All About Crypto - CTF 竞赛密码学方向指南

作者：道路结冰（roadicing）

日期：2019-12-05

0 前言

密码学不仅是 CTF 竞赛中的一个重要的独立考察方向，也经常作为考点出现在其他方向的题目当中。本文从个人作为一名 CTF 密码学方向选手的视角入手，对需要掌握的相关内容进行整理与分析，以期帮助选手更好的学习密码学，更好的进行密码学方向题目的训练，取得更好的比赛成绩。

1 扎实的数学功底

密码分析的各个环节都离不开数学分析，对于选手求解一道 CTF 中的密码学方向题目来讲也是如此。在 CTF 中，主要考察密码学选手数论和抽象代数两个方向的数学知识。

1.1 数论

在数论方向，选手应能较为熟练的了解和掌握包括：

- 整除理论（了解素数、合数、因数、倍数、整除等基本概念，掌握唯一分解定理、裴蜀定理、扩展欧几里得定理、算数基本定理等基本定理）
- 同余理论（了解同余、原根、底数、指数、平方剩余、同余式、同余方程等基本概念，掌握欧拉定理、费马小定理、中国剩余定理、二次互反律、威尔逊定理等基本定理）
- 连分数理论（了解连分数、无穷级数等基本概念，熟悉最佳有理数逼近、循环连分数展开、佩尔方程求解等运算过程）
- 不定方程（了解低次代数曲线所对应的不定方程的基本模型，熟悉二元一次不定方程、多元一次不定方程、掌握通过代数恒等变化、不等式估算、同余法、构造法、无穷递降法等常用的不定方程求解方法）
- 数论函数（了解欧拉函数、莫比乌斯函数、单位函数、恒等函数、除数函数等常用函数，掌握莫比乌斯反演、狄利克雷卷积等常用方法）

1.2 抽象代数

在抽象代数方向，选手应能较为熟练的了解和掌握包括：

- 群论（熟悉群代数结构，掌握群相关性质及其运算律）
- 环论（熟悉群代数结构，掌握环相关性质及其运算律）
- 域论（熟悉群代数结构，掌握域相关性质及其运算律）
- 格论（熟悉格代数结构，掌握格相关性质及其运算律）
- 线性代数（熟悉向量空间代数结构，掌握向量相关性质及其运算律）

除此之外，诸如逻辑学、几何学、拓扑学、泛函分析、概率论、数理统计等其他数学分支，虽然在 CTF 竞赛中密码学方向题目直接考察较少，但选手应至少对其相关基本知识有一个大致了解。

2 密码学技能树

2.1 古典密码学

古典密码学作为早期 CTF 竞赛中密码学方向的一种常见考察形式，目前已经逐渐退出国际赛的历史舞台了。

CTF 中的古典密码主要以代替（substitution）密码和置换（permutatuion）密码两种形式出现，在题目当中，出题人通常不会显式的告诉你题目所采用的加密算法，而是仅仅给出密文，预期选手通过特征检索（如密文字符集中存在标志性的特殊字符）、题目暗示（如题目名称、题目描述中出现了对加密算法的隐喻）等方式猜测出题目中可能使用的加密算法，或使用数理统计（针对密钥空间较大的代换密码，如仿射、维吉尼亚等）、爆破（针对密钥空间较小的代换或置换密码，如栅栏、移位等）等方式恢复出密钥，最后解密密文拿到 FLAG。

这类题目的难点通常不在于分析而在于猜测，因此往往难度较低，但是由于代替和置换是密码学算法中的两个最基本的操作，很多现代密码学算法中的运算都可以看作是这两种运算的复合运算，因此古典密码学题目也可作为初接触 CTF 竞赛密码学方向的选手的练习题目，有助于培养选手对基本运算操作的理解。

2.2 现代密码学

现代密码学作为目前 CTF 竞赛中密码学方向的主要考察形式，从总体上可以分为对称密码学、非对称密码学、哈希函数和数字签名四大类题目，其中每类题目在知识点层面虽互有交集，但由于考察形式各有侧重，因此本文对这四类题目类型分别进行论述。

2.2.1 对称密码学

2.2.1.1 序列密码（流密码）

流密码通常以字节或比特为基本单位来处理信息，其密钥通常派生自一个较短的种子密钥，而派生算法一般为一个伪随机数生成算法，流密码的安全性取决于密钥流的安全性，因此 CTF 中的流密码类题目也主要以伪随机数生成器部分为主，当然除此之外，题目有时也会考察选手对某一具体的流密码算法的理解和分析能力，如 A5/1、RC4 等。

对于伪随机数生成器来讲，常见的考察模型主要可以分为两类：

一类为线性同余生成器（LCG），题目要求选手通过对生成器源码审计，找出设计缺陷（针对生成器参数随机化的场景）或进行数学推导恢复未知参数（针对参数恒定不变但缺失部分参数的场景），继而连续预测出接下来产生的若干个随机数，从而达到服务器要求拿到 FLAG。

另一类为反馈移位寄存器，其中又可分为线性反馈移位寄存器（LFSR）和非线性反馈移位寄存器（NFSR）两类主要考察模型，出题人通常会根据某一初始状态采用某种生成方法生成一份输出结果，然后将生成方法和输出结果提供给选手，预期选手还原出初始状态从而作为 FLAG。

2.2.1.2 分组密码（块密码）

块密码使用某一基本块为基本单位来处理信息，在加密时需要将明文数据分成若干基本块，再针对每一块进行加密，如果最后一块的长度小于基本块的长度，还需要进行 padding。

目前 CTF 中针对块密码主要从三个角度考察：

第一类是从块密码当中的 ARX（A-有限域上的模加，R-循环移位，X-异或）三种基本操作入手，考察选手对组合运算的熟练程度和理解能力。

第二类是从具体算法角度入手，考察 AES、DES 等经典加密算法（或该加密算法的自定义修改版本）的线性攻击、差分攻击、积分攻击等攻击手法和选手做密码分析的能力。

第三类是从分组模式入手，考察算法在 padding 时（如针对 PKCS5 Padding 的 Padding Oracle 攻击）或加密模式上（如 CBC 字节翻转攻击、CFB 重放攻击）可能会出现的问题。

2.2.2 非对称密码学

CTF 竞赛中的非对称密码学主要考察三大类问题，即大整数分解问题、离散对数求解问题（包括椭圆曲线上的离散对数求解问题）和基于格（Lattice）的计算问题：

2.2.2.1 大整数分解问题

CTF 中大整数分解问题主要以 RSA 算法为模型进行考察，RSA 在目前 CTF 竞赛中考察频率可以说所有考点第一位，几乎任何一场比赛都会有一道密码学题目考察 RSA，根据题目中给定的 RSA 算法的不同场景，其攻击手法五花八门，需要选手具备很强的数论功底。

但也正是由于 RSA 知识点考察过于热门，导致网上相关资料和现成的攻击脚本较为成熟，对于一些简单的 RSA 类题目，选手往往只需判断出该题涉及到的 RSA 的哪一类场景，然后根据特征去检索攻击手法即可，但对于一些国际赛上的高质量 RSA 类题目，仍然需要选手具备极强的分析和推导能力。目前针对 RSA 的常见攻击大致包括以下几类：

- 针对模数的攻击：如 Pollard's p -1 算法（p-1 光滑）、Williams's p + 1（p+1 光滑）、试除法（|p-q|过大）、费马分解（|p-q|过小）、共模攻击、模不互素攻击等。
- 针对公钥的攻击：如小公钥指数攻击、低加密指数广播攻击等。
- 针对私钥的攻击：如 Wiener's attack（基于连分数）、私钥低位泄露攻击等。
- Coppersmith & LLL 相关攻击：如已知 m 高位攻击、已知 p 高位攻击、Coppersmith's short-pad attack、Boneh and Durfee attack 等。
- 结合 Oracle 的攻击：有时题目当中除了提供给选手参数之外还会提供一个加密/解密 Oracle，结合 Oracle 可以衍生出更多种攻击形式，如针对加密 Oracle 的公钥泄露攻击（选择明文），针对解密 Oracle 的 RSA LSB parity Oracle 攻击（选择密文）。

2.2.2.2 离散对数求解问题

CTF 中考察 DLP 类问题主要以 Diffie–Hellman 密钥交换协议和 ElGamal 算法为主，要求选手能够通过审计代码找出问题关键点，并使用攻击算法求解 DLP 问题，常用的 DLP 攻击算法包括：

- 小步大步法（Baby-step giant-step，中间相遇攻击的思想）
- Pollard's Rho algorithm（基于 Miller-Rabin 算法，递归求解）
- Pollard's Kangaroo algorithm（基于随机步）
- Pohlig-Hellman algorithm（针对阶 n 是光滑且仅有小素因子）

CTF 中考察 ECDLP 类问题主要以椭圆曲线加密（ECC）为主，其曲线有限域通常为以素数为模的域 GF(p)或特征为 2 的域 GF（$\boldsymbol{2^m}$），ECDLP 类题目的考察方式除了包括上面提到的 DLP 的一些常见模型和攻击手法的椭圆曲线版以外，也包括一些针对曲线上存在的问题的攻击形式，如：

- 针对 $E(F_p)=p$（Frobenius 轨迹 $t=p+1-E(Fp)=1$）的F_p上非正规椭圆曲线的 Smart's attack。
- 针对 $p|q+1-E(F_q)$（Frobenius 轨迹 t 是 p 的倍数）的超奇异椭圆曲线的 MOV 攻击（借助 Tate pairing 或者 Weil pairing）。

2.2.2.3 基于格的计算问题

CTF 中考察格中的计算困难问题主要包括最近向量问题（CVP）和最短向量问题（SVP）问题，其考察形式主要分为两类：

一类是利用格理论去分析已知的经典密码算法，如使用格基规约算法（LLL）来分析 DSA 签名系统（如 DSA nonce biases）、RSA 加密系统（如 RSA 的小私钥攻击、Coppersmith 相关的攻击）、背包加密系统（如 Merkle–Hellman 背包公钥加密算法）等。

另一类是分析基于格中困难问题设而设计新的后量子密码体制，如 NTRU 密码系统、GGH 密码系统、Gentry 算法的全同态加密系统、基于 LWE 问题的密码系统、Ajtai–Dwork 概率公钥密码系统等。

基于格的计算问题类题目在目前 CTF 竞赛中频繁出现，一度成为国际赛当中的主流考点，尤其是对 LLL 算法的理解和使用，成为解出许多高分值题目的关键。很多时候格相关攻击较为复杂，需要选手结合论文来进行推导，关于 CTF 竞赛中的论文问题会在后面的章节具体阐述。

2.2.3 数字签名

数字签名主要用于对数字消息进行签名，主要用于防止消息被冒名伪造、篡改，以及通信双方的身份鉴别。由于数字签名主要依赖于非对称密码算法，因此 CTF 当中考察数字签名类题目也主要依托非对称密码体系来进行设计，常见的包括 RSA 签名、ElGamal 签名、DSA 签名、针对某一特定椭圆曲线的 ECC 签名等，题目模型通常为要求我们提供某一特定字符串的签名，如果能正确通过验证则提供给选手 FLAG，针对数字签名类题目的攻击我们一般从三个角度来切入：

- 设法直接计算私钥，比如参数值过小或曲线选择不合适，导致私钥可以被直接计算出来（如 2019 ASIS CTF Quals-Halloween Party 题目，给定 y 坐标求 x 坐标，计算 2 在模 P.order()上的逆并将结果乘 2*P 直接得到 P）

- 设法恢复私钥，即通过若干特殊明文签名对，采用建立方程等方式重构出密钥（如 2019 DEFCON CTF Quals-Tania 题目，通过 LLL 算法和 Babai 最近平面算法恢复出密钥）
- 设法伪造签名，即在不知道私钥的情况下，直接构造出特定字符串的签名来拿到 FLAG（如 2019 RealWorld CTF-bank 题目，通过 rogue-key attack 伪造 Schnorr 比特币签名算法的银行提款签名）。

2.2.4 哈希函数

CTF 中考察哈希函数类问题主要以两类场景进行展开：
一类是哈希碰撞类场景，常见的攻击类型包括：

- 同谋碰撞攻击：生成任意两个不同的消息x和y，使得哈希值$f(x)$=$f(y)$
- 第二原像攻击：给定任意一个x及其哈希值$f(x)$，可以生成一个与之不同的y，使得哈希值$f(x)$=$f(y)$
- 相同前缀攻击：给定任意一个前缀p，可以生成两个不同的后缀x和y，使得哈希值$f(p+x)$=$f(p+y)$
- 选择前缀攻击：给定任意两个不同的前缀p和q，可以生成两个不同的后缀x和y，使得哈希值$f(p+x)$=$f(q+y)$

在考察形式上，题目通常会直接要求选手向服务器提交 A、B 两个输入，如果满足 A≠B 但 HASH(A)=HASH(B)，则判断碰撞成功。HASH 碰撞类题目虽然场景简单，但通常题目难度较大，而且很多时候都是出题人魔改之后的哈希算法，因此要求选手具备较强的分析能力，在攻击上多以考察代数攻击为主，如：

- 利用 small capacity parameter 误用问题构造 SHA3 Keccak 海绵结构碰撞（2019 0CTF/TCTF Quals-babysponge）
- 利用 Petit-Lauter 攻击构造超奇异同构图上 Charles-Goren-Lauter 哈希函数的弱实例碰撞（2017 HXP CTF-categorical）

另一类则是哈希伪造场景，虽然很多时候伪造哈希也可以通过寻找碰撞的方式来实现，但是它和碰撞场景下的考察侧重点通常不同，一种典型的攻击手法即哈希长度扩展攻击，即在已知$f(s+x)$、x的值和s的长度的情况下（这里s的值未知，即代表哈希时的 salt），要求选手对于任意一个y，能够计算出$f(s+x+y)$的值。对于哈希伪造的场景，我们的重点一般直接从攻击手法入手，而不需要对源码进行审计（绝大多数情况下，这类场景也不会给出哈希算法细节，而是直接调用现成的 MD5、SHA1 库函数）。

3 读 paper 的能力

自 2018 年起，大量论文题开始登上 CTF 历史舞台，然而一场标准的 CTF 国际赛通常为 48 小时，留给选手的分析时间是有限的，因此对于很多国际赛上的密码分析类题目来讲，选手是很难在短时间内提出一种可行的攻击手段的。这个时候，就需要快速提炼出题目当中的模型、场景等关键内容，然后去检索并阅读与之对应的会议、期刊论文，从中寻找攻击思想或手段。因此，选手的论文阅读理解能力（尤其是英文文献的阅读能力）对于密码学方向来讲就显得尤为重要。近年来国际赛当中 CTF 题目与学术界论文或研究成果联系紧密，如：

- 2018 PlaidCTF-braid 题目，考察选手通过阅读论文[1]，解决一个辫子群（Braid Group）下的 Conjugacy Search 问题。
- 2018 Hack.lu CTF- Escape the grid 题目，考察选手通过阅读论文[2]，攻破一个使用了非安全函数产生仿射变换的 rasta 密码系统。
- 2019 0CTF/TCTF Quals-zer0lfsr 题目，考察选手通过通过阅读论文[3]，利用 Fast Correlation Attack 恢复出一个 Meier-Staffelbach 模型下 0.75 相关性的 3 个 LFSR 的初始状态（虽然这道题在当时比赛时被 z3 非预期了）。
- 2019 PwnThyBytes CTF-Wrong Ring 题目，考察选手通过阅读论文[4]，通过将环上带误差学习（RLWE）样本转化为多项式带误差学习(PLWE)样本，攻击一个（non-dual）RLWE 的实例来求私钥。

还有很多题目直接以某一论文或研究成果为背景进行设计，如：

- 2017 ASIS CTF Final-Marijuana

2017 年 5 月，Divesh Aggarwal 等学者提出了一种基于梅森素数的公钥密码体系[5]，2 个月后，该密码系统被 Marc Beunardeau 等学者攻破，相关成果以论文形式发表[6]，同年 9 月，该事件以 CTF 形式出现在当年的 ASIS CTF 竞赛当中，预期选手通过阅读论文使用随机划分和 LLL 攻击攻破密码系统。

- 2017 HXP CTF-notsosmart

2017 年 2 月，马萨里克大学的 Matus Nemec 和 Marek Sys 等学者发现英飞凌科技提供的 RSALib 库中的密钥生成算法存在漏洞，并据此提出了一种可行的 RSA ROCA 攻击。同年 11 月，相关攻击细节以论文形式披露[7]（该攻击同步影响了 TPM 1.2, 4.3.5 版本前的 YubiKey 4 等产品，CVE 编号：CVE-2017-15361），两周后，该事件以 CTF 形式出现在当年的 HXP CTF 竞赛当中，预期选手通过阅读论文使用 ROCA 攻击攻破一个 RSA 密钥生成算法。

4 扎实的编程功底

对于绝大多数情况下，当选手通过审计源码或者阅读论文整理出一种攻击手段之后，接下来都需要通过编程的方式写一个 solver 或 exp 来计算出你所需要的数据或实现你的攻击方式，这个时候就需要选手具备良好的编程功底，因为很多时候将复杂的数学模型转化成可行的脚本并不是一件容易的事。

这里所提到的编程主要可以看成两个部分：

一类是使用 python、C/C++等这类语言进行编程，它是我们主要使用的编程语言，我们通过代码来描述我们的攻击过程，继而实现攻击。

第二类是针对某一工具的编程，常用的包括 SageMath、MatLab、Mathematica 编程等，这一类的编程往往起到的不是描述而是计算或辅助绘图的作用，帮助我们更好的完成解题过程。其中尤其以 SageMath 最为常用，如针对群、环、域等代数结构的计算，在 SageMath 中都可以很方便的进行操作，而不需要进行代数结构的二次描述，另外很多常用算法都以内置函数的形式在 SageMath 当中集成，可以很方便的供选手使用。

5 值得关注的 CTF

由于 CTF 竞赛中专职密码学方向的选手通常较少，因此一场比赛是否有较高质量的密码学题目其实往往取决于办比赛的战队中有没有一个优秀的密码学选手，目前有一些比赛的密码学题目质量是要远高于其他竞赛的，值得参赛选手去重点关注（另外有一些战队虽然拥有很强的密码学选手，但是战队办比赛较少，可以关注其每次比赛赛后发布的 writeup 来了解其做题时的一些思路和解法）：

- Teaser Dragon CTF（波兰 Google 安全团队 Dragon Sector 战队主办）
- 0CTF/TCTF （中国腾讯安全 A*0*E 战队主办）
- HXP CTF （德国慕尼黑工业大学 HXP 战队主办）
- ASIS CTF & Crypto CTF （伊朗谢里夫理工大学 ASIS 战队主办）
- CODE BLUE CTF （日本联合战队 binja 主办）
- PlaidCTF （美国卡耐基梅隆大学 PPP 战队主办）
- Midnight Sun CTF （瑞典 HackingForSoju 战队主办）
- SpamAndFlags Teaser CTF （匈牙利布达佩斯技术与经济大学!SpamAndHex 战队主办）
- PwnThyBytes CTF （罗马尼亚 PwnThyBytes 战队主办）

6 不只是 CTF

对于一名密码学方向的 CTF 选手来讲，除了通过活跃于各大 CTF 比赛和在平时保持足够的 CTF 题目训练强度来提升自己的水平之外，实际上还有一些其他的密码学相关竞赛可以用来作为训练，每年诸如 NSU CRYPTO、WhibOx Contest 等解题类密码学竞赛，也可以参加或做赛后练习用，以此来锻炼密码分析能力。

7 对密码学的热情

兴趣因素放在本文最后一个环节进行论述，但是实际上这是学习密码学的第一步，也是最重要的一步，拥有对探索密码学领域的无限热情，会使你比任何人都更有可能成为一名优秀的 CTF 密码学选手。

8 参考文献

[1] Ko, K.H., & Lee, J.W. (2006). A fast algorithm to the conjugacy problem on generic braids.

[2] Dobraunig C. et al. (2018) Rasta: A Cipher with Low ANDdepth and Few ANDs per Bit. In: Shacham H., Boldyreva A. (eds) Advances in Cryptology – CRYPTO 2018. CRYPTO 2018. Lecture Notes in Computer Science, vol 10991. Springer, Cham

[3] Meier W. (2011) Fast Correlation Attacks: Methods and Countermeasures. In: Joux A. (eds) Fast Software Encryption. FSE 2011. Lecture Notes in Computer Science, vol 6733. Springer, Berlin, Heidelberg

[4] Elias Y., Lauter K.E., Ozman E., Stange K.E. (2015) Provably Weak Instances of Ring-LWE. In: Gennaro R., Robshaw M. (eds) Advances in Cryptology -- CRYPTO 2015. CRYPTO 2015. Lecture Notes in Computer Science, vol 9215. Springer, Berlin, Heidelberg

[5] Aggarwal D., Joux A., Prakash A., Santha M. (2018) A New Public-Key Cryptosystem via Mersenne Numbers. In: Shacham H., Boldyreva A. (eds) Advances in Cryptology – CRYPTO 2018. CRYPTO 2018. Lecture Notes in Computer Science, vol 10993. Springer, Cham

[6] Beunardeau M., Connolly A., Géraud R., Naccache D. (2019) On the Hardness of the Mersenne Low Hamming Ratio Assumption. In: Lange T., Dunkelman O. (eds) Progress in Cryptology – LATINCRYPT 2017. LATINCRYPT 2017. Lecture Notes in Computer Science, vol 11368. Springer, Cham

[7] Matus Nemec, Marek Sys, Petr Svenda, Dusan Klinec, Vashek Matyas, The Return of Coppersmith's Attack: Practical Factorization of Widely Used RSA Moduli, Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, October 30-November 03, 2017, Dallas, Texas, USA [doi>10.1145/3133956.3133969]