(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108494728 B
(45)授权公告日 2021.01.26

(21)申请号 201810122846.0

(22)申请日 2018.02.07

(65)同一申请的已公布的文献号
申请公布号 CN 108494728 A

(43)申请公布日 2018.09.04

(73)专利权人 平安普惠企业管理有限公司
地址 518000 广东省深圳市前海深港合作区前湾一路1号A栋201室(入驻深圳市前海商务秘书有限公司)

(72)发明人 林泽全

(74)专利代理机构 深圳众鼎专利商标代理事务所(普通合伙) 44325
代理人 阳开亮

(51)Int.Cl.
H04L 29/06(2006.01)
H04L 29/08(2006.01)

(56)对比文件
CN 104021172 A,2014.09.03
CN 103401835 A,2013.11.20
CN 103605688 A,2014.02.26

审查员 程梦莉

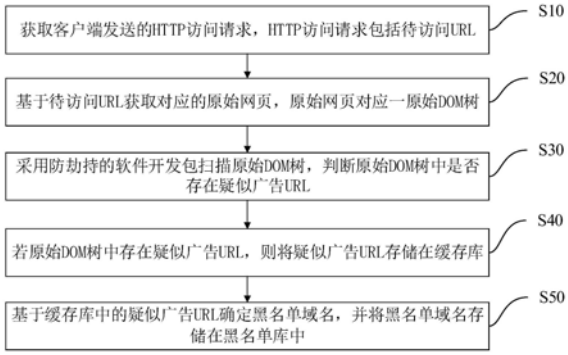
权利要求书2页 说明书10页 附图3页

(54)发明名称

防止流量劫持的黑名单库创建方法、装置、设备及介质

(57)摘要

本发明公开了一种防止流量劫持的黑名单库创建方法、装置、设备及介质。该方法包括:获取客户端发送的HTTP访问请求,HTTP访问请求包括待访问URL;基于待访问URL获取对应的原始网页,原始网页对应一原始DOM树;采用防劫持的软件开发包扫描原始DOM树,判断原始DOM树中是否存在疑似广告URL;若原始DOM树中存在疑似广告URL,则将疑似广告URL存储在缓存库;基于缓存库中的疑似广告URL确定黑名单域名,并将黑名单域名存储在黑名单库中。该方法提高了黑名单域名获取的准确性和待访问URL对应的原始网页确认网络广告资源信息的速度,优化网络广告资源信息识别的全面性。



CN 108494728 B

1.一种防止流量劫持的黑名单库创建方法,其特征在于,包括:

获取客户端发送的HTTP访问请求,所述HTTP访问请求包括待访问URL;

基于所述待访问URL获取对应的原始网页,所述原始网页对应一原始DOM树,所述原始DOM树包括至少一个DOM标签;

采用防劫持的软件开发包扫描所述原始DOM树,判断所述原始DOM树中至少一个DOM标签对应的原始URL是否存在疑似广告URL,所述防劫持的软件开发包是由JavaScript代码组成的用于检测是否存在疑似广告URL的软件开发包;所述疑似广告URL是指包含广告代码整体性特征、URL跳转特征和需要展示在网页具体位置的绝对定位特征中的至少一个的DOM标签对应的URL;

若所述原始DOM树中存在所述疑似广告URL,则将所述疑似广告URL存储在缓存库;

基于所述缓存库中的所述疑似广告URL确定符合黑名单判断方法的黑名单域名,并将所述黑名单域名存储在黑名单库中。

2.如权利要求1所述的防止流量劫持的黑名单库创建方法,其特征在于,所述采用防劫持的软件开发包扫描所述原始DOM树,判断所述原始DOM树中是否存在疑似广告URL,包括:

采用防劫持的软件开发包扫描所述原始DOM树,获取所述原始DOM树包含的原始URL;

若所述原始URL的域名与所述待访问URL的域名不匹配,则确定所述原始DOM树中存在所述疑似广告URL。

3.如权利要求1所述的防止流量劫持的黑名单库创建方法,其特征在于,所述基于所述缓存库中的所述疑似广告URL确定黑名单域名,包括:

对所述缓存库中的每一所述疑似广告URL进行域名提取,获取相应的疑似域名;

确定所述缓存库中数量达到预设值的疑似域名为黑名单域名。

4.如权利要求3所述的防止流量劫持的黑名单库创建方法,其特征在于,所述对所述缓存库中的每一所述疑似广告URL进行域名提取,获取相应的疑似域名,包括:

调用所述防劫持的软件开发包中的正则表达式对所述缓存库中的每一所述疑似广告URL进行域名提取,获取对应的所述疑似域名。

5.如权利要求1所述的防止流量劫持的黑名单库创建方法,其特征在于,在所述将所述黑名单域名存储在黑名单库中的步骤之后,所述防止流量劫持的黑名单库创建方法还包括:

获取误判恢复请求,所述误判恢复请求包括目标URL;

调用所述防劫持的软件开发包中的正则表达式对所述目标URL进行域名提取,获取目标域名;

将所述黑名单库中存储的与所述目标域名一致的黑名单域名删除,更新所述黑名单库。

6.如权利要求5所述的防止流量劫持的黑名单库创建方法,其特征在于,在所述将所述黑名单库中存储的与所述目标域名一致的黑名单域名删除的步骤之后,所述防止流量劫持的黑名单库创建方法还包括:

将所述黑名单库中存储的与所述目标域名一致的黑名单域名作为白名单域名,并将所述白名单域名存储在白名单库中;

在所述将所述疑似广告URL存储在缓存库的步骤之后,所述防止流量劫持的黑名单库

创建方法还包括:若所述疑似广告URL对应的域名存储在所述白名单库中,则将所述疑似广告URL从所述缓存库中删除。

7.一种防止流量劫持的黑名单库创建装置,其特征在于,包括:

访问请求获取模块,用于获取客户端发送的HTTP访问请求,所述HTTP访问请求包括待访问URL;

原始网页获取模块,用于基于所述待访问URL获取对应的原始网页,所述原始网页对应一原始DOM树,所述原始DOM树包括至少一个DOM标签;

疑似广告URL判断模块,用于采用防劫持的软件开发包扫描所述原始DOM树,判断所述原始DOM树中至少一个DOM标签对应的原始URL是否存在疑似广告URL,所述防劫持的软件开发包是由JavaScript代码组成的用于检测是否存在疑似广告URL的软件开发包;所述疑似广告URL是指包含广告代码整体性特征、URL跳转特征和需要展示在网页具体位置的绝对定位特征中的至少一个的DOM标签对应的URL;

缓存库存储模块,用于在所述原始DOM树中存在所述疑似广告URL时,将所述疑似广告URL存储在缓存库;

黑名单域名获取模块,用于基于所述缓存库中的所述疑似广告URL确定符合黑名单判断方法的黑名单域名,并将所述黑名单域名存储在黑名单库中。

8.如权利要求7所述的防止流量劫持的黑名单库创建装置,其特征在于,所述黑名单域名获取模块包括:

疑似域名获取单元,用于对所述缓存库中的每一所述疑似广告URL进行域名提取,获取相应的疑似域名;

黑名单域名获取单元,用于确定所述缓存库中数量达到预设值的疑似域名为黑名单域名。

9.一种终端设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现如权利要求1至6任一项所述防止流量劫持的黑名单库创建方法的步骤。

10.一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至6任一项所述防止流量劫持的黑名单库创建方法的步骤。

防止流量劫持的黑名单库创建方法、装置、设备及介质

技术领域

[0001] 本发明涉及网络安全领域，尤其涉及一种防止流量劫持的黑名单库创建方法、装置、设备及介质。

背景技术

[0002] 当用户在请求一个网页时，广告运营商会在与该网页相关的网页资源信息中插入网络广告资源信息，让客户端(通常是浏览器)展示与网页无关的资源信息，以达到广告运营商流量劫持的目的。这些网络广告资源信息通常为一些弹窗、宣传性广告或者直接显示其他网页的内容。目前针对广告运营商流量劫持的处理方法大部分是通过创建黑名单的方法实现的。但当前黑名单的创建方法通常是有开发人员通过人工写入的方式实现的，不能实现对网页中是否存在网络广告资源信息的准确识别和判断，造成网络广告资源信息识别不全面。

发明内容

[0003] 本发明实施例提供一种防止流量劫持的黑名单库创建方法、装置、设备及介质，以解决当前防止流量劫持的黑名单不能做到对网络广告资源信息全面识别的问题。

[0004] 第一方面，本发明实施例提供一种防止流量劫持的黑名单库创建方法，包括：

[0005] 获取客户端发送的HTTP访问请求，所述HTTP访问请求包括待访问URL；

[0006] 基于所述待访问URL获取对应的原始网页，所述原始网页对应一原始DOM树；

[0007] 采用防劫持的软件开发包扫描所述原始DOM树，判断所述原始DOM树中是否存在疑似广告URL；

[0008] 若所述原始DOM树中存在所述疑似广告URL，则将所述疑似广告URL存储在缓存库；

[0009] 基于所述缓存库中的所述疑似广告URL确定黑名单域名，并将所述黑名单域名存储在黑名单库中。

[0010] 第二方面，本发明实施例提供一种防止流量劫持的黑名单库创建装置，包括：

[0011] 访问请求获取模块，用于获取客户端发送的HTTP访问请求，所述HTTP访问请求包括待访问URL；

[0012] 原始网页获取模块，用于基于所述待访问URL获取对应的原始网页，所述原始网页对应一原始DOM树；

[0013] 疑似广告URL判断模块，用于采用防劫持的软件开发包扫描所述原始DOM树，判断所述原始DOM树中是否存在疑似广告URL；

[0014] 缓存库存储模块，用于在所述原始DOM树中存在所述疑似广告URL时，将所述疑似广告URL存储在缓存库；

[0015] 黑名单域名获取模块，用于基于所述缓存库中的所述疑似广告URL确定黑名单域名，并将所述黑名单域名存储在黑名单库中。

[0016] 第三方面，本发明实施例提供一种终端设备，包括存储器、处理器以及存储在所述

存储器中并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现所述防止流量劫持的黑名单库创建方法的步骤。

[0017] 第四方面，本发明实施例提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序被处理器执行时实现所述防止流量劫持的黑名单库创建方法的步骤。

[0018] 本发明实施例提供的防止流量劫持的黑名单库创建方法、装置、设备及介质，通过获取客户端发送的HTTP访问请求获取待访问URL，基于获取到的待访问URL获取对应的原始网页对应的原始DOM树。然后采用防劫持的软件开发包扫描该原始DOM树，获取该原始DOM树中存在的疑似广告URL并存储在缓存库中，有助于提高后续黑名单域名提取的效率。对缓存库中的疑似广告URL进行域名提取，获取黑名单域名，有助于提高黑名单域名确认的准确性。将该黑名单域名存储在黑名单库中，有助于提高后续对待访问URL对应的原始网页进行黑名单域名识别的准确性，提高待访问URL对应的原始网页确认网络广告资源信息的速度，优化网络广告资源信息识别的全面性。

附图说明

[0019] 为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例的描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0020] 图1是本发明实施例1中防止流量劫持的黑名单库创建方法的一流程图。

[0021] 图2是图1中步骤S30的一具体示意图。

[0022] 图3是图1中步骤S50的一具体示意图。

[0023] 图4是本发明实施例1中防止流量劫持的黑名单库创建方法的另一流程图。

[0024] 图5是本发明实施例2中防止流量劫持的黑名单库创建装置的一原理框图。

[0025] 图6是本发明实施例4中终端设备的一示意图。

具体实施方式

[0026] 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所有其他实施例，都属于本发明保护的范围。

[0027] 实施例1

[0028] 图1示出本实施例中防止流量劫持的黑名单库创建方法的流程图。该防止流量劫持的黑名单库创建方法应用在服务器中，该服务器与客户端通过网络进行信息交互，可防止广告运营商在正常的网页资源信息中插入网络广告资源信息，达到防止广告运营商流量广告劫持的目的。如图1所示，该防止流量劫持的黑名单库创建方法包括如下步骤：

[0029] S10：获取客户端发送的HTTP访问请求，HTTP访问请求包括待访问URL。

[0030] 其中，待访问URL是指用户需要访问的网页地址。具体地，与客户端通信相连的服务器会接收客户端发送的HTTP访问请求，该HTTP访问请求一般携带有网页地址URL，该URL

即为客户端发送给服务器需要访问的网页地址。

[0031] S20:基于待访问URL获取对应的原始网页,原始网页对应一原始DOM树。

[0032] 具体地,原始网页是指待访问URL对应的网页。服务器根据HTTP访问请求中的待访问URL获取该待访URL对应的原始网页,每一原始网页都对应一DOM树,该DOM树即为该原始网页对应的原始DOM树。原始DOM树是指待访问URL对应的原始网页加载的所有网页资源信息对应的DOM树。

[0033] 其中,DOM树(Document Object Model,文档对象模型)是专门适用于HTML(超级文本标记语言)的文档对象模型,该HTML是指为网页创建和其它可在网页浏览器中看到的信息设计的一种标记语言。一个网页的本质就是由一个HTML(超级文本标记语言)组成的,DOM树就是该网页对应的文档对象模型。在DOM树中,网页中的各个元素都被看作一个个对象,从而使网页中的元素也可以被计算机语言获取或者编辑。一个网页中存在至少一个元素,一个元素对应DOM树中的一个DOM标签,即一个DOM树中存在至少一个DOM标签。

[0034] S30:采用防劫持的软件开发包扫描原始DOM树,判断原始DOM树中是否存在疑似广告URL。

[0035] 其中,防劫持的软件开发包是由一套JavaScript代码组成的用于检测是否存在疑似广告URL的软件开发包,该套JavaScript代码是在浏览器中以script标签的方式引入到该软件开发包的。如JavaScript代码在该软件开发包中的表现形式为<script src=“a.js”>,其中,src后为该软件开发包的地址。软件开发包(Software Development Kit,即SDK)是指一种为软件开发提供的工具包,通常是用于为特定的软件包、软件框架、硬件平台和操作系统等建立应用软件的开发工具的集合。

[0036] 疑似URL是指符合预设特征的DOM标签对应的URL。该预设特征是指广告运营商植入的广告代码对应的DOM标签的特征。广告代码对应的DOM标签的特征包括但不限于广告代码整体性特征、URL跳转特征和需要展示在网页具体位置的绝对定位特征。其中,广告代码整体性特征是指广告运营商需要展示的完整的广告信息,该广告信息对应的广告代码就是一段完整的代码,即表现在DOM树中就是一个整体,表现形式可以是以 开始,以 结束中的一段代码。URL跳转特征是指插入一张广告图,并加了<a>的URL链接,a为一串代表该图片存放位置的字符串。绝对定位特征是指在待访问URL对应的原始网页对应的DOM树的尾部多出来很多iframe和嵌入有广告代码的div,如待访问URL对应的原始网页的最后一个元素为 ,非法插入的代码为 <script src=”a.js”>。

[0037] 具体地,待访问URL对应的原始网页加载网页资源信息,该网页资源信息在网页上可以有多种展示方式,包括但不限于图片、文字、网址和视频。这些网页资源信息,即就是网页中的元素。这些网页中的元素对于软件开发包来说,都是以DOM标签存在的。

[0038] 进一步地,获取客户端发送的HTTP访问请求后,服务器基于该HTTP访问请求获取防劫持的软件开发包。在待访问URL对应的原始网页加载完成该网页的所有网页资源信息后,该待访问URL对应的原始网页会出现一个onload的状态事件,该状态事件是指接入防劫持的软件开发包对该待访问URL对应的原始网页加载的网页资源信息进行处理的请求事件,该状态事件有一个接口,可以接入防劫持的软件开发包对DOM树进行扫描。

[0039] 防劫持的软件开发包基于该状态事件采用广度优先的扫描方式对该待访问URL对应的原始网页对应的DOM树进行扫描,从DOM树最外层的“html”标签开始进行扫描,逐层扫

描对应的DOM标签，查找符合预设特征的DOM标签是否存在疑似广告URL。采用广度优先的扫描方式对DOM树中的所有DOM标签进行遍历，可以对DOM树中每一层级包含的所有DOM标签进行扫描，扫描完一个层级的所有DOM标签再扫描下一层级的所有DOM标签，以便按出队的顺序访问同一层级该DOM标签的所有相邻DOM标签，适合对DOM树进行全面扫描。若原始DOM树中存在广告代码整体性特征、URL跳转特征和绝对定位特征这三种预设特征中的任何一种，则可以认定该原始DOM树中存在疑似广告URL，该疑似广告URL为初步确定可能为广告的URL，确定该疑似广告URL有助于确定黑名单域名，从而保证步骤S50基于该疑似广告提取域名，实现将黑名单域名存储在黑名单库中。

[0040] S40：若原始DOM树中存在疑似广告URL，则将疑似广告URL存储在缓存库。

[0041] 具体地，判断原始DOM树中是否存在符合预设特征的DOM标签，若存在符合预设特征的DOM标签，则可认定该DOM树中存在疑似广告URL，将该DOM标签也就是疑似广告URL存储在缓存库中。可以理解地，将疑似广告URL存储在缓存库中可以做到对缓存库中存储的疑似广告URL等数据进行快速处理(包括但不限于查询处理)，不需要请求服务器并获取服务器发送的处理指令进行数据处理。

[0042] 本实施例中的缓存库可以是mysql关系型数据库，mysql关系型数据库是一种开放源代码的关系型数据库管理系统，提供了面向多种编程语言的编程接口(APIs)，支持多种字段类型并且提供了完整的操作符支持查询中的SELECT和WHERE操作。mysql关系型数据库具有速度高、可靠性好和适应性强等特点，使用mysql关系型数据库进行存储疑似广告URL，可以实现主从配置和读写分离的功能，能为数据的存储提供高效的服务。

[0043] S50：基于缓存库中的疑似广告URL确定黑名单域名，并将黑名单域名存储在黑名单库中。

[0044] 其中，黑名单域名是指对疑似广告URL进行域名提取后获取的域名。黑名单库是指存储黑名单域名的数据库。具体地，对存储在缓存库中的疑似广告URL进行域名提取，若该疑似广告URL提取的域名符合预设的黑名单判断方法，则确定该疑似广告URL提取的域名确定为黑名单域名。然后，将该黑名单域名存储在预先创建的黑名单库中，以便于后续进行黑名单域名识别时，可作为参考依据。

[0045] 步骤S10-S50，可通过获取客户端发送的HTTP访问请求获取待访问URL，基于获取到的待访问URL获取对应的原始网页对应的原始DOM树。采用防劫持的软件开发包扫描该原始DOM树，获取该原始DOM树中存在的疑似广告URL，并对该疑似广告URL进行域名提取获取黑名单域名，将得到的黑名单域名存储在黑名单库中，由此确认的黑名单库有助于提高后续对待访问URL对应的原始网页进行黑名单域名识别的准确性，提高网络广告资源信息识别全面性。

[0046] 在一具体实施方式中，如图2所示，步骤S30中，采用防劫持的软件开发包扫描原始DOM树，判断原始DOM树中是否存在疑似广告URL，具体包括如下步骤：

[0047] S31：采用防劫持的软件开发包扫描原始DOM树，获取原始DOM树包含的原始URL。

[0048] 具体地，一个DOM树中包含至少一个DOM标签。采用防劫持的软件开发包采用广度优先的扫描方式对该待访问URL对应的原始网页对应的原始DOM树进行扫描，从该原始DOM树最外层的html标签开始进行扫描，逐层扫描每一层极的DOM标签，确定DOM标签中以URL形式存在的至少一个DOM标签，再查找该DOM标签中包含的URL，该URL为要获取的原始URL。

[0049] S32:若原始URL的域名与待访问URL的域名不匹配,则确定原始DOM树中存在疑似广告URL。

[0050] 其中,原始URL的域名是指对原始URL进行域名提取获得的因特网上的地址,待访问URL的域名是指对待访问URL进行域名提取获得的因特网上的地址。获取原始URL的域名与待访问URL的域名的过程通过步骤S51进行详细描述,为避免赘述,有此不一一详述。

[0051] 具体地,对获取的原始DOM树包含的原始URL的域名和待访问URL的域名进行判断,判断原始URL的域名和待访问URL的域名是否匹配。若两者匹配一致,则表示该原始URL为该用户需要访问的网页原有的网页资源信息;若匹配不一致,则表示该原始DOM树中本身存在疑似URL,该原始URL不是用户需要访问的网页原有的网页资源信息。通过对原始URL的域名和待访问URL的域名进行匹配处理,可快速有效地确定原始网页中是否存在疑似广告URL。

[0052] 在一具体实施方式中,如图3所示,步骤S50中,基于缓存库中的疑似广告URL确定黑名单域名,具体包括如下步骤:

[0053] S51:对缓存库中的每一疑似广告URL进行域名提取,获取相应的疑似域名。

[0054] 当被确定为疑似广告URL后,该疑似广告URL就会存储在缓存库中,缓存库中存储至少一个疑似URL。对缓存库中的每一个疑似URL进行域名提取,提取出来的该域名则为疑似域名。

[0055] 进一步地,调用防劫持的软件开发包中的正则表达式对缓存库中的每一疑似广告URL进行域名提取,获取对应的疑似域名。

[0056] 其中,正则表达式又称规则表达式(Regular Expression,在代码中常简写为regex、regexp或RE)。正则表达式是对字符串操作的一种逻辑公式,本实施例中,该正则表达式是用来表达对字符串的一种过滤逻辑。字符串包括普通字符(如a到z之间的字母)和特殊字符(又称为“元字符”,如“$、*、&、#、+、?”)。

[0057] 具体地,防劫持的软件开发包中有封装好的正则表达式。步骤S51具体为:采用该封装好的正则表达式对缓存库中的每一疑似广告URL进行拆分,以拆分为协议名称、域名和参数这三个部分;然后,去除协议名称和域名后面的参数部分,只保留域名,从而获取相应的疑似域名。如疑似广告URL为:http://pos.baidu.com/s?hei=250&wid=250&di=u3031286&ltu=lV-RgLBX*E5wJyFr&r=35d363d1cad5eabfcd131082d275f954#,其中,“http”对应协议名称,“pos.baidu.com”对应域名,域名后的所有内容可统称为参数。在采用正则表达式对上述疑似广告URL进行域名提取时只保留域名部分“pos.baidu.com”,则“pos.baidu.com”为疑似域名。

[0058] S52:确定缓存库中数量达到预设值的疑似域名为黑名单域名。

[0059] 其中,黑名单域名是指同一疑似域名在缓存库存储的次数达到(即大于或等于)预设值时,确定该疑似域名为黑名单域名。预设值是指预先设置的疑似域名存储在缓存库中的数量。该预设值用于判断疑似域名是否为黑名单域名。

[0060] 若该疑似域名在缓存库中出现一次,未达到预设值时,还不能确认该疑似域名就是黑名单域名,可能只是一个与待访问URL的域名不匹配的域名,当该疑似域名在缓存库中存储的数量达到预设值时,则可以确认该疑似域名为黑名单域名。可以理解地,设置疑似域名的数量达到预设值时才确定为黑名单域名,可以减少黑名单域名的误判,提高确定黑名单域名的准确性。

[0061] 在一具体实施方式中，如上所述，若缓存库中的疑似广告URL的数量达到预设值时认定其为黑名单域名，可能存在误判，会导致后续被误判的疑似广告URL进入黑名单库中，导致无法进行访问或其他操作。如图4所示，在将黑名单域名存储在黑名单库中的步骤之后，该防止流量劫持的黑名单库创建方法还包括：

[0062] S61：获取误判恢复请求，误判恢复请求包括目标URL。

[0063] 误判恢复请求是服务器接收到用户需要进行恢复查看被隐藏内容的恢复请求，该隐藏内容是指加入黑名单的黑名单域名对应的URL显示的网页资源信息的内容。目标URL是指需要恢复查看被隐藏内容对应的URL。具体地，在进行黑名域名确认的过程中，可能会存在误判情况。当用户在访问某一网页时，由于服务器已经将和待访问网页的域名不一致的疑似广告URL对应的域名判断为黑名单域名，并加入黑名单库中。因此，该网页只显示没有加入黑名单库中部分内容，加入黑名单库中的部分内容被隐藏不进行显示。在浏览器显示网页资源信息对应的网页内容时，该网页会出现一个是否查看隐藏内容的通知信息。若用户点击恢复该隐藏内容，则服务器会获取一个恢复请求，该恢复请求则为误判恢复请求。同时该误判恢复请求中包括需要恢复的隐藏内容对应的URL，该URL则为目标URL。获取误判恢复请求可以减少加入黑名单库中误存的域名，帮助用户浏览完整的网页资源信息对应的网页内容。

[0064] S62：调用防劫持的软件开发包中的正则表达式对目标URL进行域名提取，获取目标域名。

[0065] 当服务器接收到用户发送的误判恢复请求时，调用防劫持的软件开发包中的正则表达式对目标URL进行域名提取，获取该目标URL对应的目标域名，该域名提取过程如步骤S51中描述，为避免重复，不一一赘述。

[0066] S63：将黑名单库中存储的与目标域名一致的黑名单域名删除，更新黑名单库。

[0067] 基于获取到目标域名，服务器对该目标域名和黑名单库存储的黑名单域名进行比较确认，将与目标域名一致的黑名单库中存储的黑名单域名删除，更新黑名单库。步骤S63，可保证黑名单库中存储的黑名单域名可以根据实际情况进行不断的调整，降低黑名单域名的误判率，保证黑名单库中存储的黑名单的准确性。

[0068] 在该具体实施方式中，在步骤S63之后，即将黑名单库中存储的与目标域名一致的黑名单域名删除的步骤之后，该防止流量劫持的黑名单库创建方法还可以包括：

[0069] S64：将黑名单库中存储的与目标域名一致的黑名单域名作为白名单域名，并将白名单域名存储在白名单库中。

[0070] 在创建黑名单库的同时创建一白名单库，该白名单库是指存储某一网页允许用户访问的网页的URL对应的目标域名的数据库。基于目标域名对黑名单库中存储的黑名单域名进行比较判断，将与目标域名一致的黑名单域名作为白名单域名，并将该白名单域名存储在白名单库中。

[0071] 本实施例中，白名单库中还包括预先存储的白名单域名。该预先存储的白名单域名为：一些待访问网页是允许插入不属于该网页正常的网页资源信息的网络广告资源信息，这时就可以对该网络广告资源信息对应的URL采用正则表达式进行域名提取，并将提取到的域名存储在白名单库中。

[0072] 当防劫持的软件开发包对用户访问网页的原始DOM树中的所有DOM标签进行扫描

时，确定疑似广告URL并将疑似广告URL存储在缓存库之后，需对该疑似广告URL进行域名提取，以确定疑似广告URL对应的域名(即步骤S51中的疑似域名)，并在判断该疑似域名与白名单库中的白名单域名一致时，显示该疑似广告URL对应的网页资源信息。例如，百度网页中允许插入的百度推广广告，这些百度推广广告对应的URL经防劫持的软件开发包扫描确定为疑似广告URL，但经域名提取之后确定域名在白名单库中，则可显示该百度推广广告对应的URL的网页资源信息。这样可以避免将某一网页允许用户访问的网页资源信息对应的网页内容误加入黑名单中，造成不必要的网页资源信息对应的网页内容的损失，能更全面反映该网页资源信息对应的网页内容。

[0073] 在一具体实施方式中，在步骤S40之后，即在将疑似广告URL存储在缓存库的步骤之后，该防止流量劫持的黑名单库创建方法还包括：若疑似广告URL对应的域名存储在白名单库中，则将疑似广告URL从缓存库中删除。

[0074] 可以理解地，在将疑似广告URL存储在缓存库之后，需对该疑似广告URL进行域名提取，以确定疑似广告URL对应的域名(即步骤S51中的疑似域名)，并在判断该疑似广告URL对应的域名存储在白名单库中时，则表明该疑似广告URL对应的域名属于白名单库，其对应的URL的内容是需要显示的网页资源信息对应的网页内容。为了避免出现只删除黑名单库中存储的疑似广告URL对应的域名，而没有删除存储在缓存库中的疑似广告URL，从而导致该疑似广告URL对应的网页资源信息对应的网页内容仍然不能正常显示。因此，在确认疑似广告URL对应的域名存储在白名单库中后，需将疑似广告URL从缓存库中删除。

[0075] 本发明实施例提供的防止流量劫持的黑名单库创建方法，通过采用防劫持的软件开发包中的正则表达式对疑似广告URL进行域名提取，并存储在缓存库中，有助于提高后续黑名单域名提取的高效性。当该缓存库中的同一URL对应的域名数量达到预设值时，则将该域名存储在黑名单库中，使得确认的黑名单库有助于提高后续对待访问URL对应的原始网页进行黑名单域名识别的准确性，使得网络广告资源信息识别更加全面。

[0076] 应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

[0077] 实施例2

[0078] 图5示出与实施例1中防止流量劫持的黑名单库创建方法一一对应的防止流量劫持的黑名单库创建装置的原理框图。如图5所示，该防止流量劫持的黑名单库创建装置包括访问请求获取模块10、原始网页获取模块20、疑似广告URL判断模块30、缓存库存储模块40和黑名单域名获取模块50。其中，访问请求获取模块10、原始网页获取模块20、疑似广告URL判断模块30、缓存库存储模块40和黑名单域名获取模块50的实现功能与实施例中防止流量劫持的黑名单库创建方法对应的步骤一一对应，为避免赘述，本实施例不一一详述。

[0079] 访问请求获取模块10，用于获取客户端发送的HTTP访问请求，HTTP访问请求包括待访问URL。

[0080] 原始网页获取模块20，用于基于待访问URL获取对应的原始网页，原始网页对应一原始DOM树。

[0081] 疑似广告URL判断模块30，用于采用防劫持的软件开发包扫描原始DOM树，判断原始DOM树中是否存在疑似广告URL。

[0082] 缓存库存储模块40，用于在原始DOM树中存在疑似广告URL时，将疑似广告URL存储在缓存库。

[0083] 黑名单域名获取模块50，用于基于缓存库中的疑似广告URL确定黑名单域名，并将黑名单域名存储在黑名单库中。

[0084] 优选地，疑似广告URL判断模块30包括原始URL获取单元31和疑似广告URL确认单元32。

[0085] 原始URL获取单元31，用于采用防劫持的软件开发包扫描原始DOM树，获取原始DOM树包含的原始URL。

[0086] 疑似广告URL确认单元32，用于在原始URL的域名与待访问URL的域名不匹配时，确定原始DOM树中存在疑似广告URL。

[0087] 优选地，黑名单域名获取模块50包括疑似域名获取单元51和黑名单域名获取单元52。

[0088] 疑似域名获取单元51，用于对缓存库中的每一疑似广告URL进行域名提取，获取相应的疑似域名。

[0089] 黑名单域名获取单元52，用于确定缓存库中数量达到预设值的疑似域名为黑名单域名。

[0090] 优选地，疑似域名获取单元51，用于调用防劫持的软件开发包中的正则表达式对缓存库中的每一疑似广告URL进行域名提取，获取对应的疑似域名。

[0091] 优选地，防止流量劫持的黑名单库创建装置还包括误判恢复请求获取单元61、目标域名获取单元62、黑名单库更新单元63和白名单域名获取单元64。

[0092] 误判恢复请求获取单元61，用于获取误判恢复请求，误判恢复请求包括目标URL。

[0093] 目标域名获取单元62，用于调用防劫持的软件开发包中的正则表达式对目标URL进行域名提取，获取目标域名。

[0094] 黑名单库更新单元63，用于将黑名单库中存储的与目标域名一致的黑名单域名删除，更新黑名单库。

[0095] 白名单域名获取单元64，用于将黑名单库中存储的与目标域名一致的黑名单域名作为白名单域名，并将白名单域名存储在白名单库中。

[0096] 优选地，防止流量劫持的黑名单库创建装置还包括：疑似广告URL删除模块70，用于在疑似广告URL对应的域名存储在白名单库中时，将疑似广告URL从缓存库中删除。

[0097] 实施例3

[0098] 本实施例提供一计算机可读存储介质，该计算机可读存储介质上存储有计算机程序，该计算机程序被处理器执行时实现实施例1中防止流量劫持的黑名单库创建方法，为避免重复，这里不再赘述。或者，该计算机程序被处理器执行时实现实施例2中防止流量劫持的黑名单库创建装置中各模块/单元的功能，为避免重复，这里不再赘述。

[0099] 实施例4

[0100] 图6是本发明一实施例提供的终端设备的示意图。如图6所示，该实施例的终端设备80包括：处理器81、存储器82以及存储在存储器82中并可在处理器81上运行的计算机程序83，例如防止流量劫持的黑名单库创建程序。处理器81执行计算机程序83时实现上述各个防止流量劫持的黑名单库创建方法实施例中的步骤，例如图1所示的步骤S10至S50。或

者，处理器81执行计算机程序83时实现上述各装置实施例中各模块/单元的功能，例如图5所示访问请求获取模块10、原始网页获取模块20、疑似广告URL判断模块30、缓存库存储模块40和黑名单域名获取模块50的功能。

[0101] 示例性的，计算机程序83可以被分割成一个或多个模块/单元，一个或者多个模块/单元被存储在存储器82中，并由处理器81执行，以完成本发明。一个或多个模块/单元可以是能够完成特定功能的一系列计算机程序指令段，该指令段用于描述计算机程序83在终端设备80中的执行过程。例如，访问请求获取模块10、原始网页获取模块20、疑似广告URL判断模块30、缓存库存储模块40和黑名单域名获取模块50。

[0102] 终端设备80可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。终端设备可包括，但不仅限于，处理器81、存储器82。本领域技术人员可以理解，图6仅仅是终端设备80的示例，并不构成对终端设备80的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如终端设备还可以包括输入输出设备、网络接入设备、总线等。

[0103] 所称处理器81可以是中央处理单元(Central Processing Unit,CPU)，还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0104] 存储器82可以是终端设备80的内部存储单元，例如终端设备80的硬盘或内存。存储器82也可以是终端设备80的外部存储设备，例如终端设备80上配备的插接式硬盘，智能存储卡(Smart Media Card,SMC)，安全数字(Secure Digital,SD)卡，闪存卡(Flash Card)等。进一步地，存储器82还可以既包括终端设备80的内部存储单元也包括外部存储设备。存储器82用于存储计算机程序以及终端设备所需的其他程序和数据。存储器82还可以用于暂时地存储已经输出或者将要输出的数据。

[0105] 所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。

[0106] 另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

[0107] 所述集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明实现上述实施例方法中的全部或部分流程，也可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，该计算机程序在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁

碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括是电载波信号和电信信号。

[0108] 以上所述实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围,均应包含在本发明的保护范围之内。

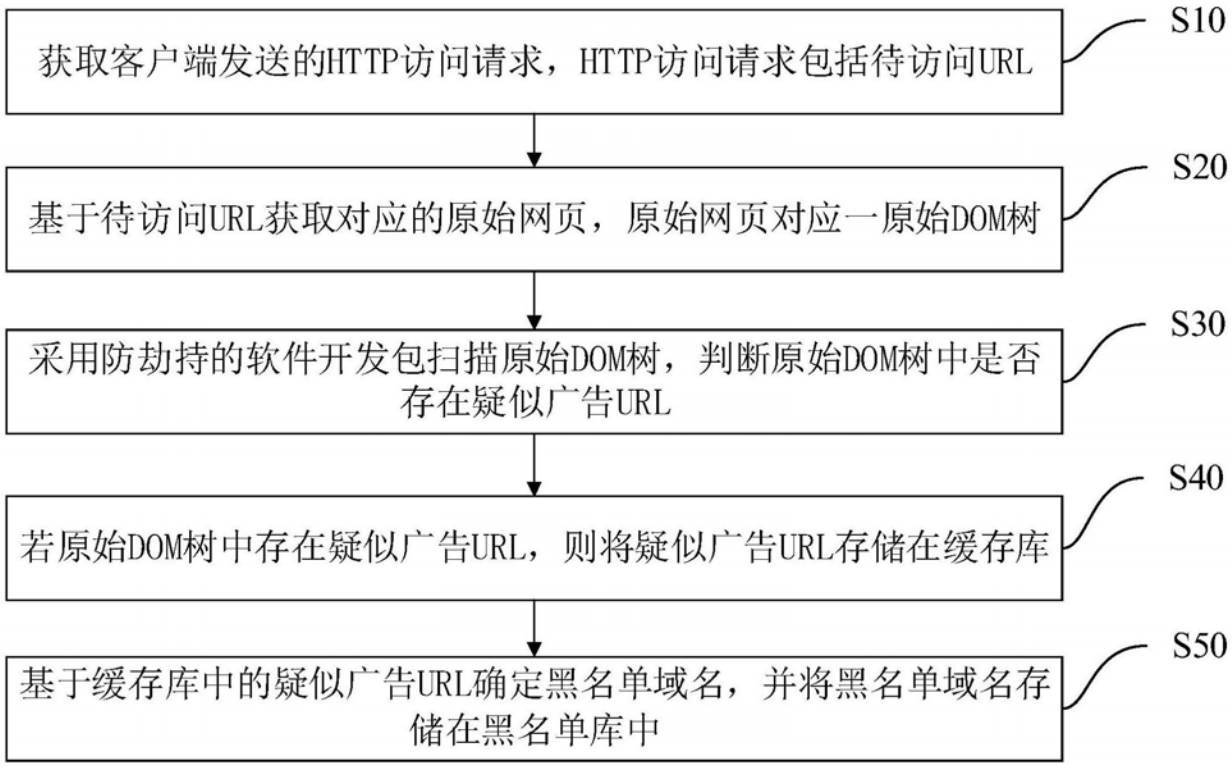

获取客户端发送的HTTP访问请求，HTTP访问请求包括待访问URL
S10
基于待访问URL获取对应的原始网页，原始网页对应一原始DOM树
S20
采用防劫持的软件开发包扫描原始DOM树，判断原始DOM树中是否存在疑似广告URL
S30
若原始DOM树中存在疑似广告URL，则将疑似广告URL存储在缓存库
S40
基于缓存库中的疑似广告URL确定黑名单域名，并将黑名单域名存储在黑名单库中
S50

图1

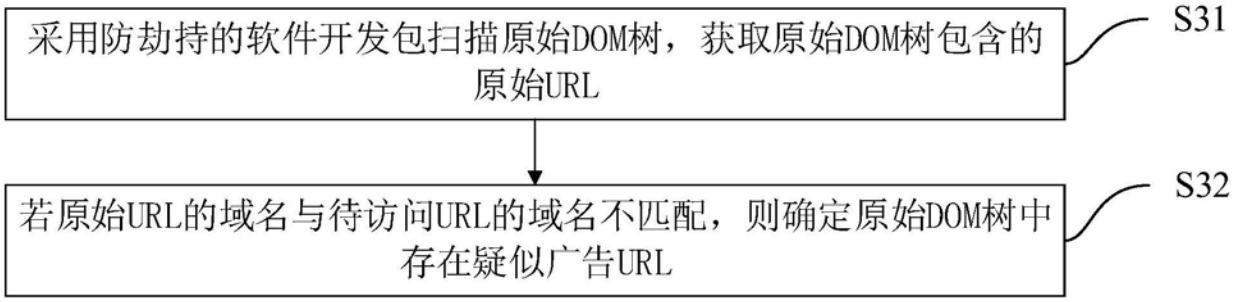

采用防劫持的软件开发包扫描原始DOM树，获取原始DOM树包含的原始URL
S31
若原始URL的域名与待访问URL的域名不匹配，则确定原始DOM树中存在疑似广告URL
S32

图2

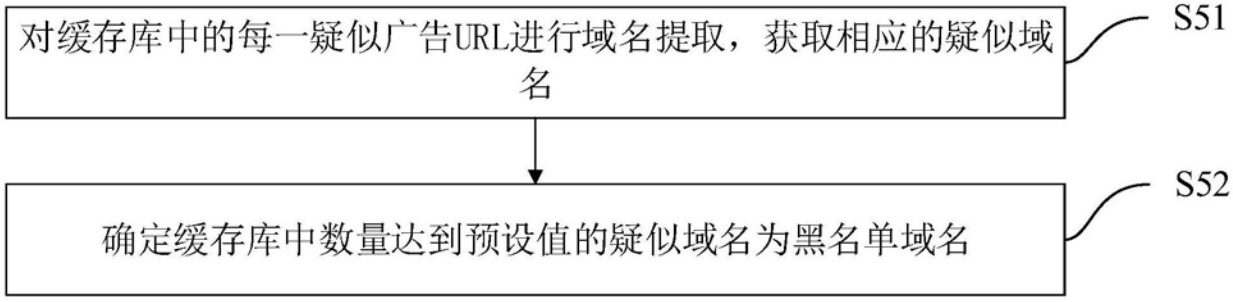

对缓存库中的每一疑似广告URL进行域名提取，获取相应的疑似域名
S51
确定缓存库中数量达到预设值的疑似域名为黑名单域名
S52

图3

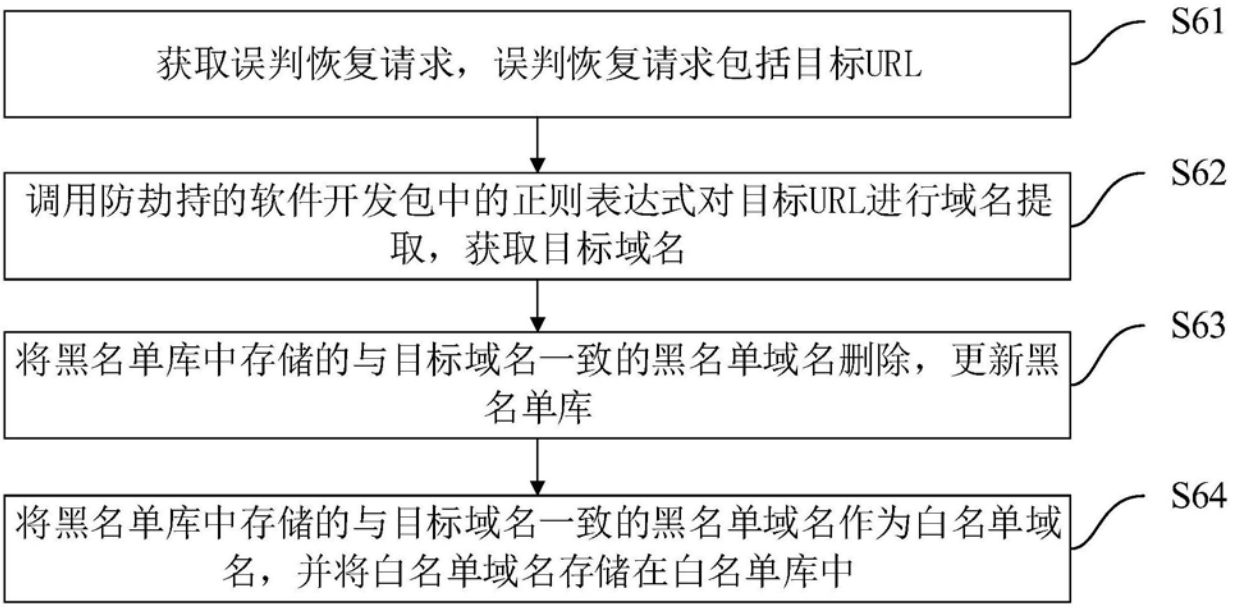

获取误判恢复请求，误判恢复请求包括目标URL
S61
调用防劫持的软件开发包中的正则表达式对目标URL进行域名提取，获取目标域名
S62
将黑名单库中存储的与目标域名一致的黑名单域名删除，更新黑名单库
S63
将黑名单库中存储的与目标域名一致的黑名单域名作为白名单域名，并将白名单域名存储在白名单库中
S64

图4

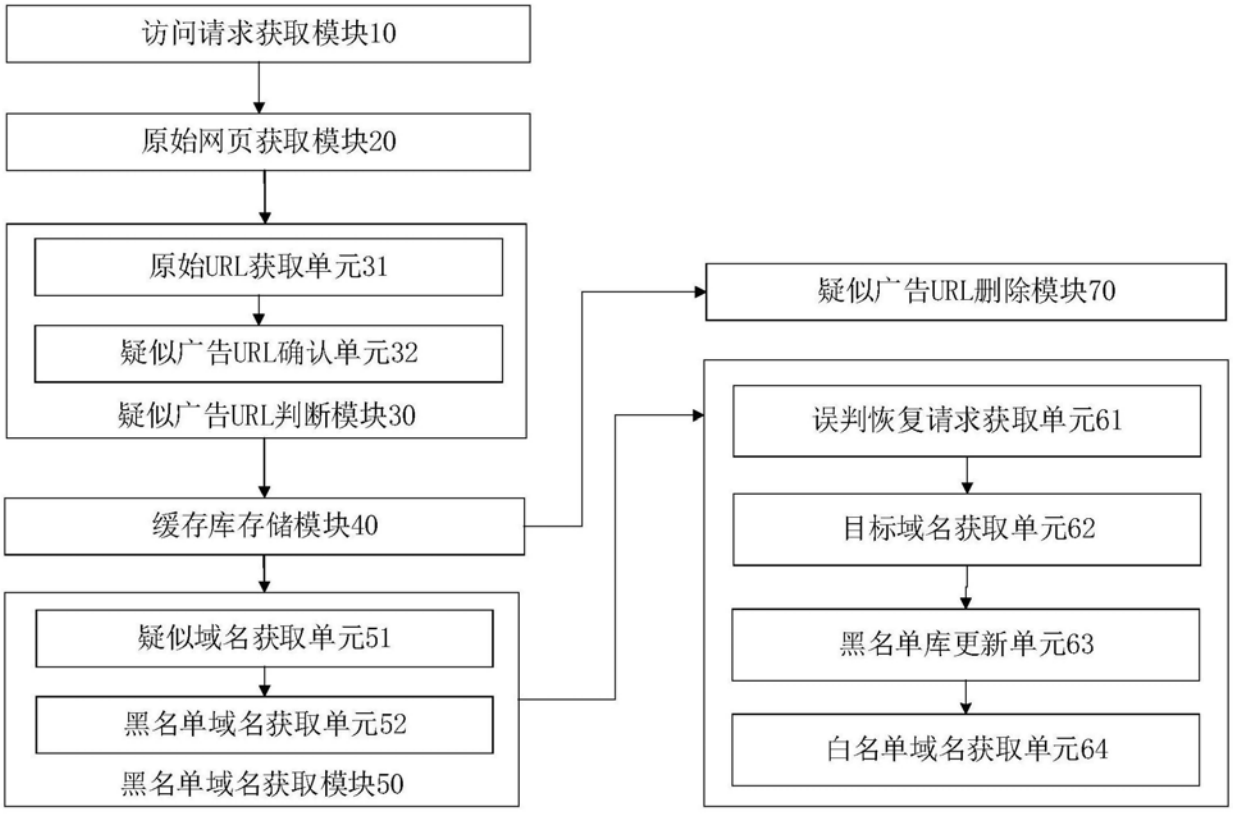

访问请求获取模块10
原始网页获取模块20
原始URL获取单元31
疑似广告URL确认单元32
疑似广告URL判断模块30
缓存库存储模块40
疑似域名获取单元51
黑名单域名获取单元52
黑名单域名获取模块50
疑似广告URL删除模块70
误判恢复请求获取单元61
目标域名获取单元62
黑名单库更新单元63
白名单域名获取单元64

图5

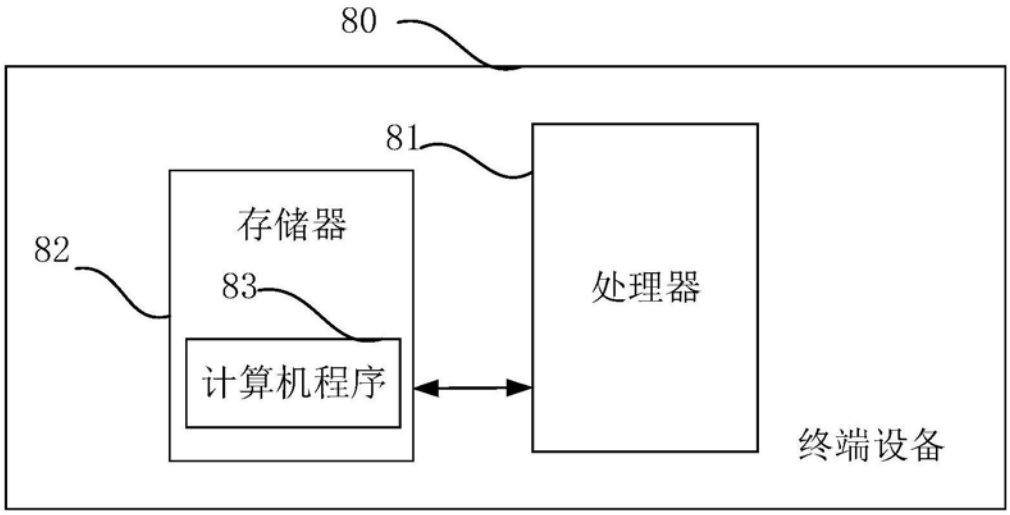


图6