密码密钥傻傻分不清？认识密码学中的最高机密

原创 严强 微众银行区块链 4月15日

来自专辑

WeDPR隐私保护周三见

严强，SMU信息安全方向博士，信息安全顶级国际学术会议最佳论文奖获得者；曾作为Google隐私保护基础技术架构部门唯一来自中国的早期核心成员，领导研发的技术方案在Android和Google Play生态各大门户产品中全面集成投产。

密码学为何称之为密码学？密码和密钥究竟有何区别？隐私保护方案中，密钥的角色是否

可以被替代？密钥在使用过程中存在哪些风险？

这里，我们将以密码学中的密码为起点，展开一系列对密码学算法核心组件的技术剖析。密码和密钥在密码学算法中有着至关重要的地位，了解密码和密钥的作用，有助于理解基于密码学的隐私保护方案是否具备有效性。密码和密钥对于用户而言，则是最终达成隐私数据『始于人、利于人、忠于人』隐私保护效果的无上法器。

密码学的英文为Cryptography，源自希腊语“κρυπτός秘密”和“γράφειν书写”。最初，其研究主要集中在『如何在攻击者存在的环境中隐秘地传输信息』，是一个关于信息编码的学科，由于其最重要研究目标之一是保密，实现敏感信息的秘密编码，所以被称之为密码学。

密码学中的密码，和我们日常生活中登录各类信息化系统所使用的密码是两个不同的概念。前者包含了信息加密编码、密文解密解码、数据完整性验证等一系列信息变换过程。而后者更多地是指代密码学信息变换过程中所使用的便于用户记忆的一类密钥，为了以示区别，在下文中称之为**用户口令**。

在密码学中，密钥的作用与现实生活中的钥匙很相似，只有掌握密钥的用户，才能解密对应的隐私数据，或进行数字签名等相关敏感操作。

为什么密钥能够有这么神奇的作用，一切要从柯克霍夫原则谈起。

柯克霍夫原则

柯克霍夫原则是现代密码学算法设计基本原则之一，最早由荷兰密码学家Auguste Kerckhoffs在1883的论文La Cryptographie Militaire（军用密码学）中提出。

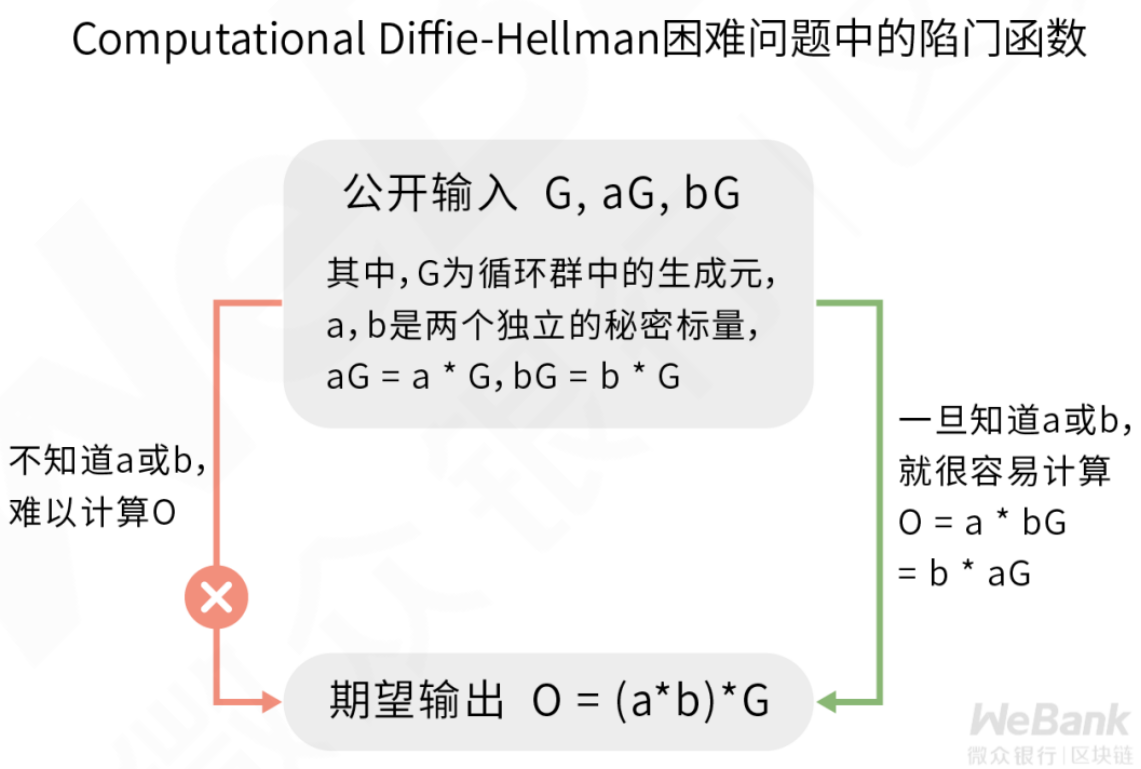
其核心思想是『**密码学算法的安全性，不应该建立在算法设计保密的基础上**』。即便算法设计是公开的，只要实际使用的密钥没有被攻击者获知，密码学算法产生的密文信息就不应该被轻易破解。

被誉为“信息论之父”的美国数学家、电子工程师、密码学家Claude Elwood Shannon后来将这一原则进一步扩展，应用到任意信息安全相关的系统，由此也奠定了密钥在现代密码学中的核心地位。

密钥具体如何使用呢？这里有必要回顾第3论中提到的，密码学算法设计所基于的计算不对称性，以及与之相关的一个重要概念——单向陷门函数。

一个单向陷门函数可以抽象为y = f(x, key)，其中，x为敏感的隐私数据输入，y为经过算法保护的不敏感输出，key就是密钥。实际情形下，根据具体密码学算法设计和实现上的差异，密钥key可以有不同的表现形式，也可以表现为多个秘密参数。

如果以上函数是一个密码学安全的单向陷门函数，在不知道密钥key的前提下，很难从输出y通过逆函数反推出输入x，由此避免了隐私数据的泄露。



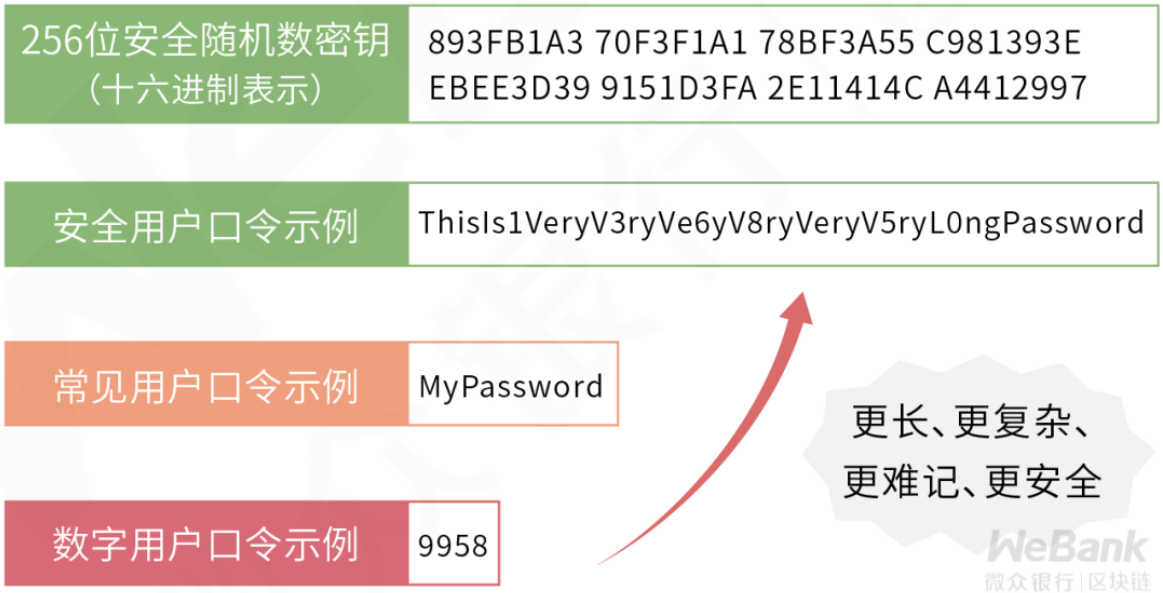
由此可见，密钥就是密码学信息变换过程中的最高机密。谁掌握了密钥，谁就掌握了隐私数据的访问权。

人类可用的密钥

一般而言，再精密的隐私保护方案，最终都需要服务于人类用户。由于密码学隐私保护方案的安全性很大程度上取决于密钥的长度和复杂性，这也为人类用户在使用密钥时带来了不小挑战。

目前业界主流推荐的密码学安全强度是256位，即密钥的信息熵至少等价为256比特的随机数。如果我们用常见的字母数字来设定用户所用的密钥，该密钥的长度至少为256/log2(26*2+10) ~= 43个随机字符。

考虑到用户通常为了便于记忆而拼接字典中的单词来构成密钥，此时为了满足密钥信息熵的随机性要求，实际可能需要使用长度更长的密钥。



相比之下，现有系统对用户口令的长度一般要求在6~20字符之间，对于部分应用4~6位数字用户口令也不少见。所以，这些用户口令的随机性和长度都不足以达到256位安全强度。

如果一个隐私保护方案所使用的密钥只源自用户口令，是无法满足隐私数据的安全性要求

的。

然而，普通人类并不具备计算机一般强大的计算和记忆能力，难以记忆和处理过长的密钥。此时，需要借助技术手段来提高人类可用密钥的信息熵，常见的解决方案有以下三类：

	方案设计	风险分析
平台全权托管	由平台为用户生成满足安全强度的密钥。每次使用时，平台服务商需要验证用户口令，确保隐私数据的访问授权真实反映了用户意愿。	难以防范内部人员或系统异常等导致的隐私数据泄露风险，需要依赖平台服务商的信誉来保障密钥不被滥用。
本地全权托管	本地可信设备为用户生成满足安全强度的密钥。每次使用时，本地可信设备需要验证用户口令，提取本地密钥与平台服务商进行认证交互。	本地可信设备有失窃的风险，失窃之后，攻击者可能通过硬件攻击技术，提取其中的密钥。
混合托管	本地可信设备为用户生成满足安全强度的密钥。每次使用时，本地可信设备需要验证用户口令，提取本地密钥，同时结合用户口令一同与平台服务商进行认证交互。	本地可信设备仍有失窃的风险，但攻击者需要同时窃取用户口令，才能破解平台服务商的认证。

三类解决方案中，平台全权托管的用户体验最好，同时也伴随着最大的隐私风险。混合托管和本地全权托管，在用户体验上差异不大，混合托管相关的隐私风险更低。

需要注意的是，这里存在一个固有的设计取舍，隐私数据的自主权与数据服务的完备性不可兼得。

平台全权托管方案中，用户隐私数据的实际控制权在平台手中，由此平台可以提供诸如用户口令重设、数据恢复等关键数据服务。

然而，在其他托管方案中，用户隐私数据的实际控制权在用户手中，一旦用户遗失密钥或用户口令，则平台无法解密对应的数据，也无法提供口令重设等相关密钥服务。

对于企业而言，具体方案的选择，需要结合用户使用习惯和行业监管要求，建议在平台全权托管和混合托管之间做选择。对于高敏感性隐私数据，酌情选择混合托管，并需要配合

密钥恢复方案使用。

密钥相关的风险

隐私数据的自主权往往是隐私保护方案强调的重点，但是为了切切实实地获得控制权，仅仅是安全地使用单个安全密钥，就可能会给用户体验方面带来显著负担，而且还需要防范其他密钥相关的泄露风险。

这些风险可以大致分为以下两类：

内在风险

这类风险与隐私保护方案的内在设计和实现有关。由于绝大部分密码学算法和协议不是信息论安全，也就说，同一个密钥使用的次数越多，理论上被破解的概率越大。

对应的常见风险分析手段是，考虑对应密码学算法和协议在选择明文攻击（Chosen-plaintext Attack, CPA）和选择密文攻击（Chosen-ciphertext Attack, CCA）下，是否依旧安全。

这两类攻击都允许攻击者获得一定数量的隐私数据明文和密文对，由此分析破解所使用的密钥。

在现实生活中，攻击者非常有可能获得这样的能力，截获明文和密文对，甚至主动注入数据，生成破解分析所需的明文和密文对，这类风险是真实存在的。

外在风险

这类风险虽然与隐私保护方案的内在设计和实现无关，但却实实在在地对方案的实际效果产生巨大威胁。

比较典型的攻击有社会工程学，具体指通过欺骗性手段，如钓鱼网站、诈骗短信等，诱导用户直接给出密钥，或者通过下载安装病毒木马，间接盗取密钥。

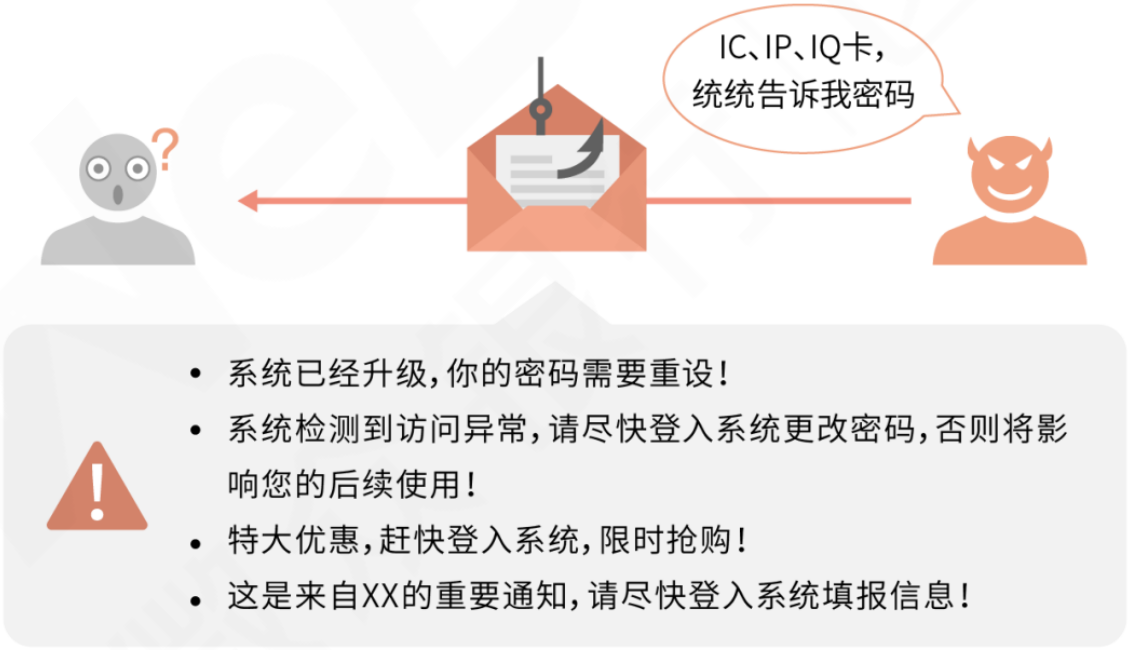
无论是哪一类风险，如果用户只有一个密钥，一旦被盗，所有的账户都有被盗的风险，后果不堪设想。

处理好这些风险的必要条件，就是产生并使用多个随机密钥，但这也为隐私保护方案的可用性带来了更大的挑战。

无论隐私保护方案设计安全性多高，如果由于用户体验差，用户难以接受，或者以不安全的变通方式使用，其真实有效性都会大打折扣。这也是学术方案向业务方案转化最常见的阻碍之一。

除了探索更优的方案设计，适当的用户教育也是非常必要的推广手段。

总体而言，同时处理好密钥使用过程的安全性和可用性，是落实隐私保护的重要前提。



社会工程学可能会利用以下因素进行攻击：

- 心理弱点：本能反应、信任、好奇、贪婪、恐惧等
- 行为相关的规则制度漏洞

正是：隐私数据控制难自主，访问密钥在手任我行！

密钥是任何基于密码学技术方案的最高机密，如何保障其安全性，并让作为隐私数据属主的人类用户方便地记忆和使用，是将隐私控制权回归属主的关键。

这个过程难免会引入数量繁多的密钥，如何实现有效的密钥管理，对于计算机系统和人类用户可以使用哪些不同的技术和策略，欲知详情，敬请关注下文分解。

---END---

《隐私保护周三见》

“科技聚焦人性，隐私回归属主”，这是微众银行区块链团队推出《隐私保护周三见》深度栏目的愿景与初衷。每周三晚8点，专家团队将透过栏目和各位一起探寻隐私保护的发展之道。

栏目内容含括以下五大模块：关键概念、法律法规、理论基础、技术剖析和案例分享，如您有好的建议或者想学习的内容，欢迎随时提出。

栏目支持单位：零壹财经、陀螺财经、巴比特、火讯财经、火星财经、价值在线、链客社区

往期集锦

长按二维码关注

微众银行区块链

白皮书下载 | 订阅干货 | 进群交流 | 合作联系