(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108647511 A
(43)申请公布日 2018.10.12

(21)申请号 201810324327.2

(22)申请日 2018.04.12

(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 吴宇

(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 21/46(2013.01)

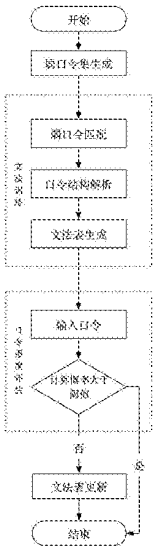
权利要求书2页 说明书5页 附图2页

(54)发明名称

基于弱口令推导的口令强度评估方法

(57)摘要

本发明公开了一种基于弱口令推导的口令强度评估方法，包括以下步骤：1)弱口令集生成：从口令样本中按照出现频率降序的方式选择排名靠前的口令作为弱口令集；2)文法训练：基于弱口令集解析训练集中的口令，生成带弱口令标签的概率上下文无关文法表；3)口令强度评估：输入口令，根据文法训练生成的文法表计算口令的概率，概率值越高表示口令强度越低；4)文法表更新：根据输入的口令，动态调整带弱口令标签的概率上下文无关文法的概率分布。本发明利用现有概率上下文无关文法推导与弱口令集中口令相似的口令，继承了传统口令强度评估方法的效率性和鲁棒性的同时，可以消除潜在的弱口令，增强口令抵抗口令猜测攻击的能力，提高用户的安全性。



CN 108647511 A

1.一种基于弱口令推导的口令强度评估方法，其特征在于，该方法包括以下具体步骤：

步骤1：弱口令集生成

从口令样本中按照出现频率降序的方式选择排名靠前的口令作为弱口令集；

步骤2：文法训练

基于弱口令集解析训练集中的口令，生成带弱口令标签的概率上下文无关文法表；

步骤3：口令强度评估

输入口令，根据文法训练生成的文法表计算口令的概率，概率值越高表示口令强度越低；

步骤4：文法表更新

根据输入的口令，动态调整带弱口令标签的概率上下文无关文法的概率分布。

2.根据权利要求1所述的基于弱口令推导的口令强度评估方法，其特征在于，步骤2具体包括：

步骤A1：弱口令匹配

将训练集中的口令或其子串与弱口令集中的口令进行相似度匹配，用于下一步的口令结构解析；

如果训练集中口令的子串与弱口令集中的口令成功匹配，继续对该口令中剩余未匹配部分继续执行上述匹配流程，直至对该口令的所有子串都执行过一次匹配，最终返回一个最优值序列；

步骤A2：口令结构解析

首先将步骤A1中返回的最优值序列用弱口令标签进行标记；剩下无法匹配的部分再用原始的概率上下文无关文法标签进行匹配，直至最终完成整个口令的解析；

步骤A3：文法表生成

当训练集中的所有口令解析完成，生成带弱口令标签的概率上下文无文法表；

其中：所述相似度匹配使用的算法包括但不局限于bk-tree。

3.根据权利要求2所述的基于弱口令推导的口令强度评估方法，其特征在于，步骤A1，具体过程为：

步骤A11：设置编辑距离阈值，相似度阈值大小；

步骤A12：获得所有编辑距离小于等于编辑距离阈值且相似度大于等于相似度阈值的待解析口令子串-对应弱口令字符串对；

步骤A13：在A12的基础上获得编辑距离最小的所有字符串对；

步骤A14：在A13的基础上获得所有相似度最大的字符串对；

步骤A15：在A14的基础上获得所有弱口令长度最大的字符串对；

步骤A16：如果A15求出的所有字符串对构成的集合为空，则表示待解析口令与弱口令集中的口令匹配失败；如果不为空，则表示待解析口令与弱口令集中的口令匹配成功，从字符串对构成的集合中随机选择一个字符串对作为最优解返回。

4.根据权利要求2所述的基于弱口令推导的口令强度评估方法，其特征在于，所述原始的概率上下文无关文法标签分为：数字、字母、特殊字符。

5.根据权利要求1所述的基于弱口令推导的口令强度评估方法，其特征在于，步骤2中所述带弱口令标签的概率上下文无关文法包括但不限于非终结字符集、终结字符集、始变

量和规则集。

6.根据权利要求5所述的基于弱口令推导的口令强度评估方法，其特征在于，非终结字符中的元素包括但不限于：字母字符、数字字符、特殊字符、键盘连续符、插入操作、删除操作、替换操作和弱口令字符串。

7.根据权利要求1所述的基于弱口令推导的口令强度评估方法，其特征在于，所述步骤4具体包括：

步骤B1：根据输入的口令确定频数加1的结构；

步骤B2：文法表中结构总数加1；

步骤B3：更新步骤B1中结构的概率；

步骤B4：依次更新文法表中其他结构的概率。

基于弱口令推导的口令强度评估方法

技术领域

[0001] 本发明属于信息安全技术领域，特别涉及一种基于弱口令推导的口令强度评估方法。

背景技术

[0002] 互联网技术的飞速发展深刻地改变着人们的学习、工作和生活方式，近年来以移动互联网，电子商务为代表的信息技术极大地便利人们的生活。而与互联网密切相关的信息安全问题也越来越受到人们的重视。身份认证作为保护用户信息安全的重要方式，被广泛的应用于互联网中的各个服务站点。

[0003] 身份认证是保护用户信息安全的主要手段。口令认证因为其部署方便、使用灵活等特点成为互联网中最广泛使用的身份认证方法。然而基于口令的认证系统却存在许多安全性和可用性问题。在口令认证系统中，系统要求用户创建一个可打印的字符串(即口令)，并将这个字符串作为验证用户身份的手段。由于人脑记忆的有限性，人类难以记忆复杂而安全的口令，而往往倾向于使用简单的口令。而简单口令的使用，可能会导致口令认证系统的脆弱性。

[0004] 一个好的口令强度评估器应该有能力刻画弱口令间的相似性。比如口令123456是公认的弱口令，口令123.456和123456有非常高的相似性，很有理由认为用户有很大的可能性根据口令123456构造出的新口令123.456。

[0005] 然而，学术界领先的基于PCFG算法的口令强度评估方法，却无法判断这一点。PCFG模型将用户口令字符分成字母(L)、数字(D)、特殊字符(S)三类，并假设用户通过“拼接”的方式来生成口令。

[0006] 因此，将传统的基于概率上下文无关文法与各网站普遍使用的弱口令结合起来，对于那些被现有的口令强度评估器评为“健壮”但实际十分不安全的弱口令评估有着十分重要的意义。

发明内容

[0007] 本发明的目的在于弥补现有口令强度评估方法的不足，将传统的概率上下文无关文法和弱口令集相结合，提供一种利用弱口令集推导的口令强度评估方法，在继承了传统口令强度评估方法的效率性和鲁棒性的同时，可以识别更多被误判为“健壮”的弱口令，增强口令抵抗口令猜测攻击的能力，提高口令的安全性。

[0008] 实现本发明目的的具体技术方案是：

[0009] 一种基于弱口令推导的口令强度评估方法，该方法包括以下具体步骤：

[0010] 步骤1：弱口令集生成

[0011] 从口令样本中按照出现频率降序的方式选择排名靠前的口令作为弱口令集；

[0012] 步骤2：文法训练

[0013] 基于弱口令集解析训练集中的口令，生成带弱口令标签的概率上下文无关文法

表；

[0014] 步骤3：口令强度评估

[0015] 输入口令，根据文法训练生成的文法表计算口令的概率，概率值越高表示口令强度越低；

[0016] 步骤4：文法表更新

[0017] 根据输入的口令，动态调整带弱口令标签的概率上下文无关文法的概率分布。

[0018] 本发明步骤2具体包括：

[0019] 步骤A1：弱口令匹配

[0020] 将训练集中的口令或其子串与弱口令集中的口令进行相似度匹配，用于下一步的口令结构解析；

[0021] 如果训练集中口令的子串与弱口令集中的口令成功匹配，继续对该口令中剩余未匹配部分继续执行上述匹配流程，直至对该口令的所有子串都执行过一次匹配，最终返回一个最优值序列；

[0022] 步骤A2：口令结构解析

[0023] 首先将步骤A1中返回的最优值序列用弱口令标签进行标记；剩下无法匹配的部分再用原始的概率上下文无关文法标签进行匹配，直至最终完成整个口令的解析；

[0024] 步骤A3：文法表生成

[0025] 当训练集中的所有口令解析完成，生成带弱口令标签的概率上下文无文法表；

[0026] 其中：所述相似度匹配使用的算法包括但不局限于bk-tree。

[0027] 所述步骤A1，具体过程为：

[0028] 步骤A11：设置编辑距离阈值，相似度阈值大小；

[0029] 步骤A12：获得所有编辑距离小于等于编辑距离阈值且相似度大于等于相似度阈值的待解析口令子串-对应弱口令字符串对；

[0030] 步骤A13：在A12的基础上获得编辑距离最小的所有字符串对；

[0031] 步骤A14：在A13的基础上获得所有相似度最大的字符串对；

[0032] 步骤A15：在A14的基础上获得所有弱口令长度最大的字符串对；

[0033] 步骤A16：如果A15求出的所有字符串对构成的集合为空，则表示待解析口令与弱口令集中的口令匹配失败；如果不为空，则表示待解析口令与弱口令集中的口令匹配成功，从字符串对构成的集合中随机选择一个字符串对作为最优解返回。

[0034] 所述原始的概率上下文无关文法标签分为：数字、字母、特殊字符。

[0035] 本发明步骤2中所述带弱口令标签的概率上下文无关文法包括但不限于非终结字符集、终结字符集、始变量和规则集。

[0036] 本发明所述非终结字符中的元素包括但不限于：字母字符、数字字符、特殊字符、键盘连续符、插入操作、删除操作、替换操作和弱口令字符串。

[0037] 本发明所述步骤4具体包括：

[0038] 步骤B1：根据输入的口令确定频数加1的结构；

[0039] 步骤B2：文法表中结构总数加1；

[0040] 步骤B3：更新步骤B1中结构的概率；

[0041] 步骤B4：依次更新文法表中其他结构的概率。

[0042] 本发明基于现有弱口令集，结合概率文上下文无关文方法，由弱口令集中口令递推出更多相似的口令，从而有效计算与弱口令集中口令相似口令的概率，在继承了传统口令强度评估方法的效率性和鲁棒性的同时，增强口令抵抗口令猜测攻击的能力，提高口令强度评估方法的精度。

附图说明

[0043] 图1为本发明流程图；

[0044] 图2为本发明训练集中口令与弱口令集中口令的匹配流程图。

具体实施方式

[0045] 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。

[0046] 实施例

[0047] 本实施例中有关的技术术语代表的含义如下：

[0048] PCFG:概率上下文无关文法(Probabilistic Context Free Grammar)

[0049] W:弱口令集

[0050] w:W中的元素

[0051] Wn:弱口令集中长度为n的口令(Lmin≤n≤Lmax)

[0052] Lmax:目标系统允许接收口令串的最大长度

[0053] Lmin:目标系统允许接收口令串的最小长度

[0054] T:训练集

[0055] OLCS(Optimal Longest Common Subsequence):最优最长公共子序列算法

[0056] pw:待解析的口令

[0057] SUB:pw所有子串的集合

[0058] SUB×W:SUB与W的笛卡尔积

[0059] sub:SUB中的元素

[0060] DT:编辑距离阈值

[0061] ST:相似度阈值

[0062] V＝{Start,A,L,U,D,S,K,insert,delete,replace,no,W1,W2,...Wn}，为非终结符集，其元素称为非终结符

[0063] Σ＝{95可打印的ASCII字符}，为与V不相交的终结符集，其元素称为终结符

[0064] Start是V的子集，称为始变量集

[0065] P是规则集，元素称为规则，形如α→β，其中α为非终结符，β由非终结符和终结符组成

[0066] A:字母字符，An代表n个连续的字母字符

[0067] L,U:字母字符掩码，其中L代表小写字母，U代表大写字母

[0068] D:数字字符，Dn代表n个连续的数字字符

[0069] S:特殊字符，S_n代表n个连续的特殊字符

[0070] K:键盘连续字符,Kn代表n个键盘连续字符(n≥4)

[0071] insert:对弱口令集中口令进行插入操作

[0072] delete:对弱口令集中口令进行删除操作

[0073] replace:对弱口令集中口令进行一般性替换操作

[0074] no:不对弱口令集中口令进行操作

[0075] 参阅图1,本实施例包括如下步骤:

[0076] 步骤1:弱口令集生成

[0077] 从口令样本中按照出现频率降序的方式选择排名靠前的口令作为弱口令集;

[0078] 步骤2:文法训练

[0079] 基于弱口令集解析训练集中的口令,生成带弱口令标签的概率上下文无关文法表;具体包括:

[0080] 步骤A1:弱口令匹配

[0081] 将训练集中的口令pw或其子串与弱口令集W中的口令进行相似度匹配,用于下一步的口令结构解析。

[0082] 如果训练集中口令的子串与弱口令集中的口令成功匹配,继续对该口令中剩余未匹配部分(pw-sub)继续执行上述匹配流程,直至对该口令的所有子串都执行过一次匹配,最终返回一个最优值序列opts={opt_1,opt_2,...opt_n}。

[0083] 步骤A2:口令结构解析

[0084] 首先将步骤A1中返回的opts标记pw的各个子串,如果pw的子串sub与W中长度为n的w匹配,则sub被标记为Wn;pw剩下的没有和W中口令匹配的子串再用原始的概率上下文无关文法的LDS标签进行匹配,直至最终完成整个口令的解析。

[0085] 步骤A3:文法表生成

[0086] 当训练集中的所有口令解析完成,生成带弱口令标签的概率上下文无关文法表。

[0087] 步骤A1中使用的相似度匹配算法包括但不局限于bk-tree。

[0088] 参阅图2,其中distance函数用于求两个字符串的编辑距离,smilarity函数用于求两个字符串的相似度,len函数用于求字符串的长度。所述步骤A1的具体过程为:

[0089] 口令pw与弱口令集W中口令的相似度匹配的流程为:

[0090] 步骤A11:设置编辑距离阈值DT,相似度阈值ST大小;

[0091] 步骤A12:获得所有编辑距离小于等于DT且相似度大于等于ST的待解析口令子串-对应弱口令字符串对(sub,w),其中(sub,w)∈SUB×W;

[0092] 步骤A13:在A12的基础上获得编辑距离最小的所有字符串对(sub,w);

[0093] 步骤A14:在A13的基础上获得所有相似度最大的字符串对(sub,w);

[0094] 步骤A15:在A14的基础上获得所有弱口令长度最大的字符串对(sub,w);

[0095] 步骤A16:如果A15求出的所有字符串对构成的集合为空,则表示口令pw与弱口令集W中的口令匹配失败;如果不为空,则表示口令pw与弱口令集W中的口令匹配成功,从字符串对构成的集合中随机选择一个字符串对(sub,w)作为最优解opt(opt=(sub,w),opt∈SUB×W返回。

[0096] 所述原始概率上下文无关文法标签分为:数字、字母、特殊字符。

[0097] 所述带弱口令标签的概率上下文无关文法G包含但不限于非终结字符集、终结字

符集、始变量和规则集。

[0098] 所述非终结字符中的元素包括但不限于：字母字符、数字字符、特殊字符、键盘连续符、插入操作、删除操作、替换操作、弱口令字符串。

[0099] 如口令avai^lable123∈T,available∈W。直接用PCFG匹配方法解析的口令结构为L4S1L5D3;avai^lable与弱口令available最相似(编辑距离最短、相似度最大、匹配长度最长)，所以(avai^lable,available)作为最优值，由于最优值序列只有一个，所以直接将(avai^lable,available)返回。

[0100] 步骤3：口令强度评估

[0101] 输入口令，根据文法训练生成的文法表计算口令的概率，概率值越高表示口令强度越低；

[0102] 如果输入口令是123.456且123456∈W,W6的概率为0.28,W_6→123456的概率为0.4,W_6→insert的概率为0.3,insert→S1的概率为0.11,S_1→.的概率为0.52,则123.456会被确认为由123456(结构为W_6)插入一个特殊字符“.”生成，所以口令123.456的概率为：P(123.456)＝P(Start→W6)*P(W6→123456)*P(W6→insert)*(insert→S1)*P(S1→.)

[0103] ＝0.28*0.4*0.3*0.11*0.52

[0104] ≈0.00192。

[0105] 步骤4：文法表更新

[0106] 根据输入的口令，动态调整带弱口令标签的概率上下文无关文法的概率分布；具体包括：

[0107] 步骤B1：根据输入的口令确定频数加1的结构；

[0108] 步骤B2：文法表中结构总数加1；

[0109] 步骤B3：更新步骤B1中结构的概率；

[0110] 步骤B4：依次更新文法表中其他结构的概率。

[0111] 设文法表中总共有n种结构，所有结构出现的总数为N。则第i个结构出现的概率P_i＝f_i/N,其中f_i为第i个结构出现的频数。当新注册一个口令后，假设注册口令后i对应结构的频数加1，结构总数N也加1。

[0112] 则第i个结构的概率更新为

[0113] $P_i'=(f_i+1)/(N+1)$ (1)

[0114] 其他结构的概率也依次更新为

[0115] $P_j'=f_j/(N+1)$,$j\neq i$ (2)

[0116] 如当用户注册一个口令123.456abc后，判断口令123.456abc中的部分123.456与弱口令123456相似，并将123.456标记为W6。剩余部分abc没有相应的弱口令可以匹配，按照PCFG分段方法，会被标记为L3，所以口令123.456abc会被识别标记为W_6L_3。

[0117] 则与口令相关的结构W_6L_3、L_3、W_6，终结符串123.456abc，以及规则W_6→insert，insert→.W_6→123456的概率依据公式(1)进行更新，其他结构的概率依据公式(2)进行相应调整。

[0118] 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。

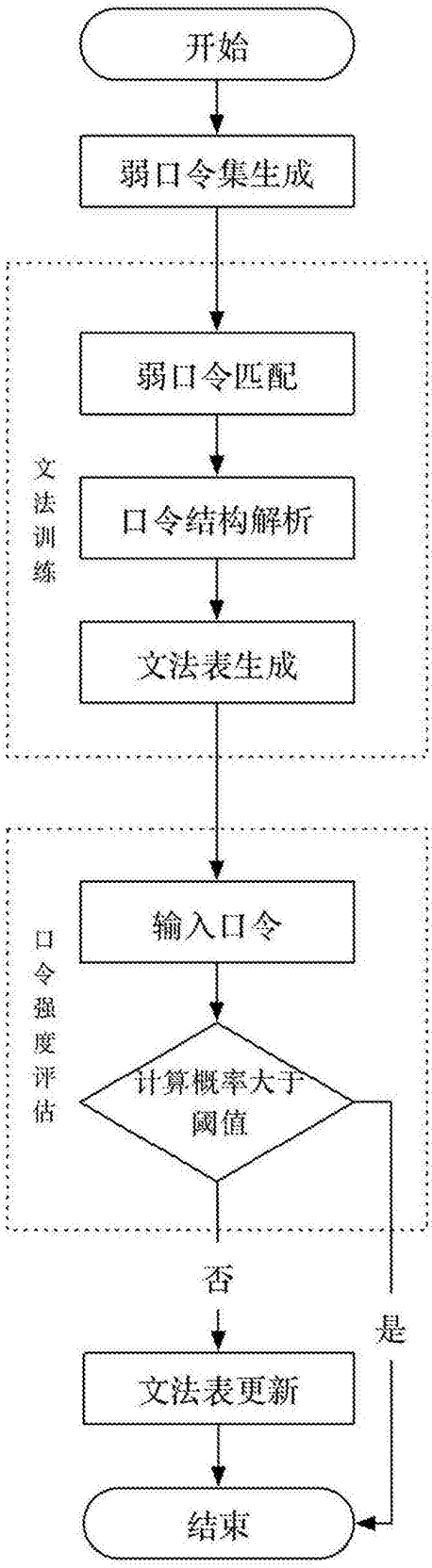

开始
弱口令集生成
文法训练
弱口令匹配
口令结构解析
文法表生成
口令强度评估
输入口令
计算概率大于阈值
否
是
文法表更新
结束

图1

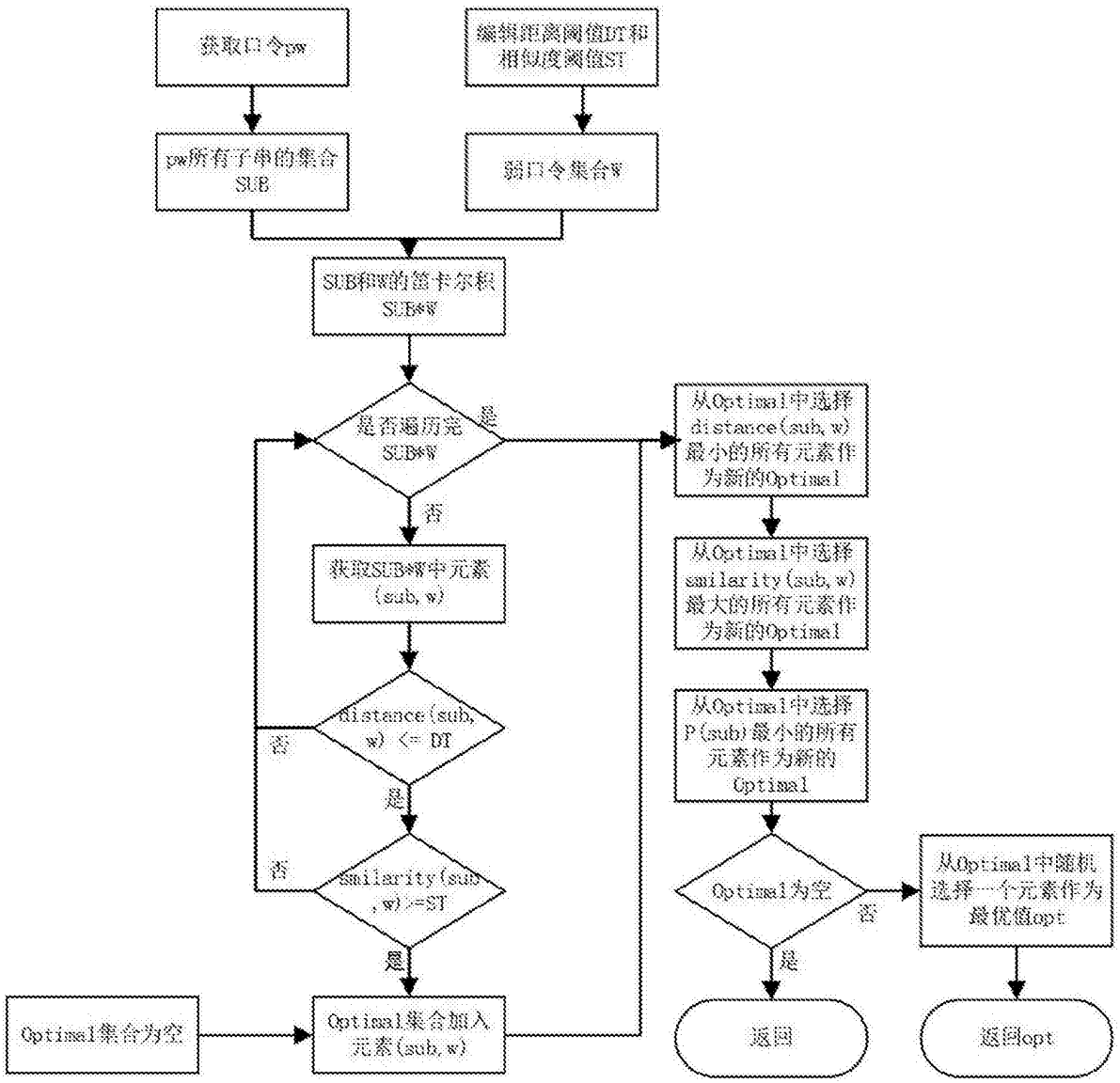

获取口令pw
编辑距离阈值DT和相似度阈值ST
pw所有子串的集合SUB
弱口令集合W
SUB和W的笛卡尔积SUB*W
是否遍历完SUB*W
是
否
从Optimal中选择distance(sub,w)最小的所有元素作为新的Optimal
从Optimal中选择smilarity(sub,w)最大的所有元素作为新的Optimal
从Optimal中选择P(sub)最小的所有元素作为新的Optimal
获取SUB*W中元素(sub,w)
distance(sub,w) <= DT
否
是
smilarity(sub,w)>=ST
否
是
Optimal为空
否
是
从Optimal中随机选择一个元素作为最优值opt
Optimal集合为空
Optimal集合加入元素(sub,w)
返回
返回opt

图2