



(12)发明专利申请

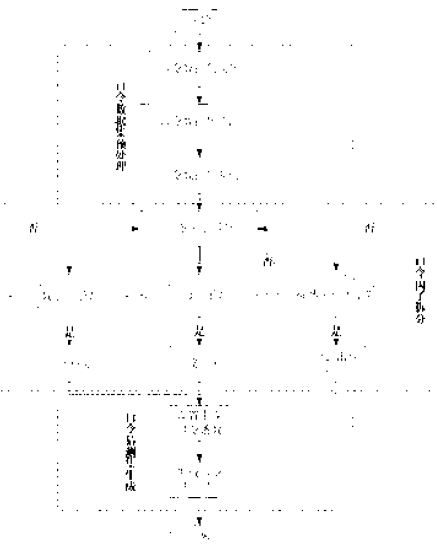
(10)申请公布号 CN 111191008 A
(43)申请公布日 2020. 05. 22

(21)申请号 201911407189.5
(22)申请日 2019.12.31
(71)申请人 华东师范大学
地址 200241 上海市闵行区东川路500号
(72)发明人 何道敬 周贝贝 陆城
(74)专利代理机构 上海蓝迪专利商标事务所
(普通合伙) 31215
代理人 徐筱梅 张翔
(51)Int.Cl.
G06F 16/33(2019.01)
G06F 16/36(2019.01)

权利要求书1页 说明书3页 附图1页

(54)发明名称
一种基于数字因子逆序的口令猜测方法

(57)摘要
本发明公开了一种基于数字因子逆序的口令猜测方法,包括数据清洗阶段、口令因子拆分阶段和口令猜测阶段;其中,口令数据集预处理:收集不同口令集,对口令集数据进行清洗及编码;口令因子拆分:将口令按照数字、字母和特殊字符分段为数字因子、字母因子和特殊字符因子,并统计各因子包含的内容出现的频率;口令猜测集生成:基于口令因子分析阶段得到的各口令因子出现频率,设置生成口令的条数,生成口令猜测集合。本发明对口令因子拆分分析,并逆序分析数字因子,实验结果显示逆序猜测的口令准确度较正序更好,使用本方法可以生成离线口令字典库,用于离线口令猜测。



1. 一种基于数字因子逆序的口令猜测方法, 其特征在于, 该方法包括以下具体步骤:

步骤1: 口令数据集预处理

收集不同口令集, 对口令集数据进行清洗, 完成口令集数据预处理;

步骤2: 口令因子拆分

将口令按照数字、字母和特殊字符分别拆分为数字因子、字母因子和特殊字符因子, 并统计各因子中包含的内容出现的频率及因子间组合的频率;

步骤3: 口令猜测集生成

基于口令因子拆分步骤得到的各口令因子出现频率及因子间组合的频率, 设置生成口令的条数, 生成口令猜测集合。

2. 根据权利要求1所述的基于数字因子逆序的口令猜测方法, 其特征在于, 所述口令集数据预处理为: 剔除无效口令及对口令数据集进行one-hot编码。

3. 根据权利要求1所述的基于数字因子逆序的口令猜测方法, 其特征在于, 所述步骤2具体包括:

步骤A1: 将口令拆分为数字因子、字母因子及特殊字符因子;

步骤A2: 逆序猜测数字, 并统计数字出现频率;

步骤A3: 中英文语料库提取语义匹配字母因子, 并统计字母因子出现频率;

步骤A4: 统计特殊字符出现频率。

4. 根据权利要求3所述的基于数字因子逆序的口令猜测方法, 其特征在于, 所述逆序猜测数字采用PCFG、Markov或神经网络方法。

5. 根据权利要求3所述的基于数字因子逆序的口令猜测方法, 其特征在于, 所述英文语料库为COCA语料库、中文语料库为sougou语料库。

一种基于数字因子逆序的口令猜测方法

技术领域

[0001] 本发明属于信息安全技术领域,特别涉及一种基于数字因子逆序的口令猜测方法。

背景技术

[0002] 在互联网技术普及的今天,出现了多种多样的用户身份认证方式,口令认证以其简单易用性作为网络用户身份认证的一种主流方式,并且在可预见的未来仍然如此,随着数据泄露事件的不断发生,对口令猜测的研究越来越多,按照口令猜测的环境一般分为在线口令猜测和离线口令猜测。在线口令猜测需要与网络认证服务器进行交互,通过向服务器提交用户设置的口令来获得服务系统的使用权。一般的认证服务器都会在一定时间内限制用户口令的认证次数,目的是为了防止暴力猜测算法,服务器同时会设置一些验证码来防止机器通过不断尝试从而成功认证系统。而离线口令猜测则是首先获取加密之后的口令文件,然后将已有字典库的口令逐条进行哈希运算,并与目标口令的哈希散列进行对比。相对于在线口令猜测,离线口令猜测拥有更宽松的计算环境,近几年来,离线口令猜测也从原先的漫无目的、基于简单变换的算法演进到基于口令频率模型的系统方法,在猜测效率上获得了提高,因此针对离线的口令猜测方法提出的也比较多,但是传统的口令猜测方法都是从前向后的,没有人从后向前进行分析,但是在对口令中的数字因子进行分析时,数字因子从后向前分析比从前向后分析更加准确。

发明内容

[0003] 本发明的目的是提供一种基于数字因子逆序的口令猜测方法,该方法能够增加口令猜测的准确率和效率,指导离线口令字典的生成。

[0004] 实现本发明目的的具体技术方案是:

一种基于数字因子逆序的口令猜测方法,该方法包括以下具体步骤:

步骤1:口令数据集预处理

收集不同口令集,对口令集数据进行清洗,完成口令集数据预处理;

步骤2:口令因子拆分

将口令按照数字、字母和特殊字符分别拆分为数字因子、字母因子和特殊字符因子,并统计各因子中包含的内容出现的频率及因子间组合的频率;

步骤3:口令猜测集生成

基于口令因子拆分步骤得到的各口令因子出现频率及因子间组合的频率,设置生成口令的条数,生成口令猜测集合。

[0005] 所述口令集数据预处理为:剔除无效口令及对口令数据集进行one-hot编码。

[0006] 所述步骤2具体包括:

步骤A1:将口令拆分为数字因子、字母因子及特殊字符因子;

步骤A2:逆序猜测数字,并统计数字出现频率;

步骤A3:中英文语料库提取语义匹配字母因子,并统计字母因子出现频率;

步骤A4:统计特殊字符出现频率。

[0007] 所述逆序猜测数字采用PCFG、Markov或神经网络方法。

[0008] 所述英文语料库为COCA语料库、中文语料库为sougou语料库。

[0009] 本发明的目的在于弥补现有口令猜测技术的不足,将口令分为数字因子、字母因子和特殊字符因子,将数字因子进行逆序分析,并且把字母因子与中英文语料库中的语义进行匹配,特殊字符则按照出现频率进行统计,可以增加口令猜测的准确率和效率,指导离线口令字典的生成。

附图说明

[0010] 图1是本发明的流程图。

具体实施方式

[0011] 结合以下具体实施例和附图,对本发明作进一步的详细说明。实施本发明的过程、条件、实验方法等,除以下专门提及的内容之外,均为本领域的普遍知识和公知常识,本发明没有特别限制内容。

实施例

[0012] 本实施例中有关的技术术语代表的含义如下:

D:口令中包含的数字因子

L:口令中包含的字母因子

S:口令中包含的特殊字符因子

L_n:口令中字母因子出现频率最高的前n条

n:设置口令猜测集中生成口令的条数。

[0013] 如图1所示,本实施例包括如下三个阶段:

第一阶段:口令数据集预处理:对收集到的12306口令数据集,对其进行清洗,剔除所有的无效口令,然后对口令数据进行编码;

第二阶段:口令因子拆分,以口令“password123123!”为例:将口令按照数字D、字母L和特殊字符S分别拆分为数字因子“123123”、字母因子“password”和特殊字符因子“!”:

1)对数字采取逆序猜测方法:若按照正序顺序,按照数字出现频率,前三位为“123”,则根据用户设置口令的习惯后三位可能为“123”也可能为“456”,但是按照数字逆序,后三位为“123”则前三位出现“123”的频率最大。

[0014] 2)字母语料库匹配。由于12306是中文口令库,所以选择sougou语料库进行匹配,统计的“password”单词的出现次数加1,并且更新L_n,即字母因子出现频率最高的前n个;

3)统计特殊字符“!”出现次数,并更新各特殊字符出现的频率及出现频率排名。

[0015] 第三阶段:口令猜测集生成:基于口令因子分析阶段得到的口令因子出现频率及口令因子组合频率,设置生成口令的条数n,生成口令猜测集合。

[0016] 本发明的保护内容不局限于以上实施例。在不背离发明构思的精神和范围下,本领域技术人员能够想到的变化和优点都被包括在本发明中,并且以所附的权利要求书为保

护范围。

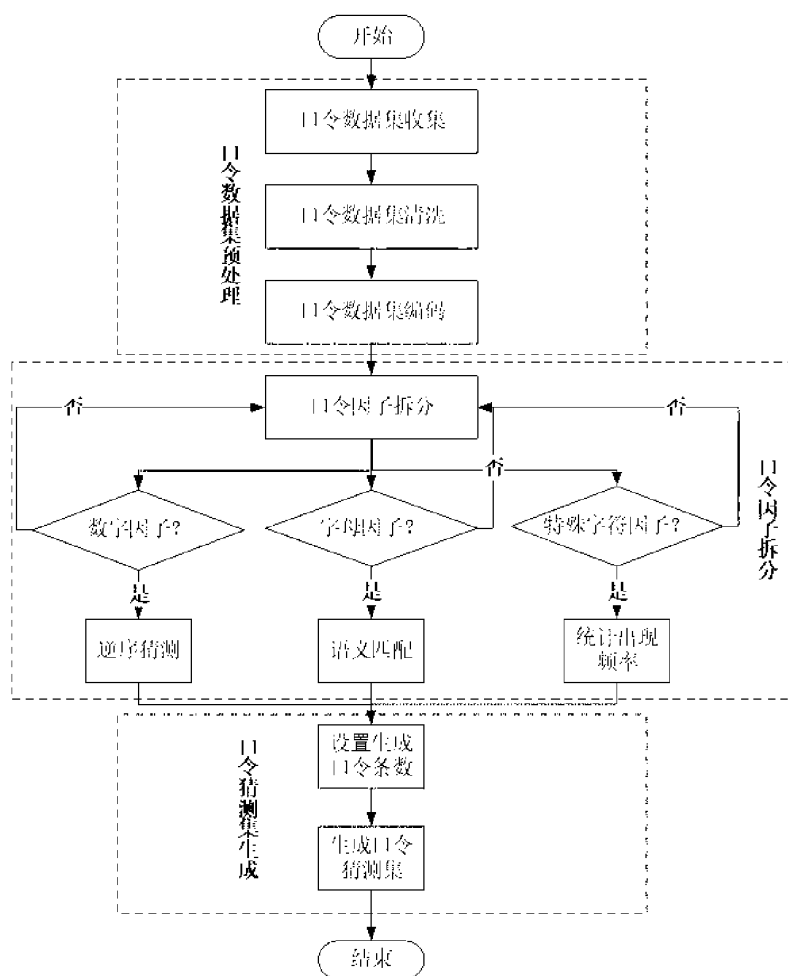


图1