(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2019 年 7 月 4 日 (04.07.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/128529 A1

(51) 国际专利分类号:
G06F 21/55 (2013.01)

(21) 国际申请号: PCT/CN2018/116100

(22) 国际申请日: 2018 年 11 月 19 日 (19.11.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201711463325.3 2017年12月28日 (28.12.2017) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (**ALIBABA GROUP HOLDING LIMITED**) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 李龙飞 (**LI, Longfei**); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (**BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION**); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

(54) **Title:** URL ATTACK DETECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

(54) 发明名称: URL攻击检测方法、装置以及电子设备

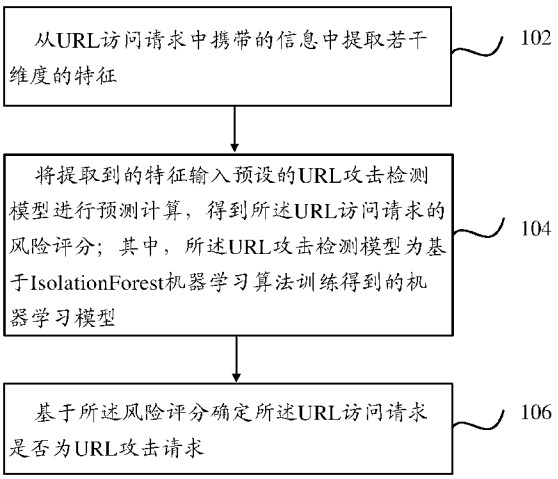

图 1

(57) **Abstract:** A URL attack detection method, comprising: extracting several dimensions of features from information carried in a URL access request (102); inputting the extracted features to a preset URL attack detection model for performing forecast calculation to obtain a risk score of the URL access request, the URL attack detection model being a machine learning model obtained by training on the basis of an Isolation Forest machine learning algorithm (104); and determining, on the basis of the risk score, whether the URL access request is a URL attack request (106).

(57) 摘要: 一种URL攻击检测方法，包括：从URL访问请求中携带的信息中提取若干维度的特征（102）；将提取到的特征输入预设的URL攻击检测模型进行预测计算，得到所述URL访问请求的风险评分；其中，所述URL攻击检测模型为基于Isolation Forest机器学习算法训练得到的机器学习模型（104）；基于所述风险评分确定所述URL访问请求是否为URL攻击请求（106）。

WO 2019/128529 A1

[见续页]

PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) **指定国**(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

URL 攻击检测方法、装置以及电子设备

技术领域

本说明书涉及计算机应用领域，尤其涉及一种 URL 攻击检测方法、装置、以及电子设备。

背景技术

在互联网的应用场景中，每天都会产生大量的对于网址的 URL 访问请求。在这些大量的 URL 访问请求中，也不乏不法分子试图通过不合法的 URL 访问请求而发起的 URL 攻击；例如，常见的 URL 攻击如木马攻击、SQL 注入攻击、跨站脚本攻击（XSS）等。这一类非法的 URL 访问请求，通常会与普通的 URL 访问请求存在一定的区别；因此，在构建线上系统的同时，通过一些安全手段对非法用户发起的 URL 攻击进行快速的识别检测是不可忽视的问题。

发明内容

本说明书提出一种 URL 攻击检测方法，所述方法包括：

从 URL 访问请求中携带的信息中提取若干维度的特征；

将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型；

基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

可选的，所述方法还包括：

从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征；其中，所述若干 URL 访问请求样本均未被标记样本标签。

基于提取到的特征构建若干训练样本；

基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型。

可选的，所述 URL 攻击检测模型包括基于 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树；

所述基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型，包括：

基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训练样本子集；

从所述若干维度的特征中为各训练样本子集随机选择一分类特征作为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间中，为各训练样本子集随机选取一分类临界值；

将各训练样本子集中所述分类特征的取值大于所述分类临界值的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，分别分类为所述根节点的叶节点；以及，

将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

可选的，所述将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分，包括：

基于提取到的特征构建预测样本；

基于所述预测样本中各特征的取值，从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点；

计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值，并对所述平均值进行归一化处理，得到所述 URL 访问请求的风险评分。

可选的，所述信息包括：域名信息，和/或 URL 参数；所述若干维度的特征包括：从 URL 访问请求中携带的域名信息中提取出的特征；和/或从 URL 访问请求中携带的 URL 参数中提取出的特征。

可选的，所述特征包括以下特征中的多个的组合：字符总数、字母总数、

数字总数、符号总数、不同字符数、不同字母数、不同数字数、不同符号数。

本说明书还提出一种 URL 攻击检测装置，所述装置包括:

第一提取模块，从 URL 访问请求中携带的信息中提取若干维度的特征;

计算模块，将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型;

确定模块，基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

可选的，所述装置还包括:

第二提取模块，从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征；其中，所述若干 URL 访问请求样本均未被标记样本标签。

构建模块，基于提取到的特征构建若干训练样本;

训练模块，基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型。

可选的，所述 URL 攻击检测模型包括基于 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树;

所述训练模块:

基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训练样本子集;

从所述若干维度的特征中为各训练样本子集随机选择一分类特征作为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间中，为各训练样本子集随机选取一分类临界值;

将各训练样本子集中所述分类特征的取值大于所述分类临界值的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，分别分类为所述根节点的叶节点；以及，

将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

可选的，所述计算模块：

基于提取到的特征构建预测样本；

基于所述预测样本中各特征的取值，从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点；

计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值，并对所述平均值进行归一化处理，得到所述 URL 访问请求的风险评分。

可选的，所述信息包括：域名信息，和/或 URL 参数；所述若干维度的特征包括：从 URL 访问请求中携带的域名信息中提取出的特征；和/或从 URL 访问请求中携带的 URL 参数中提取出的特征。

可选的，所述特征包括以下特征中的多个的组合：字符总数、字母总数、数字总数、符号总数、不同字符数、不同字母数、不同数字数、不同符号数。

本说明书还提出一种电子设备，包括：

处理器；

用于存储机器可执行指令的存储器；

其中，通过读取并执行所述存储器存储的与 URL 攻击检测的控制逻辑对应的机器可执行指令，所述处理器被促使：

从 URL 访问请求中携带的信息中提取若干维度的特征；

将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型；

基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

本说明书实施例提供的技术方案，通过将从 URL 访问请求中提取出的特征输入至基于 Isolation Forest 机器学习算法训练出的 URL 攻击检测模型进行预测计算，来对 URL 访问请求进行攻击检测，可以提前发现潜在的 URL 攻击，从而有助于对潜在的异常 URL 访问及时的进行安全防护。

附图说明

图 1 是本说明书一实施例示出的 URL 攻击检测方法的流程图;

图 2 是本说明书一实施例示出的一种构建训练样本集训练 Isolation Forest 模型的流程图;

图 3 是本说明书一实施例提供的承载一种 URL 攻击检测装置的电子设备所涉及的硬件结构图;

图 4 是本说明书一实施例提供的一种所述 URL 攻击检测装置的逻辑框图。

具体实施方式

本说明书旨在提出一种基于 Isolation Forest（孤立森林）机器学习算法对均未被标记风险标签的 URL 访问请求样本进行机器学习训练，来构建 URL 攻击检测模型，并使用该 URL 攻击检测模型对正常的 URL 访问请求进行攻击检测，来发现潜在的 URL 攻击的技术方案。

在实现时，可以预先准备若干 URL 访问请求样本；其中，这些 URL 访问请求样本均未被标记风险标签。然后，可以对这些 URL 访问请求样本进行数据切分，从这些 URL 访问请求样本中携带的信息中提取出若干维度的特征;

例如，在实际应用中，上述信息具体可以包括域名信息、URL 参数，在这种情况下，可以对 URL 访问请求样本进行数据切分，提取出 URL 访问请求与样本中携带的域名信息（比如主域名和对应的域名后缀）、URL 参数（比如 URL 参数名和对应的参数取值），然后从提取出的域名信息、URL 参数中提取出若干个维度的特征。

进一步，当从 URL 访问请求样本中，分别提取出若干个维度的特征后，可以对这些特征进行归一化处理，然后将归一化处理后的特征作为建模特征来构建训练样本。

当训练样本构建完成后，可以基于 Isolation Forest 机器学习算法对这些训练样本进行训练，来构建 URL 攻击检测模型；例如，可以采用 Isolation

Forest 机器学习算法对训练样本进行二叉树分类，构建出多颗随机二叉树。

最后，当 URL 攻击检测模型训练完成后，可以按照相同的方式，从需要进行攻击检测的 URL 访问请求携带的信息中分别提取出若干维度的特征，并基于提取出的特征来构建预测样本，将构建完成的预测样本输入至上述 URL 攻击检测模型中进行预测计算，得到该 URL 访问请求的风险评分，然后可以基于该风险评分来确定该 URL 访问请求是否为 URL 攻击请求。

在以上技术方案中，通过将从 URL 访问请求中提取出的特征输入至基于 Isolation Forest 机器学习算法训练出的 URL 攻击检测模型进行预测计算，来对 URL 访问请求进行攻击检测，可以提前发现潜在的 URL 攻击，从而有助于对潜在的异常 URL 访问及时的进行安全防护。

下面通过具体实施例并结合具体的应用场景对本说明书进行描述。

请参考图 1，图 1 是本说明书一实施例提供的一种 URL 攻击检测方法，执行以下步骤：

步骤 102，从 URL 访问请求中携带的信息中提取若干维度的特征；

步骤 104，将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型；

步骤 106，基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

在本说明书中，建模方可以预先收集大量的未进行标记的 URL 访问请求作为无标记样本，并基于收集到的这些无标记样本来构建训练样本集，然后基于 Isolation Forest 机器学习算法对该训练样本集进行无监督的机器学习训练，来构建上述 URL 攻击检测模型。

请参见图 2，图 2 为本说明书示出的一种构建训练样本集训练 Isolation Forest 模型的流程图。

如图 2 所示，首先，可以对收集到的这些未进行标记的原始的 URL 访问请求样本分别进行数据切分，提取出这些 URL 访问请求样本中携带的信息。

其中，上述 URL 访问请求中携带的信息是指那些能够从中提取出，可以反映 URL 访问请求是否存在风险的特征的信息。

在示出的一种实施方式中，上述信息具体可以包括 URL 访问请求中携带的 URL 参数和域名信息等。上述 URL 参数，可以包括 URL 参数名（ParamName）以及对应的参数取值（ParamValue）；而上述域名信息，可以包括主域名和与主域名对应的域名后缀。

例如，以上述信息为 URL 访问请求中携带的 URL 参数为例，可以对原始的 URL 访问请求样本进行数据切分，提取出这些 URL 访问请求样本中携带的 URL 参数名（ParamName）以及对应的参数取值（ParamValue）；

又如，以上述信息为 URL 访问请求中携带的信息为例，可以对原始的 URL 访问请求样本进行数据切分。提取出 URL 访问请求中携带的主域名和与主域名对应的域名后缀。当提取出这些 URL 访问请求样本中携带的信息后，可以从这些信息中筛选出已知的 URL 攻击请求中较为常见的那一部分信息，用以构建机器学习模型。即筛选出最能够表征 URL 攻击请求的特征的信息，来参与建模。

例如，以上述信息为 URL 访问请求中携带的 URL 参数为例，对于部分只在个别的 URL 访问请求中出现的特殊 URL 参数，由于这部分 URL 参数并不能真实反映出 URL 攻击请求的特征，因此对于这部分 URL 参数可以进行过滤。

又如，以上述信息为 URL 访问请求中携带的信息为例，对于部分只在个别的 URL 访问请求中出现的特殊信息，由于这部分信息并不能真实反映出 URL 攻击请求的特征，参与建模会对模型的结果造成干扰，因此对于这部分信息可以进行过滤处理。

进一步的，对于筛选出的信息，可以从这些信息中分别提取出若干个维度的特征，来作为建模特征。

其中，需要说明的是，建模方在建模时，从 URL 访问请求样本中提取出的信息具体可以采用 URL 访问请求样本中携带的域名信息和 URL 参数中

的其中一个，也可以同时采用上述域名信息和 URL 参数作为信息。

因而，在这种情况下，建模方在从信息中提取到的特征，则可以包括以下示出的三种情况:

在一种情况下，如果建模方采用 URL 访问请求样本中携带的域名信息作为上述信息，那么最终提取到的特征，可以仅包括从 URL 访问请求样本中携带的域名信息中提取出的若干维度的特征;

在另一种情况下，如果建模方采用 URL 访问请求样本中携带的 URL 参数作为上述信息，那么最终提取到的特征，可以仅包括从 URL 访问请求样本中携带的 URL 参数中提取出的若干维度的特征;

在第三种情况下，如果建模方同时采用 URL 访问请求样本中携带的 URL 参数和域名信息作为信息，此时上述 URL 参数和上述域名信息将同时参与建模，那么最终提取到的特征，可以同时包括从 URL 访问请求样本中携带的 URL 参数和域名信息中分别提取出的若干维度的特征；其中，从这些信息中提取出的特征，在本说明书中不进行特殊限定，在实际应用中，任意形式的能够表征 URL 攻击请求中携带的信息的特征以及规律的特征，都可以被选定作为建模特征。

例如，在实际应用中，参与建模的本领域技术人员，可以基于经验从这些信息中提取出若干个维度的特征，然后基于这些特征进行尝试建模，并对建模结果进行评估，来从中筛选出对模型的贡献度最高的若干个维度的特征作为建模特征。

在示出的一种实施方式中，从这些信息中提取出的特征，可以包括信息的字符总数、信息的字母总数、信息的数字总数、信息的符号总数、信息的不同字符数、信息的不同字母数、信息的不同数字数、信息的不同符号数等 8 个维度。

例如，如果建模方采用 URL 访问请求样本中携带的域名信息作为上述信息，最终提取到的特征可以包括域名信息的字符总数、域名信息的字母总数、域名信息的数字总数、域名信息的符号总数、域名信息的不同字符数、域名

信息的不同字母数、域名信息的不同数字数、域名信息的不同符号数等 8 个维度；

如果建模方采用 URL 访问请求样本中携带的 URL 参数作为上述信息，最终提取到的特征可以包括 URL 参数的字符总数、URL 参数的字母总数、URL 参数的数字总数、URL 参数的符号总数、URL 参数的不同字符数、URL 参数的不同字母数、URL 参数的不同数字数、URL 参数的不同符号数等 8 个维度；

如果建模方同时采用 URL 访问请求样本中携带的 URL 参数和域名信息作为信息，最终提取到的特征可以包括 URL 参数的字符总数、URL 参数的字母总数、URL 参数的数字总数、URL 参数的符号总数、URL 参数的不同字符数、URL 参数的不同字母数、URL 参数的不同数字数、URL 参数的不同符号数、域名信息的字符总数、域名信息的字母总数、域名信息的数字总数、域名信息的符号总数、域名信息的不同字符数、域名信息的不同字母数、域名信息的不同数字数、域名信息的不同符号数等 16 个维度。

其中，需要说明的是，在实际应用中，本领域技术人员可以将以上 8 个基础维度进行组合作为建模特征，或者从以上 8 个基础维度中进一步筛选出多个维度进行组合作为建模特征，在本说明书中不进行特别限定。

当然，以上示出的 8 个维度的特征仅为示例性的；显然，在实际应用中，本领域技术人员也可以从这些信息中提取出以上 8 个维度以外的其它维度的特征作为建模特征，在本说明书中不再进行一一列举。

请继续参见图 2，当从筛选出的信息中分别提取出若干个维度的特征后，由于不同的特征的取值范围可能并不统一，因此还可以对这些维度的特征进行归一化处理，将不同的特征的取值范围归一化到一个统一的数值区间，从而来消除由于特征的取值范围不同对建模精度造成的影响。

当对提取出的特征归一化处理完成之后，可以基于从各 URL 访问请求样本携带的信息中提取出的特征，为各 URL 访问请求样本分别创建一个对应的特征向量作为训练样本；其中，创建的特征向量的维度，与提取出的特征的

维度相同。

当为各 URL 访问请求样本构建了对应的特征向量后，此时可以基于为各 URL 访问请求样本构建的特征向量，创建一个目标矩阵；例如，假设共计收集到 N 条 URL 访问请求样本，从每一个 URL 访问请求样本提取出 M 维的特征，那么该目标矩阵具体可以是一个 N*M 维的目标矩阵。

此时，创建的该目标矩阵，即为最终参与机器学习模型训练的训练样本集。

请继续参见图 2，当训练样本集训练完毕，可以基于 Isolation Forest 机器学习算法对这些训练样本进行训练，来构建上述 URL 攻击检测模型。其中，Isolation Forest 算法是一种通过构建多个随机二叉树，从原始的数据集中挖掘出异常数据样本的算法。所谓随机二叉树，是指基于随机生成的分类特征，以及随机生成的与分类特征的取值对应的分类临界值构建而成的二叉树。即在构建随机二叉树时，所使用的分类特征以及与分类特征的取值对应的分类临界值均为随机生成的。

而利用 Isolation Forest 算法对构建完成的训练样本集进行训练，来构建 URL 异常检测模型的过程，即为利用 Isolation Forest 算法对训练样本集中的训练样本进行分类，构建 M 棵随机二叉树的过程。

在初始状态，建模方在基于 Isolation Forest 算法对上述训练样本集进行训练之前，需要对 Isolation Forest 算法进行参数配置，为 Isolation Forest 算法配置需要构建的随机二叉树个数 M，以及在构建单棵随机二叉树时需要从训练样本集中抽样的训练样本数 N。

其中，上述 M 和 N 的取值，可以采用工程经验值，或者基于建模方实际的需求进行自定义设置；例如，Isolation Forest 算法默认需要构建的随机二叉树个数为 100，每一刻随机二叉树需要采样的训练样本数为 256。

当建模方完成对 Isolation Forest 算法的参数配置后，建模方可以通过在搭建的计算平台（比如服务器集群）中运行 Isolation Forest 算法，对构建完成的训练样本集进行训练，来构建最终的 URL 异常检测模型。

以下对利用 Isolation Forest 算法对训练样本集中的训练样本进行分类，来构建随机二叉树的流程，进行详细描述。

首先，可以基于建模方配置的上述 N 值，针对训练样本集进行 M 次的均匀抽样。其中，所述均匀抽样，是指在执行的 M 次抽样中，每一次从训练样本集中抽样出的训练样本集的数量都相同。

当完成训练样本的均匀抽样后，可以基于抽样出的训练样本，来构建出 M 个训练样本子集， 然后针对每一个训练样本子集中的训练样本分别进行分类，来构建出 M 棵随机二叉树。

进一步的，在针对一个训练样本子集中的训练样本进行分类，来构建随机二叉树时，首先可以从构成训练样本的若干维度的特征中，为该训练样本子集随机选择一个特征作为分类特征，并将该分类特征作为根节点；以及，确定该分类特征当前在该训练样本子集中的最大取值和最小取值，然后在该最大取值和最小取值构成的取值区间中，为该训练样本集随机选取一分类临界值。

当选定了作为根节点的分类特征以及分类临界值后，此时可以针对该训练样本子集执行第一级的分类，将该训练样本子集中各个训练样本的上述分类特征的取值，分别与上述分类临界值进行比较，然后基于比较结果将该训练样本子集中的训练样本分类为，上述分类特征的取值大于上述分类临界值的训练样本，和上述分类特征小于上述分类临界值的训练样本两类，并将分类出的这两类训练样本，分别作为上述根节点的叶节点。

例如，在实现时，可以将该训练样本子集中上述分类特征的取值小于上述分类临界值的训练样本，分类到二叉树的左树分支，将这一类训练样本作为上述根节点在二叉树上的左树叶节点；而将该训练样本子集中上述分类特征的取值大于上述分类临界值的训练样本，分类到二叉树的右树分支，将这一类训练样本作为上述根节点在二叉树上的右树叶节点。

此时针对该训练样本子集的第一级分类完成。

进一步，当第一级分类完成后，可以继续完成针对上述训练样本子集的

第二级分类。

此时，可以将已经分类得到的两个叶节点中的训练样本，分别作为新的训练样本子集，然后针对上述新的训练样本子集，来迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止；

例如，仍然可以采用相同的方式，为各新的训练样本子集随机选择分类特征以及分类临界值，然后将各新的训练样本子集中的训练样本分类为，上述分类特征的取值大于上述分类临界值的训练样本，和上述分类特征小于上述分类临界值的训练样本两类，并将分类出的这两类训练样本，分别作为上一级的叶节点的下一级叶节点，以此类推，直到在执行某一级的分类后，得到的下一级的叶节点中的训练样本不可再分时停止；比如，叶节点中只剩一个训练样本，或者叶节点中的训练样本完全相同，表明得到的叶节点中的训练样本已经不可以再继续分类。

其中，需要说明的是，为根节点以及各级子节点随机选择的分类特征，需要保持不同；例如，在一种实现方式中，在将某一个特征选择为随机二叉树中某一节点的分类特征后，可以将该特征移除，后续在为其它节点选择分类特征时，将在该特征以外的其它特征中过来进行随机选择。

另外，以上示出的 Isolation Forest 算法的迭代分类的停止条件，默认情况下可以是得到的叶节点中的训练样本已经不可以再继续分类，在实际应用中，建模方也可以在为 Isolation Forest 算法配置算法参数时，可以为得到的随机二叉树配置一个最大的二叉树深度(即从根节点开始节点的最大层数)。在这种情况下，上述停止条件，也可以是当通过上述迭代分类的过程，得到的随机二叉树的深度达到为算法配置的最大的二叉树深度时，算法可以立即停止（此时得到的各叶节点中的训练样本可能仍然可以再继续分类）。

以上示出的为针对其中一个训练样本子集中的训练样本进行迭代分类，构建单棵随机二叉树的过程。

相似的，可以针对每一个训练样本子集重复执行以上分类过程，最终可以基于上述 M 个训练样本子集，构建出 M 棵随机二叉树，此时针对上述的

训练样本集的训练完成，得到的上述 M 棵随机二叉树，即为最终构建出的 URL 异常该检测模型。

在本说明书中，当上述 URL 攻击检测模型训练完毕后，可以按照如图 2 示出的相同的特征提取方式，从需要进行攻击检测的 URL 访问请求提取信息，从提取到的信息中筛选信息、从筛选出的信息中提取若干个维度的特征（与模型训练阶段的特征一致），然后基于提取到的特征构建预测样本，并将预测样本输入至上述 URL 攻击检测模型进行预测计算，得到该 URL 访问请求的风险评分。

以下对利用训练完成的 URL 攻击检测模型对 URL 访问请求进行风险评分的流程，进行详细描述。

在计算构建出的预测样本的风险评分时，首先需要估算出该预测样本在每颗随机二叉树中的路径深度 h(x);

具体的，可以基于该预测样本中各特征的取值，从各棵随机二叉树的根节点开始，按照由上至下的顺序遍历整棵随机二叉树，来查找该预测样本在随机二叉树中对应的叶节点;

例如，首先可以确定该预测样本中与根节点的分类特征对应的取值，然后基于该取值，来查找该预测样本所在的第一级叶节点。在查找到第一级叶节点后，可以继续确定该预测样本中与该第一级叶的分类特征对应的取值，然后基于该取值，继续查找该预测样本所在的第二级叶节点，以此类推，可以通过逐级遍历，直到查找到该预测样本对应的叶节点时停止。

当查找到与上述预测样本对应的叶节点后，此时可以记录在遍历随机二叉树的过程中，从根节点到查找到的该叶节点之间一共经过的边的数目 e，以及与上述预测样本对应的叶节点中的训练样本数 n。

此时最终得到的路径深度 h(x)，可以用如下公式来表征:

$$h(x)= e+C（n）$$

其中，C（n）为修正值，可以用如下公式来表征:

$$C(n) = 2H(n-1) - \frac{2(n-1)}{n}$$

其中，H(n-1) 可用 ln(n-1)+0.5772156649 估算，这里的常数是欧拉常数。

当通过以上公式，估算出该预测样本在每颗随机二叉树中的路径深度 h(x)后，可以进一步求解该预测样本在每颗随机二叉树的路径深度的平均值，然后对得到的平均值进行归一化处理，将计算结果量化到 0~1 之间，得到该 URL 访问请求的风险评分；

最终得到的风险评分可以用如下公式进行表征：

$$Score(x) = 2^{-\frac{E\{h(x)\}}{C(\phi)}}$$

其中，Score(x)表示预测样本 X 最终的风险评分；E{h(x)}表示预测样本在每颗随机二叉树中的路径深度 h(x)；φ表示单棵随机二叉树的训练样本数；C(φ)表示用φ条训练样本构建的二叉树的平均路径长度，在上述公式中用来对计算结果作归一化处理。

当通过上述 URL 攻击检测模型预测出该 URL 访问请求的风险评分后，可以进一步基于该 URL 风险评分，来确定该 URL 访问请求是否为 URL 攻击请求；

例如，在一种实现方式中，可以将该风险评分与预设的风险阈值进行比较，来确定该 URL 访问请求的具体类型；如果该风险评分大于或者等于预设的风险阈值，则表明该 URL 访问请求为 URL 攻击请求；反之，如果该风险评分小于该预设的风险阈值，则表明该 URL 访问请求为正常 URL 访问请求。

通过以上实施例可知，在本说明书中，通过将从 URL 访问请求中提取出的特征输入至基于 Isolation Forest 机器学习算法训练出的 URL 攻击检测模型进行预测计算，来对 URL 访问请求进行攻击检测：

一方面，通过这种方式，可以提前发现潜在的 URL 攻击，从而有助于对潜在的异常 URL 访问及时的进行安全防护。

另一方面，由于 Isolation Forest 算法是一种无监督的机器学习算法，在训练模型时所需的训练样本可以不再需要标记样本标签，因此对于建模方而

言，可以省去为训练样本打标而造成的大量人力成本。

与上述方法实施例相对应，本说明书还提供了一种 URL 攻击检测装置的实施例。本说明书的 URL 攻击检测设备的实施例可以应用在电子设备上。装置实施例可以通过软件实现，也可以通过硬件或者软硬件结合的方式实现。以软件实现为例，作为一个逻辑意义上的装置，是通过其所在电子设备的处理器将非易失性存储器中对应的计算机程序指令读取到内存中运行形成的。从硬件层面而言，如图 3 所示，为本说明书的 URL 攻击检测装置所在电子设备的一种硬件结构图，除了图 3 所示的处理器、内存、网络接口、以及非易失性存储器之外，实施例中装置所在的电子设备通常根据该电子设备的实际功能，还可以包括其他硬件，对此不再赘述。

图 4 是本说明书一示例性实施例示出的一种 URL 攻击检测装置的框图。

请参考图 4，所述 URL 攻击检测装置 40 可以应用在前述图 3 所示的电子设备中，包括有：第一提取模块 401、计算模块 402 和确定模块 403。

第一提取模块 401，从 URL 访问请求中携带的信息中提取若干维度的特征；

计算模块 402，将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型；

确定模块 403，基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

在本例中，所述装置 40 还包括：

第二提取模块 404（图 4 中未示出），从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征；其中，所述若干 URL 访问请求样本均未被标记样本标签。

构建模块 405（图 4 中未示出），基于提取到的特征构建若干训练样本；

训练模块 406（图 4 中未示出），基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型。

在本例中，所述 URL 攻击检测模型包括基于 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树；

所述训练模块 406：

基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训练样本子集；

从所述若干维度的特征中为各训练样本子集随机选择一分类特征作为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间中，为各训练样本子集随机选取一分类临界值；

将各训练样本子集中所述分类特征的取值大于所述分类临界值的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，分别分类为所述根节点的叶节点；以及，

将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

在本例中，所述计算模块 403：

基于提取到的特征构建预测样本；

基于所述预测样本中各特征的取值，从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点；

计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值，并对所述平均值进行归一化处理，得到所述 URL 访问请求的风险评分。

在本例中，所述信息包括：域名信息，和/或 URL 参数；所述若干维度的特征包括：从 URL 访问请求中携带的域名信息中提取出的特征；和/或从 URL 访问请求中携带的 URL 参数中提取出的特征。

在本例中，所述特征包括以下特征中的多个的组合：字符总数、字母总数、数字总数、符号总数、不同字符数、不同字母数、不同数字数、不同符号数。

上述装置中各个模块的功能和作用的实现过程具体详见上述方法中对应步骤的实现过程，在此不再赘述。

对于装置实施例而言，由于其基本对应于方法实施例，所以相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本说明书方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机，计算机的具体形式可以是个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件收发设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任意几种设备的组合。

与上述方法实施例相对应，本说明书还提供了一种电子设备的实施例。该电子设备包括：处理器以及用于存储机器可执行指令的存储器；其中，处理器和存储器通常通过内部总线相互连接。在其他可能的实现方式中，所述设备还可能包括外部接口，以能够与其他设备或者部件进行通信。

在本实施例中，通过读取并执行所述存储器存储的与 URL 攻击检测的控制逻辑对应的机器可执行指令，所述处理器被促使：

从 URL 访问请求中携带的信息中提取若干维度的特征；

将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型；

基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

在本例中，通过读取并执行所述存储器存储的 URL 攻击检测的控制逻辑对应的机器可执行指令，所述处理器还被促使：

从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征；其中，

所述若干 URL 访问请求样本均未被标记样本标签。

基于提取到的特征构建若干训练样本;

基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型。

在本实施例中，所述 URL 攻击检测模型包括基于 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树;

通过读取并执行所述存储器存储的 URL 攻击检测的控制逻辑对应的机器可执行指令，所述处理器还被促使:

基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训练样本子集;

从所述若干维度的特征中为各训练样本子集随机选择一分类特征作为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间中，为各训练样本子集随机选取一分类临界值;

将各训练样本子集中所述分类特征的取值大于所述分类临界值的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，分别分类为所述根节点的叶节点；以及，

将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

在本例中，通过读取并执行所述存储器存储的 URL 攻击检测的控制逻辑对应的机器可执行指令，所述处理器还被促使:

基于提取到的特征构建预测样本;

基于所述预测样本中各特征的取值，从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点;

计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值，并对所述平均值进行归一化处理，得到所述 URL 访问请求的风险评分。

在本例中，所述信息包括：域名信息，和/或 URL 参数；所述若干维度的特征包括：从 URL 访问请求中携带的域名信息中提取出的特征；和/或从

URL 访问请求中携带的 URL 参数中提取出的特征。

在本例中，提取出的所述若干维度的特征包括以下特征中的多个的组合：信息的字符总数、信息的字母总数、信息的数字总数、信息的符号总数、信息的不同字符数、信息的不同字母数、信息的不同数字数、信息的不同符号数。

本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本说明书的其它实施方案。本说明书旨在涵盖本说明书的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本说明书的一般性原理并包括本说明书未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本说明书的真正范围和精神由下面的权利要求指出。

应当理解的是，本说明书并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本说明书的范围仅由所附的权利要求来限制。

以上所述仅为本说明书的较佳实施例而已，并不用以限制本说明书，凡在本说明书的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本说明书保护的范围之内。

权利要求书

1. 一种 URL 攻击检测方法，所述方法包括:

从 URL 访问请求中携带的信息中提取若干维度的特征;

将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于孤立森林 Isolation Forest 机器学习算法训练得到的机器学习模型;

基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

2. 根据权利要求 1 所述的方法，所述方法还包括:

从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征;其中，所述若干 URL 访问请求样本均未被标记样本标签;

基于提取到的特征构建若干训练样本;

基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型。

3. 根据权利要求 2 所述的方法，所述 URL 攻击检测模型包括基于 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树;

所述基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型，包括:

基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训练样本子集;

从所述若干维度的特征中为各训练样本子集随机选择一分类特征作为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间中，为各训练样本子集随机选取一分类临界值;

将各训练样本子集中所述分类特征的取值大于等于所述分类临界值的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，分别分类为所述根节点的叶节点；以及，

将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

4. 根据权利要求 3 所述的方法，所述将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分，包括：

基于提取到的特征构建预测样本；

基于所述预测样本中各特征的取值，从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点；

计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值，并对所述平均值进行归一化处理，得到所述 URL 访问请求的风险评分。

5. 根据权利要求 1 所述的方法，所述信息包括：域名信息，和/或 URL 参数；所述若干维度的特征包括：从 URL 访问请求中携带的域名信息中提取出的特征；和/或从 URL 访问请求中携带的 URL 参数中提取出的特征。

6. 根据权利要求 5 所述的方法，所述特征包括以下特征中的多个的组合：字符总数、字母总数、数字总数、符号总数、不同字符数、不同字母数、不同数字数、不同符号数。

7. 一种 URL 攻击检测装置，所述装置包括：

第一提取模块，从 URL 访问请求中携带的信息中提取若干维度的特征；

计算模块，将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型；

确定模块，基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

8. 根据权利要求 7 所述的装置，所述装置还包括：

第二提取模块，从若干 URL 访问请求样本携带的信息中分别提取若干维度的特征；其中，所述若干 URL 访问请求样本均未被标记样本标签；

构建模块，基于提取到的特征构建若干训练样本；

训练模块，基于 Isolation Forest 机器学习算法对所述若干训练样本进行训练得到所述 URL 攻击检测模型。

9. 根据权利要求 8 所述的装置，所述 URL 攻击检测模型包括基于 Isolation Forest 机器学习算法训练得到的 M 棵随机二叉树；

所述训练模块：

基于从所述若干训练样本中均匀抽样出的训练样本构建出 M 个训练样本子集；

从所述若干维度的特征中为各训练样本子集随机选择一分类特征作为根节点，以及在所述分类特征的最大取值和最小取值构成的取值区间中，为各训练样本子集随机选取一分类临界值；

将各训练样本子集中所述分类特征的取值大于所述分类临界值的训练样本，和所述分类特征的取值小于所述分类临界值的训练样本，分别分类为所述根节点的叶节点；以及，

将各叶节点中的训练样本作为新的训练样本子集，迭代执行以上分类过程，直到得到的各叶节点中的训练样本不可再分类时停止。

10. 根据权利要求 9 所述的装置，所述计算模块：

基于提取到的特征构建预测样本；

基于所述预测样本中各特征的取值，从根节点开始遍历各棵随机二叉树查找与所述预测样本对应的叶节点；

计算查找到的叶节点在各棵随机二叉树中的路径深度的平均值，并对所述平均值进行归一化处理，得到所述 URL 访问请求的风险评分。

11. 根据权利要求 7 所述的装置，所述信息包括：域名信息，和/或 URL 参数；所述若干维度的特征包括：从 URL 访问请求中携带的域名信息中提取出的特征；和/或从 URL 访问请求中携带的 URL 参数中提取出的特征。

12. 根据权利要求 11 所述的装置，所述特征包括以下特征中的多个的组合：字符总数、字母总数、数字总数、符号总数、不同字符数、不

同字母数、不同数字数、不同符号数。

13. 一种电子设备，包括:

处理器;

用于存储机器可执行指令的存储器;

其中，通过读取并执行所述存储器存储的与 URL 攻击检测的控制逻辑对应的机器可执行指令，所述处理器被促使:

从 URL 访问请求中携带的信息中提取若干维度的特征;

将提取到的特征输入预设的 URL 攻击检测模型进行预测计算，得到所述 URL 访问请求的风险评分；其中，所述 URL 攻击检测模型为基于 Isolation Forest 机器学习算法训练得到的机器学习模型;

基于所述风险评分确定所述 URL 访问请求是否为 URL 攻击请求。

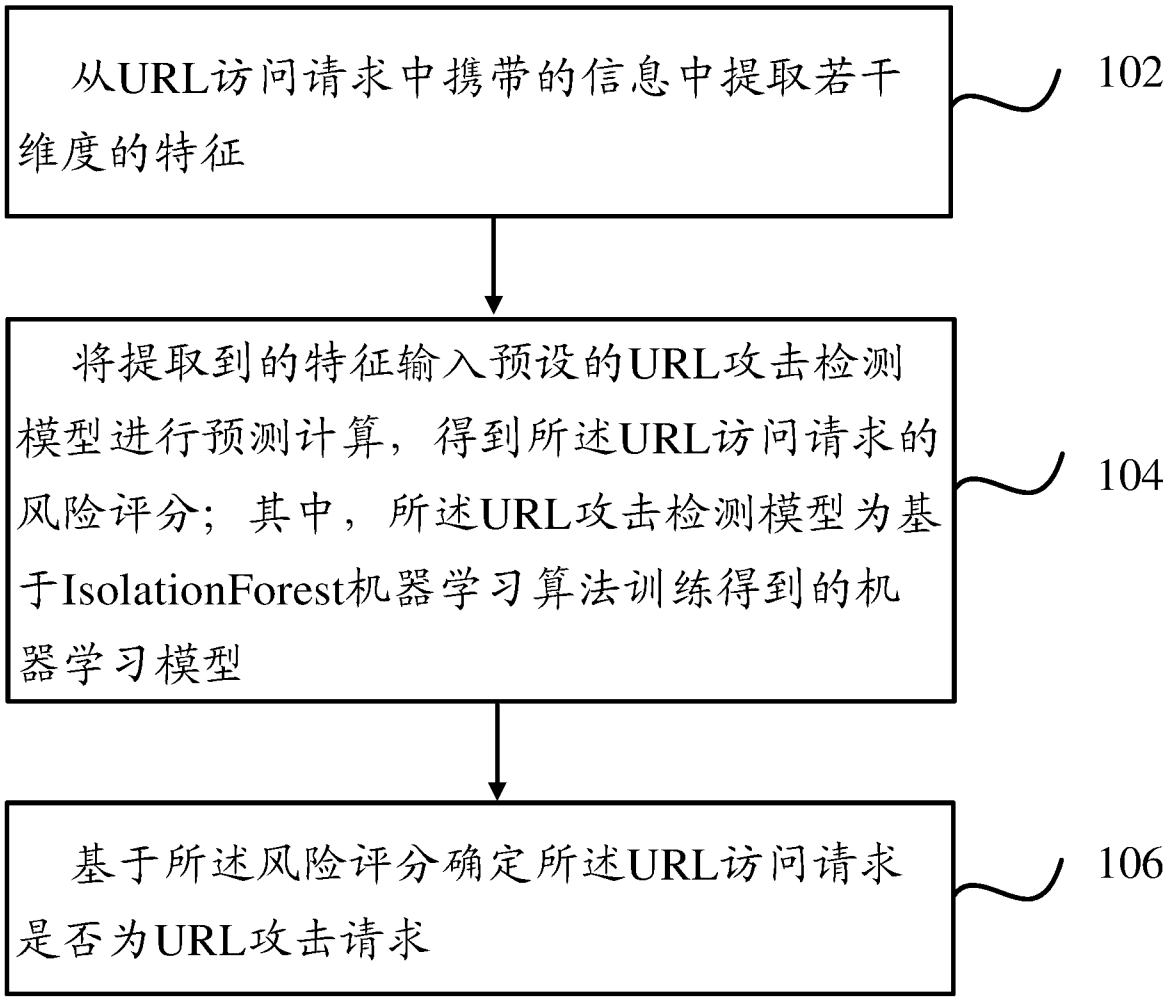
从URL访问请求中携带的信息中提取若干维度的特征
102
将提取到的特征输入预设的URL攻击检测模型进行预测计算，得到所述URL访问请求的风险评分；其中，所述URL攻击检测模型为基于IsolationForest机器学习算法训练得到的机器学习模型
104
基于所述风险评分确定所述URL访问请求是否为URL攻击请求
106

图 1

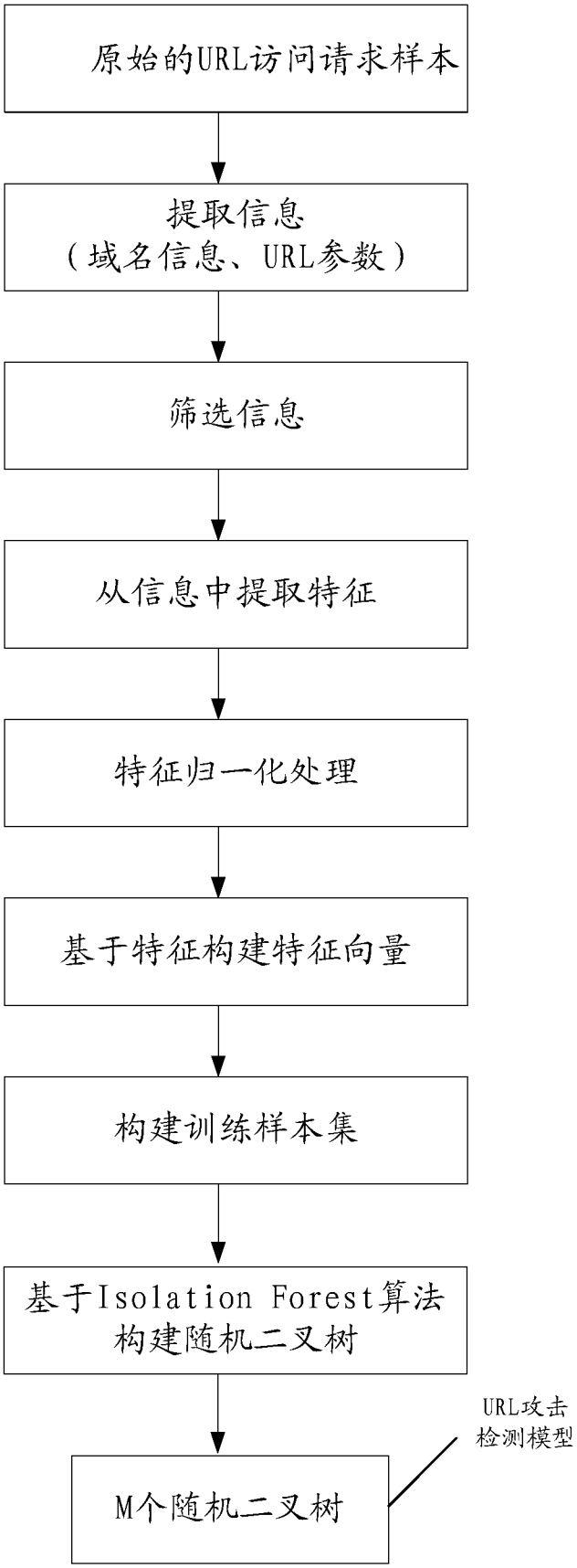
原始的URL访问请求样本
提取信息
（域名信息、URL参数）
筛选信息
从信息中提取特征
特征归一化处理
基于特征构建特征向量
构建训练样本集
基于Isolation Forest算法
构建随机二叉树
M个随机二叉树
URL攻击
检测模型

图 2

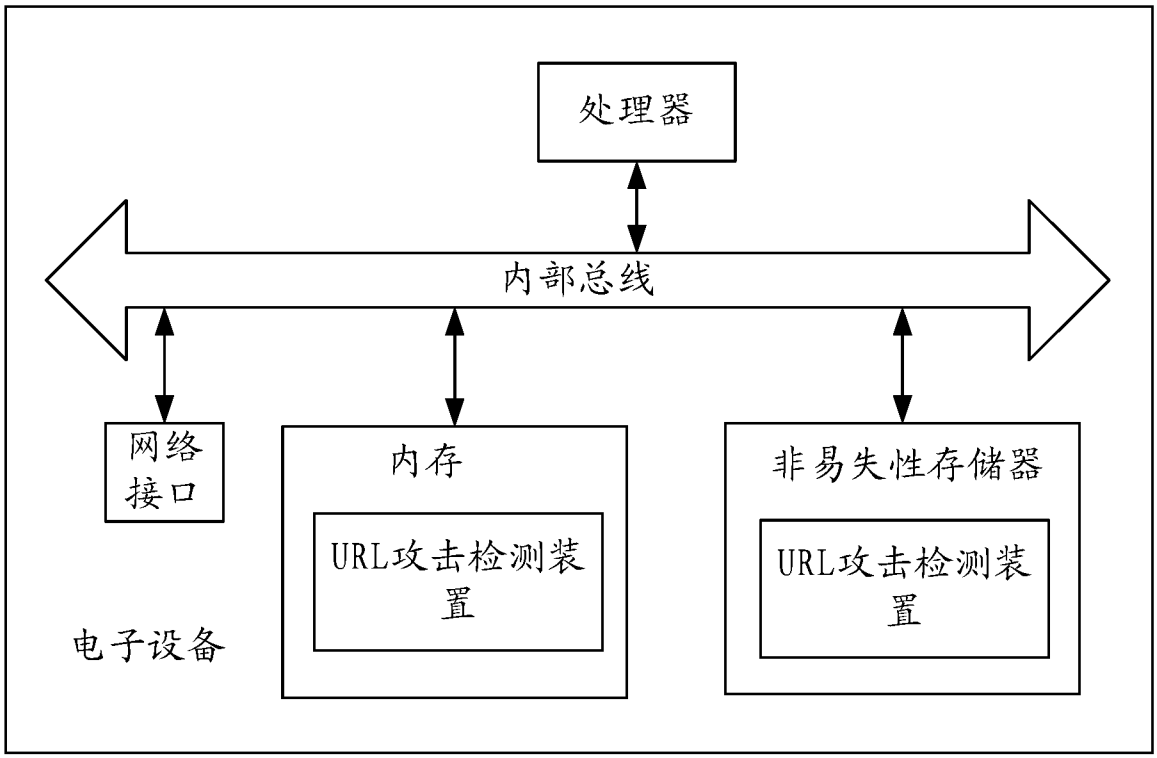
处理器
内部总线
网络接口
内存
URL攻击检测装置
非易失性存储器
URL攻击检测装置
电子设备

图 3

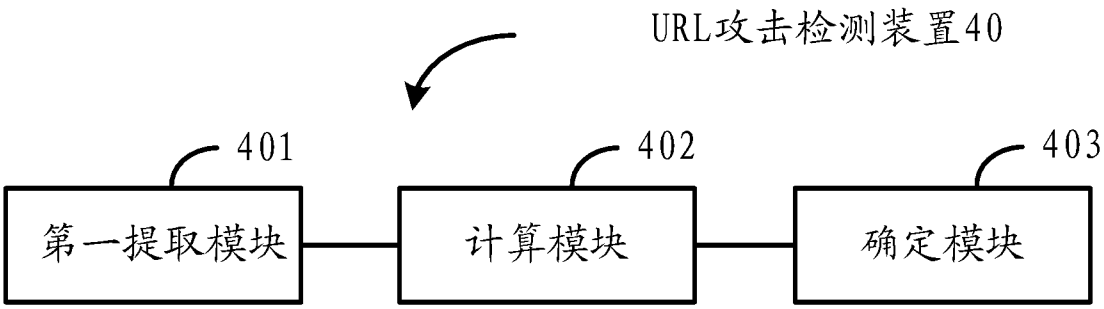
URL攻击检测装置40
401
402
403
第一提取模块
计算模块
确定模块

图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/116100

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/55(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F, H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI, GOOGLE: 攻击, 检测, 风险, 钓鱼, 恶意, 木马, URL, HTTP, 地址, 网址, 统一资源定位符, 访问请求, 模型, 训练, 学习, 网络, 二叉树, attack, isolation forest, iforest, fishing, detect, network, access, address, risk, model, training, learn, malcious

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107346388 A (SICHUAN SILENCE INFORMATION TECHNOLOGY CO., LTD. ET AL.) 14 November 2017 (2017-11-14) description, paragraphs 0029 and 0035-0099, and figures 1-7	1-13
PX	CN 108229156 A (ALIBABA GROUP HOLDING LIMITED) 29 June 2018 (2018-06-29) claims 1-13	1-13
PX	CN 107992741 A (ALIBABA GROUP HOLDING LIMITED) 04 May 2018 (2018-05-04) description, paragraphs 0052-0122, and figures 1-3	1-13
PX	CN 108111489 A (ALIBABA GROUP HOLDING LIMITED) 01 June 2018 (2018-06-01) description, paragraphs 0055-0123, and figures 1-4	1-13
PX	CN 107577945 A (ALIBABA GROUP HOLDING LIMITED) 12 January 2018 (2018-01-12) description, paragraphs 0055-0124, and figures 1-4	1-13
A	CN 104537303 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 22 April 2015 (2015-04-22) entire document	1-13

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 January 2019	**18 February 2019**

Name and mailing address of the ISA/CN	Authorized officer
National Intellectual Property Administration, PRC (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011185425 A1 (NATIONAL TAIWAN UNIVERSITY OF SCIENCE & TECHNOLOGY) 28 July 2011 (2011-07-28) entire document	1-13
A	CN 104735074 A (SUZHOU PAYEGIS INFORMATION TECHNOLOGY CO., LTD.) 24 June 2015 (2015-06-24) entire document	1-13

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/116100

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 107346388 A	14 November 2017	None	
CN 108229156 A	29 June 2018	None	
CN 107992741 A	04 May 2018	None	
CN 108111489 A	01 June 2018	None	
CN 107577945 A	12 January 2018	None	
CN 104537303 A	22 April 2015	None	
US 2011185425 A1	28 July 2011	TW 201126983 A	01 August 2011
CN 104735074 A	24 June 2015	None	

国际检索报告

国际申请号
PCT/CN2018/116100

A. 主题的分类

G06F 21/55(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F，H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT,WPI,EPODOC,CNKI,GOOGLE： 攻击，检测，风险，钓鱼，恶意，木马，URL，HTTP，地址，网址，统一资源定位符，访问请求，模型，训练，学习，网络，二叉树，attack, isolation forest, iforest, fishing, detect, network, access, address, risk, model, training, learn，malcious

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CN 107346388 A (四川无声信息技术有限公司 等) 2017年 11月 14日 (2017-11-14) 说明书第0029，0035-0099段，图1-7	1-13
PX	CN 108229156 A (阿里巴巴集团控股有限公司) 2018年 6月 29日 (2018-06-29) 权利要求1-13	1-13
PX	CN 107992741 A (阿里巴巴集团控股有限公司) 2018年 5月 4日 (2018-05-04) 说明书第0052-0122，图1-3	1-13
PX	CN 108111489 A (阿里巴巴集团控股有限公司) 2018年 6月 1日 (2018-06-01) 说明书第0055-0123段，图1-4	1-13
PX	CN 107577945 A (阿里巴巴集团控股有限公司) 2018年 1月 12日 (2018-01-12) 说明书第0055-0124段，图1-4	1-13
A	CN 104537303 A (中国科学院深圳先进技术研究院) 2015年 4月 22日 (2015-04-22) 全文	1-13
A	US 2011185425 A1 (NATIONAL TAIWAN UNIVERSITY OF SCIENCE & TECHNOLOGY) 2011年 7月 28日 (2011-07-28) 全文	1-13

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2019年 1月 29日	2019年 2月 18日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 范玉霞 电话号码 86-(10)-53961331

国际检索报告

国际申请号
PCT/CN2018/116100

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 104735074 A (江苏通付盾信息科技有限公司) 2015年 6月 24日 (2015 - 06 - 24) 全文	1-13

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/116100

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 107346388 A	2017年 11月 14日	无	
CN 108229156 A	2018年 6月 29日	无	
CN 107992741 A	2018年 5月 4日	无	
CN 108111489 A	2018年 6月 1日	无	
CN 107577945 A	2018年 1月 12日	无	
CN 104537303 A	2015年 4月 22日	无	
US 2011185425 A1	2011年 7月 28日	TW 201126983 A	2011年 8月 1日
CN 104735074 A	2015年 6月 24日	无	