



(12) 发明专利申请

(10) 申请公布号 CN 103685228 A

(43) 申请公布日 2014. 03. 26

(21) 申请号 201310476173. 6

(22) 申请日 2013. 10. 12

(71) 申请人 北京奇虎科技有限公司

地址 100088 北京市西城区新街口外大街  
28 号 D 座 112 室(德胜园区)

申请人 奇智软件(北京)有限公司

(72) 发明人 龙专 李纪峰 赵武

(74) 专利代理机构 北京思睿峰知识产权代理有  
限公司 11396

代理人 赵爱军

(51) Int. Cl.

H04L 29/06(2006. 01)

G06F 21/55(2013. 01)

G06F 17/30(2006. 01)

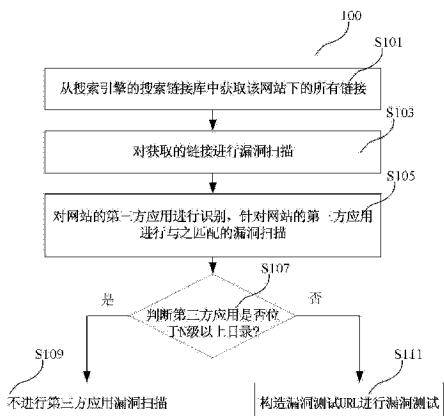
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种网站漏洞快速扫描方法及设备

(57) 摘要

本发明公开了一种网站漏洞快速扫描方法及设备，其中该方法包括：从搜索引擎的搜索链接库中获取该网站下的所有链接；对获取的链接进行漏洞扫描；对网站的第三方应用进行识别，针对网站的第三方应用进行与之匹配的漏洞扫描；以及判断第三方应用是否位于 N 级目录以上，若是，则不进行第三方应用漏洞扫描；若第三方应用位于 N 级目录以下，则构造漏洞测试统一资源定位符 URL 进行漏洞测试，其中 N 为大于 1 的整数。本发明使得首次使用漏洞扫描平台的用户感受到极速的漏洞扫描效果，用户体验得到极大的提高。



1. 一种网站漏洞快速扫描方法,包括:

从搜索引擎的搜索链接库中获取所述网站下的所有链接;

对所述获取的链接进行漏洞扫描;

对网站的第三方应用进行识别,针对所述网站的第三方应用进行与之匹配的漏洞扫描;以及

判断第三方应用是否位于 N 级目录以上,若是,则不进行第三方应用漏洞扫描;若所述第三方应用位于 N 级目录以下,则构造漏洞测试统一资源定位符 URL 进行漏洞测试,其中 N 为大于 1 的整数。

2. 根据权利要求 1 所述的方法,其中,

所述从搜索引擎的搜索链接库中获取该网站下的所有链接的步骤之前,还包括:

判断针对所述网站的域名的扫描是否为首次扫描,若是,则判断搜索链接库中该网站的爬虫接口是否可用,若是,则从搜索引擎的搜索链接库中获取该网站下的所有链接,否则,则启动蜘蛛或爬虫程序获取该网站下的所有链接。

3. 根据权利要求 1 或 2 所述的方法,其中,所述对网站的第三方应用进行识别,针对所述网站的第三方应用进行与之匹配的漏洞扫描的步骤包括:

获取第三方应用的名称;并且

根据预置的第三方应用与漏洞扫描方法的对应关系,执行对第三方应用匹配的漏洞扫描。

4. 根据权利要求 3 所述的方法,其中,

所述获取第三方应用的名称的步骤包括:

根据网站的版权信息来提取第三方应用程序的名称;或者

根据第三方应用特有的级联样式表 CSS 或 JavaScript 文件的 MD5 来判断、识别网站使用的是第三方应用程序的名称;或者

根据第三方应用的登录文件来判断网站使用的是第三方应用程序的名称。

5. 一种网站漏洞快速扫描设备,其包括:

爬虫优化模块,其适于从搜索引擎的搜索链接库中获取该网站下的所有链接;

扫描模块,其适于对所述获取的链接进行漏洞扫描,对网站的第三方应用进行识别,针对所述网站的第三方应用进行与之匹配的漏洞扫描;

漏洞测试模块,其适于判断第三方应用是否位于 N 级以上目录,若是,则不进行第三方应用漏洞扫描,其中 N 为大于 1 的整数,若所述第三方应用位于 N 级目录以下,则构造漏洞测试 URL 进行漏洞测试。

6. 根据权利要求 5 所述的设备,所述设备还包括:

首次扫描判断模块,其适于判断对所述网站的域名的扫描是否为首次扫描,如果不是首次扫描则直接退出;

爬虫接口判断模块,其适于在首次扫描判断模块判断对所述网站的域名的扫描为首次扫描时,判断搜索链接库中该网站的爬虫接口是否可用;

普通爬虫模块,其适于在爬虫接口判断模块判断搜索引擎的爬虫接口在所述网站不可用时,启用蜘蛛或爬虫程序获取该网站下的所有链接;

其中,在爬虫接口判断模块判断搜索链接库中该网站的爬虫接口可用时,所述爬虫优

化模块从搜索引擎的搜索链接库中获取该网站下的所有链接。

7. 根据权利要求 5 或 6 所述的设备, 其中,

所述扫描模块包括 :

获取单元, 其适于获取第三方应用的名称 ; 以及

漏洞扫描单元, 其适于根据预置的第三方应用与漏洞扫描方法的对应关系, 执行对第三方应用匹配的漏洞扫描。

8. 根据权利要求 7 所述的设备, 其中,

所述获取单元通过如下方式获取第三方应用的名称 :

根据网站的版权信息来提取第三方应用程序的名称 ; 或者

根据第三方应用特有的级联样式表 CSS 或 JavaScript 文件的 MD5 来判断、识别网站使用的是第三方应用程序的名称 ; 或者

根据第三方应用的登录文件来判断网站使用的是第三方应用程序的名称。

## 一种网站漏洞快速扫描方法及设备

### 技术领域

[0001] 本发明涉及计算机技术领域，尤其涉及一种网站漏洞快速扫描方法及设备。

### 背景技术

[0002] 用户第一次使用某种产品的时候会有一种新鲜感，都希望尽快体验产品的功能。对于网站安全扫描产品也是如此，现有技术中用户首次对网站进行扫描时需要耗费大量的时间，主要是因为扫描平台并没有针对新用户做任何特殊优化，往往会接踵而至的逐项进行常规扫描。然而，用户刚开始接触此类产品时，对于网站安全扫描产品的各项功能可能并不熟悉，希望尽快的体验一下产品的各项功能，但极其耗时的常规扫描使得用户短时间内无法顺利进行其他操作，最终导致极大的降低了用户的体验。

### 发明内容

[0003] 鉴于上述问题，提出了本发明以便提供一种克服上述问题或至少部分地解决上述问题的一种网站漏洞快速扫描方法及设备，能够加快漏洞扫描的速度。

[0004] 本发明的实施例提供了一种网站漏洞快速扫描方法，包括：从搜索引擎的搜索链接库中获取该网站下的所有链接；对所述获取的链接进行漏洞扫描；对网站的第三方应用进行识别，针对所述网站的第三方应用进行与之匹配的漏洞扫描；以及判断第三方应用是否位于 N 级目录以上，若是，则不进行第三方应用漏洞扫描；若所述第三方应用位于 N 级目录以下，则构造漏洞测试统一资源定位符 URL 进行漏洞测试，其中 N 为大于 1 的整数。

[0005] 可选地，所述从搜索引擎的搜索链接库中获取该网站下的所有链接的步骤之前，还包括：判断针对所述网站的域名的扫描是否为首次扫描，若是，则判断搜索链接库中该网站的爬虫接口是否可用，若是，则从搜索引擎的搜索链接库中获取该网站下的所有链接，否则，则启动蜘蛛或爬虫程序获取该网站下的所有链接。

[0006] 可选地，所述对网站的第三方应用进行识别，针对所述网站的第三方应用进行与之匹配的漏洞扫描的步骤包括：获取第三方应用的名称；并且根据预置的第三方应用与漏洞扫描方法的对应关系，执行对第三方应用匹配的漏洞扫描。

[0007] 可选地，所述获取第三方应用的名称的步骤包括：根据网站的版权信息来提取第三方应用程序的名称；或者根据第三方应用特有的级联样式表 CSS 或 JavaScript 文件的 MD5 来判断、识别网站使用的是第三方应用程序的名称；或者根据第三方应用的登录文件来判断网站使用的是第三方应用程序的名称。

[0008] 根据本发明的另一个方面，还提供了一种网站漏洞快速扫描设备，其包括：爬虫优化模块，其适于从搜索引擎的搜索链接库中获取该网站下的所有链接；扫描模块，其适于对所述获取的链接进行漏洞扫描，对网站的第三方应用进行识别，针对所述网站的第三方应用进行与之匹配的漏洞扫描；漏洞测试模块，其适于判断第三方应用是否位于 N 级以上目录，若是，则不进行第三方应用漏洞扫描，其中 N 为大于 1 的整数，若所述第三方应用位于 N 级目录以下，则构造漏洞测试 URL 进行漏洞测试。

[0009] 可选地，设备还包括：首次扫描判断模块，其适于判断对所述网站的域名的扫描是否为首次扫描，如果不是首次扫描则直接退出；爬虫接口判断模块，其适于在首次扫描判断模块判断对所述网站的域名的扫描为首次扫描时，判断搜索链接库中该网站的爬虫接口是否可用；普通爬虫模块，其适于在爬虫接口判断模块判断搜索引擎的爬虫接口在所述网站不可用时，启用蜘蛛或爬虫程序获取该网站下的所有链接；其中，在爬虫接口判断模块判断搜索链接库中该网站的爬虫接口可启用时，所述爬虫优化模块从搜索引擎的搜索链接库中获取该网站下的所有链接。

[0010] 可选地，所述扫描模块包括：获取单元，其适于获取第三方应用的名称；以及漏洞扫描单元，其适于根据预置的第三方应用与漏洞扫描方法的对应关系，执行对第三方应用匹配的漏洞扫描。

[0011] 可选地，所述获取单元通过如下方式获取第三方应用的名称：根据网站的版权信息来提取第三方应用程序的名称；或者根据第三方应用特有的 CSS 或 JAVASCRIPT 文件的 MD5 来判断、识别网站使用的是第三方应用程序的名称；或者根据第三方应用的登录文件来判断网站使用的是第三方应用程序的名称。

[0012] 由上述本发明的实施例的技术方案可知，本发明的实施例具有如下有益效果：使用从搜索的链接库里面获取网站的链接来替换原有的爬虫程序，扫描速度可提高 10 倍以上，即原来需要蜘蛛或爬虫程序 30 分钟左右的网站，可在 1 ~ 2 分钟内完成扫描。由于本发明的实施例有效地降低了扫描时间，使第一次使用漏洞扫描平台的用户感受到极速的漏洞扫描效果，用户体验得到极大的提高。

## 附图说明

[0013] 图 1 示出了根据本发明的一个实施方式的网站漏洞快速扫描方法 100 的流程图；以及

[0014] 图 2 示出了根据本发明的另一个实施方式的网站漏洞快速扫描设备 200 的结构框图。

## 具体实施方式

[0015] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术大员。

[0016] 参考图 1，其中示意性地示出了根据本发明的实施例的网站漏洞快速扫描方法 100 的流程图。如图 1 所示，所示方法 100 包括：步骤 S101、S103、S105、S107、S109 和 S111，方法 100 始于步骤 S101，其中，从搜索引擎的搜索链接库中获取该网站下的所有链接。

[0017] 可选地，在步骤 S101 中，搜索引擎的搜索链接库中保存了之前爬虫或蜘蛛获得的各种链接，可以从中提取出所需网站的相关链接，而无需启动爬虫重新获得数据。

[0018] 可选地，也可以通过其他搜索引擎获得与该网站相关的链接，例如在搜索引擎的搜索栏中输入网站域名（如 webscan.360.cn），获取该网站下的所有链接（如得到 webscan.360.cn/news/、webscan.360.cn/task/ 等链接）。该搜索引擎可以是谷歌搜索引

擎、百度搜索引擎。当然可以理解的是，在本发明的实施例中并不限定搜索引擎的具体类型。

[0019] 随后，在步骤 S103 中，对获取的链接进行漏洞扫描。

[0020] 漏洞扫描平台的漏洞库分服务器、目录、文件、参数和内容几个级别。可选地，在本发明的实施例中，可以对获取的链接进行目录级别的漏洞扫描，通过加快目录级别的漏洞扫描速度，即可减少漏洞测试时间，达到漏洞测试优化的目的。

[0021] 随后，在步骤 S105 中，对网站的第三方应用进行识别，针对网站的第三方应用进行与之匹配的漏洞扫描。

[0022] 在本发明的实施例中，通过加快第三方应用的漏洞扫描，即可加快漏洞扫描的时间，而使用第三方应用的漏洞匹配扫描和三级以上目录不执行第三方应用的漏洞扫描是加快扫描速度的最有效的方式。

[0023] 可选地，在本发明的实施例中，可以采用以下方式对第三方应用进行漏洞扫描，首先获取第三方应用的名称；然后根据预置的第三方应用与漏洞扫描方法的对应关系，执行对第三方应用匹配的漏洞扫描。预置的第三方应用与漏洞扫描方法的对应关系可记录下在表中，参见下表：

[0024]

第三方应用的名称 1	漏洞扫描方法 1
第三方应用的名称 2	漏洞扫描方法 2
第三方应用的名称 3	漏洞扫描方法 3

[0025] 其中，获取第三方应用的名称的步骤包括：根据网站的版权信息来提取第三方应用程序的名称；或者根据第三方应用特有的 CSS(Cascading Style Sheet, 级联样式表) 或 JAVASCRIPT 文件的 MD5(消息摘要算法第五版) 来判断、识别网站使用的是第三方应用程序的名称；或者根据第三方应用的登录文件来判断网站使用的是第三方应用程序的名称。

[0026] 具体地，在本发明的实施例中，可以采用以下方式获取第三方应用的名称：

[0027] 方式一、根据 Powered By 来提取第三方应用的名称；

[0028] 如果提取不到，则可根据各种第三方应用程序的 CSS 或 JAVASCRIPT 文件的 MD5 来判断。如某网站上的 abc.js 的 MD5 编码与 Discuz 的 aaa.js 的 MD5 编码一致，就识别为这个网站使用的是 Discuz。

[0029] 方式二、根据第三方应用程序的一些特有文件来判断，如特有文件可以是登录文件。

[0030] 当不存在与之匹配的第三方应用漏洞扫描方法时，进行下面的步骤：在步骤 S107 中，判断第三方应用是否位于 N 级以上目录，其中 N 为大于 1 的整数。

[0031] 可选地，在本发明的实施例中，N 级可以为 3 级，例如：webscan.360.cn/aa/bb/cc/dd/ 这种为 3 级以上目录。当然可以理解的是，在本发明的实施例中并不限定 N 的取值。

[0032] 如果第三方应用位于 N 级目录以上，则进入步骤 S109，在步骤 S109 中，不进行第三方应用漏洞扫描。

[0033] 如果第三方应用位于 N 级目录以下，则进入步骤 S111，在步骤 S111 中，构造漏洞测

试 URL(统一资源定位符)进行漏洞测试。

[0034] 可选地,在步骤 S111 中,通过抓取到的 URL,然后构造漏洞测试 URL 去测试,如果构造的 URL 符合某种漏洞判断规则(如 XSS、SQL 注入等),则表示发现了一个漏洞。

[0035] 根据本发明的实施例,所述网站漏洞快速扫描方法 100 还可以包括一个或者多个可选步骤,以实现额外或者附加的功能,然而这些可选步骤对于实现本发明的目的而言并非是不可或缺的,根据本发明的实施例的网站漏洞快速扫描方法 100 完全可以在没有这些可选步骤的情况下,实现本发明的目的。这些可选步骤未在图 1 中示出,但它们与上述各步骤之间的先后执行可以由本领域技术人员根据下述教导而容易地得出。需要指出的是,只要没有特别说明,这些可选步骤连同上述步骤的执行顺序可以根据实际需要进行选择。

[0036] 可选地,在本发明的实施例中,在步骤 S101 之前,方法 100 还包括:步骤 S113、步骤 S115、步骤 S117、步骤 S119 和步骤 S121。

[0037] 在步骤 S113 中,判断针对网站的域名的扫描是否为首次扫描。当然可以理解的是,在本发明的实施例中并不限定判断的具体方式。

[0038] 如果在步骤 S113 中判断出针对网站的域名的扫描不是首次扫描,则进入步骤 S115,在步骤 S115 中,启动蜘蛛或爬虫程序获取该网站下的所有链接。

[0039] 如果在步骤 S113 中判断出针对网站的域名的扫描是首次扫描,则进入步骤 S117,在步骤 S117 中,判断搜索链接库中该网站的爬虫接口是否可用。

[0040] 如果在步骤 S117 中判断搜索链接库中该网站的爬虫接口可用,则进入步骤 S119,在步骤 S119 中,从搜索引擎的搜索链接库中获取该网站下的所有链接。

[0041] 可选地,在本发明的实施例中,爬虫接口目前可以直接从 HBase 集群中获取数据,输入一个网站的域名(如 webscan.360.cn),输出这个网站的下的所有链接(webscan.360.cn/news/、webscan.360.cn/task/ 等)。

[0042] 如果在步骤 S117 中判断搜索链接库中该网站的爬虫接口不可用,则进入步骤 S121,在步骤 S121 中启动蜘蛛或爬虫程序获取该网站下的所有链接。

[0043] 在本发明的实施例中使用从搜索的链接库里面获取网站的链接来替换原有的爬虫程序,扫描速度可提高 10 倍以上,即原来需要蜘蛛或爬虫程序 30 分钟左右的网站,在本发明的实施例中可在 1~2 分钟内完成扫描。由于本发明的实施例有效地降低了扫描时间,使第一次使用漏洞扫描平台的用户感受到极速的漏洞扫描效果,用户体验得到极大的提高。

[0044] 如图 2 所示,根据本发明的实施例的网站漏洞快速扫描设备 200 可以主要包括:爬虫优化模块 210、扫描模块 230、漏洞测试模块 250。应当理解,图 2 中所表示的各个模块的连接关系仅为示例,本领域技术人员完全可以采用其它的连接关系,只要在这样的连接关系下各个模块也能够实现本发明的功能即可。

[0045] 在本说明书中,各个模块的功能可以通过使用专用硬件、或者能够与适当的软件相结合来执行处理的硬件来实现。这样的硬件或专用硬件可以包括专用集成电路(ASIC)、各种其它电路、各种处理器等。当由处理器实现时,该功能可以由单个专用处理器、单个共享处理器、或者多个独立的处理器(其中某些可能被共享)来提供。另外,处理器不应该被理解为专指能够执行软件的硬件,而是可以隐含地包括、而不限于数字信号处理器(DSP)硬件、用来存储软件的只读存储器(ROM)、随机存取存储器(RAM)、以及非易失存储设备。

[0046] 在本发明的实施例中,爬虫优化模块 210,其适于从搜索引擎的搜索链接库中获取该网站下的所有链接;可选地,搜索引擎的搜索链接库中保存了之前爬虫或蜘蛛获得的各种链接,可以从中提取出所需网站的相关链接,而无需启动爬虫重新获得数据。

[0047] 在本发明的实施例中,扫描模块 230,其适于对所述获取的链接进行漏洞扫描,对网站的第三方应用进行识别,针对所述网站的第三方应用进行与之匹配的漏洞扫描;

[0048] 在本发明的实施例中,漏洞测试模块 250,其适于判断第三方应用是否位于 N 级以上目录,若是,则不进行第三方应用漏洞扫描,其中 N 为大于 1 的整数,若所述第三方应用位于 N 级目录以下,则构造漏洞测试 URL 进行漏洞测试。

[0049] 可选地,在本发明的实施例中,设备 200 还包括:

[0050] 首次扫描判断模块,其适于判断对所述网站的域名的扫描是否为首次扫描,如果不是首次扫描则直接退出;

[0051] 爬虫接口判断模块,其适于在首次扫描判断模块判断对所述网站的域名的扫描为首次扫描时,判断搜索链接库中该网站的爬虫接口是否可用;

[0052] 普通爬虫模块,其适于在爬虫接口判断模块判断搜索引擎的爬虫接口在所述网站不可用时,启用蜘蛛或爬虫程序获取该网站下的所有链接;

[0053] 其中,在爬虫接口判断模块判断搜索链接库中该网站的爬虫接口可用时,所述爬虫优化模块从搜索引擎的搜索链接库中获取该网站下的所有链接。

[0054] 可选地,在本发明的实施例中,所述扫描模块 230 包括:

[0055] 获取单元,其适于获取第三方应用的名称;以及

[0056] 漏洞扫描单元,其适于根据预置的第三方应用与漏洞扫描方法的对应关系,执行对第三方应用匹配的漏洞扫描。

[0057] 可选地,在本发明的实施例中,所述获取单元通过如下方式获取第三方应用的名称:

[0058] 根据网站的版权信息来提取第三方应用程序的名称;或者

[0059] 根据第三方应用特有的 CSS 或 JAVASCRIPT 文件的 MD5 来判断、识别网站使用的是第三方应用程序的名称;或者

[0060] 根据第三方应用的登录文件来判断网站使用的是第三方应用程序的名称。

[0061] 尽管已经结合特定的实施例描述了本发明,但是并不是限定于此处描述的特定形式。而是,本发明的范围仅仅由后附的权利要求限制。在权利要求中,术语“包括”不排除存在其它部件或步骤。此外,尽管各个特征可以包括在不同的权利要求中,但是这些特征可以被有利地组合,且在不同权利要求中包含的内容不意味着特征的组合是不可行和 / 或不利的。此外,单个的含义不排除多个。因此,“个”、“第一”、“第二”等的含义不排除多个。此外,权利要求中的附图标记不应被解释为对范围的限制。

[0062] 以上所述仅是本发明的具体实施方式,应当指出的是,对于本领域的普通技术人员来说,在不脱离本发明精神的前提下,可以作出若干改进、修改、和变形,这些改进、修改、和变形都应视为落在本申请的保护范围内。

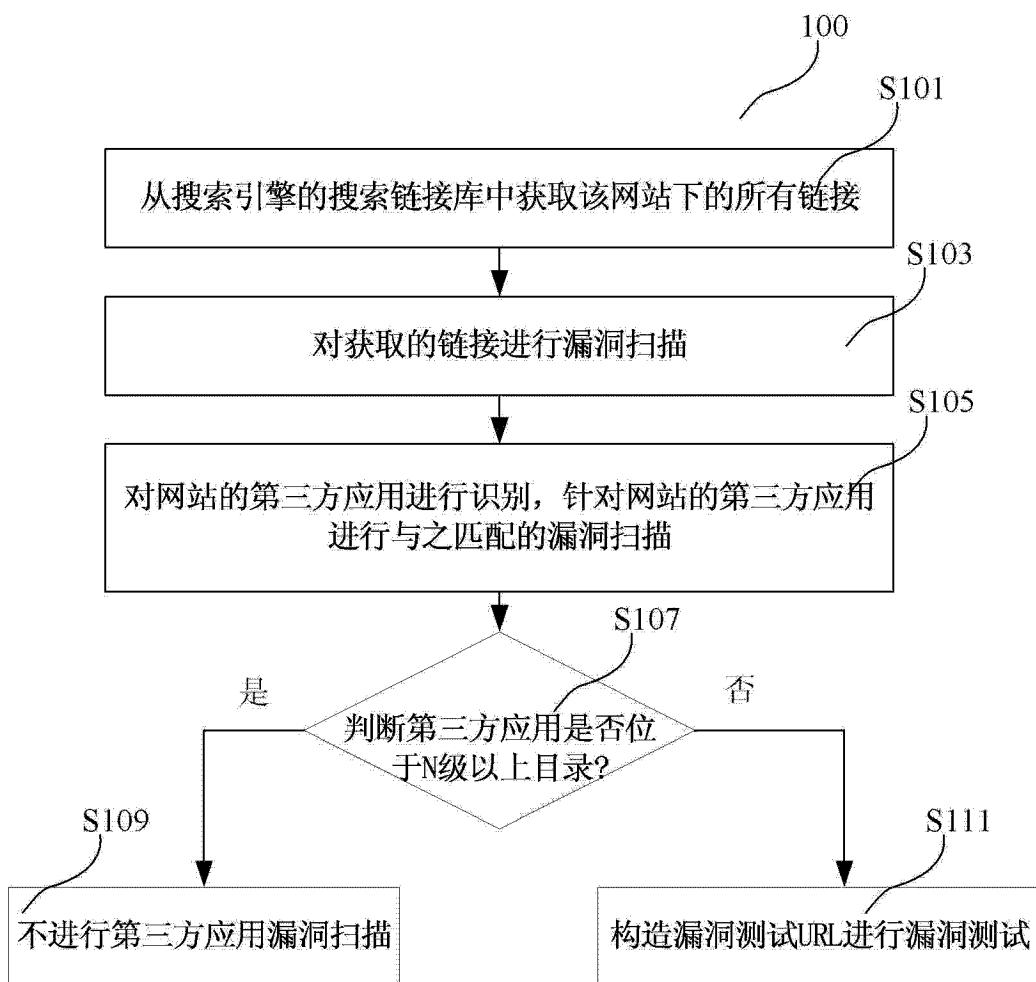


图 1

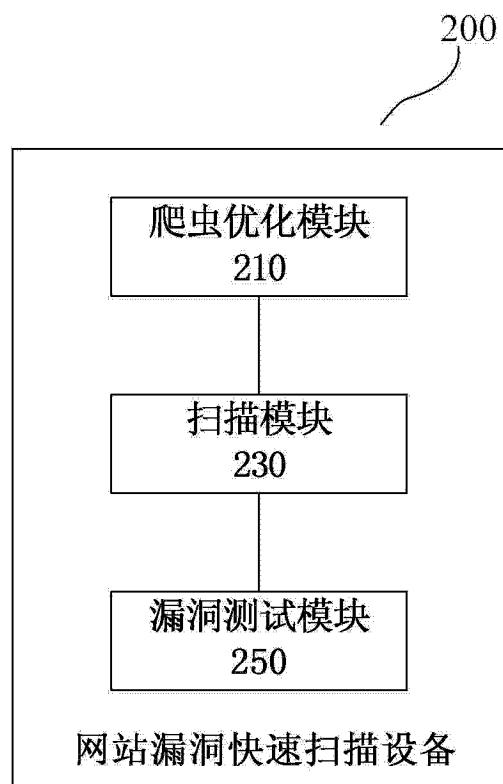


图 2