

密码学与复杂性理论

G. RUGGIU

摘要

最近，复杂性理论被提出作为评估密码机性能的基础。它们与香农模型进行了比较，并对随机性概念提出了新的见解。但需要强调的是，统计学的观点仍然是更可靠的。

密码学与复杂性理论

最近，复杂性理论被提出作为评估密码系统性能的基础。在这篇简短的综述中，我们将介绍连接这两个概念的不同方法。

复杂性理论相对较新，其动机是分析算法的效率。它们的主要特点是，它们是处理非常通用算法的非常通用的理论；它们最具体的成果提供了一些关于算法渐近行为的信息。

密码学的核心问题是评估保密系统的安全性，即系统对密码分析的免疫力。当密码分析可行时，这种评估必须衡量找到解决方案需要多少时间和信息。

1. 香农模型

这个问题的第一个数学处理是由香农¹在四十年代完成的。他的理论使得问题得以正确形式化。因此，他可以给出设计保密系统的一些指导原则。

香农的方法基于一个概率模型，该理论的核心是评估明文概率。主要有两个参数：

- 明文的先验概率： $P(m)$
- 当截获密文 c 时，明文 m 的条件概率 $P(m/c)$ 。

香农定义的主要概念是"完全保密"：当对于所有 m, c ，有 $P(m) = P(m/c)$ 时，一个密码系统是完全保密的。因此，密文的知识不提供关于明文的任何信息：此类系统的密码分析是不可能的。

但完全保密有局限性：它要求密钥数量至少与明文消息数量一样多。这意味着密钥必须与消息一样长。因此，显然这些系统除了在特定情况下外是不实用的，因为密钥必须通过安全信道交换。

在实践中，大多数系统具有有限长度的密钥。如何表征这些系统的安全性？香农表明，对于这些系统，存在一个消息的最小长度，称为"唯一解距离"，超过这个距离，密码分析就有唯一解。这个距离的存在是因为明文所属语言的冗余性。

在这种情况下，可以通过尝试所有不同的密钥来找到解决方案：能给出合理明文的密钥就是正确的密钥。如果尝试次数太多，这种穷举搜索就必须被视为不可能：密码分析者预计没有足够的时间来找到解决方案。

但是如何确保所有这些尝试都是必要的呢？算法的复杂性理论试图回答这个问题。

2. 算法的复杂性

该理论试图给出解决一个问题的难度的度量。通常，解决一个问题的算法定义一个需要两种类型资源的计算：时间（或计算步骤数）和空间（或用于存储计算中使用的信息的内存）。这些定义了复杂性度量。复杂性是计算输入长度的函数。

让我们回顾一下复杂性理论的主要结果。在一个通用的计算模型（例如图灵机）中，根据时间复杂性（即计算步骤数 $f(n)$ ）定义了一个函数层次结构。复杂性类根据输入长度 n 的增长速度来定义。

例如，有以下几类：

线性： $f(n) = O(n)$

多项式： $f(n) = O(n^a)$

指数： $f(n) = O(2^n)$

等等。

符号 $O(\alpha)$ 表示渐近值与 α 成正比。

人们通常认为，复杂性至少是指数级的问题是难解的，意即没有实用的算法来解决它。另一方面，多项式时间复杂性通常等同于实际可计算性。（对于多项式时间界限的次数没有明确的截止点）。因此，区分多项式时间算法和指数时间算法很重要。

需要一个新的概念：多项式时间可归约性。问题A可多项式时间归约到问题B，如果存在一个总可计算函数 f ，其计算时间以输入 x 长度的多项式为界，使得：

$$A(x) = B(f(x)), \forall x.$$

A 已经被多项式归约到 B。另一个概念是相对完全性：设 B 是问题集合 C 中的一个问题。如果 C 中的每个 A 都可以多项式归约到 B，则称 B 是 C-难的；如果 B 也属于 C，则称 B 是 C-完全的。因此，在某种意义上，C-完全问题是 C 中最难或最困难的问题。

3. NP完全性

为了在多项式类和指数类之间寻找中间类，人们考虑了非确定性算法。在这些算法中，在计算的任何一点都可能有多条指令适用。可以选择其中任何一条指令。

因此，非确定性算法定义了尽可能多的计算路径（对应于可能的选择），并且至少有一条路径通向解决方案。因此，如果机器“猜出”解决方案，它就选择了正确的计算路径；如果机器无法猜出解决方案，它必须尝试所有可能的计算路径，而计算路径的数量通常是指数级的。

可由多项式时间算法解决的问题类称为 \mathbb{P} ；NP类由可由非确定性算法在多项式时间内解决的问题组成（假设机器能猜出解决方案）。

了解 P 和 NP 之间的关系非常重要。这个问题是计算理论中最重要的问题之一。

目前，情况还不是很清楚。人们普遍认为 P 真包含于 NP。如果是这样，NP 应该是 P 和困难问题之间的一个良好中间类。另一个类非常有趣：CO-NP。它由那些补问题在 NP 中的问题组成（假设这些问题属于“是-否”类型，而补问题是“否-是”。不知道是否 $NP = CO - NP$ ）。在 $CO-NP \neq NP$ 的假设下，NP-完全问题不在 NP 和 CO-NP 的交集中。因此，它们比 $NP \cap CO-NP$ 中的问题更难。例如，合数问题属于 $NP \cap CO-NP$ 。但如果任何 NP-完全问题在 NP 和 CO-NP 的交集中，那么 $NP = CO-NP$ 。

G. Brassard¹ 证明，如果存在某个单向函数 f ，那么 P 真包含于 N 和 NP 的交集中，并且如果 f^{-1} 是 NP-难的，那么 $NP = CO - NP$ 。一个函数是单向的，如果它易于计算 ($f \notin P$) 且 f^{-1} 难以计算 ($f^{-1} \notin P$)。

现在很明显，加密和解密操作属于 \mathbb{P} ，因为它们通常在线性时间内完成。但是解密（在没有密钥的情况下）是一种非确定性的密码分析，因为需要猜出正确的密钥。

现在我们遇到了主要问题：密码分析问题是 NP-完全的吗？如果是这样，就会有证据表明它是难解的。

从一个非常普遍的观点来看，密码分析问题归结为解决一个布尔方程，其未知数是密钥的比特。这个问题是 NP-完全的。

当然，特定密码机的密码分析不是 NP-完全的，因为它是一个特定的布尔方程。但没有理由为这台机器寻找特定的算法。这意味着该密码机具有某些可供特定算法利用的特性。因此，设计密码系统的第一条指导原则是避免任何逻辑上的特殊性。

然而，必须强调，复杂性理论应用于密码分析时必须非常谨慎：

- 计算理论处理的是最坏情况，而一个高度复杂的函数几乎总是可能容易计算的。
- 在密码学中，并不需要精确解，而且已知一些 NP-完全问题有好的近似解可以计算。
- 密码分析者可能有足够的辅助信息，使得即使问题是 NP-完全的，他也能解决问题⁵。

4. 序列的复杂性

让我们考察另一个观点。不分析机器本身，而是分析机器产生的输出序列，我们能说什么？

机器缺乏逻辑特殊性这一点必须在输出序列的结构中体现出来，即输出必须看起来像一个随机序列。

根据 Kolmogorov⁶ 和 Chaitin⁷，序列 S 的复杂性 I 是能够产生 S 作为输出的最短程序 P 的长度，其中计算机 C 接受 P 作为输入。可以证明这种复杂性与 C 无关。

这种复杂性度量具有一些重要的性质：

- 序列 S 的复杂性最多是 S 的长度，因为总是可以通过直接展示 S 来描述它；这样的程序长度就是 S 的长度。
- 长度为 k 的大多数序列的复杂性大约为 k 。例如，对于足够大的 n ，所有长度为 n 的序列中，有 99.8% 的序列其复杂性大于 $n - 10$ 。

现在我们可以定义一个算法随机序列。记作：A-随机。

粗略地说，一个序列是 A-随机的，如果它的复杂性大约等于其长度。更精确地说，一个长度为 n 的序列 S 是 t -A-随机的，如果它的复杂性大于 $n - t$ 。

但是没有算法可以判定一个序列是否是 A-随机的。然而，当 n 足够大时，长度为 n 的序列是 A-随机的概率接近于 1。所以，如果一个序列是通过抛硬币定义的，那么它是 A-随机的概率接近于 1。

该理论的主要兴趣在于建立复杂性和随机性之间的联系。因此，它证明了如果密码机的输出是 A-随机的，那么该机器没有逻辑特殊性，并且密码分析可能是困难的。

幸运的是，A-随机性与概率定义是一致的：如果一个序列是 A-随机的，那么它在统计上是随机的。但反之则不成立：一些在统计上随机的序列并不是 A-随机的⁸。这意味着统计检验虽然不能判定一个序列是否是 A-随机的，但却是判定随机性的良好近似算法：如果一个序列在统计上不随机，那么它也不是 A-随机的。

5. 表观复杂性

但实际上，已知密码机产生的序列 S 具有较低的复杂性，其数量级与密钥 K 的长度相当：对于每个明文 m ，我们都有方程：

由于只需考虑长度达到唯一解距离的明文 m ， S 的复杂性就等同于 K 的复杂性。但对于每个 m ， f_m^{-1} 必须是困难的（每个 f_m 都是单向的），以便解出这个关于 K 的方程是不可行的。找到 K 等价于找到一个能生成 S 的程序。这引出了一个新概念：表观复杂性 I_A ，其目的是衡量完成 f_m^{-1} 的难度。已经提出了不同的 I_A 度量方法，这些方法是从序列本身的结构推导出来的。我们现在可以定义表观随机性：如果一个序列的表观复杂性是最大的（通常与其长度同阶），则称该序列是表观随机的。

让我们注意，通常如果 S 是 A-随机的，那么它也是表观随机的。假设 $I_A(S)$ 由最短程序 P_S 定义，使得在计算机 C 上， P_S 的输出是 K ：

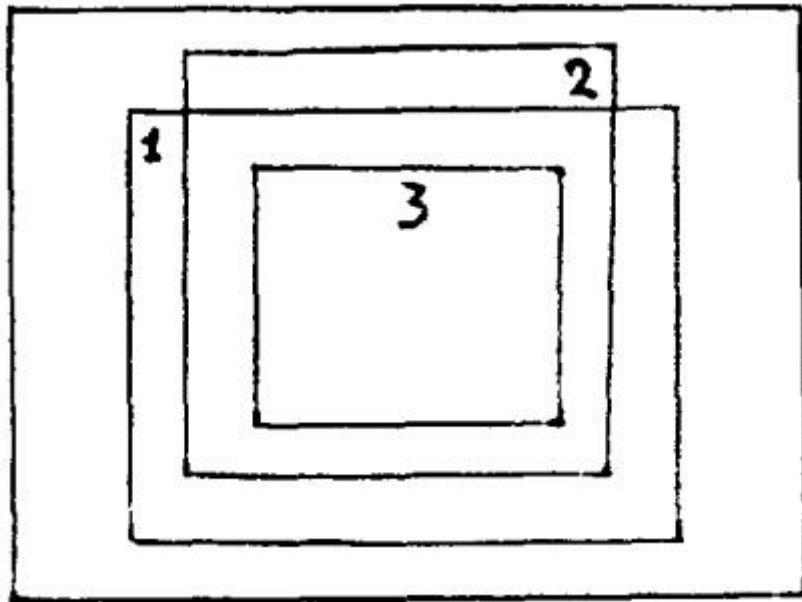
如果 S 不是表观随机的，那么 P_S 的长度 $\ell(P_S)$ 必须比 S 的长度短得多。

计算机 C 可以从 f_m 和 K 的程序计算出 S ：

设 $X(S)$ 为 S 的复杂性。

那么 $X(S) = \ell(f_m) + \ell(K) \simeq \ell(K)$ （对于足够大的 K ）。所以： $\ell(K) \simeq \ell(S)$ 。但是 $S = C(f_m, C(P_S))$ ；所以： $X(S) \simeq \ell(P_S) << \ell(S)$ ，如果 S 是 A-随机的，这不可能。

因此，复杂性概念可以通过算法、统计和表观复杂性来近似。相应的随机性概念在以下关系中相互关联：



图中：

- 1: 表观随机序列
- 2: 统计随机序列
- 3: A-随机序列。

唯一有效的算法是统计检验，问题在于定义适当的随机性统计检验；而这个问题至今还远未明确。

6. 结论

总之，将复杂性理论应用于密码机评估还有很多不足之处。每种理论都有其缺陷和不足。要实现这一目标，还有很多工作要做。

然而，每种理论都提供了对该主题的互补观点。但迄今为止，没有一种理论提供任何有用的工具来评估密码机的安全性，到目前为止，统计检验仍然是最值得信赖的评估方法。

参考文献

1. C. SHANNON - Communication Theory of Secrecy Systems B.S.T.J. Vol. 28, October, 1949, p. 656.
2. M. MACHTEY, P. YOUNG - An introduction to the general Theory of algorithms - North-Holland, 1978.
3. G. BRASSARD - A note on the complexity of cryptography - I E E E Trans. on I.T., Vol. IT-25, n° 2, March 1979, p. 232.
4. W. DIFFIE, M. HELLEMAN - New directions in cryptography, I E E E Trans. on I.T., Vol. IT-22, n° 6, November 1976, p. 644.
5. A. LEMPEL - Cryptology in transition, Computing Surveys, Vol. 11, n° 4, December 1979, p. 285 (Example, p. 300).
6. A. KOLMOGOROV - Three approaches to the quantitative definition of information. Problemy Peredaci Informaci 1, 4-7, 1965.
7. G. CHAITIN - Algorithmic Information Theory - IBM J. Res. Dev. Vol. 21, July 1977, p. 350.

8. T. FINE - Theories of Probability. Academic Press, 1973 (Chapter V).
9. A. LEMPEL, J. ZIV - On the complexity of sequences - I E E E Trans. on I.T., Vol. IT-22, n° 1, January 1976, p. 75.
10. E. FISHER - Measuring Cryptographic performance with production Processes. Cryptologia, Vol. 5, n° 3, July 1981, p. 158.
-

专业术语中英文对照表

英文术语	中文翻译
Cryptology	密码学
Complexity Theory	复杂性理论
Cryptographic system	密码系统
Cryptanalysis	密码分析
Shannon's model	香农模型
Perfect secret	完全保密
A priori probability	先验概率
Conditional probability	条件概率
Cryptogram	密文
Unicity distance	唯一解距离
Redundancy	冗余性
Exhaustive search	穷举搜索
Algorithm complexity	算法复杂性
Time complexity	时间复杂性
Space complexity	空间复杂性
Turing machine	图灵机
Complexity class	复杂性类
Polynomial time	多项式时间
Exponential time	指数时间
Polynomial time reducibility	多项式时间可归约性
C-hard	C-难的
C-complete	C-完全的
Non-deterministic algorithm	非确定性算法
P (complexity class)	P (复杂性类)
NP (complexity class)	NP (复杂性类)
NP-complete	NP-完全的
CO-NP	补NP类
One-way function	单向函数
Boolean equation	布尔方程
Algorithmic Information Theory	算法信息论
Kolmogorov complexity	柯尔莫哥洛夫复杂性
Algorithmic randomness (A-random)	算法随机性 (A-随机)
Statistical randomness	统计随机性

英文术语	中文翻译
Apparent complexity	表观复杂性
Apparent-randomness	表观随机性
Statistical tests	统计检验