建立在比 SV 源更切合实际的源上? 由于 BCL 源和 Block 源比 SV 源更切合实际, 因此, 能否利用 BCL 源、Block 源等来构造差分隐私体制是一个很有意义的课题。

4.3 最新进展

以该问题为部分动机, Dodis 和 Yao[24] 在 CRYPTO 2015 中引入了一个直观的、模块化的框架来研究关于传统隐私和差分隐私的不可能性结果, 这些结果是基于一般的非完美随机源的。该结果统一并强化了观点: 多数情况下, 对于非完美随机源来说, 隐私是不可能的, 除非该源是(几乎)确定性可抽取的[24]。

整体思路

Dodis 和 Yao [24] 引入了源的可表达性和可分离性的概念来衡量秘密的“非完美性”。从高层面上说, 文献[24]借鉴文献[20](那里只集中于研究 SV 源)中的思想, 不过以一种更模块化和量上最优化的方式来研究, 从而使得其证明在某种程度上更富有启发性[24]。从本质上说, 这些结果利用 3 步得到了基于非完美随机源 $\mathcal{R}$ 的一个给定的隐私体制 P 的不可能性结果:

步骤 1 基于源 R 的隐私的不可能性问题 → 源 R 的可表达性。 直观地讲, R 的可表达性的意思是 R 足以“区分”任意两个不是几乎处处相等的函数 f 和 g: 存在 R 中的分布使得 SD($f(R)$,$g(R)$) 是“显著的”, 其中 SD 为两个概率分布的统计距离。

在这一抽象下, 容易得出: 多数隐私体制 P(例如抽取器、加密、秘密分享、承诺)的某种程度的安全性与 R 的可表达性相矛盾。例如, 当 P 是加密方案时, $f(r)$ 和 $g(r)$ 理解为密钥 r 下的两个不同明文的加密函数。对于抽取、秘密分享及承诺有相似的讨论。

更有趣的是, 文献[24]得出结论: 若源是某种程度上可表达的, 则不可实现基于该源的某种程度的差分隐私。该证明与隐私体制的不可能性结果的证明有相似处, 但前者包含的思想更丰富。这是因为差分隐私性仅仅限制了“相接近的”数据库的安全性, 而效用性则仅对于相对“远”的数据库是有意义的。特别地, 正因如此, 源 R 上的可表达性的要求对于排除差分隐私来说, 与传统隐私相比要稍微强一些。除了这点量上的不同, 在文献[24]的讨论中传统隐私和差分隐私没有质上的不同。

总之, 看似简单的“把隐私归约为可表达性”的讨论正是文献[24]的框架的一个特征, 仅仅在这一步涉及应用 P 的具体细节。接着将集中于研究源 R 。

步骤 2 源 R 的可表达性 → 源 R 的可分离性。 直观地讲, R 的可分离性的含义是: R 足以“分离”任意充分大的不交集合 G 和 B。不失一般性, 假设 $|G| \geqslant |B|$, 则存在 R 中的分布 R, 使得 $|\Pr[R \in G] - \Pr[R \in B]|$ 是“显著的”。不难发现, 可分离性与可表达性密切相关, 不过前者限制为其支撑集互不相交的布尔函数 f 和 g(即 G 和 B 的特征函数), 这就使得处理起来更容易些。

文献[24]证明了由可分离性一般可推导出可表达性, 前后两者的参数几乎相同。这恰恰是文献[24]不同于[20]且在量上改进文献[20]的地方: 文献[20]用了位到位的混合讨论来阐明 SV 源的可表达性, 而文献[24]则采用了更聪明的“universal hashing trick”来证明, 从而使得其结果与函数 f 和 g 的值域无关(此值域对应于密文、承诺及秘密分享等的规模)。(a) 被抽取的位可保证是几乎无偏的, (b) 尽管抽取器可能输出 $\perp$, 但它至少在均匀分布上将以足够大的概率输出 0 或 1。

与步骤 1 相结合, 文献[24]得到以下两个结果。首先, 把基于源 $\mathcal{R}$ 的几种隐私体制 P 的不可能性结果归约为一个更简单的 $\mathcal{R}$ 的可分离性。第二, 把基于 $\mathcal{R}$ 的 P 的可行性结果归约为基于 $\mathcal{R}$ 的确定性弱位抽取的存在性。回顾 Bosley 和 Dodis 的结果: 由几种传统的隐私原语(仅包括多位的加密和承诺, 但不包括秘密分享)的安全性可推出多位确定性抽取器的存在性[5]。从而文献[24]不可比拟地对上述结果进行了补充。从积极方面来说, Dodis 和 Yao [24] 的结果可应用于更广泛的隐私原语(例如秘密分享、一位加密和承诺)。从消极方面来说, Dodis 和 Yao [24] 仅讨论一种相对弱的一位抽取器, 这里的抽取器可能输出, 而 Bosley 和 Dodis[5]则得到了传统的可能多位的抽取器。

步骤 3　几种源 R 的可分离性。与文献[1,20,37]中的结果不同, 上述所有结果对于任意非完美随机源来说都是正确的。为了得到基于自然源的隐私的不可能性结果, 必须确定具体的源的比较好的可分离性界。 由文献[20]可得 SV 源和一般的弱源的可分离性界, 文献[24]证明了 Block 源[13] 和 BCL 源[19]的新的可分离性界。特别地, Block 源的可分离性界是不容易得到的, 是文献[24]的亮点之一。

这些 Block 源和 BCL 源的新的可分离性结果除了其自身的价值之外, 对差分隐私的研究也是很有意义的(见下面)。事实上, 这两者都可被看做是对结构化很强的(且不切实际的！)SV 源的一种切合实际的放松, 不过不如弱源更一般/非结构化。既然已经得出对于 SV 源来说, 差分隐私是可行的, 那么一个自然的问题是研究当源慢慢地变得更切合实际/非结构化, 且在变成一般的弱源之前, 能以多快的速度回归为不可能性结果。

把新的和旧的不可能性结果综合起来

把步骤 1-3 应用于具体的源(即弱源、Block 源、SV 源及 BCL 源), 我们立刻得到关于传统隐私的各种不可能性结果。虽然这些结果主要为关于差分隐私的(全新的)不可能性结果的基础, 它们也对文献[20]的结果在量上进行了改进(源于从更强的可表达性到可分离性的归约)。例如, 它们甚至排除了常数(与可忽略的值构成对比)安全的加密/承诺/秘密分享, 且与密文/承诺/分享的规模无关。相关地, 我们自然地得到了关于 Block/BCL 源的比 SV 源更强的不可能性结果。

更有价值的是, 文献[24]得到了第一个基于非完美随机源的关于差分隐私的不可能性结果。受文献[17]的正面结果的启发, 文献[24]的关于非完美随机源的可分离性结果(仅仅)对基于 SV 源的差分隐私体制的不可能性结果无效。正如文献[24]所解释的, 导致这一失败的原因不是文献[24]的框架太弱而不能应用于 SV 源或差分隐私, 而是由于差分隐私中隐私和效用之间的“部分与全体之间的差距”。

然而, 一旦我们考虑一般的弱源, 或结构化更强的 Block/BCL 源, 将很快地得到相应的不可能性结果！例如, 当研究效用为 ρ 的 ε-差分隐私时, 最小熵为 k(其中 $k = n - \log(\varepsilon\rho) - \mathrm{O}(1)$)的 n 位弱源将被排除, 更一般地, 不管块数 n/m 为多少, 每块长度为 m, 且每块的最小熵为 k(其中 $k=m-\log(\varepsilon\rho)-\mathrm{O}(1)$)的 n-位 Block 源也将被排除。

当 $b = \Omega(\log(\varepsilon\rho)/\gamma)$ 时, BCL 源也被排除。由于 $\varepsilon\rho$ 一般为常数, 故 $\log(\varepsilon\rho)$ 为更小的常数, 这就意味着文献[24]甚至排除了常数熵缺损 n-k(或 m-k, 对于 Block 源)或常数 b。文献[24]把关于传统隐私和差分隐私的不可能性结果进行对比, 观察到后者只比前者稍微弱一点。从而得出结论: 基于切合实际的非完美随机源的差分隐私的实现仍是相当苛刻的。

不过 Dodis 和 Yao[24] 只考虑了加密体制中的 1 位对称加密体制, 且密钥取自非完美随机源的情形。公钥加密体制中,为了实现其标准的 IND-CPA 安全性, 不仅要求公钥和消息是随机的, 对于每个和每次加密来说, 还需引入新鲜独立的随机字串。Bellare 等人[4]定义并设计了新型的公钥加密体制(即 hedged 公钥加密),该体制只需消息和其它随机变量构成的联合分布有足够的信息熵, 就能保证其体制的安全性。可见, 若把密钥之外的其他随机因素考虑进去, 有望得到隐私体制的正面的安全性结果。

5　计算理论意义下基于弱随机密钥的密码学原语的安全性研究

密码学原语的安全性[25]可形式化地定义如下:

令 T 表示由运行时间、电路规模、预言机的询问数等构成的元组。假设敌手 A 知道 T。对于任意 $r \in \{0,1\}^m$, 令 $f(r)$ 表示当密钥为 r 时敌手 A 的攻击优势。密码学原语 P 在理想模型(或现实模型)下是 (T, ε)-安全的, 若对于知道 T 的任意敌手 A 来说, $f(U_m)$(或 $f(R)$)的期望的上界为 ε, 这里 U_m 表示 $\{0,1\}^m$ 上的均匀分布(R 表示某非均匀分布)。$f(R)$的期望称为弱期望。例如, 对于 CPA-安全的对称加密方案来说, 敌手的“资源”T=(t,q), 这里 t 为敌手的运行时间, q 为敌手 A 的总的加密询问次数。特别地, 允许敌手 A(适应性地)向挑战者 C(r)对任意 q-1 条消息

$s_1,...,s_{q-1}$ 进行密钥为 r 的加密询问, (在任意时刻)进行一次“挑战询问”(s_0^* , s_1^*)。对于挑战询问来说, 挑战者 C (r) 均匀随机地选取 $b\in\{0,1\}$, 然后返回 s_b^* 的加密。 最后, 敌手输出一位“b'”, 若 $b'=b$, 则敌手赢得了该游戏。敌手在密钥 r 上的攻击优势 $f(r)$ 为 Pr[$b'=b$]-1/2。

根据安全模型的选取特点, 我们把常用的密码学原语分为不可预测性应用(例如单向函数、消息认证码及签名方案)和不可区分性应用(例如 CPA-安全的对称加密方案、弱伪随机函数、抽取器以及非延展抽取器)。“不可预测性”描述在安全游戏中敌手难以预测新的合理的“对”这一性质,“不可区分性”则表示敌手难于区分“挑战对”。

Dodis 和 Yu[25] 得到了关于 $f(R)$的不等式, 该不等式把$f(R)$的弱期望的上界表示为两部分的积: 第一部分只依赖于熵缺陷(即弱随机密钥的长度 m=length(R) 和熵之间的差), 第二部分依赖于 $f(U_m)$ 或 $f(U_m)^2$ 的期望,即敌手在服从均匀分布的理想密钥下的平均攻击优势。得出结论: 当敌手的攻击能力有限时, 在密钥的熵缺损足够小的前提下, 若一个密码体制(包括不可预测性体制和“square-friendly”的不可区分性体制)在均匀随机密钥源下是计算理论意义上安全的, 则当其中的密钥用弱密钥来代替时, 其安全性不会受到很大影响。更具体地, 有下述两个结果:

结果 1: 若不可预测性应用 P 在理想模型下是 (T, ε)-安全的, 则应用 P 在密钥 R 的最小熵满足 $H_\infty(R)\geqslant m-d$ 的现实模型下是 $(T,2^d\varepsilon)$-安全的。

结果 2: 若“square-friendly”不可区分性应用 P 在理想模型下是 (T',ε) -安全和 (T',T,γ) -可模拟的, 则应用 P 在密钥 R 的碰撞熵满足 $H_2(R)\geqslant m-d$ 的现实模型下是($T,\sqrt{2^{d-1}(\varepsilon+\gamma)}$) -安全的。

与之前针对具体的密码学原语进行研究(例如, 文献 [16,38] 巧妙地分析了基于弱源的(信息论上)的一次消息认证码的安全性)不同, 文献[25]对多种密码学原语的安全性进行抽象和概括, 从统一的角度去研究。不过结果 1 只研究了最小熵时的情形, 结果 2 只研究了碰撞熵时的情形。 而 Rényi 熵作为最一般的熵概念(它涵盖了最小熵、碰撞熵、Shannon 熵和其它一些熵[39]), 为我们提供了一个新的更一般的对密钥随机性的测量方法。Yao 和 Li[45] 以 Hölder 不等式为基本工具, 通过挖掘 Rényi 熵与碰撞熵之间的联系和概率函数 $f(R)$的不同阶矩之间的关系, 把上述结论扩充到了 Rényi 熵时的情形, 得出类似的结论。然而, 文献[25]和[45]仅研究了当密钥的熵给定时的情形, 且其结果不能涵盖一次性密钥加密算法、伪随机数生成器(PRG)等非“square-friendly”的不可区分性体制。若密钥取自其它非完美随机源(例如 Block 源、SV 源、偏差控制受限源)时, 文献[25,45]却没法充分利用源的信息来分析密码体制的安全性, 更不涉及明文等也取自非完美随机源的情形。

Backes 等人[9] 研究了基于最大熵和最小熵被限定的秘密源(例如 SV 源)的密码学原语的安全缺损的量化方法, 得出结论: 若秘密服从均匀随机分布时, 两个体制 X_0 和 X_1 是不可区分的, 则当秘密用满足该限定条件的弱随机秘密来代替时, 这两个体制在某种程度上是差分不可区分的。这里的秘密包括密钥和加密过程中引入的随机值等。该结论为我们研究基于非完美随机源的密码学原语的安全性提供了潜在的工具——“差分不可区分”模型, 但离解决“若秘密取自非完美随机源时,体制 X_0 和 X_1 是否仍为不可区分的?”这一问题还有很大差距。

6 结束语

基于非完美随的密码学原语的安全性研究是一个较新的研究课题。对该课题的研究方兴未艾, 尽管已取得了一些成果, 但仍有许多值得探索的问题, 下面列举几个, 希望能达到抛砖引玉的目的。

(1) 研究更复杂场景下的隐私问题的不可能性结果;

(2) 挖掘并形式化新型的切合实际的非完美随机源, 并研究关于隐私问题的不可能性结果;

(3) 构造基于新型的非完美随机源的差分隐私机制;

(4) 在弱随机密钥服从 SV 分布、Block 分布或偏差控制受限分布, 且敌手的攻击能力有限的前提下, 系统深入地研究基于这种弱随机密钥的密码学原语与基于均匀随机密钥的密码学原语之间的内在联系, 探索影响密码学原语的安全性的本质因素, 最终得出基于弱随机密钥的密码学原语的安全性的一般性结论。

致 谢 本课题得到国家 863 计划(No.2015AA016004), 国家自然科学基金(Nos.61170189,61370126,61202239), 北航软件开发环境国家重点实验室探索性自选课题以及校级基本科研业务费项目(No.30486301)资助。

参考文献

[1] Austrin P., Chung K.M., Mahmoody M., et al. On the Impossibility of Cryptography with Tamperable Randomness. *CRYPTO 2014*, vol.8616 of LNCS, pp. 462-479

[2] Andreev A.E., Clementi A.E.F., Rolim J.D.P., et al. Weak random sources, hitting sets, and BPP simulations. *SIAM J. Comput.*, 1999, 28(6): 2103-2116

[3] Alwen J., Dodis Y., and Wichs D. Survey: Leakage Resilience and the Bounded Retrieval Model. *ICITS 2009*, pp. 1-18

[4] Bellare M., Brakerski Z., Naor M., et al. Hedged public-key encryption: How to protect against bad randomness. *ASIACRYPT 2009*, pp. 232–249

[5] Bosley C. and Dodis Y. Does privacy require true randomness? *TCC 2007*, vol. 4392 of LNCS, pp. 1-20

[6] Boyen X., Dodis Y., Katz J., et al. Secure remote authentication using biometric data. *EUROCRYPT 2005*, vol. 3494 of LNCS, pp. 147-163

[7] Brakerski Z. and Goldwasser S. Circular and Leakage Resilient Public-Key Encryption under Subgroup Indistinguishability -(or: Quadratic Residuosity Strikes Back). *CRYPTO 2010*, pp.1-20

[8] Barak B. and Halevi S. A model and architecture for pseudo- random generation with applications to /dev/random. *ACM Conference on Computer and Communications Security 2005*, pp. 203-212

[9] Backes M., Kate A., Meiser S., and Ruffing T. Secrecy Without Perfect Randomness: Cryptography with (Bounded) Weak Sources. *ACNS 2015*, pp. 675-695

[10] Blum M. Independent unbiased coin-flips from a correclated biased source-a finite state Markov chain. *Combinatorica*, 1986, 6(2): pp. 97-108

[11] Barak B., Shaltiel R., and Tromer E. True random number generators secure in a changing environment. In: *Proceedings of the 5th Cryptographic Hardware and Embedded Systems, 2003*: pp. 166-180

[12] Chor B., Friedman J., Goldreich O., et al. The Bit Extraction Problem or t-resilient Functions. In *Proc. of 26th FOCS*, 1985: pp. 396-407

[13] Chor B., Goldreich O. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 1988, 17(2): 230-261

[14] Clément Louis Canonne, Venkatesan Guruswami, Raghu Meka, and Madhu Sudan. Communication with Imperfectly Shared Randomness. *ITCS 2015*: 257-262

[15] Chen R., Mu Y., Yang G., et al. Strongly Leakage-Resilient Authenticated Key Exchange. *CT-RSA 2016*, pp. 19-36

[16] Dodis Y., Katz J., Reyzin L., and Smith A. Robust fuzzy extractors and authenticated key agreement from close secrets. *CRYPTO 2006*, vol. 4117 of LNCS, pp. 232-250

[17] Dodis Y., López-Alt A., Mironov I., and Vadhan S.P. Differential Privacy with Imperfect Randomness. *CRYPTO 2012*, pp. 497-516

[18] Dwork C., McSherry F., Nissim K., and Smith A. Calibrating noise to sensitivity in private data analysis. *TCC 2006*, vol. 3876 of LNCS, pp. 265-284

[19] Dodis Y. New Imperfect Random Source with Applications to Coin-Flipping. *ICALP 2001*, pp. 297-309

[20] Dodis Y., Ong S.J., Prabhakaran M., and Sahai A. On the (im)possibility of cryptography with imperfect randomness. *FOCS 2004*, pp. 196-205

[21] Dodis Y., Ostrovsky R., Reyzin L., Smith A. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 2008, 38(1): 97-139

[22] Deng H., Qin B., Du R.Y., Zhang H.G., et al. Finding Key Leakage in Hierarchical Distribution of Encrypted Data. *INCoS 2013*, pp. 780-785

[23] Dodis Y., Spencer J. On the (non) Universality of the One-Time Pad. *FOCS 2002*, pp. 376-385

[24] Dodis Y. and Yao Y.Q. Privacy and Imperfect Randomness. *CRYPTO (2) 2015*, vol. 9216 of LNCS, pp. 463-482

[25] Dodis Y. and Yu Y. Overcoming Weak Expectations. *TCC 2013*, pp. 1-22

[26] Gennaro R., Krawczyk H., and Rabin T. Secure hashed diffie-hellman over non-ddh groups. *EUROCRYPT 2004* , vol. 3027 of LNCS, pp. 361-381

[27] Ghosh A., Roughgarden T., and Sundararajan M. Universally utility maximizing privacy mechanisms. *STOC 2009*, pp. 351-360

[28] Hardt M. and Talwar K. On the geometry of differential privacy. *STOC 2010*, pp. 705-714

[29] Hu C.Y., Yu Z.X., Yang R.P., Xu Q.L., et al. Weak leakage resilient extractable hash proof system and construction for weak leakage resilient CCA-secure public-key encryption. *IJES 2015*, 7(3/4), pp. 216-229

[30] Ishai Y., Weiss M., and Yang G. Making the Best of a Leaky Situation: Zero-Knowledge PCPs from Leakage- Resilient Circuits. *TCC (A2) 2016*, pp. 3-32

[31] Krawczyk H. Cryptographic Extraction and Key Derivation: The HKDF Scheme. *CRYPTO 2010*, vol. 6223 of LNCS, pp. 631-648

[32] Kamp J. and Zuckerman D. Deterministic extractors for bit-fixing sources and exposure resilient cryptography. In *Proc. of the 44th Symposium on Foundations of Computer Science (FOCS'03)*, pp. 92-101

[33] Liu S.L. Biometric-based key generation. *World science*, no 12, 2009.
(刘胜利. 基于生物特征的密钥提取. *世界科学*, 2009 年第 12 期.)

[34] Lichtenstein D., Linial N., Saks M.E. Some extremal problems arising from discrete control processes. *Combinatorica*, 1989, 9(3), 269-287

[35] Liu S.L., Weng J., Zhao Y.L. Efficient Public Key Cryptosystem Resilient to Key Leakage Chosen Ciphertext Attacks. *CT-RSA 2013*, pp. 84-100

[36] Li Z.Q., Zhang B., Yao Y., Lin D.D. Cube Cryptanalysis of LBlock with Noisy Leakage. *ICISC 2012*, pp. 141-155

[37] James L. McInnes and Benny Pinkas. On the impossibility of private key cryptography with weakly random keys. *CRYPTO 1990*, vol. 537 of LNCS, pp. 421-435

[38] Maurer U.M., Wolf S. Privacy amplification secure against active adversaries. *CRYPTO 1997*, vol. 1294 of LNCS, pp. 307-321

[39] Rényi, A. On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547-561, 1960

[40] Standaert F.X., Pereira O., Yu Y., et al. Leakage Resilient Cryptography in Practice. *Towards Hardware-Intrinsic Security*, pp.

99-134 , 2010

[41] Santha M. and Vazirani U.V. Generating quasi-random sequences from semirandom sources. *J. Comput. Syst. Sci.*, 1986, 33(1): 75-87

[42] von Neumann J. Various techniques used in connection with random digits. *National Bureau of Standards, Applied Math. Series*, 1951, 12: 36-38

[43] Vazirani U.V. and Vazirani V.V. Random polynomial time is equal to slightly random polynomial time. *FOCS 1985*, pp. 417-428

[44] 姚燕青. 基于非完美随机源的密码学原语的安全性研究 [博士学位论文]. 学位授予单位地点: 北京航空航天大学, 2015 年

[45] Yao Y.Q. and Li Z.J. Overcoming Weak Expectations via the Rényi Entropy and the Expanded Computational Entropy. *Information Theoretic Security-7th International Conference (ICITS 2013)*, vol. 8317 of LNCS, pp. 162-178

[46] Zuckerman D. Simulating BPP using a general weak random source. *Algorithmica*, 1996, 16(4/5): 367-391

[47] Zhang H., Zhou Y., and Feng D.G. An Efficient Leakage Characterization Method for Profiled Power Analysis Attacks. *ICISC 2011*, pp. 61-73

姚燕青 于 2015 年在北京航空航天大学计算机学院获得博士学位。现任北京航空航天大学计算机学院讲师。主要研究领域为基于非完美随机源的密码学、弹性泄漏非延展编码等。Email: yaoyanqing1984@buaa.edu.cn

李舟军 于 1999 年在国防科学技术大学计算机学院获得博士学位。现为北京航空航天大学计算机学院教授, 博士生导师, 信息安全系主任。主要研究领域为网络与信息安全技术、数据挖掘与人工智能。Email: lizj@buaa.edu.cn.