第九届（2024）全国高校密码数学挑战赛

赛题一

一、赛题名称：RSA 密码系统的特定密钥泄露攻击

二、赛题描述

作为当前应用最为广泛的公钥密码体制之一, RSA 系统的密码分析颇受关注. 设定 RSA 密码的公开密钥为加密指数e及模数N, 其中$N = p \times q$是两个大素数的乘积, 相关参数定义详见 2.2 节**赛题所使用的 RSA 密码系统描述**部分. 从数学的角度讲,该密码体制破译相当于计算 RSA 函数$f(x) \equiv x^e \bmod N$在$\mathbb{Z}_N^*$的逆问题（设定$f(x)$的定义域和值域均为$\mathbb{Z}_N^* = \{a \in \{1,2,\dots,N-1\} | \mathrm{GCD}(a,N) = 1\}$, 可以证明该函数为置换）, 即已知$y \equiv x^e \bmod N$的取值, 在乘法群$\mathbb{Z}_N^*$中求解整数$y$的$e$次根这一数论问题. 目前模数$N$规模为 1024 比特的 RSA 密码系统一般情况下认为是安全的, 但是如果参数选取不当, 或者特定私钥信息发生泄漏, 同样存在被破译的可能.

本赛题中用户使用 RSA 加密软件发送多组明文字符串消息m, 假定所有加密数据c都已经被截获, 并且公钥证书中e和N的值均已知, 此外还额外已知用户私钥d的部分信息. 要求选手尽可能多的破解明文消息, 并进一步恢复该加密软件的密钥参数信息.

2.1 符号说明

（1）本赛题中的整数a若无特殊说明均为十进制表示. 以$a = 65537$为例, 其对应的二进制表示为$0b10000000000000001$, 相应的十六进制表示为$0x10001$;

（2）符号$\| a \|$表示正整数a表示成二进制形式时的比特长度. 设$\| a \| = n$,即$a = \sum_{i=0}^{n-1} a_i 2^i$,注意此时最高比特位(Most Significant Bit, MSB):$a_{n-1} = 1$并且满足$\log_2 a < \| a \| \leq \log_2 a + 1$;

（3）符号$\mathrm{GCD}(a,b)$表示整数a和b的最大公因子. 若a和b互素, 则$\mathrm{GCD}(a,b) = 1$.

2.2 赛题所使用的 RSA 密码系统描述

（1）密钥生成算法

- Step1:用户使用随机数发生器(Random Number Generator, RNG)选取合适规模的素数p和q, 计算模数$N = p \times q$;
- Step2:令$\varphi(N) = (p-1) \times (q-1)$,用户选取合适的加密指数$e$满足$\mathrm{GCD}(e, \varphi(N)) = 1$,并计算其逆元$d$, 即$ed \equiv 1 \mathrm{mod} \varphi(N)$;
- Step3:用户公布其公钥为(e, N), 秘密保存其私钥d.

（2）加密算法

假定 Bob 想要发送某保密的明文字符串消息M给用户 Alice，Bob 首先将消息M编码(具体编码规则 2.3 节**加解密过程示范**详细介绍)为不超过模数N的整数m, 进而通过公开渠道查找到 Alice 的公钥(e, N), 之后计算密文$c \equiv m^e \mathrm{mod} N$, 并将$c$的值发送用户 Alice.

（3）解密算法

接收到密文c后, Alice 使用其私钥d计算$m \equiv c^d \mathrm{mod} N$, 并利用编码规则将整数$m$解码为字符串消息$M$.

2.3 加解密过程示范

为了更好理解加密算法, 提供如下具体实例供参赛选手理解.

（1）编码规则说明

假定 Bob 要发送给 Alice 的明文消息M是字符串"HelloWorld2024", 通过查表可知该 14 个字符对应的 ASCII 码依次为：0x48,0x65, 0x6c,0x6c,0x6f,0x57,0x6f,0x72,0x6c,0x64,0x32,0x30,0x32,0x34,将其设定为整数m：

$$m = 0x34323032646c726f576f6c6c6548$$

（2）加密计算密文说明

假定 Alice 的公钥(e, N)取值如下：

e=0x10001;

N=0x781e760887ad042c97ff8991da8a46e1fea82c0ab1800f8a3a3432f742ef768803d6e3d4b58ef5b8efcf26df95c57ffece3750f5614364a16128882c7ab2aad7904a4c207c1747939fbc507a3415598d08e025f687f2ebbc548686cf9bf32a912fa194ebbab5af41a4209132a87f226318e6b9eb9be64b374f57f146118ed185.

Bob 计算密文$c \equiv m^e \mathrm{mod} N$的值如下:

c=0x3966b3b12045320b01cd076bc4d16c7866fa79946da45166a68ad31b91895437192da2f250d7712b22a0a37622892a4d46a94e42d78335635ef55b23c109e305a2ca99210121318228179eef02d9fc09df3dbb7d1db77552b5d83f0c0365d2bcece711a1f7cbe4958cfe21ea90950aa24443e83b6bc329c

766ea37b5ac14ce78.

（3）解密恢复明文说明

Alice 使用私钥d计算$m \equiv c^d \bmod N$可恢复整数m:

$$m = 0x34323032646c726f576f6c6c6548.$$

进一步利用上述编码规则可将整数m解码为明文消息M：
"HelloWorld2024".

需要说明的是, 此处 Alice 的模数N是两个随机选取的 512 比特素数的乘积, 具体取值如下:

p=0x9dc9138c9acf35c15f3ea8298c35ab0f8d5706b5c6b9f7086944e2d08249459ad0084c41fce59c4a69edc02d63982a4aa21f6bf2a7f410c7a42a48cd0ad02847;

q=0xc2e3474251ce262bb613411fe3aa69816efc6aa3859c1e9eefb5c79078aa8109817f5aabc8f8b2425559623a066c2cc06edcb29c981e2b892ac2d3e7b183e9d3.

利用模数N的分解或私钥d均可恢复出密文所对应的明文信息, 参赛选手可自行验证.

2.4 补充说明

对于本次赛题的 RSA 密码体制, 有以下事项需要说明:

(1)本次挑战赛共 12 组数据, 题目$i(1 \le i \le 12)$除了已知公钥(e_i, N_i)以及加密明文字符串M_i产生的密文c_i外, 每组数据均额外已知私钥d_i的部分信息.

(2)每组数据的明文消息M_i的字符数不超过 128 个.

(3)每组数据的公钥模数N_i均是由两个长为 512 比特的素数p_i与q_i乘积得到, 其中素数p_i是由某个特定的随机数发生器 RNG 产生,而 q_i可看作是随机产生的 512 比特长的素数.

(4)生成素数p_i的随机数发生器 RNG 进一步描述如下：输入一个长度较短的种子密钥s_i（对应的整数不超过 2048），该随机数发生器经一系列简单变换后扩展为具有一定周期结构并且长度为 512 的比特串a_i. 将比特串a_i视作二进制串，相应的整数记作b_i.进一步，将其转化为整数c_i. 注意，由二进制串a_i转化为整数c_i的过程中，较少部分位置的比特可能会发生翻转，即由 0 翻转为 1，或将 1 翻转为 0. 最后, 在整数c_i附近选取素数即为 512 比特长的整数p_i.

2.5 成绩评判标准

本竞赛成绩分为两大部分, 总分共计 **500** 分:

- 第一部分（**400** 分）：根据提供的公钥(e_i, N_i)、密文c_i及私钥d_i的部分信息(详见附件), 要求恢复出密文c_i所对应的**有意义的明文消息**M_i. 每道题目按照难度区分为 25 分、30 分、40 分,共计 400 分.
- 第二部分（**100** 分）：根据公钥模数N_i的分解信息, 给出素数p_i的生成方法及相应的种子密钥s_i, 描述该随机数发生器存在的规律, 并用程序代码验证结果的合理性.

上述两部分要求选手给出正确计算结果的同时, 简述求解原理、算法步骤和实现效率. 若程序未能给出正确的计算结果,如果求解原理正确并能给出合理的计算估计（所需要的时间和空间等）, 也可酌情给分. 如果使用他人理论方法或程序代码必须在报告中给出明确引用, 否则该部分报告内容作废. 此外, 如果求解算法中有一定理论创新并正确阐述的, 也可酌情加分, 但总分不超过 **500** 分.

三、密码学背景及相关问题的研究进展

整体而言, 对 RSA 密码系统的分析工作可分为两类：基于数学的RSA攻击和基于实现的RSA攻击. 基于数学的RSA攻击可看作是使用计算数论的方法通过分解模数N来实现 RSA 攻击. 在 2020 年美密会上, 来自法国和美国的研究团队基于数域筛法(Number Field Sieve)成功将一个 829 比特长的模数 RSA-250 分解, 这是公开领域分解的最大规模的 RSA 模数.

现实中, 攻击者通过能量分析、时间攻击等非数学手段的侧信道技术可获取私钥d的部分比特位, 因此基于实现的RSA攻击可看作是在弱化的模型下对RSA问题的求解. 1998 年亚密会上, Boneh, Durfee 等[1]首次提出了部分私钥泄露攻击, 这类攻击关注在泄露私钥d特定比例的信息后, 能否在多项式时间内破解 RSA 密码体制, 其中基于格基约化算法的 Coppersmith 方法在 RSA 密码的部分私钥泄露攻击中发挥了重要作用.

低解密指数攻击可看作部分私钥泄露攻击的一种特殊情形. 1990 年, Wiener 利用初等数论连分数的方法证明, 如果私钥$d < N^{0.25}$, 利用公开的模数N及加密指数e可多项式时间计算出私钥d. 这意味着若占比 75%的私钥d高比特位已知且其全为 0 的情形下可多项式时间

完整恢复私钥d. 这个结果在 1998 年被 Boneh 和 Durfee 基于 Coppersmith 方法改进为$d < N^{1-\sqrt{2}/2} \approx N^{0.292}$, 这也是至今为止低解密指数攻击最好的理论上界. Boneh 等密码学家猜测, 在 RSA 密码体制中若私钥$d < N^{0.5}$极有可能是不安全的, 但该公开问题长久以来仍然未能解决.

针对一般情形的部分私钥泄露攻击, 该类密码问题解决的关键在于如何把 RSA 私钥d泄露的比特信息转化为整系数模多项式方程求小根的数学问题, 进而可利用 LLL 算法基于格中的近似最短向量问题求解. 目前国际密码学界非常重视 Coppersmith 方法的应用, 并将其由单变元多项式方程的小值解推广至双变元及多变元情形,相关方法的巧妙运用及示例可参考文献[2]和[3].

四、参考文献

[1]Boneh D., Durfee G., Frankel Y.. An attack on RSA given a small fraction of the private key bits[C]. In: Advances in Cryptology—ASIACRYPT'98. Springer Berlin Heidelberg, 1998: 25–34.

[2]Micheli G. and Heninger N.. Recovering cryptographic keys from partial information, by example. IACR Cryptology ePrint Archive, 2020:1506, 2020.

[3]王世雄, 屈龙江, 李超, 付绍静. 私钥低比特特定泄露下的 RSA 密码分析[J]. 密码学报，2015,2(5):390-403.