(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114726823 B
(45)授权公告日 2022.08.30

(21)申请号 202210537109.3

(22)申请日 2022.05.18

(65)同一申请的已公布的文献号
申请公布号 CN 114726823 A

(43)申请公布日 2022.07.08

(73)专利权人 北京金睛云华科技有限公司
地址 100088 北京市海淀区北三环中路44号58号1层21号
专利权人 金睛云华(沈阳)科技有限公司

(72)发明人 胡文友 杨润峰 曲武 胡永亮

(74)专利代理机构 成都华复知识产权代理有限公司 51298
专利代理师 任丽娜

(51)Int.Cl.
H04L 61/3015(2022.01)
H04L 9/40(2022.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

(56)对比文件
CN 113709152 A,2021.11.26
CN 109617909 A,2019.04.12
CN 110830490 A,2020.02.21
US 2018288086 A1,2018.10.04

审查员 毕雅超

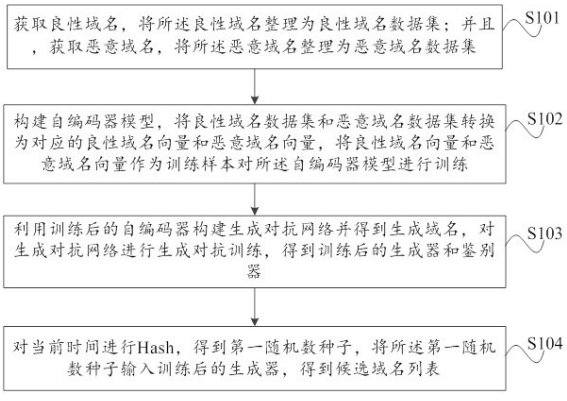
权利要求书3页 说明书8页 附图4页

(54)发明名称

一种基于生成对抗网络的域名生成方法、装置和设备

(57)摘要

本发明的实施例提供了一种基于生成对抗网络的域名生成方法、装置和设备。所述方法包括获取良性域名和恶意域名,得到良性域名数据集和恶意域名数据集;构建自编码器模型,将良性域名数据集和恶意域名数据集转换为对应的良性域名向量和恶意域名向量,作为训练样本对自编码器模型进行训练;构建生成对抗网络并得到生成域名,对生成对抗网络进行生成对抗训练,得到训练后的生成器和鉴别器;将所述第一随机数种子输入训练后的生成器,得到候选域名列表。以此方式,能够在模拟良性域名隐藏特征的同时规避掉已被检测恶意域名中的隐藏特征,使得生成的域名能躲避域名检测器的检测,拥有较高的抗检测能力,同时加快对抗生成网络的对抗速度。



CN 114726823 B

1.一种基于生成对抗网络的域名生成方法,其特征在于,包括:

获取良性域名,将所述良性域名整理为良性域名数据集;以及,获取恶意域名,将所述恶意域名整理为恶意域名数据集;

构建自编码器模型,将所述良性域名数据集和恶意域名数据集对应转换为良性域名向量和恶意域名向量,将所述良性域名向量和恶意域名向量作为训练样本对所述自编码器模型进行训练;

利用训练后的自编码器构建生成对抗网络并得到生成域名,对所述生成对抗网络进行生成对抗训练,得到训练后的生成器和鉴别器;

对当前时间进行Hash,得到第一随机数种子,将所述第一随机数种子输入训练后的生成器,得到候选域名列表;

所述自编码器模型包括编码器和解码器,所述编码器用于输入良性域名向量和恶意域名向量,依次通过卷积层、最大池化层、LSTM和高速网络层,输出域名特征向量;所述解码器用于输入所述域名特征向量,依次通过高速网络层、LSTM、最大池化层和卷积层,输出重构域名向量;所述自编码器模型的损失函数为:

$$CONTRANSTIVE(Y_i - \hat{Y}_i) = \frac{1}{2N}\sum_{i=1}^{N} T\left\|Y_i - \hat{Y}_i\right\|_2^2 + (1-T)\,max\left(M - \left\|Y_i - \hat{Y}_i\right\|_2^2, 0\right)^2$$

其中,$CONTRANSTIVE(*)$表示对比损失函数;Y_i代表原始域名向量;$\hat{Y}_i$代表编码重构后的域名向量;$\|*\|$为欧式距离;T表示原始域名是否为良性域名,当Y_i为良性域名时T取1,当Y_i为恶意域名时T取0;M代表阈值,表示恶意域名编码重构后的域名向量$\hat{Y}_i$与原始域名向量Y_i最大距离;N为域名特征向量的维度;

所述利用训练后的自编码器构建生成对抗网络并得到生成域名,包括:

构建生成网络和鉴别网络,并将训练后的自编码器拆分为编码器和解码器;

将冻结参数的解码器与所述生成网络组成生成器;将冻结参数的编码器与所述鉴别网络组成鉴别器;

对当前时间进行Hash,得到第二随机数种子,并将所述第二随机数种子输入所述生成器中的生成网络,得到域名特征向量;

将所述域名特征向量输入所述生成器中的解码器,输出生成域名。

2.根据权利要求1所述的方法,其特征在于,所述将所述良性域名整理为良性域名数据集,包括:

从所述良性域名中提取每个良性域名的二级域名作为良性域名字符串,得到良性域名字符串列表;

对所述良性域名字符串列表中的良性域名字符串进行随机排序,将排序后的良性域名字符串列表作为良性域名数据集。

3.根据权利要求1所述的方法,其特征在于,所述将所述恶意域名整理为恶意域名数据集,包括:

从所述恶意域名中提取每个恶意域名的二级域名作为恶意域名字符串,得到恶意域名字符串列表;

对所述恶意域名字符串列表中的恶意域名字符串进行随机排序,将排序后的恶意域名字符串列表作为恶意域名数据集。

4.根据权利要求1所述的方法,其特征在于,所述对所述生成对抗网络进行生成对抗训练,包括:

将所述生成域名作为输入数据输入所述鉴别器中编码器;

所述编码器将所述生成域名映射为域名特征向量,输入到所述鉴别网络中,输出域名类型鉴别结果。

5.根据权利要求1所述的方法,其特征在于,还包括:

对所述候选域名列表进行筛选,得到DGA生成域名。

6.根据权利要求5所述的方法,其特征在于,所述对所述候选域名列表进行筛选,包括:

删除所述候选域名列表中不符合RFC 1035规范的域名;和/或

删除所述候选域名列表中二级域名长度小于3个字符的域名。

7.一种基于生成对抗网络的域名生成装置,其特征在于,包括:

获取模块,用于获取良性域名,将所述良性域名整理为良性域名数据集;并且,获取恶意域名,将所述恶意域名整理为恶意域名数据集;

第一训练模块,用于构建自编码器模型,将所述良性域名数据集和恶意域名数据集转换为对应的良性域名向量和恶意域名向量,将所述良性域名向量和恶意域名向量作为训练样本对所述自编码器模型进行训练;所述自编码器模型包括编码器和解码器,所述编码器用于输入良性域名向量和恶意域名向量,依次通过卷积层、最大池化层、LSTM和高速网络层,输出域名特征向量;所述解码器用于输入所述域名特征向量,依次通过高速网络层、LSTM、最大池化层和卷积层,输出重构域名向量;所述自编码器模型的损失函数为:

$$CONTRANSTIVE(Y_i-\hat{Y}_i)=\frac{1}{2N}\sum_{i=1}^{N}T\left\|Y_i-\hat{Y}_i\right\|_2^2+(1-T)\,max(M-\left\|Y_i-\hat{Y}_i\right\|_2^2,0)^2$$

其中,$CONTRANSTIVE(*)$表示对比损失函数;Y_i代表原始域名向量;$\hat{Y}_i$代表编码重构后的域名向量;$\|*\|$为欧式距离;T表示原始域名是否为良性域名,当Y_i为良性域名时T取1,当Y_i为恶意域名时T取0;M代表阈值,表示恶意域名编码重构后的域名向量$\hat{Y}_i$与原始域名向量Y_i最大距离;N为域名特征向量的维度;

第二训练模块,用于利用训练后的自编码器构建生成对抗网络并得到生成域名,对所述生成对抗网络进行生成对抗训练,得到训练后的生成器和鉴别器;所述利用训练后的自编码器构建生成对抗网络并得到生成域名,包括:构建生成网络和鉴别网络,并将训练后的自编码器拆分为编码器和解码器;将冻结参数的解码器与所述生成网络组成生成器;将冻结参数的编码器与所述鉴别网络组成鉴别器;对当前时间进行Hash,得到第二随机数种子,并将所述第二随机数种子输入所述生成器中的生成网络,得到域名特征向量;将所述域名特征向量输入所述生成器中的解码器,输出生成域名;

输出模块,用于对当前时间进行Hash,得到第一随机数种子,将所述第一随机数种子输入训练后的生成器,输出候选域名列表。

8.一种电子设备,包括至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其特征在于,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-6中任一项所述的方法。

一种基于生成对抗网络的域名生成方法、装置和设备

技术领域

[0001] 本发明一般涉及域名生成领域，并且更具体地，涉及一种基于生成对抗网络的域名生成方法、装置和设备。

背景技术

[0002] 随着恶意木马产业的发展，很多木马早已摆脱了过去“单打独斗”的作战方式，而是通过网络相互关联起来，通过指挥大量受到感染的计算机共同行动，进而发挥出协同效果。这样既可以集中起来同时对某个目标进行打击，也可以互相分散各自所承受的风险。在木马攻击过程中进行指挥的关键节点便是命令及控制服务器(Command and Control Server，C&C服务器)。受到感染的计算机通过生成域名与C&C服务器建立连接。域名生成算法(Domain generation algorithms，DGA)可以快速产生大量生成域名的算法。

[0003] 传统基于黑名单的防护手段无法有效应对DGA生成的域名，一方面，黑名单的更新速度远远赶不上DGA域名的生成速度；另一方面问题是防御者必须阻断所有的DGA域名才能阻断C&C服务器通信。

[0004] 近几年，研究人员对DGA域名的检测进行了大量的研究。这些方法主要分为两类，一类是基于域名状态进行检测，一类是对域名名称进行分析检测。对于域名状态的检测主要是通过域名在商业平台注册的情况和流量分析来进行评判，本质是通过域名的一些行为特征指标判别域名性质，但是这些性质常在僵尸控制者操作受感染电脑后才会表现出来；而对域名名称进行的分析检测可以拥有更好的实时效果，尤其借助深度学习对于域名数据进行表征学习，可以更快的适应不断变化的DGA生成方法，也大大减少了人力物力的巨大投入。

[0005] 基于对抗样本的域名生成方式可以使得生成的对抗域名具备很高的抗检测能力，能够误导DGA域名检测器做出错误的分类，提高DGA的抗检测性能，但如果对样本的生成方式不加以限制，可能会导致生成的对抗样本过于自由化，无法最大限度的提升对抗样本的抗检测能力。在基于生成对抗网络的域名生成方式在生成对抗的过程中，由于生成器生成对抗样本的过程只考虑了良性域名特征，未考虑已经被检测器检测出的恶意域名的特征，会导致生成对抗网络的训练往往耗时较长，严重影响对抗网络训练的速度和效率，而且会在生成的样本中出现恶意域名的隐藏特征。

发明内容

[0006] 根据本发明的实施例，提供了一种基于生成对抗网络的域名生成方案。本方案能够在模拟良性域名隐藏特征的同时规避掉已被检测恶意域名中的隐藏特征，使得生成的域名能躲避域名检测器的检测，拥有较高的抗检测能力，同时加快对抗生成网络的对抗速度。

[0007] 在本发明的第一方面，提供了一种基于生成对抗网络的域名生成方法。该方法包括：

[0008] 获取良性域名，将所述良性域名整理为良性域名数据集；以及，获取恶意域名，将

所述恶意域名整理为恶意域名数据集；

[0009] 构建自编码器模型，将所述良性域名数据集和恶意域名数据集对应转换为良性域名向量和恶意域名向量，将所述良性域名向量和恶意域名向量作为训练样本对所述自编码器模型进行训练；

[0010] 利用训练后的自编码器构建生成对抗网络并得到生成域名，对所述生成对抗网络进行生成对抗训练，得到训练后的生成器和鉴别器；

[0011] 对当前时间进行Hash，得到第一随机数种子，将所述第一随机数种子输入训练后的生成器，得到候选域名列表。

[0012] 进一步地，所述将所述良性域名整理为良性域名数据集，包括：

[0013] 从所述良性域名中提取每个良性域名的二级域名作为良性域名字符串，得到良性域名字符串列表；

[0014] 对所述良性域名字符串列表中的良性域名字符串进行随机排序，将排序后的良性域名字符串列表作为良性域名数据集。

[0015] 进一步地，所述将所述恶意域名整理为恶意域名数据集，包括：

[0016] 从所述恶意域名中提取每个恶意域名的二级域名作为恶意域名字符串，得到恶意域名字符串列表；

[0017] 对所述恶意域名字符串列表中的恶意域名字符串进行随机排序，将排序后的恶意域名字符串列表作为恶意域名数据集。

[0018] 进一步地，所述自编码器模型包括编码器和解码器，所述编码器用于输入良性域名向量和/或恶意域名向量，输出域名特征向量；所述解码器用于输入所述域名特征向量，输出重构域名向量；

[0019] 所述自编码器模型的损失函数为：

[0020] $$CONTRANSTIVE(Y_i - \hat{Y}_i) = \frac{1}{2N}\sum_{i=1}^{N} T\left\|Y_i - \hat{Y}_i\right\|_2^2 + (1-T)\,max\left(M - \left\|Y_i - \hat{Y}_i\right\|_2^2, 0\right)^2$$

[0021] 其中，$CONTRANSTIVE(*)$表示对比损失函数；Y_i代表原始域名向量；$\hat{Y}_i$代表编码重构后的域名向量；$\|*\|$为欧式距离；T表示原始域名是否为良性域名，当Y_i为良性域名时T取1，当Y_i为恶意域名时T取0；M代表阈值，表示恶意域名编码重构后的域名向量$\hat{Y}_i$与原始域名向量Y_i最大距离；N为域名特征向量的维度。

[0022] 进一步地，所述利用训练后的自编码器构建生成对抗网络并得到生成域名，包括：

[0023] 构建生成网络和鉴别网络，并将训练后的自编码器拆分为编码器和解码器；

[0024] 将冻结参数的解码器与所述生成网络组成生成器；将冻结参数的编码器与所述鉴别网络组成鉴别器；

[0025] 对当前时间进行Hash，得到第二随机数种子，并将所述第二随机数种子输入所述生成器中的生成网络，得到域名特征向量；

[0026] 将所述域名特征向量输入所述生成器中的解码器，输出生成域名。

[0027] 进一步地，所述对所述生成对抗网络进行生成对抗训练，包括：

[0028] 将所述生成域名作为输入数据输入所述鉴别器中编码器；

[0029] 所述编码器将所述生成域名映射为域名特征向量，输入到所述鉴别网络中，输出

域名类型鉴别结果。

[0030] 进一步地,该方法还包括:

[0031] 对所述候选域名列表进行筛选,得到DGA生成域名。

[0032] 进一步地,所述对所述候选域名列表进行筛选,包括:

[0033] 删除所述候选域名列表中不符合RFC 1035规范的域名;和/或

[0034] 删除所述候选域名列表中二级域名长度小于3个字符的域名。

[0035] 在本发明的第二方面,提供了一种基于生成对抗网络的域名生成装置。该装置包括:

[0036] 获取模块,用于获取良性域名,将所述良性域名整理为良性域名数据集;并且,获取恶意域名,将所述恶意域名整理为恶意域名数据集;

[0037] 第一训练模块,用于构建自编码器模型,将所述良性域名数据集和恶意域名数据集转换为对应的良性域名向量和恶意域名向量,将所述良性域名向量和恶意域名向量作为训练样本对所述自编码器模型进行训练;

[0038] 第二训练模块,用于利用训练后的自编码器构建生成对抗网络并得到生成域名,对所述生成对抗网络进行生成对抗训练,得到训练后的生成器和鉴别器;

[0039] 输出模块,用于对当前时间进行Hash,得到第一随机数种子,将所述第一随机数种子输入训练后的生成器,输出候选域名列表。

[0040] 在本发明的第三方面,提供了一种电子设备。该电子设备至少一个处理器;以及与所述至少一个处理器通信连接的存储器;所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行本发明第一方面的方法。

[0041] 在本发明的第四方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,所述计算机指令用于使所述计算机执行本发明第一方面的方法。

[0042] 应当理解,发明内容部分中所描述的内容并非旨在限定本发明的实施例的关键或重要特征,亦非用于限制本发明的范围。本发明的其它特征将通过以下的描述变得容易理解。

附图说明

[0043] 结合附图并参考以下详细说明,本发明各实施例的上述和其他特征、优点及方面将变得更加明显。在附图中,相同或相似的附图标记表示相同或相似的元素,其中:

[0044] 图1示出了根据本发明的实施例的基于生成对抗网络的域名生成方法的流程图;

[0045] 图2示出了根据本发明的实施例的自编码器结构图;

[0046] 图3示出了根据本发明的实施例的生成网络结构图;

[0047] 图4示出了根据本发明的实施例的鉴别网络结构图;

[0048] 图5示出了根据本发明的实施例的基于生成对抗网络的域名生成装置的方框图;

[0049] 图6示出了能够实施本发明的实施例的示例性电子设备的方框图;

[0050] 其中,600为电子设备、601为CPU、602为ROM、603为RAM、604为总线、605为I/O接口、606为输入单元、607为输出单元、608为存储单元、609为通信单元。

具体实施方式

[0051] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的全部其他实施例，都属于本发明保护的范围。

[0052] 另外，本文中术语“和/或”，仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

[0053] 本发明中，能够在模拟良性域名隐藏特征的同时规避掉已被检测恶意域名中的隐藏特征，使得生成的域名能躲避域名检测器的检测，拥有较高的抗检测能力，同时加快对抗生成网络的对抗速度。

[0054] 图1示出了本发明实施例的基于生成对抗网络的域名生成方法的流程图。

[0055] 该方法包括：

[0056] S101、获取良性域名，将所述良性域名整理为良性域名数据集。

[0057] 作为本发明的一种实施例，所述获取良性域名，可以通过下载Alexa数据集，从所述Alexa数据集中整理出良性域名信息。所述Alexa数据集是亚马逊提供的全球排名TOP一百万的网站域名集合，文件是CSV格式，以排名、域名组成。Alexa数据集中网站均为良性域名，只需读取数据集并去除域名排名就可作为良性域名信息使用。

[0058] 作为本发明的一种实施例，所述将所述良性域名整理为良性域名数据集，包括：

[0059] 从所述良性域名中提取每个良性域名的SLD(Second Level Domain，二级域名)作为良性域名字符串，得到良性域名字符串列表；

[0060] 可以使用Numpy(Numerical Python)对所述良性域名字符串列表中的良性域名字符串进行随机排序，将排序后的良性域名字符串列表作为良性域名数据集。Numpy是Python的一种开源的数值计算扩展。

[0061] S101中还包括获取恶意域名，将所述恶意域名整理为恶意域名数据集。

[0062] 作为本发明的一种实施例，所述获取恶意域名可以下载360DGA数据集，并从360DGA数据集中整理出恶意域名信息。360DGA数据集是360netlab实验室公布的已经检测出的DGA恶意域名集合。由于数据集中包含域名、检测日期、所属家族等信息，需要将其他无用信息进行去除仅保留域名信息，所以称为整理出恶意域名信息。

[0063] 作为本发明的一种实施例，所述将所述恶意域名整理为恶意域名数据集，包括：

[0064] 从所述恶意域名中提取每个恶意域名的SLD作为恶意域名字符串，得到恶意域名字符串列表；

[0065] 使用Numpy对所述恶意域名字符串列表中的恶意域名字符串进行随机排序，将排序后的恶意域名字符串列表作为恶意域名数据集。

[0066] S102、构建自编码器模型，将所述良性域名数据集和恶意域名数据集转换为对应的良性域名向量和恶意域名向量，将所述良性域名向量和恶意域名向量作为训练样本对所述自编码器模型进行训练。

[0067] 作为本发明的一种实施例，如图2所示，所述自编码器模型包括编码器和解码器。

[0068] 在本实施例中，所述编码器包括卷积层、最大池化层、LSTM和高速网络。所述编码

器用于输入良性域名向量和/或恶意域名向量，输出域名特征向量。输入的良性域名向量和/或恶意域名向量依次通过卷积层、最大池化层、LSTM和高速网络输出域名特征向量。LSTM(Long Short-Term Memory，长短期记忆网络)是一种时间循环神经网络。

[0069] 在本实施例中，所述解码器包括高速网络、LSTM、最大池化层和卷积层；所述解码器用于输入所述域名特征向量，依次通过高速网络、LSTM、最大池化层和卷积层，输出重构域名向量。

[0070] 所述自编码器模型的损失函数为：

[0071] $$CONTRANSTIVE(Y_i-\hat{Y}_i)=\frac{1}{2N}\sum_{i=1}^{N}T\left\|Y_i-\hat{Y}_i\right\|_2^2+(1-T)max(M-\left\|Y_i-\hat{Y}_i\right\|_2^2,0)^2$$

[0072] 其中，$CONTRANSTIVE(*)$表示对比损失函数；Y_i代表原始域名向量；$\hat{Y}_i$代表编码重构后的域名向量；$\|*\|$为欧式距离；T表示原始域名是否为良性域名，当Y_i为良性域名时T取1，当Y_i为恶意域名时T取0；M代表阈值，表示恶意域名编码重构后的域名向量$\hat{Y}_i$与原始域名向量Y_i最大距离；N为域名特征向量的维度。

[0073] 作为本发明的一种实施例，将所述良性域名数据集和恶意域名数据集转换为对应的良性域名向量和恶意域名向量包括：

[0074] 良性域名字符串列表使用One-hot编码技术进行编码，获得良性域名向量；以及恶意域名字符串列表使用One-hot编码技术进行编码，获得恶意域名向量。One-hot编码技术是将英文域名转换为纯数字向量的常用技术。

[0075] 每个经过One-hot编码后的域名向量Y_i输入搭建的自编码器模型，得到一个编码重构后的域名向量$\hat{Y}_i$。通过对比损失函数训练自编码器，确保良性域名编码前后的欧式距离更小，恶意域名编码前后的欧式距离更大。

[0076] S103、利用训练后的自编码器构建生成对抗网络并得到生成域名，对所述生成对抗网络进行生成对抗训练，得到训练后的生成器和鉴别器。

[0077] 作为本发明的一种实施例，所述利用训练后的自编码器构建生成对抗网络并得到生成域名，包括：

[0078] 首先，构建生成网络和鉴别网络，并将训练后的自编码器拆分为编码器和解码器；其中，如图3所示，生成网络依次包括两个LSTM和两个全连接层。如图4所示，鉴别网络依次包括两个全连接层和两个LSTM。

[0079] 其次，将冻结参数的解码器与所述生成网络组成生成器G；将冻结参数的编码器与所述鉴别网络组成鉴别器D。

[0080] 再次，对当前时间进行Hash，得到第二随机数种子，并将所述第二随机数种子输入所述生成器中的生成网络，得到域名特征向量。

[0081] 最后，将所述域名特征向量输入所述生成器中的解码器，输出生成域名。

[0082] 作为本发明的一种实施例，所述对所述生成对抗网络进行生成对抗训练，包括：

[0083] 将所述生成域名作为输入数据输入所述鉴别器中编码器；

[0084] 所述编码器将所述生成域名映射为域名特征向量，输入到所述鉴别网络中，输出域名类型鉴别结果，完成对生成器和鉴别器的训练。

[0085] S104、对当前时间进行Hash，得到第一随机数种子，将所述第一随机数种子输入训

练后的生成器,得到候选域名列表。

[0086] 在本实施例中,提取生成对抗网络的生成器G作为域名生成器;对当前时间进行Hash,得到第一随机数种子;将所述第一随机数种子输入所述域名生成器中,输出候选域名列表。

[0087] 由于输出的候选域名列表中可能会存在不符合的域名,故需要对所述候选域名列表进行筛选。

[0088] 作为本发明的一种实施例,可以通过以下两种筛选方式进行筛选:

[0089] (1)删除所述候选域名列表中不符合RFC 1035规范的域名;

[0090] (2)删除所述候选域名列表中SLD(Second Level Domain,二级域名)长度小于3个字符的域名。

[0091] 通过上述筛选后,剩余的候选域名列表中即为最终生成的域名。

[0092] 本方法将对比损失函数引入到生成对抗域名生成方法之中,使用对比损失函数训练自动编码器结构来学习良性域名的特征并规避已被检测的恶意域名特征,并将训练后的自编码器重新组合成生成对抗网络,使得生成的域名仅包含良性域名的特征,而不会有任何域名检测器检测的恶意域名特征,避免样本中包含恶意域名的隐藏特征。

[0093] 本发明提出的生成对抗域名生成方法能够在很好模拟良性域名隐藏特征的同时规避掉已被检测恶意域名中的隐藏特征,使得生成的域名仅包含良性域名的特征,而不会有任何域名检测器检测的恶意域名特征,拥有较高的抗检测能力,同时防止对抗生成网络过于自由化导致的训练过程缓慢的问题。

[0094] 本发明同时结合生成对抗网络的高泛化性能,将冻结参数的自编码器组成生成对抗网络,通过互相对抗后,使得生成器生成的域名不具备随机性,更接近人类命名域名的习惯,达到充分模仿良性域名的目的,提高本域名生成方法的抗检测能力。

[0095] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明并不受所描述的动作顺序的限制,因为依据本发明,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于可选实施例,所涉及的动作和模块并不一定是本发明所必须的。

[0096] 以上是关于方法实施例的介绍,以下通过装置实施例,对本发明所述方案进行进一步说明。

[0097] 如图5所示,装置500包括:

[0098] 获取模块510,用于获取良性域名,将所述良性域名整理为良性域名数据集;并且,获取恶意域名,将所述恶意域名整理为恶意域名数据集;

[0099] 第一训练模块520,用于构建自编码器模型,将所述良性域名数据集和恶意域名数据集转换为对应的良性域名向量和恶意域名向量,将所述良性域名向量和恶意域名向量作为训练样本对所述自编码器模型进行训练;

[0100] 第二训练模块530,用于利用训练后的自编码器构建生成对抗网络并得到生成域名,对所述生成对抗网络进行生成对抗训练,得到训练后的生成器和鉴别器;

[0101] 输出模块540,用于对当前时间进行Hash,得到第一随机数种子,将所述第一随机数种子输入训练后的生成器,输出候选域名列表。

[0102] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，所述描述的模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0103] 本发明的技术方案中，所涉及的用户个人信息的获取，存储和应用等，均符合相关法律法规的规定，且不违背公序良俗。

[0104] 根据本发明的实施例，本发明还提供了一种电子设备。

[0105] 图6示出了可以用来实施本发明的实施例的电子设备600的示意性框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例，并且不意在限制本文中描述的和/或者要求的本发明的实现。

[0106] 设备600包括计算单元601，其可以根据存储在只读存储器(ROM)602中的计算机程序或者从存储单元608加载到随机访问存储器(RAM)603中的计算机程序，来执行各种适当的动作和处理。在RAM 603中，还可存储设备600操作所需的各种程序和数据。计算单元601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0107] 设备600中的多个部件连接至I/O接口605，包括：输入单元606，例如键盘、鼠标等；输出单元607，例如各种类型的显示器、扬声器等；存储单元608，例如磁盘、光盘等；以及通信单元609，例如网卡、调制解调器、无线通信收发机等。通信单元609允许设备600通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0108] 计算单元601可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元601的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元601执行上文所描述的各个方法和处理，例如方法S101~S104。例如，在一些实施例中，方法S101~S104可被实现为计算机软件程序，其被有形地包含于机器可读介质，例如存储单元608。在一些实施例中，计算机程序的部分或者全部可以经由ROM 602和/或通信单元609而被载入和/或安装到设备600上。当计算机程序加载到RAM 603并由计算单元601执行时，可以执行上文描述的方法S101~S104的一个或多个步骤。备选地，在其他实施例中，计算单元601可以通过其他任何适当的方式(例如，借助于固件)而被配置为执行方法S101~S104。

[0109] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括：实施在一个或者多个计算机程序中，该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释，该可编程处理器可以是专用或者通用可编程处理器，可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令，并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0110] 用于实施本发明的方法的程序代码可以采用一个或多个编程语言的任何组合来

编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器，使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行，作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0111] 在本发明的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0112] 为了提供与用户的交互，可以在计算机上实施此处描述的系统和技术，该计算机具有：用于向用户显示信息的显示装置(例如，CRT(阴极射线管)或者LCD(液晶显示器)监视器)；以及键盘和指向装置(例如，鼠标或者轨迹球)，用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互；例如，提供给用户的反馈可以是任何形式的传感反馈(例如，视觉反馈、听觉反馈、或者触觉反馈)；并且可以用任何形式(包括声输入、语音输入、触觉输入)来接收来自用户的输入。

[0113] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如，作为数据服务器)、或者包括中间件部件的计算系统(例如，应用服务器)、或者包括前端部件的计算系统(例如，具有图形用户界面或者网络浏览器的用户计算机，用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如，通信网络)来将系统的部件相互连接。通信网络的示例包括：局域网(LAN)、广域网(WAN)和互联网。

[0114] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器，也可以为分布式系统的服务器，或者是结合了区块链的服务器。

[0115] 应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本发明中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本发明的技术方案所期望的结果，本文在此不进行限制。

[0116] 上述具体实施方式，并不构成对本发明保护范围的限制。本领域技术人员应该明白的是，根据设计要求和其他因素，可以进行各种修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等，均应包含在本发明保护范围之内。

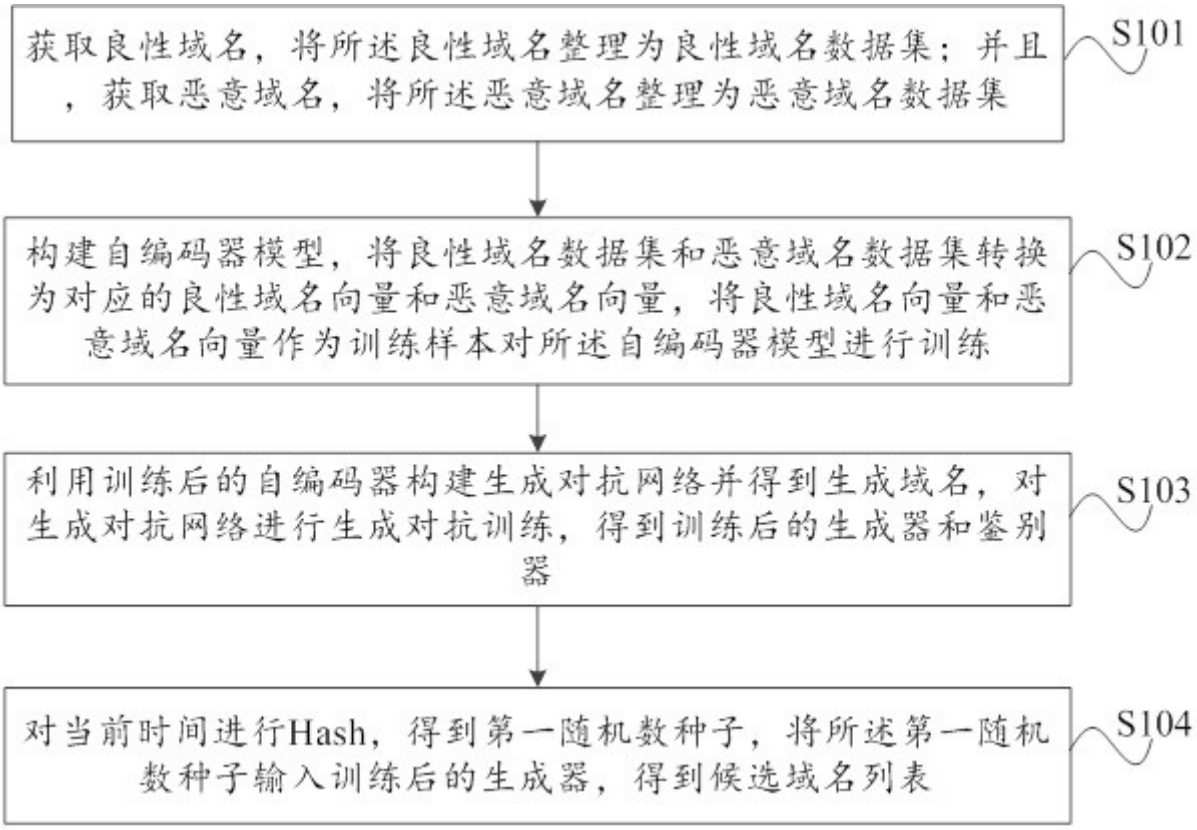

获取良性域名，将所述良性域名整理为良性域名数据集；并且，获取恶意域名，将所述恶意域名整理为恶意域名数据集
S101
构建自编码器模型，将良性域名数据集和恶意域名数据集转换为对应的良性域名向量和恶意域名向量，将良性域名向量和恶意域名向量作为训练样本对所述自编码器模型进行训练
S102
利用训练后的自编码器构建生成对抗网络并得到生成域名，对生成对抗网络进行生成对抗训练，得到训练后的生成器和鉴别器
S103
对当前时间进行Hash，得到第一随机数种子，将所述第一随机数种子输入训练后的生成器，得到候选域名列表
S104

图1

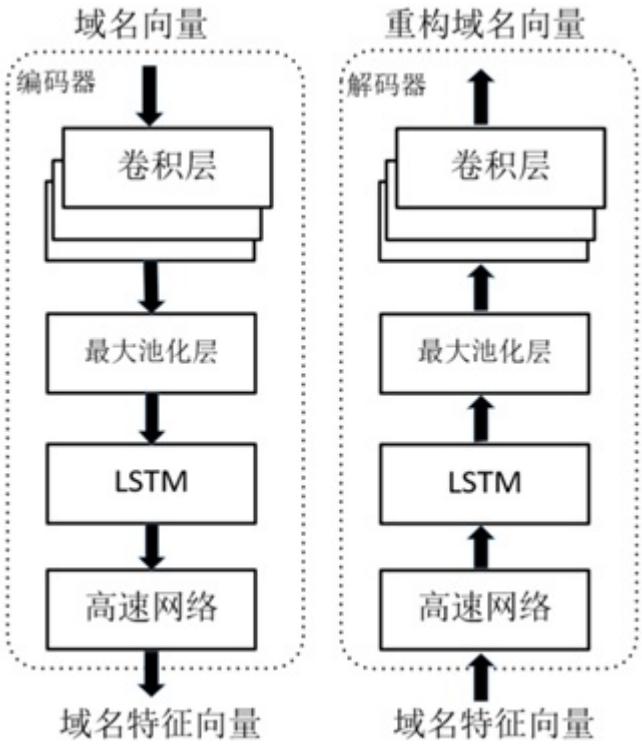

域名向量
编码器
卷积层
最大池化层
LSTM
高速网络
域名特征向量
重构域名向量
解码器
卷积层
最大池化层
LSTM
高速网络
域名特征向量

图2

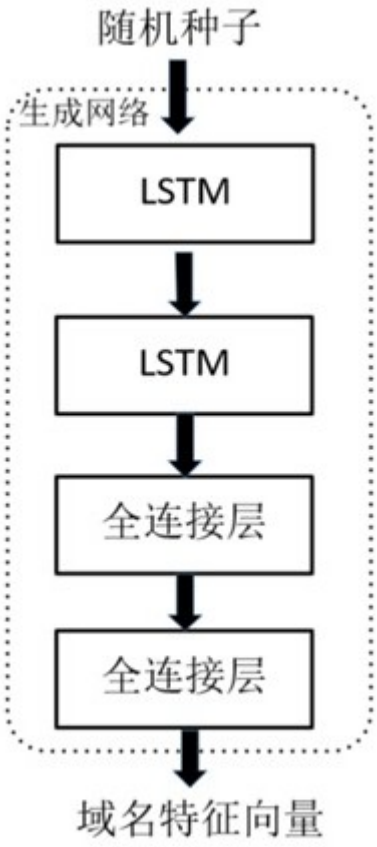


图3

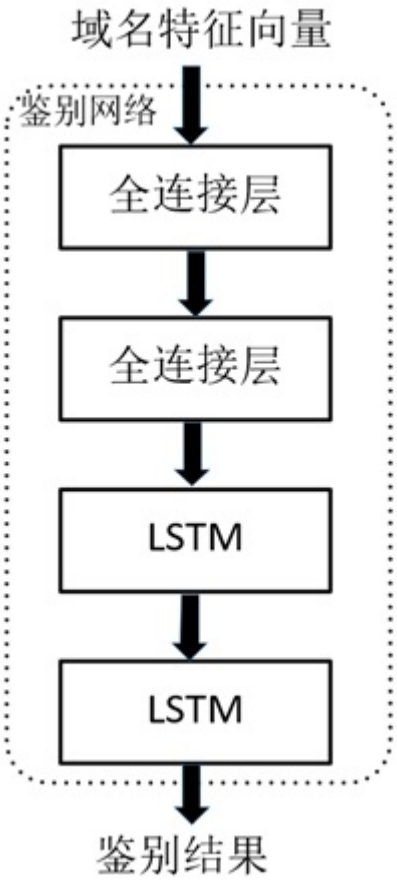


图4

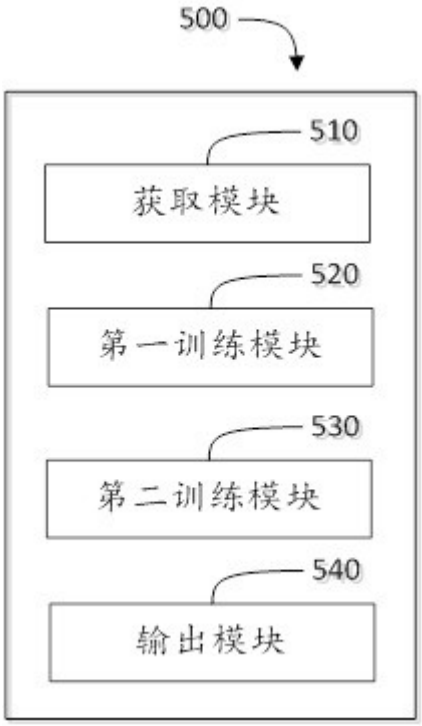

500
510
获取模块
520
第一训练模块
530
第二训练模块
540
输出模块

图5

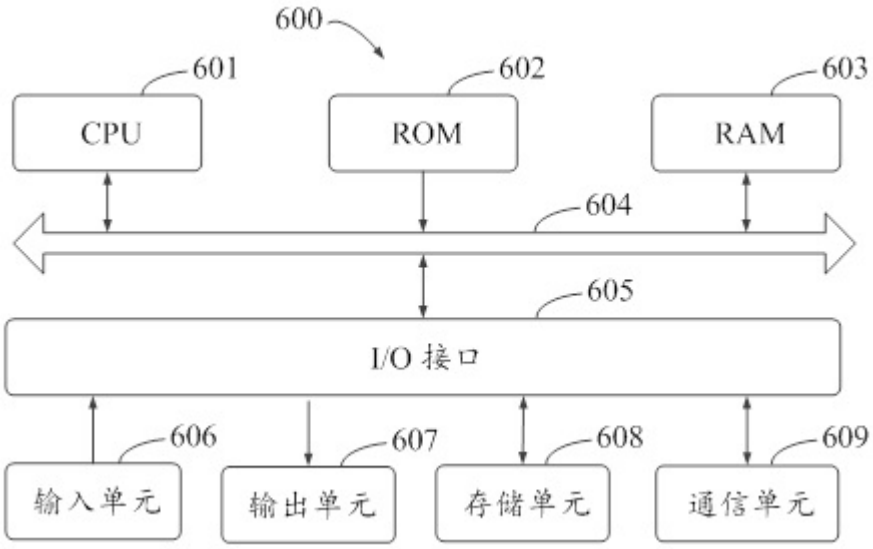

600
601
CPU
602
ROM
603
RAM
604
605
I/O 接口
606
输入单元
607
输出单元
608
存储单元
609
通信单元

图6