doi :10.3969/j.issn.1671-1122.2019.08.001

基于主题 PCFG 的口令猜测模型研究

毕红军[1]，谭儒[1,2]，赵建军[2,3]，李昱甫[2]

（1. 北京交通大学电子信息工程学院，北京 100044；2. 中国科学院信息工程研究所，北京 100093；
3. 中国科学院大学网络空间安全学院，北京 100049）

摘　要：口令是一种重要的身份认证方式，用户为了能够方便记住口令，常把一些与人相关的要素信息加入口令中。传统基于概率上下文无关文法（PCFG）算法进行的口令安全评估，并没有关注用户兴趣爱好、文化背景等与人相关的主题因素。文章基于传统 PCFG 算法，重点针对口令字母字段进行分析研究，通过对所收集的数据库字母字段的对比，提取用户口令与主题的关系，进而提出基于主题 PCFG 的口令猜测模型——T-PCFG 模型。文章围绕收集的 7 个数据库中的 3300 万口令数据集进行实验，结果显示，主题为兴趣爱好时口令的猜测成功率比普通口令的猜测成功率高 2.37~8.2 个百分点。

关键词：概率上下文无关文法；口令；主题；口令猜测；口令安全

中图分类号：TP309　**文献标识码：**A　**文章编号：**1671-1122（2019）08-0001-07

中文引用格式：毕红军，谭儒，赵建军，等 · 基于主题 PCFG 的口令猜测模型研究 [J]· 信息网络安全，2019，19（8）：1–7·

英文引用格式：BI Hongjun, TAN Ru, ZHAO Jianjun, et al· Research on Password Guessing Model Based on Theme PCFG[J]· Netinfo Security, 2019, 19(8): 1–7·

Research on Password Guessing Model Based on Theme PCFG

BI Hongjun[1], TAN Ru[1,2], ZHAO Jianjun[2,3], LI Yufu[2]

(1· *School of Electronic and Information Engineering*, *Beijing Jiaotong University*, *Beijing* 100044, *China*; 2· *Institute of Information Engineering*, *Chinese Academy of Sciences*, *Beijing* 100093, *China*; 3· *School of Cyber Security*, *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*)

Abstract: Password is an important method of identity authentication· In order to be able to remember passwords conveniently, users often add some related information about people to passwords· Traditional password security assessment based on probabilistic context free grammar(PCFG) does not pay attention to user-related subject factors such as user hobbies and cultural backgrounds· Based on the traditional PCFG algorithm, this paper focuses on the analysis of the password letter field· By comparing the collected database letter fields, the relationship between the user password and the subject is extracted, and then the password guessing model based on the theme PCFG is proposed T-PCFG model· The article carried out experiments on the 33 million passwords collected from the seven databases· The results show that when the subject is a hobby, the success rate of password guessing is 2·37~8·2 percentage points higher than the normal one·

Key words: PCFG; password; theme; password guess; password security

收稿日期：2019-5-18

基金项目：北京市科技计划 [D181100000618002]

作者简介：毕红军（1965—），男，北京，副教授，硕士，主要研究方向为网络安全、信息网络、电子与通信工程；谭儒（1992—），男，湖南，硕士研究生，主要研究方向为网络安全；赵建军（1990—），男，甘肃，博士研究生，主要研究方向为 Web 安全、口令猜解；李昱甫（1997—），男，河南，本科，主要研究方向为 Web 安全、大数据分析。

通信作者：赵建军 zhaojianjun@iie.ac.cn

0 引言

随着网络的发展，网络安全和网络威胁问题日趋严重。为了防止系统被攻击，越来越多的身份认证方法被提出，如多因子认证[1]、生物认证[2]等。虽然这些认证方式具有一定的可用性，但是由于用户自身原因，文本口令因更具实用性而被广泛使用。

由于人的大脑一般只能记忆5~7个口令[3]，使得一些用户选择了一些容易被猜测的口令[4]，这种情况增加了攻击者破解口令的成功率。为了解决用户使用弱口令的行为，不同的系统会根据系统特性来强迫用户制定口令，这样的方法称为口令生成策略，不同系统的口令生成策略是不同的，具体根据该系统的口令强度评测器[5]（Password Strength Meters, PSM）决定。PSM在口令生成策略中具有检测作用，帮助用户识别口令的安全等级。用户可以根据PSM的分析结果检测自己制定的口令的安全等级。然而，因为数据不足、安全等级划分标准不统一等问题，PSM也存在很多缺陷。深入了解用户如何构造口令可以更好地评估口令强度。

随着时间的推移，系统制定的口令生成策略越来越严格，并且要求用户制定强口令。例如，必须是数字+特殊符号+字母的组合 ；必须有大小写字母交替出现 ；不允许使用当前系统认定的弱口令。

一个口令的强弱，通常由该口令使用的次数决定。例如，弱口令“ 123456 ”，因其字符较短，成为了很多人使用的口令。但是成为众人使用的口令不仅仅是因为它简短、可记忆性强等原因，还与个人习惯如习俗、文化、兴趣等有关。因此，一个复杂的口令也可能成为弱口令。一项研究[6]表明，复杂口令“ ji32k7au4a83 ”也是一个弱口令。研究结果表明，在中国台湾键盘上键入字母J，然后键入字母I，将添加两个符号（ㄨ+ㄛ），发音为u和o（显示在键的右上方），形成wo。然后用户输入字符的第三声音调，因此是3。“ ji3 ”代表“ wǒ ”，也就是“ 我 ”。以此类推，“ 2k7 ”代表“ 的 ”，“ au4 ”代表“ 密 ”，“ a83 ”代表“ 碼 ”。所以，“ ji32k7au4a83 ”这个看起来安全的“ 强 ”口令在中国台湾网友的习惯中就是“ 我的密码 ”的意思，被用户作为口令广泛使用。

随着口令研究的发展，越来越多的口令猜测模型被提出，如*N*阶马尔科夫模型（Markov *N*-grams）[6,7]、概率上下文无关文法模型（Probabilistic Context Free Grammar，PCFG）[8,9]等。这些模型打破了传统的启发式的方式，将口令研究带入了一个新的阶段，为近几年口令研究模型奠定了基础。虽然这些口令模型在实用性上比传统启发式模型的猜测方式更好，但是依然存在很多缺陷。这些口令猜测模型在评估过程中忽略了用户在构造口令时的一些因素，如习俗、文化、兴趣等，且在实际应用中使用的统计方法需要花费大量的时间和存储空间。所以这些模型在使用过程中的可用性有待提高。

为了更好地解决这些问题，本文从用户个人兴趣爱好出发，在大规模数据集上对个人兴趣与口令的相关性进行分析，提出了基于主题PCFG的口令猜测模型——T-PCFG模型。

1 相关工作

口令的安全性能大致分为两个类别[10-12]，第一类是口令分布的安全性，第二类是口令自身的安全性。第一类可以通过攻击算法和统计学方法评定，第二类只能通过攻击算法评定。当前破解成功的猜测次数是最好的评估指标。

现阶段口令猜测攻击算法分为漫步口令猜测攻击算法和定向口令猜测攻击算法，区别在于攻击时是否利用个人信息。下面介绍当前最流行的两个漫步口令猜测攻击算法。

1）基于PCFG的算法

WEIR[13]等人于2009年提出了基于PCFG的漫步口令猜测攻击算法，将一个口令分成了*L*字段（字母）、*D*字段（数字）、*S*字段（特殊字符），对每一个字段的长度进行计数表示，且假设各字段之间是相互独立的，最后得到一张结构表和一张字段表。例如，口

令“ zhu$*000 ”被分成$L_3$（该字符有3个字母“ zhu ”）、$S_2$（该字符有2特殊字符“ $* ”）、$D_3$（该字符有3个数字“ 000 ”），则该口令的结构就是$L_3S_2D_3$。

该算法分为训练阶段和猜测阶段，训练阶段通过统计计算得到的各种口令结构对应的频率及各字段组成对应的频率，得到相应的结构频率表和结构中字段L、S、D的频率表。例如，对于口令“ zhu$*000 ”，统计出训练集中全部口令结构中$L_3S_2D_3$的频率，以及“ zhu ”在长度为3的$L$字段中的频率，“ 000 ”在长度为3的$D$字段中的频率，“ $* ”在长度为2的$S$字段中的频率。假设频率$P(L_3S_2D_3)=0{\cdot}3$，$P(L_3\rightarrow zhu)=0{\cdot}1$，$P(D_3\rightarrow 000)=0{\cdot}2$，$P(S_2\rightarrow \$*)=0{\cdot}1$，那么口令“ zhu$*000 ”的可猜测率$P(zhu\$*000)=P(L_3S_2D_3)\times P(L_3\rightarrow zhu)\times P(D_3\rightarrow 000)\times P(S_2\rightarrow \$*)=0{\cdot}0006$，表示“ zhu$*000 ”可被破解的概率为0·0006。

通过该方法可以计算出训练集中每个口令的频率，然后按频率大小进行排序，得到一个概率递减的猜测集。最后在测试阶段使用这个猜测集对测试集进行破解测试，得到口令破解的成功率，从而对模型的有效性进行评估。

2）基于Markov的算法

NARAYANAN[14]等人于2005年将Markov链技术引入口令的研究中。基于Markov算法的漫步猜测攻击模型认为用户会按照一定的先后顺序构造口令，以口令前后字符的关联性为基础，通过统计的方法得到口令对应的概率值。该算法对整个口令进行训练，通过字符之间的联系计算破解口令的概率。Markov有阶的概念，N阶Markov模型需要统计出N个字符之后紧跟着的字符频数。在测试阶段，Markov模型可以计算出每个口令的频率，然后按频率大小进行排序，得到一个猜测集。最后使用猜测集对测试集进行口令破解实验，得到口令猜测成功概率。

本文基于主题的漫步口令猜测模型，通过对L字段的研究，将传统方法中L字段提取方法进行修改，组成新的猜测集，然后通过得到的猜测集进行实验。

2 T-PCFG

2.1 理论基础

2.1.1 口令强弱相对性

口令由人生成，与人的行为习惯产生直接关系，每个人的行为习惯根据内因和外因的不同千差万别。一个口令，在一个系统中是强口令，但是在另一个系统中可能就是弱口令。例如，一个人很喜欢访问天涯论坛，则访问论坛的口令中可能存在“ tianya ”字段，那么该口令在天涯系统中可能就是一个弱口令，但是在Acfun中很可能是一个强口令。

一个用户往往需要管理几十个甚至上百个帐户口令[15,16]，其对于系统安全不够重视[17]可能导致存在一些弱口令行为。本文使用的7个数据库及简介如表1所示。

表 1 7 个数据库及简介

数据库	类别	数据量	数据类别	简介
游戏库 2	兴趣库（游戏）	23855063	明文	游戏论坛，主要内容有魔兽世界等
游戏库 1	兴趣库（游戏）	436346	明文	官方网站，主要内容有守望先锋等
动漫库 2	兴趣库（动漫）	1507023	MD5	动漫网站，主要内容有恶魔岛等
动漫库 1	兴趣库（动漫）	152788	明文	动漫网站，主要内容有动漫视频等
普通库 1	普通库	129303	明文	网络订票网站
普通库 2	普通库	700023	明文	网盘系统
普通库 3	普通库	6871375	明文	婚恋网站

2.1.2 常用口令分析

大量研究[18-20]表明，用户除了使用单词作为口令外，还常常将单词进行简单的变换，以满足网站口令设置策略的要求。例如，“ zhoulan123 ”可以满足“ 字母＋数字 ”的要求。本文对表2的6个数据库中常用的口令进行分析（动漫库2的数据已经过哈希处理，所以不参与比较）。由表2可以看出 ：1）大多数口令包含数字信息，且大部分口令由纯数字组成。2）“ 123456 ”作为最习惯使用的口令，依然居于弱口令之首，大部用户还是习惯于使用该口令。如果一个攻击者单纯使用“ 123456 ”攻击一个系统，也会对该系统造成一定影响，给系统的安全性带来很大挑战。3）大部分用户习惯使用类似于关于情感的口

令，如“ woaini ”、“ woaini1314 ”、“ 5201314 ”等。4）有些用户习惯使用“ 字母＋数字排序 ”的方式，如“ a123456 ”、“ qq123456 ”。5）一些用户习惯按照键盘的顺序进行口令设置，如“ 1qaz2wsx ”、“ asdasd ”。

表 2 数据库中常用的前 10 个口令

排名	动漫库 1	游戏库 1	普通库 1	游戏库 2	普通库 2	普通库 3
1	123456	woaini1314	123456	123456	123456	123456
2	111111	a123456789	a123456	123456789	111111	123456789
3	000000	caonima	5201314	111111	123456789	111111
4	123123	asd8814520	123456a	123123	000000	000000
5	123456789	123456789a	111111	000000	123123	5201314
6	lucifer	abcd1234	woaini	5201314	5201314	123123
7	5201314	qq123456	123123	12345678	123	1314520
8	1234567	qqq7758521	000000	12345	123321	123321
9	123321	1q2w3e4r	qq123456	123321	12345678	1234567
10	asdasd	Woaini521	1qaz2wsx	1314520	7758521	12345678
占比	5·80%	0·61%	1·79%	8·80%	7·87%	6·59%

对表2中6个数据库进行分析，约有0·61%~8·80%的用户使用排名前10的口令，如果使用这些口令破解这些数据库，则有0·61%~8·80%的成功率。口令的字符组成结构和长度受口令生成策略影响。例如，CSDN要求口令长度不少于8位，Myspace要求口令必须由“ 字母＋数字 ”组成。对于普通网站来说，如果没有设置长度要求，大约90%的口令长度在6~11位之间[21]。

2.1.3 兴趣爱好对口令的影响

用户的口令设置受到生活经历的影响。本文从字母字段（*L*字段）入手，对比了3个兴趣库和3个普通库的*L*字段。首先将所有数据库的*L*字段提取出来分别放入不同的文档中，接着将其中一个兴趣库的*L*字段和3个普通库的*L*字段分别进行比较，最后提取出兴趣库独有的*L*字段。结果如表3所示。

通过对表3的分析可以看出，用户的兴趣爱好对其设置口令具有一定的影响，为接下来制定T-PCFG模型奠定了理论基础。

2.2 T–PCFG 模型

本文专注于个人兴趣爱好对口令结构影响的研究，在传统漫步猜测攻击算法的基础上，提出了一个关于兴趣信息的漫步猜测攻击算法。

首先将一个数据库中的数据随机打乱，均分成

表 3 部分与兴趣有关的 *L* 字段

字段	解释	数据库
dearbook	CSDN 自由品牌“ 第二书店 ”默认账号	游戏库 1
dedewang	CSDN 博主	游戏库 1
ARPG	Action Role Playing Game，游戏术语，动作角色扮演类游戏，角色的动作（特别是攻击动作）与操作（如点击鼠标）相关	游戏库 2
SCV	Space Construction Vehicle，空间建筑工程车，俗称农民、工人或工程兵，是星际争霸系列游戏中兵种的名称	游戏库 2
acfun	A 站，一个关于动漫的视频网站	动漫库 1
qrst	游戏论坛	游戏库 2
Heman	Alborz Haidarian，欧洲最好的 War3 选手之一	游戏库 1
moe	动画片《辛普森一家》中的人物，荷马的朋友，经营一个小酒吧	动漫库 1
onep	一个关于海贼王动漫中的词	动漫库 1
xiaoy	魔兽争霸 WAR3 游戏解说	游戏库 2

训练集和测试集两部分。在训练集中，按照PCFG算法提取*L*字段、*S*字段和*D*字段，分别代表字母、特殊符号和数字，并且提取出*LSD*结构频率表、*S*字段频率表和*D*字段频率表。然后根据兴趣库的类别将兴趣库进行划分，取其中一类兴趣库加入一个实验组，并提取其*L*字段。将从兴趣库提取的L_T字段（为了将T-PCFG模型和传统PCFG模型进行区分，本文将提取的*L*字段称为L_T字段）作为字典，替换训练集中的*L*字段。

对于生成的猜测集，假设数据库1生成的结构表中按频率大小排列为L_1+“ 123 ”、L_3+“ 520 ”等。根据频率最大的组合L_1+“ 123 ”在生成的字典中寻找L_1段所有字符。例如，L_1的字符是“ a ”“ n ”等。那么生成的猜测集为“ a123 ”“ n123 ”等，只有当L_1+“ 123 ”组合完后，才会轮到L_3+“ 520 ”组合，以此类推，最后生成猜测集。

本文构造的T-PCFG模型主要分为数据准备阶段、训练阶段、猜测阶段3部分。将其中一个库分成两部分，一部分数据进行训练，另一部分作为测试集，得到结构频率表、*D*字段频率表、*S*字段频率表和测试集。兴趣库提供L_T字段。根据结构频率表、*D*字段频率表、*S*字段频率表、L_T字段表生成猜测集。用生成的猜测集与测试集进行验证，最后得到结果。T-PCFG模型如图1所示。

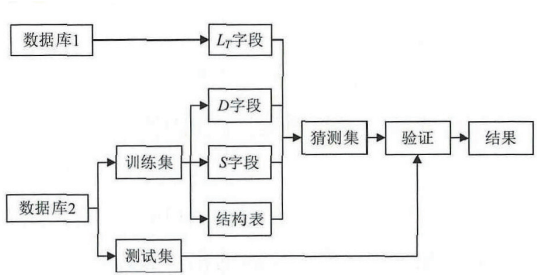


图 1 T–PCFG 模型

3 实验及分析

3.1 实验准备

本文使用两类兴趣库，一类是网络游戏类，一类是动漫类。由图1可知，一个实验组需要两个数据库，一个是提取L_T字段的兴趣库，一个是提取结构频率表和D、S字段频率表的对照库。通过大量的实验发现，口令生成规则不同，不同的数据库生成的结构频率表也不同。基于此，本文对不同的实验组进行不同的实验，然后将实验进行对比。

本文进行两个实验，第一个实验针对游戏类进行，用游戏库1提取L_T字段信息。第一个实验有4组，第1组将游戏库1提取的L_T字段作为字典信息，将游戏库2分成两部分，根据T-PCFG模型进行实验。第2组是游戏库1和普通库2的组合，第3组是游戏库1和普通库3的组合，第4组是游戏库1和普通库1的组合。

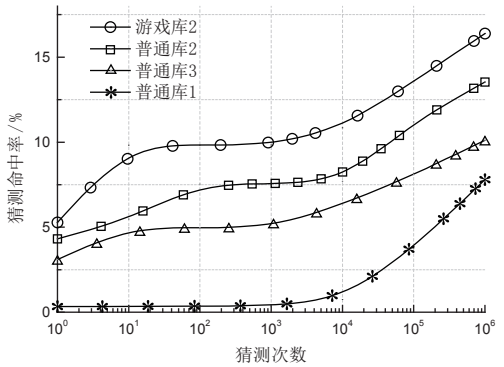
第二个实验针对动漫类进行。因为动漫库2数据库被MD5处理过，所以用动漫库1的L_T字段提取字典，分别和普通库1、普通库2、普通库3进行组合。虽然动漫库2数据被MD5处理无法获得L_T字段，但依然将其列入每个实验组。各组数据如表4所示。

表 4 实验组数据

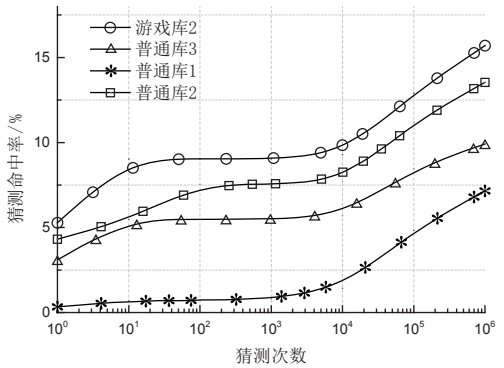
	提取 L_T 字段	D、S 字段频率表和结构频率表	测试集
a	游戏库1（游戏类）	游戏库 2（训练集，游戏类）	普通库 1（测试集，普通类），普通库 2（测试集，普通类），普通库 3（测试集，普通类），游戏库 2（测试集，游戏类）
b		普通库 1（训练集，普通类）	
c		普通库 2（训练集，普通类）	
d		普通库 3（训练集，普通类）	
e	动漫库1（动漫类）	普通库 1（训练集，普通类）	普通库 1（测试集，普通类），普通库 2（测试集，普通类），普通库 3（测试集，普通类），动漫库 2（动漫类）
f		普通库 2（训练集，普通类）	
g		普通库 3（训练集，普通类）	

3.2 实验结果

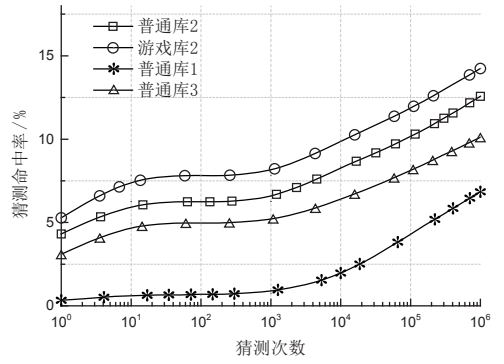
将7个数据库划分为两组进行实验。划分的依据是个人兴趣，因为不同的兴趣会对口令生成产生不同影响。为了评估T-PCFG的性能，根据表4分别对不同数据库进行实验，实验结果如图2所示。由图2可知，本文提出的T-PCFG模型在相同的猜测条件下，兴趣库的成功率高于普通库0·16~1·02倍，说明用户兴趣爱好对用户的口令构造有影响，也证明了T-PCFG模型的有效性。



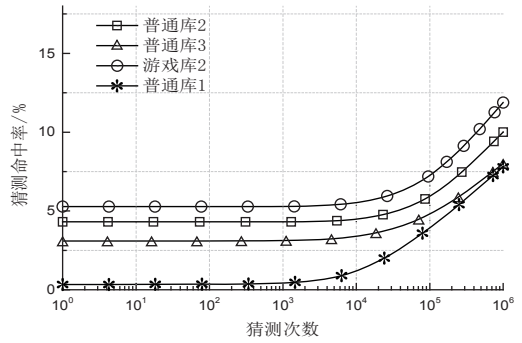
a）游戏库 1 和游戏库 2 生成的猜测集



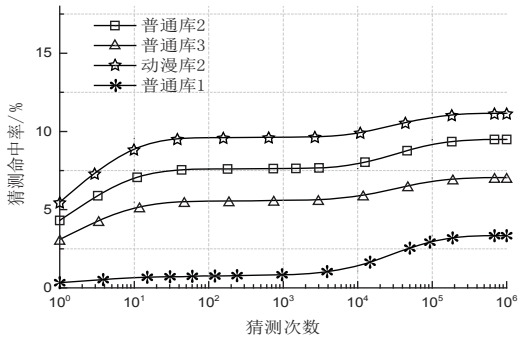
b）游戏库 1 和普通库 2 生成的猜测集



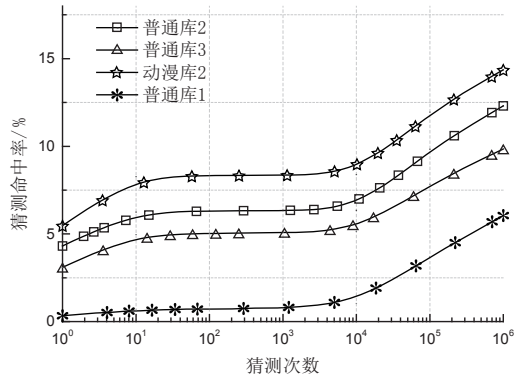
c）游戏库 1 和普通库 3 生成的猜测集



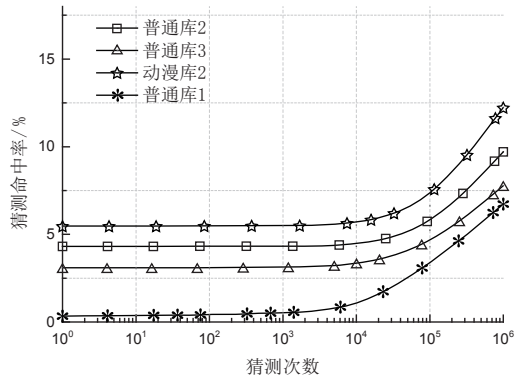
d）游戏库 1 和普通库 1 生成的猜测集



e）动漫库 1 和普通库 2 生成的猜测集



f）动漫库 1 和普通库 3 生成的猜测集



g）动漫库 1 和普通库 1 生成的猜测集

图 2 实验结果

3.3 实验分析

本文选取WEIR[13]等人漫步猜测攻击算法中的一组实验与本文算法进行对比，该组实验生成的字典口令数68611个，生成的口令猜测集300万个左右，使用的测试集33481个，猜测成功率14%左右。表5为本文实验与WEIR[13]等人实验的相关数据。

表 5 WEIR 等人实验与本文实验相关数据

实验		字典口令 / 个	猜测集 / 万个
WEIR[13] 等人		68611	300
实验一	a	72890	
	b		
	c		
	d		
实验二	e	78003	
	f		
	g		

由表5可知，在对兴趣库测试集进行实验时，最高可比PCFG模型高2个百分点，验证了T-PCFG模型的实用性。

图3为WEIR[13]等人实验结果与本文实验结果的对比。本文7组实验结果中只取主题和兴趣相关的数据。

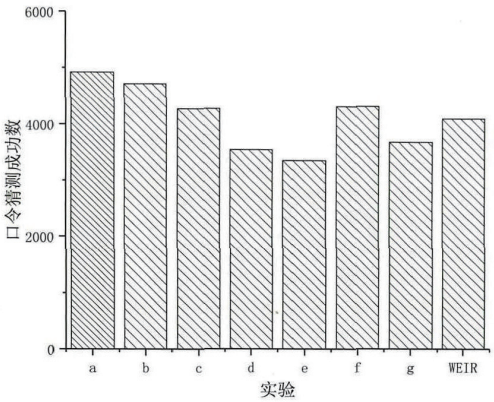


图 3 WEIR[13] 等人实验结果与本文实验结果对比

由图3可知，在本文算法中，当普通库2和普通库1作为对照库时，口令猜测成功率比WEIR[13]等人算法低，可能是因为普通库2和普通库1的口令生成规则相对严格。

4 结束语

本文从公开渠道获得7个数据库，对其中6个数据库进行数据分析和处理，经过常用口令分析，发

现兴趣爱好对口令有影响。通过对所收集的数据库字母字段的对比，提取用户口令与主题的关系，进而提出T-PCFG模型。为了验证T-PCFG模型的有效性，本文进行了两组实验。对比兴趣库和普通库在相同情况下猜测成功概率，实验结果显示，在主题为兴趣爱好时，口令的猜测成功率比普通的高2·37~8·2个百分点。为了验证实验的可靠性，本文抽取了有关兴趣爱好的实验数据与PCFG模型进行比较，结果显示，在相同条件下，最高可比PCFG模型高2个百分点，验证了模型的有效性和可行性。接下来将基于本文的研究成果，将用户兴趣爱好信息应用于定向口令猜测算法中，构造基于兴趣爱好信息的定向口令猜测模型。（责编 潘海洋）

参考文献：

[1] WANG Ding, WANG Nan, WANG Ping, et al. Preserving Privacy for Free: Efficient and Provably Secure Two-factor Authentication Scheme with User Anonymity[EB/OL]. https://www.sciencedirect.com/science/article/pii/S0020025515002431, 2018-12-11.

[2] BIDDLE R, CHIASSON S, VAN OORSCHOT P C. Graphical Passwords: Learning from the First Twelve Years[J]. ACM Computing Surveys, 2012, 44(4): 1-41.

[3] KEITH M, SHAO B, STEINBART P J. The Usability of Passphrases for Authentication: An Empirical Field Study[J]. International Journal of Human-computer Studies, 2007, 65(1): 17-28.

[4] BONNEAU J. The Science of Guessing: Analyzing an Anonymized Corpusof 70 Million Passwords[C]//IEEE. 2012 IEEE Symposium on Security and Privacy, May 20-23, 2012, San Francisco, CA, USA. New Jersey: 2012: 538-552.

[5] UR B, KELLY P G, KOMANDURI S, et al. How Does Your PasswordMeasure Up? The Effect of Strength Meters on Password Creation[C]//ACM. 21st USENIX Conference on Security symposium, August 8-10, 2012, Bellevue, CA, USA. New York: ACM, 2012: 65-80.

[6] MELICHER W, UR B, SEGRETI S M, et al. Fast, Lean, and Accurate: Modeling Password Guessability Using Neural Networks[J]. Journal of Networks, 2013, 8(6): 1278-1284.

[7] VAITHYASUBRAMANIAN S, CHRISTY A. A Scheme to Create Secured Random Password Using Markov Chain[J]. Artificial Intelligence and Evolutionary Algorithms in Engineering Systems, 2014, 2(11): 809-814.

[8] VERASR, COLLINS C, THORPE J. On the Semantic Patterns of Passwords and Their Security Impact[EB/OL]. https://www.docin.com/p-1542143434.html, 2018-12-12.

[9] NEDERHOF M J, SATTA G. Computing Partition Functions of PCFGs[J]. Research on Language & Computation, 2009, 7(2-4): 233-233.

[10] MA J, YANG Weining, LUO Min, et al. A Study of Probabilistic Password Models[C]//IEEE. 2014 IEEE Symposium on Security and Privacy, May 18-21, 2014, San Jose, CA, USA. New Jersey: IEEE, 2014: 689-704.

[11] WAGN Ding, CHENG H, GU Q, et al. Understanding Passwords of Chinese Users: Characteristics, Security, and Implications[EB/OL]. http://wangdingg.weebly.com/uploads/2/0/3/6/20366987/chinesepassword_full_v7.pdf, 2018-12-22.

[12] SHAY R, BAUER L, CHRISTIN N, et al. A Spoonful of Sugar?: The Impact of Guidance and Feedback on Password-creation Behavior[C]//ACM. 33rd Annual ACM Conference on Human Factors in Computing Systems, April 18-23, 2015, Seoul, Republic of Korea. New York: ACM, 2015: 2903-2912.

[13] WEIR M, AGGARWAL S, MEDEIROS B D, et al. Password Cracking Using Probabilistic Context-free Grammars[C]//IEEE. 30th IEEE Symposium on Security and Privacy, May 17-20, 2009, Berkeley, CA, USA. New Jersey: IEEE, 2009: 391-405.

[14] NARAYANAN A, SHMATIKOV V. Fast Dictionary Attacks on Passwords Using Time-space Tradeoff[C]//ACM. 12th ACM Conference on Computer and Communications Security, November 7-11, 2005, Alexandria, VA, USA. New York: ACM, 2005: 364-372.

[15] FLORENCIO D, HERLEY C. A Large-scale Study of Web Password Habits[C]//ACM. 16th International Conference on World Wide Web, May 8-12, 2007, Banff, Alberta, Canada. New York: 657-666.

[16] WANG Ding, CHENG Haibo, WANG Ping. Understanding Passwords of Chinese Users: A Data-driven Approach[EB/OL]. http://bit.ly/2maZLCd, 2019-1-10.

[17] FURNELL S. An Assessment of Website Password Practices[J]. Comput Secur, 2007, 26(7): 445-451.

[18] LIU Gongshen, QIU Weidong, MENG Kui, et al. Password Vulnerability Assessment and Recovery Based on Real Data Mining[J]. Chinese Journal of Computers, 2016, 39(3): 454-467.
刘功申，邱卫东，孟魁，等. 基于真实数据挖掘的口令脆弱性评估及恢复 [J]. 计算机学报，2016，39（3）：454-467.

[19] LI Zhigong, HAN Weili, XU Wenyuan. A Large-scale Empirical Analysis on Chinese Web Passwords[C]//ACM. 23rd USENIX conference on Security Symposium, August 20-22, 2014, San Diego, CA. New York: ACM: 559-574.

[20] DAILEY D V, MARKUS DÜRMUTH, PAAR C. Statistics on Password Re-use and Adaptive Strength for Financial Accounts[C]// Springer. 2014 International Conference on Security and Cryptography for Networks, September 3-5, 2014, Amalfi, Italy. Heidelberg: Springer, 2014: 218-235.

[21] YOU Lin, LIANG Jiahao. Research on Secure Identity Authentication Based on Homomorphic Encryption and Biometrics[J]. Netinfo Security, 2018, 18(4): 7-14.
游林，梁家豪. 基于同态加密与生物特征的安全身份认证研究 [J]. 信息网络安全，2018，18（4）：7-14.