文章编号：2095-2163(2020)10-0059-07 中图分类号：TP309.7 文献标志码：A

社会工程学视角下密码的设置与保管研究

张 硕，吴 瑕
（重庆邮电大学 网络空间安全与信息法学院，重庆 400065）

摘 要：随着互联网各平台相继以形式化手段“禁止”设置弱口令，令不法分子暴力破解密码的难度再度升级。因各平台密码的要求并非统一，加剧了用户进行密码设置和记忆的难度；但若降低记忆难度，范用一个“健壮”密码，则会导致撞库风险的存在。本文从社会工程学的角度分析密码破解问题，进而分析目前密码设置所面临的风险和密码保管的困境。提出基于场景、树状结构、编码和扩散混淆的四种密码设置方法，并研究出一套自建密码设置模型和二次处理加记录的密码保管方法。

关键词：社会工程学；密码；暴力破解

Research on the setting and keeping of personal password from the perspective of social engineering

ZHANG Shuo, WU Xia
(School of Cyber Security and Information Law, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

【Abstract】 As various platforms on the Internet successively "forbid" setting weak passwords by formal means, the difficulty of violent cracking of passwords by criminals has once again escalated. Because the password requirements of each platform are not uniform, it makes it more difficult for users to set and remember passwords. However, if the difficulty of memory is reduced, using a "robust" password will lead to the risk of colliding with the library. Now we analyze the problem of password cracking from the perspective of social engineering, and then analyze the risks faced by the current password setting and the dilemma of password storage. Four password setting methods based on scene, tree structure, encoding and diffusion ambiguity are proposed, and a self-built password setting model and a password storage method of secondary processing record are developed.

【Key words】 Social engineering; Password; Brute force attacks

0 引 言

在日常生活中，每个人都会面对诸多（例如购物、学术、考试、聊天、娱乐、办公等系统或平台）的密码进行设置和记忆。从学童的密码文具盒到银行中的保险柜，从手机屏幕锁到办公系统的登录，处处需要密码，密码的重要性不言而喻。如果个人密码遭到破解，轻则个人信息被泄露或被盗用，重则导致个人财产损失甚至企业的安全事故[1]。

对于密码的设置是方法论的问题，其趋同于“网络协议（在计算机网络中双方实现通信，必须遵循一些事先制定好的规则和标准；这些为进行数据交换而建立的规则、语义或标准称为网络协议）[2]”。一般而言，日常遇到的密码（只考虑密码构成，暂不考虑密码长度）有纯数字组合、混合字符组合和全字符组合，且在设置密码的时候应避开“弱口令”。综合而言，无论是密码的设置还是保管都是一个较为棘手的事情。

1 研究前提

1.1 研究对象之密码的界定

密码在中文里是“口令”（password）的通称[3]。“密码”一词从汉语语意上分析是指秘密的代码。密码从无到有直到发展至现在，其包含了二种语义：

（1）密码学中所指的密码（密码体制）；

（2）日常生活中的密码（口令）。

1.1.1 密码学中的密码

在几千年前密码就以行帮暗语和文字猜谜的方式使用，后来将密码应用到战争当中，用来传递战事信息，密码发挥了关键性的作用。随着历史发展，对于“密码”的研究，逐渐演变为密码学，其目的就是让通信双方能够在一个不安全的信道上进行保密通信[4]。密码学研究的是密码编码和破译的技术与方法，通过研究密码变化的客观规律，将其应用于编制密码，实现保密通信的技术被称为编码学；通过研究密码变化的客观规律，并将其应用于破译密码，实

基金项目：重庆市社会科学规划项目（2014YBFX103）；重庆市渝北区法学会项目（E2018-88）；重庆市创业训练项目（S201910617048X）。

作者简介：张 硕（1996-），男，硕士研究生，主要研究方向：电子数据取证；吴 瑕（1999-），女，本科生，主要研究方向：信息安全专业。

收稿日期：2020-05-22

现获取通信信息的技术被称为破译学。编码学和破译学统称为密码学[5]。密码学中的密码,是指通信双方按约定的法则进行信息特殊变换的一种重要保密手段。依照这些法则,将明文变换为密文,称为加密变换;将密文变换为明文,称为解密变换[6]。

1.1.2 本文研究的密码

本文研究的“密码”,指的用户打开设备或者登录系统的字符串或称为口令,密码的使用是进行身份的认证。用户对密码的设置,主要是为了用户对所拥“资源”的掌控,“密码”则是资源的钥匙。对于密码的设置,根据学者的统计情况见表 1[7]。

由表 1 可见,密码设置无外乎包括以下 4 个方面:

(1)与用户自身相关的信息;

(2)与平台(系统/设备)相关的信息;

(3)自定义的特殊规则;

(4)前三者的组合。

表 1 密码内容统计表[7]

Tab. 1 Password content statistics table

代号	说明	合计	占样本总数百分比/%
B	生日等日期相关	68	12.25
E	有规律的字母	16	2.88
M	有规律的数字	106	19.10
N	名字相关	25	4.5
S	与用户名相同	8	1.44
T	电话号码	122	21.98
D	学校或班级	23	4.1
A	学号	1	0.18
Y	本地邮编	1	0.18
其他特征	上述各特征的混合体	34	6.12
非特征统计	无明显特征的字符串	151	27.12
总计	样板数量总计	555	100

1.2 研究背景

1.2.1 社会工程学视野下的密码破解

社会工程学包含两层含义,广义的含义是利用社会中的各个方面要素,去解决复杂问题的方法论;狭义的含义则是针对互联网领域中“安全”的一种攻击手段。广义的“社会工程学”是建立理论并通过利用自然的、社会的和制度上的途径来逐步解决各种复杂的社会问题[8]。狭义的“社会工程学”是一种针对受害者的心理弱点、本能反应、好奇心、信任、贪婪等心理陷阱,实施诸如欺骗、伤害等危害的方法。密码中的社会工程学(攻击)是一种利用上述心理陷阱获取用户个人信息、系统/平台信息、用户的惯例/规则等信息的攻击方法[9]。社会工程学属于非传统的信息安全范畴,随着网络技术、产品和服务的日趋成熟,很多攻击手段难以快速实现,此时社会工程学攻击对于攻击者来说凸显重要,而对于防御者来说更要重视[10]。

对于社会工程学的应用如图 1 所示[11]。针对本研究而言,仅涉及信息收集获取和密码破解攻击两个技术,两者相互交叉、递进发展。信息收集获取,是将收集的信息进行分析,以备用来暴力破解密码;密码破解攻击之后,则又会被再次收集更多的信息;两者相互依赖共存。社会工程学中的信息收集分析,对密码的安全设置具有较大的安全威胁。因此,本研究从社会工程学视角去思考密码“数据”的来源,分析用户可能设置的密码。使用上述方法进行密码破解攻击,再进行密码设置和保管的剖析,以提出合理的方案进行设置和保管。

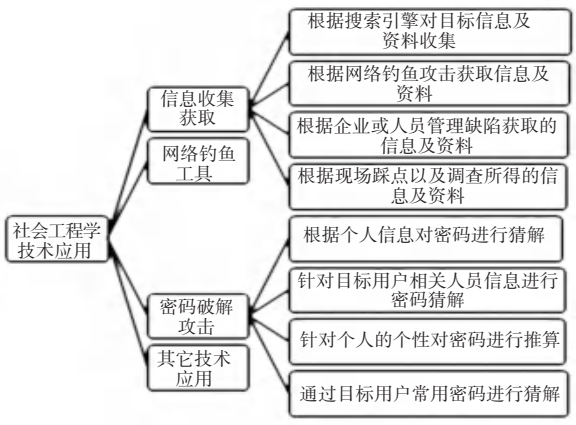

图 1 基于社会工程学的网络安全技术应用梳理[11]

Fig. 1 Carding of network security technology based on social engineering

1.2.2 密码的使用场景

为了更好地把握和界定密码的适用的场景,从物理设备和虚拟系统设置密码的二个角度分析,密码的使用存有 3 种情况:

(1)纯物理介质/机械的设备所使用的密码。例如传统密码锁、保险柜等;

(2)物理介质和“系统”相结合/混合所使用的密码。例如校园卡系统、门禁系统、银行卡系统以及个人电子设备(例如手机、电脑)等;

(3)虚拟系统/网络空间中所使用的密码。例如电子邮箱、即时通信账号、娱乐系统账号、电子商务以及不常用的网站和应用服务等。

综上所述,为了更好地把握和界定密码的适用场合,可将前二种合并,从现实生活和网络空间(系统)中使用的密码二种情况进行分析。

1.2.3 密码设置的分类

网络时代,密码的使用充满了生活的方方面面。

通过对密码进行分类设置，便可有效地应对“撞库”风险，要想很好的对密码掌控，有必要在不同的应用场景中对密码设置的方法进行分类。

（1）从空间的角度分析，将现实生活中的密码（即纯物理介质及其同系统相结合）和网络空间（纯系统）中的密码分开设置；

（2）从资金的角度分析，将涉财事务密码和非财务事务密码分开设置；

（3）从职业角度分析，将办公事务密码和生活应用密码分开设置；

（4）从密码场景的重要性来分析，将不同级别的密码（以复杂度）分级设置；

（5）从平台要求密码的简易程度的角度来分，将不同长度、不同要求的密码分开设置。

1.2.4　密码设置的挑战

1.2.4.1　弱口令

弱口令（weak password）没有严格和准确的定义，通常认为容易被人猜测到或被破解工具破解的口令均为弱口令[12]。通常弱口令包含 3 种情况：

（1）个人信息弱口令，例如身份证后 6 位、生日日期、手机号、门牌号和车牌号等；

（2）传统弱口令，例如 123456、admin、平台名称等；

（3）上述个人信息弱口令和传统弱口令的交叉、排列和组合等方法形成的密码字符串，例如 zhangsan888、1996admin 等。

1.2.4.2　暴力破解

暴力破解（exhaustive attack）或称为“穷举法”，是一种针对密码的破译方法，即将密码进行逐个推算直到找出真正的密码为止[13]。暴力破解依赖于“字典”，字典内含二种数据：

（1）弱口令数据；

（2）纯暴力数据。

纯暴力数据是指，从 1 位到 N 位数据，每一位可由数字、字符（包括大小写）或特殊符号等构成，与用户无关联性。纯暴力数据，常规手段是先生成 6 位以内的数字进行暴力测试，再根据需要生成其他数据。在暴力破解中，弱口令成功的概率远大于纯暴力数据。

1.2.4.3　撞库

撞库，即黑客通过收集网上已泄露的用户名和密码信息，尝试批量登录其他网站，得到一系列可以登录的用户账号[14]。通俗而言，撞库即通过已知的一个平台的账号和密码，利用该账号和密码去尝试登录其他平台。撞库是黑客的一种惯用手段（同撞库相关的还有洗库和脱库），其依赖手中的“社工库”。

1.2.4.4　社工库

社会工程学数据库（Social Engineering Database），简称社工库[15]。社工库是黑客与大数据方式进行结合的一种产物。黑客们将泄露的用户数据整合分析，然后进行集中归档，整理成库。黑客通过入侵有价值的网络站点，盗走用户数据库，这个过程在地下产业术语里被称为“拖库”；黑产人员把多个不同类型的数据库整合成社工库[16]。社工库不仅包括用户的账密信息，还包括相应的额外信息（用户的数据信息和应用信息），通过额外信息可以用来辅助生成弱口令数据。

2　问题提出

2.1　密码设置面临的风险

研究发现，用户在密码设置方面经常犯的两类错误，让很多互联网用户面临风险：

（1）多个账户使用同一个密码。这意味着，一旦有一个账户的密码泄露，则多个账户都可能被破解；

（2）使用容易被破解的弱密码[17]。密码的设置主要为了保护自己所拥有的“资源”，因此密码的设置一定不能设置为弱口令，且必须要足够“健壮”才能防御暴力破解，否则会给试图获取资源的恶意登录者带来可乘之机。同时，密码的设置还要考虑密码泄露后所存在的撞库问题。

2.2　密码保管存在的困境

当今在互联网的大环境下，密码应用环境复杂多变，各类平台设置密码的标准不统一，设置诸多密码难以记忆。研究显示，用户有时候会同他人分享密码，并且使用不安全的手段存放这些密码；有近三分之一（28%）的用户会同家庭成员分享密码，还有十分之一（11%）的用户会同朋友分享自己的密码，很容易造成密码被无意间泄露。超过五分之一（22%）的用户承认自己会在记事本上写下自己的密码，以便记住。这样做的话，即便密码很强，也很容易让用户面临攻击[14]。不安全地密码保管，失去了使用密码的意义。

目前，对于密码的保管，存在以下问题：

（1）记录在物质载体上。例如写在纸上、墙上等，安全系数较低，容易丢失或损害；

（2）纯记忆。除神志不清/不理性状态下，密码绝对安全，但记忆负担太重；

(3)借助于密码管理软件。例如 Clipperz、ncryptr、KeePass 等软件,安全程度适中,但依赖软件平台。

3 解决措施

3.1 基于场景的密码设置方法

欲获得比较健壮的密码,可以将个人信息弱口令作为“原始密码”,进而对其进行加工处理,便可到的较为健壮的密码。

通常情况下,基于场景设置密码时:

(1)要考虑密码的设置规则;

(2)要考虑密码设置的位数限制;

(3)要因不同场景考虑密码设置的“特”点。

前二者为密码设置规则的共性,而第 3 点则是特性。设置规则应与场景相结合。设置对于数字或字母所对应的场景,其可以对自身设定的数字或字母等采取逆序、增加特殊字符等方式进行设定。以 QQ 办公邮箱密码为例,则可以设置为办公电话(假设为 123456)的逆序加‘@’再加“QQ”,最后可设定为“654321@ QQ”。

3.2 基于树状结构的混杂密码设置方法

混杂密码的设置是定义密码设置的结构。对于混杂密码的设置灵感,来源于计算机编程技术“数据结构”中的“树形”结构。树型结构在现实世界中广泛存在,如社会组织机构关系图就可以用树来表示[18]。以淘宝账号密码的设置为例,笔者曾经使用过“@ NM0Ltaobao”作为淘宝密码。其密码设置依据为:重要度用“.(低)”、“!(中)”和“@(高)”表示,空间用“N(network)”和“l(life)”表示,钱财用“M(money)”和“n(null)”表示,是否涉及职业则用“0(生活)”和“1(工作)”表示,简易程度用“L(long)”和“s(short)”表示。为了加强与使用“环境”的相关性,以类型进行区分“P(platform)”和“e(equipment)”,再依据平台,根据其具体功能进行划分“L(login)”、“p(payment)”和“o(other)”;最终加上平台/设备名,即上述密码中的最后 6 位“taobao”,密码构成结构如图 2 所示。

依据该方法设置密码,增加了密码的长度及关联性(便于记忆),从根本上提高了密码的安全层次。但上述结构并非完美,例如对于存在位数限制或者特殊字符限制等情况,图示“结构”则需要进行一些变动。针对位数限制,则需要压缩“层数”;而对于特殊字符限制,则需要重新定义每个“节点”所对应的加密“字符”。针对于此结构给出新的启示,即“树”的左边部分(@ NM0L)可以将每个分支化为“0”和“1”的二进制形式(从上向下 11111),最后转化成一个十进制数(31),这也是一种新的密码设置方法,具体结构如图 3 所示。

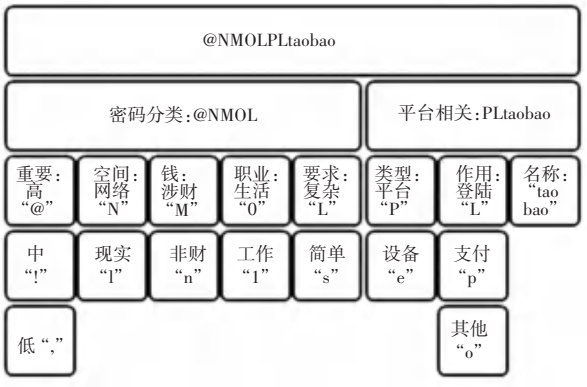

图 2 混杂密码设置结构示例

Fig. 2 An example of a mixed password setting structure

图 3 以二进制形式描述密码分类结构

Fig. 3 Describe password classification structure in binary form

图 3 类似于“摩尔斯电码对照表”。从这一角度来看,混杂密码设置方法,给密码安全设置了两重防线:一是密码本身,二是“密码对照表”的存在。利用该方法设置密码,密码安全等级高,暴力破解难度大,从根本上断绝了撞库风险。由于该方法需要进行一定的规划,且需要设置一个“结构”,同时该结构需要单独保存,因此使用较为复杂。

3.3 基于编码的密码设置方法

基于编码(强度)的密码设置指不依赖使用场景,单纯从复杂(安全)程度来设置密码。基于编码的密码设置涉及位数和字符限制问题,其暴力破解难度为指数级。其中“位数”限定了指数,而“字符限制”则限定了底数。例如 4 位纯数字,暴力破解次数为 10^4($10*10*10*10=10\,000$,每一位有 10 种可能,则 4 位密码有10 000种可能);而对于全字符(假设 10 位数字、52 位字母、5 位特殊字符)4 位密码来说,暴力破解次数为$(10+52+5)^4$($67*67*67*67=20\,151\,121$,每一位有 67 种可能,则 4 位密码有20 151 121种可能),其复杂程度超过纯数字次数约2 015 倍。

上述剖析了位数和字符数的设置所带来的暴力破解的难度。基于此,从密码本身的编码安全来分

析，在考虑密码长度的情况下，分为纯数字密码和组合型密码二种。目前，除了“特殊领域（银行卡、校园卡、门禁等）”使用纯数字，其它使用密码的情形几乎全为组合密码。

3.3.1　纯数字密码

此处的纯数字是指有位数限制、特殊领域的纯数字。例如 4 位、6 位、8 位的密码箱、银行卡、校园卡门禁等情况。对于纯数字密码而言，可对个人常用的密码数字进行以下处理：

（1）逆向设置；

（2）奇偶交换设置；

（3）左右交换方式。

此处提供参考思路，仅列举了 3 种密码设置方法提升密码的安全性。

以校园卡密码设置为例，假设卡号为 19960810 作为原始密码进行变换。采用方法如下：

（1）进行密码设置，则为 01806991；

（2）进行密码设置，则为 91698001；

（3）进行密码设置，则为 08101996。

以上 3 种方法也可以“混用”。具体实例操作见表 2。

表 2　纯数字密码设置示例

Tab. 2　Examples of pure numeric password settings

原始密码	处理方法	最终密码
（八位）19960810	①逆向	01806991
	②奇偶交换	91690801
	③左右交换	08101996
（六位）191215	①处理后 512191； ②处理后 151219； ③处理后	219151

由此得出，采用该方法设置密码，不仅具有关联性（便于记忆），同时极大的提升了密码的安全性。

3.3.2　组合型密码

组合型定义了密码基础强度，包括全字符组合和混合字符组合。全字符组合包括数字、大小写英文、特殊符号；混合字符组合则是全字符组合的一部分。数字（0-9，10 个）、大小写英文（a-z\A-Z，5 二个）、特殊字符（33 个）等；显然易见，暴力破解组合型字符相对于纯数字难度（底数增加）要大的多。

由于常用平台密码设置要求的非统一性，笔者对目前常用平台进行了密码要求的统计，见表 3。其中，要求最简单的是中国铁路 12306 和 163 邮箱等，要求最复杂的是中国及多国专利审查信息查询（其后简称专利查询）。可以看出，只要密码复杂度满足专利查询，即可满足表中所有平台的要求。因此，可以将专利查询的密码要求（密码长度在 8-18 之间，至少有一个数字、小写字母、大写字母、特殊字符）作为组合型密码的基本要求，即密码设置的最低标准，方可满足所有情况下的密码要求。以设置淘宝密码为例，可以设置为“10@ Taobao”，1 是涉财、0 是登录、@ 特殊符号间隔、Taobao 对应平台，4 个部分可以打乱排序；再以设置校园卡密码和学生系统密码为例，可以分别设置为“School @ 1”、“School@ 0”等。

此处的组合型密码不同于混杂密码设置，组合型密码注重的是满足平台要求，而混杂密码则注重的是按分类设置，二者都应在满足各自特点的情况下提升密码强度并提高可记忆性。

表 3　常用平台密码设置要求统计

Tab. 3　Statistics of common platform password setting requirements

平台名称	密码设置要求
中国铁路 12306	6-20 位字母、数字或符号
淘宝	6-20 位字符；只能包含大小写字母、数字以及标点符号（除空格）；大写字母、小写字母、数字和标点符号至少包含二种
京东	建议使用字母、数字和符号两种以上的组合，8-20 个字符
中国知网	请输入 6-20 位数字、字母或常用符号，字母区分大小写
读秀	密码要求 6-16 位，至少包含数字、字母、符号两种元素
QQ	不能包括空格，长度为 8-16 个字符，必须包含字母、数字、符号中至少二种
163 邮箱	6-16 个字符，区分大小写
中国及多国专利审查信息查询	密码长度在 8-18 之间，至少有一个数字、小写字母、大写字母、特殊字符（~，!，@，#，$，%，^，&，*）

3.4　基于扩散和混淆的密码设置方法

以上方法可以帮助人们理清思路，以较好的方式解决日常生活中设置密码的问题。但是，对于密码的设置需要一个科学的理论依据和指导。1949 年美国数学家、信息论的创始人克劳德·艾尔伍德·香农（Claude Elwood Shannon）发表了《Communication Theory of Secrecy Systems》（保密系统的通信理论），提出了混淆（confusion）和扩散（diffusion）两大设计原则，为对称密码学（发送者的加密密钥和接收者的解密密钥相同或容易相互导出的密码体制）建立了理论基础。扩散和混淆的目的是为了对抗对手对密码体

质的统计分析[3]。扩散是指:明文中的每一位影响密文中的许多位,或者说让密文中的每一位受明文中的许多位的影响,以屏蔽明文的统计特性;混淆是指:将密文与密钥之间的统计关系变得尽可能复杂,使得对手即使获得了关于密文的一些统计特性,也无法推测密钥。通过对上述概念的分析,笔者认为,对于“口令”而言,可以借鉴扩散和混淆并适应到“口令”的设置之中,以防范撞库攻击或社工攻击中的密码分析。

无论是基于场景还是基于编码,皆是对于密码字符串的设置,因此可以借鉴混淆和扩散两大原则,加之同时采用古典密码体系中的方法,则可以实现产生较好的密码设置效果。对于密码的设置研究,主要是脱离弱口令、增加暴力破解难度,防止密码分析带来的撞库问题;通过上述扩散和混淆的概念,我们可以利用扩散增加位数、字符数,以脱离弱口令、增加暴力破解难度;以混淆来减小不同平台下的密码之间的统计关系,见表 4。

表 4 基于扩散和混淆的密码设置示例

Tab. 4 Example of password setting based on spreading and obfuscation

原始密码	121722	
扩散	方法一:122177222 (偶数位重复;增加位数)	方法二:e@ e7@ @ (代换密码:1 代换成 e,2 代换成@;替换符号增加字符量)
混淆	122 177 222 (一行三位)	e@ e7 @ @ (一行两位)
最终密码	112272272(混淆竖排写出)	ee@ @ 7@(混淆竖排写出)

3.5 自建密码设置模型

自建密码设置模型兼顾场景和强度。先以场景进行分类,标注不同程度的密码层级,进而设置不同强度的密码。按场景重要程度可分为 3 类:涉财类、常用类、普通类。为了便于记忆,3 种类型皆以最高强度设置作为密码设置要求,即“密码长度在 8-18 之间,至少有一个数字、小写字母、大写字母、特殊字符。

首先规范密码格式:以平台名称+平台功能+特殊符号+数字格式为例。为便于记忆,平台名称应明确界定中文字数。例如,阿里、奇艺、付宝,也可以自定义汉字个数(按重要程度设置平台字数,若平台名称不足的可以自定义汉字进行填充)。平台名称可以用“驼峰命名法”进行拼音、英文拼写,也可以自定义拼写方法(例如汉字拼音的最后一个字母大写或者只是最后一个字母小写)。平台功能一般分为二种:

(1)功能(登录、支付、独立密码等);

(2)类型(例如网易邮箱、网易云音乐、网易游戏等)。

特殊符号和数字则可以自定义。以支付宝、网易和百度 3 个平台密码设置为例,见表 5。

从表 5 可以看出,3 种重要程度不同的平台,为了便于记忆格式上相同,但是各平台的最终密码却大相径庭。使用该方法设置密码,只需要牢记自定义的规则便可轻松实现对各平台密码的“掌控”。

表 5 自建密码模型样例

Tab. 5 Sample self-built password model

平台	程度	名称	功能	特殊字符	数字	最终密码
支付宝	涉财	fuBao	L(登陆)	$	88	fuBaoL $ 88
网易	普通	WangYi	G(游戏)	*	66	WangYiG * 66
百度	常用	baIdU	C(云盘)	^	55	baIdU^55

3.6 二次处理加记录的密码保管方法

对于密码的保管,笔者认为,较合适的方法是二次处理+平台/设备记录。二次处理类似于加密,但又不同于加密;以密码 taoBaoL $ 6991 为例,描述详见表 6。

表 6 密码保存之二次处理示例

Tab. 6 Example of secondary processing for password saving

原始密码	二次处理后	变化方式和特点
taoBaoL $ 6991	1996 $ LtaoBao	日期正序,整体逆序;位置变化
	taoL $ 19	平台名称和数字简写;简写密码
	taobaologin $ 1996	大小写有变化,登陆英文全写,日期正序,格式不变;扩充密码描述

对于平台/设备记录,则是直接将二次处理后的密码写入到移动设备(手机、平板)中的便签或记事本等相似的系统中,亦或者写入到网络平台中。采用上述方法保管/记忆密码,不仅可以加强密码设置的记忆,同时提升密码存储层级(不泄露、便于记忆)。

在此,不建议浏览器插件记录密码,虽然快捷,但不便于跨设备,且存在他人利用该设备登陆平台的风险。

4 结束语

密码的设置看似简单,实则是一门方法论,设置的手段和方法至关重要。本文首先是对研究对象进行阐述加以界定,并描述其研究背景。从社会工程学的视角出发,分析密码设置的挑战和保管的困境。根据上述问题,提出 3 种密码设置方法,并建立密码

(下转第 70 页)