计算机科学2000Vol.27№.5

基于平均复杂性的公钥密码研究

Public Key Cryptography Based on Average-Case Complexity

赵一鸣　鲍振东
（复旦大学计算机系　上海 200433）

Abstract In this survey we review the most recent result, from which we can construct a cryptosystem whose security is based on average-case complexity. Using lattice rounding technique, we analyze the hardness of computing the most significant bits of key and the entire secret. And we can construct a public key cryptosystem that is secure unless the problem which found the shortest nonzero vector in a lattice L can be solved in polynomial time. Based on the relation between the worst-case complexity and the average-case one, we introduce the problem finding large cliques in random graphs. Assuming the hardness of finding large cliques in random graphs, we can state that when a clique of sufficiently large size is randomly inserted into a random graph G, yielding graph G', finding any large clique in G' is still hard. The result can be constructed a new one-way function.

Keywords Average-case complexity, Lattices, Public key cryptosystem

1 引言

在近代密码学特别是公钥密码系统的研究中，密码系统的安全性都是基于难解的可计算问题的，如大数分解问题、计算有限域的离散对数问题、平方剩余问题以及椭圆曲线的对数问题等。必须指出的是，关于平均复杂性的研究因远比最坏情况下的复杂性研究要难得多，所以目前密码系统的安全性都是建立在最坏情况下的复杂性基础上的，而不是平均复杂性。从原理上说，一个密码系统的安全性至少应建立在它所基于的数学问题的平均复杂性仍是困难的基础之上，即是否是 NP 难的[1]。因此研究问题的平均复杂性与最坏复杂性的关系已成为密码学领域的一个重要方向。如果能将问题的平均复杂性与最坏复杂性的研究相关联，那么密码学的研究和计算复杂性的研究都将会有一番新天地。近几年，国际上计算复杂性和密码学的研究都将注意力集中到问题的平均复杂性与最坏复杂性的关系上来，引发这一研究高潮的是 Ajtai 对于格问题的研究，他首次明确给出了格最短向量问题的最坏复杂性与平均复杂性之间的关系[2]。在确定的格中，把一个向量归约到最近似的闭向量首先是由 Babai[3]提出的；而 Boneh 和 Venkatesan 则用格的归约技术研究 Differ-Hellman 协议中密钥最有效位的难计算性[4]。Goldreich, Goldwasser 和 Halevi 基于格的归约问题提出了一个公钥密码系统[5]，Ajtai 和 Dwork 则用格的理论研究公钥密码系统最坏情况和平均情况的等价性。

作为最坏复杂性与平均复杂性之间关系的另一个研究内容是在随机图中找最大团问题。当随机图中的最大团可能为$2\log_2 n$大小时，普遍推测对任何$\varepsilon \geqslant 0$，不存在多项式时间算法以不可忽略概率查找大小$\geqslant(1+\varepsilon)\log_2 n$的团。A. Jules 和 M. Peinado 在文[6]中证明了如果上述推测是正确的，则在随机图中随机插入大小为$(1+\varepsilon)\log_2 n$的团时，要找大小$\geqslant(1+\varepsilon)\log_2 n$团仍将是难的。因此，这就可以作为一个单向函数，由此构造公钥密码系统。

本文将介绍格的归约技术及在密码学中的应用，讨论最大团问题以及与公钥密码系统的联系，最后简单总结了基于计算复杂性理论研究公钥密码系统及有待进一步研究的问题。

2 格归约及其在密码中的应用

所谓格 L 是指一个 R^n 的离散可加子群，其中：$L=L(b_1,\cdots,b_n)=\{\sum\lambda_i b_i \mid \lambda\in Z, i=1,\cdots,n\}$，$b_1,\cdots,b_n\in R^n$ 是 R^n 的一组基，也称为格 L 的基。向量 $x=(x_1,\cdots,x_n)\in R^n$ 的长度是指 $\|x\|=(x_1^2+\cdots+x_n^2)$。基$(b_1,\cdots,b_n)$的长度是指 $\max\|b_i\|$。

Ajtai 给出了一个 Z_n 中的格的随机类，该类的元素中有一个长度短的向量[1]，如果能够以至少1/2的概率找到随机格的一个短的向量，那么就有一个多项式

时间的算法能够对于每个 Z_n 中的格以指数趋向1的概率解决以下三个著名的最坏情况的格归约问题：

1)以仅相差多项式因子来近似地在 n-维格中找到向量长度最短的非零向量。

2)在 n-维格中找到唯一的最短非零向量，使得任何一个长度为 $n^c\|v\|$ 的向量与 v 平行(c 是一个任意小的正数)。

3)在 n-维格中找到一组仅相差多项式因子的最短基 $b_1,\cdots,b_n$。

对第一个问题，即在随机格的集合类中找最短向量的难度等价于最坏情况下在 n-维格中计算出最短非零向量的难度，它们之间仅相差一个多项式因子，估计值是 n^8。J. Y. C. 将 Ajtai 所给出结果从 n^{24} 降到 $n^{3.5}$[7]。这使基于这一问题的密码系统的关键性能得到了很大提高。这个结果尽管已经非常漂亮，但还没有能够给出其下界。

Ajtai 和 Dwork 进一步设计了一个最坏复杂性与平均复杂性等价的公钥密码系统[3]：在假设找到 n 维格的最短非零向量是难的情况下(即不能在多项式时间内完成)，所构造的密码系统是安全的，即对于该系统的一个随机实例攻击的难度与该系统的建立所基于的格问题的最坏实例的难度是一样的。下面简单介绍基于格归约理论建立的公钥密码系统。

密钥对的产生

1)随机生成基为 $(b_1,\cdots,b_{n-1})$ 的 n－1维格 L'，使得 $\|b_i\|\leqslant M$。例如可以用文[6]给出的随机类。令 H 是包含 L' 的 n－1维子空间。

2)选择 $d\geqslant n^cM$。

3)在大立方体上随机选择向量 b_n，使得与 H 的距离满足 $d\leqslant d_L\leqslant 2d$。

4)秘密密钥为 $L'^{(d,m)}$ 的任何基。

5)对 $L=L(b_1,\cdots,b_n)$ 构造随机基 B'，(B',M)即作为公钥。

加解密过程

对于 $R\in\mathbb{R}^n$ 和 $m\in\mathbb{Z}$，令扰乱因子(R,m)为值是向量 m 和的随机变量，这里向量 m 是独立的，并且在以0为原点 R 为半径的 n 维球体中均匀分布。

为了加密0，首先在立方体 KU^n 中选择随机格点 v，这里 U^n 为 n 维单元立方体，$K\geqslant 2^nd$。对于 $m=c_0n$，$c_0\geqslant 4$，$R=n^3M$，选择(R,m)的值 w，密文为 v＋w。为了加密1，在 KU^n 中均匀随机选择一点作为密文。

令 v_H 为子空间 H 的单位正交向量，d_L 为两个超平面的距离。为了解密密文 z，接收者计算 $(v_H\cdot z)/d_L$ 的小数部分，如果是在 mR/d_L 之内，则 z 解密为0，否则解密为1。

现在考察该密码系统的安全性。应证明：若规定(d,M)格上的分布 $\mathscr{L}$ 满足对每个(d,M)格 $L\in\mathscr{L}$ 有：L 能够用长度超过 d_L 的基表示(这里 d_L 最多是 n 次多项式的因子)，那么由密文解出0或1的能力等价于解隐蔽超平面问题，若这个结论成立，则攻破密码系统的唯一方法是找到秘密密钥。

由文[1]，假定对于 L 给定了基 Y(例如格点接近于指数式均匀分布在边长至少为 $n^c\|Y\|$ 的立方体中)。显然可以在 n^cY 上定义指数式分布，这是因为平行六面体格的点构成了有限 Abelian 群[2]，然后从 $S^n(R)$ 选择均匀向量 t，则根据文[2]的结论、容易得到上述结论。

进一步，可以构造一个新的公钥密码系统。

假设 $u\in\mathbb{R}^n$，$0<\|u\|\leqslant 1$，$R>0$，并且 m 为非负整数，Ω 为 n 维立方体 $KU^{(n)}$。按如下方式定义随机变量 $H'(u,R,m)$：$X=\{x\mid x\cdot u$ 为整数，$x\in\Omega\}$，在 X 上取随机点 y，和(R,m)的值 z，$H'(u,R,m)$ 的值就是 y＋z。设 $H=\mathrm{round}_2\circ\alpha(H')$，对于 $y\in I$ 和 $\alpha>0$，$\mathrm{round}_\alpha(y)=i\alpha$(i 是满足 $i\alpha\leqslant y$ 的最大整数)；对于 $x=(x_1,\cdots,x_n)\in\mathbb{R}^n$，$\mathrm{round}_\alpha(x)=(\mathrm{round}_\alpha(x_1),\cdots,\mathrm{round}_\alpha(x_n))$。

秘密密钥是一个随机向量 u，它在集合 $\{x\in\mathbb{R}^n\mid\|x\|\leqslant 1\}$ 上是均匀分布的。公钥则是一组随机变量 $H_{u,n^{-D_1},n}$，有 $m=n^{D_3}$ 个独立值 $v_1,\cdots,v_m$，v_i 是由 u 导出的超平面中最小的点扰乱。加解密方法可用文[4]中的算法。对于这样的系统，有结论：

结论1 对于所有的 $c_1,c_2,c_3,c_4>0$，存在 c_5 和概率算法 B，使得对任意足够大的 n，条件(1)蕴涵条件(2)。

(1)u 和 $v_1,\cdots,v_m$ 是按前面方式产生的秘密密钥和公钥，A 具有概率至少是 n^{-c_2} 的规模为 n^{c_3} 的概率循环；而对于给定的 $v_1,\cdots,v_m$，A 区分随机变量 $S_{v_1,\cdots,v_m}$ 和 $E_{v_1,\cdots,v_m}$ 的概率至少是 $\frac{1}{2}+n^{-c_4}$。

(2)B 可以在 n^{c_5} 时间内用概率至少是 $1-2^{-n}$ 解规模最多是 n^{c_4} 的 n^{D_2} 的实例。

由此结论，可以知道，公钥密码系统是安全的，除非能够在多项式时间内从 n 维格 L 中找到最短的非零向量。

格归约理论在密码学中的另一个应用是研究密钥有效位问题。Okamoto 协议与 Differ-Hellman 协议一样，能使 A 和 B 交换秘密。A 随机选择自己的密钥 a，并计算 $x=g^a\bmod p$ 送给 B，然后 B 随机选择自己的密钥 b，并计算 $y=x^b\bmod p$ 并送回给 A。A 只要计算 $y^{a^{-1}}=g^b\bmod p$ 作为双方的密钥(这里 p 为公开的大素数)，第三者想通过 g^a 和 x^b 求出 g^b 在计算上是难的。但在实际使用时，双方是用 g^{ab} 的部分位作为 DES 或 IDEA 的密钥。这就存在一个问题：计算 g^b 是困难的，

但由 g^a 和 x^b 求出 g^b 的部分位是否也是困难的?这就是所谓的密钥最有效位问题。用格的归约方法可以证明 Okamoto 协议的密钥最有效位是2loglog$p$。

结论2 设 p 是素数，g 是 Z_p^* 上的生成元，$k=\lceil 2\log\log p\rceil$。对于隐蔽数 a 定义函数 $f(r)=MSB_k(ag^r \bmod p)$。则对于给定的 ORACLE，存在一个多项式时间算法，求出多项式个仅依赖于 p，g 的信息位，并且只要给定函数 f(r)就可恢复隐蔽信息 a。

该结论证明主要是利用格归约解决隐蔽数问题。首先考察均匀和独立选择输入 $r_1,\cdots,r_A$ 下的 f(r)值，其中 d 的值将在后面说明。作为约定，假设 f(r)的输出是自动转化成数 a，并且满足：

$$|(ag^r \bmod p)-a|<p/2^{k+1}$$

这通过变换 $f(r)\to f(r)\frac{p}{2^k}+\frac{p}{2^{k+1}}$ 即可做到。

设 $t_i=g^{r_i} \bmod p$，ORACLE 的回答将是转换成整数的 $a_1,\cdots,a_d,a_{d+1}$，它们满足：

$$|(at_i \bmod p)-a_i|<R(i=1,\cdots,n)$$

$$|(a \bmod p)-a_{d+1}|<R$$

这里 $R=p/2^{k+1}$，而 a_{d+1}的值则是在 r=0时 f(r)的自动转换数。现在的目标是根据这些信息恢复 a。现在考虑格 L 用矩阵的行来估量：

$$L=\begin{pmatrix} p & 0 & \cdots & 0 & 0\\ 0 & p & \cdots & 0 & 0\\ & \vdots & & & \vdots\\ 0 & 0 & \cdots & p & 0\\ t_0 & t_1 & \cdots & t_d & 1 \end{pmatrix}$$

设 $u=(a_1,\cdots,a_d,a_{d+1})$，并定义格向量 $v_a\in L$，

$v_a=(t_1a \bmod p,t_2a \bmod p,\cdots,t_da \bmod p,a)$

根据定义有 $\|u-v_a\|\leqslant R\sqrt{d+1}$，即向量 u 关于格点 v_a 是闭的，容易证明由 v_a 可以复原 a。而根据给定向量 u 求 v_a，由文[8]的归约来设计算法。这样该结论即证得。

Okamoto 协议中密钥共享的产生用函数 $Ok_g(g^{ab},g^a)=g^b$，显然有 $Ok_g(g^{r(x+s)},g^r)=g^xg^s$，因此由第2节的结论2即可得到 Okamoto 协议的密钥最有效位是 2loglog$p$。

3 团问题与密码研究

格归约理论的研究成果使得密码系统安全性有望建立在平均复杂性基础之上。由于目前密码系统所依赖的难解问题多数限于数论中因子分解和离散对数问题，人们一直希望把思路放宽，基于格归约是一个方向，而利用图论中的难解问题构造单向函数也是密码学者所期盼的。把团问题的难解性应用于密码则是这种期盼的尝试。

设 G(V,E)为图，其大小为 k 的团是指图中形成一个完全图的 k 个顶点，用 $\omega(G)$表示 G 的最大团的大小。

作为密码系统希望基于平均情况下的难问题，因此平均情况下的团问题就引起了人们的广泛兴趣。大多数的研究主要关注有 n 个顶点的 Erd os-Renyi 随机图模型 $G_{n,p}(0\leqslant p\leqslant 1)$的实例。对于图 G 以概率 p 独立插入 C(n,2)条边后即可导出这样的分布。通常考虑的情况是 p=1/2，即均匀分布。绝大多数这类图的最大团大小为$2\log_2 n-O(\log\log n)$。存在大量的比最大团小的团，对于 $k=c\log_2 n$(c 为常量且$0<c<2$)的团，估计有 $n^{\Omega(\log n)}$个。对于大小为 $\log_2 n$ 的团可以用随机贪心算法在多项式时间内找到。而对于稍大一些的团却没有找到多项式时间算法。现在猜测对任意 $\epsilon>0$，不可能存在多项式时间算法在随机图中以不可区分的概率找大小为$(1+\epsilon)\log_2 n$ 的团。这就给我们应用在密码上带来了希望。

团问题应用于密码最初是由 Ku $\check{c}$era 提出的[5]，他定义了一种广义加密模式，用团问题(更确切地说是独立集问题)来实现。所用图是模型 $G_{n,p}$类的随机图，它们嵌入了大小为 $k=n^{\kappa}(0<\kappa<1/2)$的团，但 Ku $\check{c}$era 没有严格证明其方案的安全性。

事实上，团问题应用于密码，其实质就是把一个大团嵌入到随机图中，而这个大团就是秘密，因此必须保证其他人要在这样构成的图中找到大团是难的。由于普遍猜测在随机图中找大团(大小为$(1+\epsilon)\log_2 n,\epsilon>0$)是难的，因此密码的安全性归结为数学问题就是：必须证明在一个嵌入大团的随机图中找大团与在一个随机图中找大团是一样难的。

对于给定的图 G，用 $C_k(G)$表示 G 中大小为 k 的不同(但可能重叠)团的数目。因此当 G 是根据分布随机生成时，$C_k=C_k(G)$就是一个随机变量。G 中大小为 k 的团的预期数量称为 G 的期望数，用 $E_k=EC_k(G)$表示。因此要解决前面的问题，首先应证明：当图 G 中团的数量接近于期望数 E_k 时，在随机图中插入大小为$(2-\delta)\log_2 n$ 的团的进程将产生 G，且概率类似于简单生成随机图进程的概率。换句话说，应证明当 $C_k(G)$接近 E_k 时，分布 p'_k 里的图 G 的概率也将接近分布 p 里的图的概率。这由文[6]中的引理3.1结论即可得。然后应证明的是 C_k 的变化是低的。这隐含两个内容：首先大多数的图是好的，即 $p'_k(G)/p(G)$比相应的小多项式小；而 $p_k'(G)/p(G)$较大的图即为坏图，它将占 p_k'的小因子 Δ。我们可以使 Δ 任意小。因此一个能在分布 p'_k 的图 G 的 $\frac{1}{poly}$因子成功找出一个大团的算法必须能在一个好图的集合 M 找出这样的大团，其中 M 应

满足 $p'_k(M)=\frac{1}{poly}-\Delta=\frac{1}{poly}$。由于 M 中的图都是好的，故 $p'_k(M)=\frac{1}{poly}$就意味着 $p(M)=\frac{1}{poly}$，所以算法 A 能在分布 p(图上的均匀分布)的图 G 的$\frac{1}{poly}$因子成功找出大团。由此可得下面的结论：

结论3 假设对 $\delta>0$，$k\leqslant(2-\delta)\log_2 n$，存在一个确定性多项式时间算法 A，在由分布 p'_k 导出的图中找 k-团，且概率为$\frac{1}{q(n)}$(q(n)为多项式)，则必存在多项式 q'(n)，使得算法 A 在由分布 p 导出的图中找 k-团，且概率为$\frac{1}{q'(n)}$。

在随机图中找大团是难的这一假设下，上述结论表明当团 K 的大小足够大时，例如把$\frac{3}{2}\log_2 n$ 大小的团随机插入到随机图 G 产生 G'，在 G' 中找任何大团仍是难的。由此我们就可构造单向函数 $f: f(G\swarrow K)\to G$，这里 G 是 n 个顶点的图结合，而 K 则是所有 k 个顶点的集合(即{1,2,…,n})的所有子集全体)。f(G,K)是对图 G 的改动，使得所插入的子图完全由 K 导出，显然由 K 是容易得到新图 $G'=f(G,K)$，而由 f(G,K)要高概率地求出 K，则在随机图中找大团是难的假设下，将是难的，因此 f 符合单向函数的准则。当然由 G 和 G' 可容易得到 K，利用这一特性我们就可构造一个具有分级权限的秘密密钥管理模式。其依据就是在随机图中随机插入任意常量个团 $K_1,K_2,\cdots,K_n$ 生成 G'，把 $\{K_i\}$作为秘密密钥，而 G'作为公开密钥。P_1 生成随机图 G 并在 G 上随机插入团 K_1，产生图 G_1送给 P_2，P_2在 G_1上随机插入团 K_2产生图 G_2送给 P_3，继续这个过程直到 P_n，P_n 在 G_{n-1}上随机插入团 K_n 产生图 G_n 作为公开密钥。由于 P_1知道 G，故 P_1可以得到 $P_2,P_3,\cdots,P_n$ 的秘密密钥(虽然 P_1不知道这些密钥是哪个成员)，而其他成员则不可能得到 P_1的秘密密钥。一般地，P_i 可以求出所有 $P_j(j>i)$的秘密密钥。由此即得分级密钥管理模式。

结语 用格的归约理论可以分析研究计算部分密钥位与整个密钥位的计算复杂性，研究公钥密码系统的安全性，而团问题应用于密码系统更是为计算密码研究拓展了方向。但目前还存在这样一些问题：怎样建立一个实用的基于格归约密码系统，是否能利用团问题直接建立一个公钥密码系统。所有这些都有待于我们不断研究和探讨。

参 考 文 献

1 Ajtai M. Generating Hard Instances of Lattice Problems. STOC, 1996 99～108

2 Ajtai M, Dwork C. A public-key cryptosystems with worst-case/average-case equivalence STOC, 1997. 284～293

3 Babai L. On Lovasz' lattice reduction and the nearest lattice point problem. Combinatorica, 1986, 6: 1～3

4 Boneh D, Venkatesan R. Hardness of computing the most significant bits of secret keys in Diffe-Hellman and related schemes. Proc. of Crypto, 1996

5 Goldreich O, et al. Public key cryptosystems from lattice reduction problems. Electronic Colloquium on Computational Complexity Tr96～056

6 Jules A, Peinado M. Hiding Cliques for Cryptographic Security. STOC, 1998. 678～684

7 J. Y. C An Improved Worst-Case to Average-Case Connection for Lattice Problems. Proc, 38th FOCS, 1997, 468～477

8 Boneh D, Venkatesan R. Rounding in lattices and its cryptographic applications. SODA, 1997. 675～681

9 L. Ku čera. A generalized encryption scheme based on random graphs. In Graph-Theoretic Concepts in Computer Science, WG'91, Lecture Notes in Computer Science 570, Springer-Verlag, 1991. 180～186

(上接第67页)

自然语言是非常复杂的，其处理也是十分困难的。虽然在理论上，HNC 已经显示出了强大的解模糊能力，但是否有真正的实用价值，需要做大量的验证工作。值得庆幸的是，由国家语委和"95"重大攻关课题所资助的汉语音-字流转换系统，经过测试，与 HNC 理论上的预期基本吻合。这使得我们更加满怀信心去实现"973"资助的"英-汉"互译系统的研究。

本文所给的许多思想已经在"语音-字流转换"系统中实现。因此，本文的形成也是 HNC 联合攻关组共同努力的结果。

参 考 文 献

1 黄曾阳. HNC(概念层次网络)理论——计算机理解语言研究的新路. 清华大学出版社, 1998

2 鲁川. 自然语言处理的语言模型. 见：1998年中文信息处理国际会议论文集. 清华大学出版社, 1998

3 李临定. 现代汉语句型. 商务印书馆, 1986

4 罗振声, 郑碧霞. 汉语句型自动分析和分布统计算法与策略的研究. 中文信息学报, 1994(2)

5 罗振声, 孙长健, 孙才. 汉语句型成分自动分析中谓语识别策略的研究. 计算语言学进展与应用, 清华大学出版社, 1995

6 刘志文. 自然语言语句的 HNC 表示. 语言文字应用, 1998(2)

7 陶明扬. 特征语义块的构成与复合. 中国人民大学96级硕士论文, 1999

8 张艳红. HNC 理论的主辅语义块及其变换. 北京语言文化大学96级硕士论文, 1999