



(12) 发明专利申请

(10) 申请公布号 CN 104935603 A

(43) 申请公布日 2015. 09. 23

(21) 申请号 201510352970. 2

(22) 申请日 2015. 06. 24

(71) 申请人 郑州悉知信息技术有限公司

地址 450000 河南省郑州市高新区科学大道
169 号 1 幢 1 号楼

(72) 发明人 李国强 周坤朋 孟伟 王路

(74) 专利代理机构 北京集佳知识产权代理有限
公司 11227

代理人 王宝筠

(51) Int. Cl.

H04L 29/06(2006. 01)

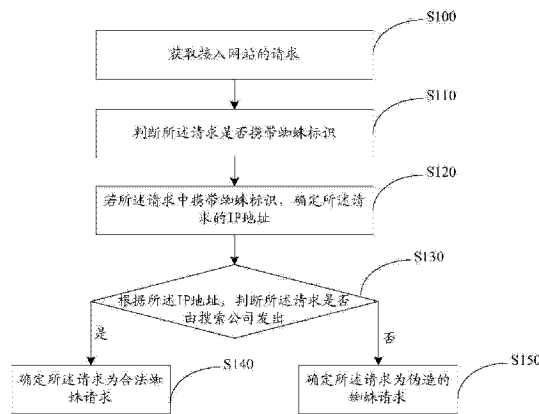
权利要求书3页 说明书10页 附图6页

(54) 发明名称

一种识别请求的方法及网站服务器

(57) 摘要

本发明实施例提供一种识别请求的方法及网站服务器,其中方法包括:获取接入网站的请求;判断所述请求是否携带蜘蛛标识;若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址;根据所述 IP 地址,判断所述请求是否由搜索公司发出;若所述请求由搜索公司发出,则确定所述请求为合法蜘蛛请求;若所述请求不是由搜索公司发出,则确定所述请求为伪造的蜘蛛请求。本发明可从接入网站的请求中有效识别出伪造的蜘蛛请求,为后续屏蔽伪造的蜘蛛请求,保证网站信息的安全提供了可能。



1. 一种识别请求的方法,其特征在于,包括:
 - 获取接入网站的请求;
 - 判断所述请求是否携带蜘蛛标识;
 - 若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址;
 - 根据所述 IP 地址,判断所述请求是否由搜索公司发出;
 - 若所述请求由搜索公司发出,则确定所述请求为合法蜘蛛请求;
 - 若所述请求不是由搜索公司发出,则确定所述请求为伪造的蜘蛛请求。
2. 根据权利要求 1 所述的识别请求的方法,其特征在于,还包括:
 - 放行所述合法蜘蛛请求,以便所述合法蜘蛛请求抓取网站信息;
 - 屏蔽所述伪造的蜘蛛请求。
3. 根据权利要求 1 或 2 所述的识别请求的方法,其特征在于,所述根据所述 IP 地址,判断所述请求是否由搜索公司发出包括:
 - 判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址;
 - 若是,则确定所述请求由搜索公司发出;
 - 若否,则确定所述请求不是由搜索公司发出。
4. 根据权利要求 3 所述的识别请求的方法,其特征在于,所述判断所述 IP 地址是否为合法 IP 地址包括:
 - 调取预设的合法 IP 地址,判断所述 IP 地址是否与预设的合法 IP 地址对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址;
 - 或,确定所述 IP 地址对应的域名信息,判断所述域名信息是否与预设的搜索公司的域名信息对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址。
5. 根据权利要求 1 或 2 所述的识别请求的方法,其特征在于,所述根据所述 IP 地址,判断所述请求是否由搜索公司发出包括:
 - 确定所述 IP 地址的归属地址;
 - 判断所述归属地址与设定的搜索公司的所在地址是否对应;
 - 若是,则确定所述请求由搜索公司发出;
 - 若否,则确定所述请求不是由搜索公司发出。
6. 根据权利要求 1 或 2 所述的识别请求的方法,其特征在于,所述根据所述 IP 地址,判断所述请求是否由搜索公司发出包括:
 - 判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址;
 - 若否,则确定所述请求不是由搜索公司发出;
 - 若是,确定所述 IP 地址的归属地址,并判断所述归属地址与设定的搜索公司的所在地址是否对应;
 - 若所述归属地址与设定的搜索公司的所在地址相对应,则确定所述请求由搜索公司发出;
 - 若所述归属地址与设定的搜索公司的所在地址不对应,则确定所述请求不是由搜索公司发出。
7. 一种网站服务器,其特征在于,包括:

请求获取模块,用于获取接入网站的请求 ;
标识判断模块,用于判断所述请求是否携带蜘蛛标识 ;
地址确定模块,用于若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址 ;
判断模块,用于根据所述 IP 地址,判断所述请求是否由搜索公司发出 ;
第一结果确定模块,用于若所述请求由搜索公司发出,则确定所述请求为合法蜘蛛请求 ;
第二结果确定模块,用于若所述请求不是由搜索公司发出,则确定所述请求为伪造的蜘蛛请求。

8. 根据权利要求 7 所述的网站服务器,其特征在於,所述判断模块包括 :

第一地址合法判断单元,用于判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址 ;

第一合法确定单元,用于若第一地址合法判断单元的判断结果为是,则确定所述请求由搜索公司发出 ;

第二合法确定单元,用于若第一地址合法判断单元的判断结果为否,则确定所述请求不是由搜索公司发出 ;

所述第一地址合法判断单元包括 :

第一判断执行子单元,用于调取预设的合法 IP 地址,判断所述 IP 地址是否与预设的合法 IP 地址对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址 ;

或,第二判断执行子单元,用于确定所述 IP 地址对应的域名信息,判断所述域名信息是否与预设的搜索公司的域名信息对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址。

9. 根据权利要求 7 所述的网站服务器,其特征在於,所述判断模块包括 :

第一归属地址判断单元,用于确定所述 IP 地址的归属地址,判断所述归属地址与设定的搜索公司的所在地址是否对应 ;

第一地址确定单元,用于若所述第一归属地址判断单元的判断结果为是,则确定所述请求由搜索公司发出 ;

第二地址确定单元,用于若所述第一归属地址判断单元的判断结果为否,则确定所述请求不是由搜索公司发出。

10. 根据权利要求 7 所述的网站服务器,其特征在於,所述判断模块包括 :

第二地址合法判断单元,用于判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址 ;

第一确定单元,用于若所述第二地址合法判断单元的判断结果为否,则确定所述请求不是由搜索公司发出 ;

第二归属地址判断单元,用于若第二地址合法判断单元的判断结果为是,确定所述 IP 地址的归属地址,并判断所述归属地址与设定的搜索公司的所在地址是否对应 ;

第二确定单元,用于若所述第二归属地址判断单元的判断结果为是,则确定所述请求由搜索公司发出 ;

第三确定单元,用于若所述第二归属地址判断单元的判断结果为否,则确定所述请求

不是由搜索公司发出。

一种识别请求的方法及网站服务器

技术领域

[0001] 本发明涉及搜索技术领域,具体涉及一种识别请求的方法及网站服务器。

背景技术

[0002] 蜘蛛请求(又被称为网页蜘蛛,网络机器人)是一种按照一定规则,自动的抓取互联网信息的程序或者脚本;蜘蛛请求一般由提供搜索服务的搜索公司发出,搜索公司通过蜘蛛请求可定向抓取相关的网站信息(如网站的网页信息等),并将抓取到的网站信息整合到搜索引擎中,提升搜索引擎输出搜索结果的全面性和准确性;由于网站信息被蜘蛛请求抓取后,能够被搜索公司整合入搜索引擎中,提升网站被用户搜索到的可能性,因此大多数的网站对于由搜索公司所发出的蜘蛛请求是持欢迎态度的,并不对其进行屏蔽。

[0003] 本发明的发明人在研究过程中发现,目前一些非搜索公司出于窃取网站信息的目的,存在伪造蜘蛛请求,恶意采集网站信息的情况;出于保证网站信息的安全,对于伪造的蜘蛛请求进行识别,从而在后续加以屏蔽,显得尤为必要;然而,伪造的蜘蛛请求与由搜索公司发出的合法蜘蛛请求在内容上基本一致,目前并没有一种能够有效识别伪造的蜘蛛请求的技术。

[0004] 因此,如何提供一种能够识别伪造的蜘蛛请求的方法,成为本领域技术人员需要考虑的问题。

发明内容

[0005] 有鉴于此,本发明实施例提供一种识别请求的方法及网站服务器,以实现有效识别伪造的蜘蛛请求的目的。

[0006] 为实现上述目的,本发明实施例提供如下技术方案:

[0007] 一种识别请求的方法,包括:

[0008] 获取接入网站的请求;

[0009] 判断所述请求是否携带蜘蛛标识;

[0010] 若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址;

[0011] 根据所述 IP 地址,判断所述请求是否由搜索公司发出;

[0012] 若所述请求由搜索公司发出,则确定所述请求为合法蜘蛛请求;

[0013] 若所述请求不是由搜索公司发出,则确定所述请求为伪造的蜘蛛请求。

[0014] 其中,所述方法还包括:

[0015] 放行所述合法蜘蛛请求,以便所述合法蜘蛛请求抓取网站信息;

[0016] 屏蔽所述伪造的蜘蛛请求。

[0017] 其中,所述根据所述 IP 地址,判断所述请求是否由搜索公司发出包括:

[0018] 判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址;

[0019] 若是,则确定所述请求由搜索公司发出;

- [0020] 若否,则确定所述请求不是由搜索公司发出。
- [0021] 其中,所述判断所述 IP 地址是否为合法 IP 地址包括:
- [0022] 调取预设的合法 IP 地址,判断所述 IP 地址是否与预设的合法 IP 地址对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址;
- [0023] 或,确定所述 IP 地址对应的域名信息,判断所述域名信息是否与预设的搜索公司的域名信息对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址。
- [0024] 其中,所述根据所述 IP 地址,判断所述请求是否由搜索公司发出包括:
- [0025] 确定所述 IP 地址的归属地址;
- [0026] 判断所述归属地址与设定的搜索公司的所在地址是否对应;
- [0027] 若是,则确定所述请求由搜索公司发出;
- [0028] 若否,则确定所述请求不是由搜索公司发出。
- [0029] 其中,所述根据所述 IP 地址,判断所述请求是否由搜索公司发出包括:
- [0030] 判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址;
- [0031] 若否,则确定所述请求不是由搜索公司发出;
- [0032] 若是,确定所述 IP 地址的归属地址,并判断所述归属地址与设定的搜索公司的所在地址是否对应;
- [0033] 若所述归属地址与设定的搜索公司的所在地址相对应,则确定所述请求由搜索公司发出;
- [0034] 若所述归属地址与设定的搜索公司的所在地址不对应,则确定所述请求不是由搜索公司发出。
- [0035] 本发明实施例还提供一种网站服务器,包括:
- [0036] 请求获取模块,用于获取接入网站的请求;
- [0037] 标识判断模块,用于判断所述请求是否携带蜘蛛标识;
- [0038] 地址确定模块,用于若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址;
- [0039] 判断模块,用于根据所述 IP 地址,判断所述请求是否由搜索公司发出;
- [0040] 第一结果确定模块,用于若所述请求由搜索公司发出,则确定所述请求为合法蜘蛛请求;
- [0041] 第二结果确定模块,用于若所述请求不是由搜索公司发出,则确定所述请求为伪造的蜘蛛请求。
- [0042] 其中,所述判断模块包括:
- [0043] 第一地址合法判断单元,用于判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址;
- [0044] 第一合法确定单元,用于若第一地址合法判断单元的判断结果为是,则确定所述请求由搜索公司发出;
- [0045] 第二合法确定单元,用于若第一地址合法判断单元的判断结果为否,则确定所述请求不是由搜索公司发出;
- [0046] 所述地址合法判断单元包括:

[0047] 第一判断执行子单元,用于调取预设的合法 IP 地址,判断所述 IP 地址是否与预设的合法 IP 地址对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址;

[0048] 或,第二判断执行子单元,用于确定所述 IP 地址对应的域名信息,判断所述域名信息是否与预设的搜索公司的域名信息对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址。

[0049] 其中,所述判断模块包括:

[0050] 第一归属地址判断单元,用于确定所述 IP 地址的归属地址,判断所述归属地址与设定的搜索公司的所在地址是否对应;

[0051] 第一地址确定单元,用于若所述第一归属地址判断单元的判断结果为是,则确定所述请求由搜索公司发出;

[0052] 第二地址确定单元,用于若所述第一归属地址判断单元的判断结果为否,则确定所述请求不是由搜索公司发出。

[0053] 其中,所述判断模块包括:

[0054] 第二地址合法判断单元,用于判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址;

[0055] 第一确定单元,用于若所述第二地址合法判断单元的判断结果为否,则确定所述请求不是由搜索公司发出;

[0056] 第二归属地址判断单元,用于若第二地址合法判断单元的判断结果为是,确定所述 IP 地址的归属地址,并判断所述归属地址与设定的搜索公司的所在地址是否对应;

[0057] 第二确定单元,用于若所述第二归属地址判断单元的判断结果为是,则确定所述请求由搜索公司发出;

[0058] 第三确定单元,用于若所述第二归属地址判断单元的判断结果为否,则确定所述请求不是由搜索公司发出。

[0059] 基于上述技术方案,本发明实施例提供的识别请求的方法,包括:获取接入网站的请求;判断所述请求是否携带蜘蛛标识;若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址;根据所述 IP 地址,判断所述请求是否由搜索公司发出;若所述请求由搜索公司发出,则确定所述请求为合法蜘蛛请求;若所述请求不是由搜索公司发出,则确定所述请求为伪造的蜘蛛请求。可以看出,本发明实施例通过蜘蛛标识,可从接入网站的请求中确定出请求的识别范围,即合法蜘蛛请求和伪造的蜘蛛请求;从而根据携带有蜘蛛标识的请求的 IP 地址,判断所述请求是否由搜索公司,当确定所述请求由搜索公司发出时,则确定所述请求为合法蜘蛛请求,当确定所述请求不是由搜索公司发出时,则确定所述请求为伪造的蜘蛛请求;进而实现从接入网站的请求中有效识别伪造的蜘蛛请求的目的,为后续屏蔽伪造的蜘蛛请求,保证网站信息的安全提供了可能。

附图说明

[0060] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术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的前提下,还可以根据

提供的附图获得其他的附图。

- [0061] 图 1 为本发明实施例提供的识别请求的方法的流程图；
- [0062] 图 2 为本发明实施例提供的识别请求的方法的另一流程图；
- [0063] 图 3 为本发明实施例提供的判断请求是否由搜索公司发出的方法流程图；
- [0064] 图 4 为本发明实施例提供的判断 IP 地址是否为合法 IP 地址的方法流程图；
- [0065] 图 5 为本发明实施例提供的判断 IP 地址是否为合法 IP 地址的另一方法流程图；
- [0066] 图 6 为本发明实施例提供的判断请求是否由搜索公司发出的另一方法流程图；
- [0067] 图 7 为本发明实施例提供的判断请求是否由搜索公司发出的再一方法流程图；
- [0068] 图 8 为本发明实施例提供的判断请求是否由搜索公司发出的又一方法流程图；
- [0069] 图 9 为本发明实施例提供的网站服务器的结构框图；
- [0070] 图 10 为本发明实施例提供的判断模块的结构框图；
- [0071] 图 11 为本发明实施例提供的第一地址合法判断单元的结构框图；
- [0072] 图 12 为本发明实施例提供的第一地址合法判断单元的另一结构框图；
- [0073] 图 13 为本发明实施例提供的判断模块的另一结构框图；
- [0074] 图 14 为本发明实施例提供的判断模块再一结构框图。

具体实施方式

[0075] 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,所描述的实施例仅仅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做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。

[0076] 图 1 为本发明实施例提供的识别请求的方法的流程图,该方法可应用于网站中接入请求的设备中,如可应用于网站服务器中,由网站服务器从接入网站的请求中识别出合法蜘蛛请求和伪造的蜘蛛请求;参照图 1,该方法可以包括:

[0077] 步骤 S100、获取接入网站的请求;

[0078] 接入网站的请求中包括:请求网站信息(如网页信息)的用户访问请求,抓取网站信息(如网页信息)的由搜索公司发出的合法蜘蛛请求,和恶意采集网站信息的伪造的蜘蛛请求等;

[0079] 本发明实施例需从接入网站的请求中确定出蜘蛛请求,从而识别出该蜘蛛请求是合法蜘蛛请求,还是伪造的蜘蛛请求。

[0080] 步骤 S110、判断所述请求是否携带蜘蛛标识;

[0081] 在接入网站的请求中判断出携带有蜘蛛标识的请求后,可将识别请求的范围进行缩小,排除出用户访问请求;

[0082] 搜索公司发出的蜘蛛请求中均会携带蜘蛛标识,该蜘蛛标识可表示当前请求为由搜索公司发出的抓取网站信息的蜘蛛请求;现有技术中,网站服务器在判断出请求携带蜘蛛标识时,一般均会放行,不加以屏蔽;然而,蜘蛛标识极易被伪造,存在蜘蛛标识的请求可能是合法蜘蛛请求也可能是伪造的蜘蛛请求,现有技术对存在蜘蛛标识的请求均进行放行,极有可能将伪造的蜘蛛请求也放行了,导致网站信息被恶意采集;

[0083] 因此,本发明实施例从接入网站的请求中,识别出携带蜘蛛标识的请求,将识别请

求的范围进行缩小后,还需在后续对存在蜘蛛标识的请求作进一步处理识别。

[0084] 步骤 S120、若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址;

[0085] 本发明实施例可对携带有蜘蛛标识的请求进行反解析,确定出携带有蜘蛛标识的请求的 IP 地址。

[0086] 步骤 S130、根据所述 IP 地址,判断所述请求是否由搜索公司发出,若是,执行步骤 S140,若否,执行步骤 S150;

[0087] 合法蜘蛛请求和伪造的蜘蛛请求的发出方并不相同,合法蜘蛛请求一般由搜索公司发出,而伪造的蜘蛛请求并不是由搜索公司发出的;因此,本发明实施例可通过携带有蜘蛛标识的请求的 IP 地址,反向追踪到携带有蜘蛛标识的请求的发出方,进而判断该追踪到的发出方是否为搜索公司;

[0088] 若追踪到的发出方为搜索公司,则可确定携带有蜘蛛标识的请求为合法蜘蛛请求;若追踪到的发出方不为搜索公司,则可确定携带有蜘蛛标识的请求为伪造的蜘蛛请求。

[0089] 步骤 S140、确定所述请求为合法蜘蛛请求;

[0090] 步骤 S150、确定所述请求为伪造的蜘蛛请求。

[0091] 本发明实施例提供的识别请求的方法,包括:获取接入网站的请求;判断所述请求是否携带蜘蛛标识;若所述请求中携带蜘蛛标识,确定所述请求的 IP 地址;根据所述 IP 地址,判断所述请求是否由搜索公司发出;若所述请求由搜索公司发出,则确定所述请求为合法蜘蛛请求;若所述请求不是由搜索公司发出,则确定所述请求为伪造的蜘蛛请求。可以看出,本发明实施例通过蜘蛛标识,可从接入网站的请求中确定出请求的识别范围,即合法蜘蛛请求和伪造的蜘蛛请求;从而根据携带有蜘蛛标识的请求的 IP 地址,判断所述请求是否由搜索公司,当确定所述请求由搜索公司发出时,则确定所述请求为合法蜘蛛请求,当确定所述请求不是由搜索公司发出时,则确定所述请求为伪造的蜘蛛请求;进而实现从接入网站的请求中有效识别伪造的蜘蛛请求的目的,为后续屏蔽伪造的蜘蛛请求,保证网站信息的安全提供了可能。

[0092] 可选的,在确定出合法蜘蛛请求后,本发明实施例可放行合法蜘蛛请求(即对合法蜘蛛请求不加以屏蔽),以便合法蜘蛛请求抓取网站信息;在确定出伪造的蜘蛛请求后,可屏蔽所述伪造的蜘蛛请求,以保证网站信息的安全;

[0093] 图 2 示出了本发明实施例提供的识别请求的方法的另一流程图,结合图 1 和图 2 所示,图 2 所示方法在图 1 所示的基础上,还包括:

[0094] 步骤 S160、放行所述合法蜘蛛请求,以便所述合法蜘蛛请求抓取网站信息;

[0095] 其中,步骤 S160 在步骤 S140 之后执行。

[0096] 步骤 S170、屏蔽所述伪造的蜘蛛请求。

[0097] 其中,步骤 S170 在步骤 S150 之后执行。

[0098] 可选的,在确定出携带有蜘蛛标识的请求,判断所述请求是否由搜索公司发出时,本发明实施例可通过判断请求的 IP 地址是否为合法 IP 地址实现;图 3 示出了本发明实施例提供的判断请求是否由搜索公司发出的方法流程图,参照图 3,该方法可以包括:

[0099] 步骤 S200、判断所述 IP 地址是否为合法 IP 地址,若是,执行步骤 S210,若否,执行步骤 S220;

[0100] 其中,所述合法 IP 地址为搜索公司对应的 IP 地址;

[0101] 可选的,本发明实施例可预设搜索公司对应的 IP 地址集合,通过判断携带有蜘蛛标识的请求的 IP 地址是否在该 IP 地址集合中,实现所述 IP 地址是否为合法 IP 地址的判断;

[0102] 可选的,由于搜索公司的域名是可确定的,如中国的搜索公司百度的域名为“baidu.com”,搜索公司谷歌的域名为“google.com”等;因此,本发明实施例也可通过预置搜索公司的域名集合,通过解析携带有蜘蛛标识的请求的 IP 地址的域名信息,判断该域名信息是否在该域名集合中,从而实现所述 IP 地址是否为合法 IP 地址的判断。

[0103] 步骤 S210、确定所述请求由搜索公司发出;

[0104] 对应的,所述请求为合法蜘蛛请求。

[0105] 步骤 S220、确定所述请求不是由搜索公司发出。

[0106] 对应的,所述请求为伪造的蜘蛛请求。

[0107] 图 4 示出了本发明实施例提供的判断 IP 地址是否为合法 IP 地址的方法流程图,参照图 4,该方法可以包括:

[0108] 步骤 S300、调取预设的合法 IP 地址;

[0109] 合法蜘蛛请求的 IP 地址可能是 1 对多模式,即一个搜索公司(如百度)发出的蜘蛛请求会有多个,而该搜索公司对应的 IP 地址也可能有多个,但这些 IP 地址均属于合法 IP 地址(合法 IP 地址为搜索公司对应的 IP 地址);

[0110] 而每一个合法蜘蛛请求对应的 IP 地址是唯一可确定的,基于此,本发明实施例可统计出各搜索公司对应的 IP 地址,从而收集到搜索公司对应的 IP 地址集合并预设在网络服务器中;当网络服务器需要判断携带有蜘蛛标识的请求的 IP 地址是否为合法 IP 地址时,可调取该 IP 地址集合记录的合法 IP 地址。

[0111] 步骤 S310、判断所述 IP 地址是否与预设的合法 IP 地址对应,若是,执行步骤 S320,若否,执行步骤 S330;

[0112] 可选的,可判断所述 IP 地址是否在所述 IP 地址集合中记录(即所述 IP 地址是否与所述 IP 地址集合中记录的任一合法 IP 地址相一致);若是,则可确定所述 IP 地址与合法 IP 地址对应;若否,则可确定所述 IP 地址与合法 IP 地址不对应。

[0113] 步骤 S320、确定所述 IP 地址为合法 IP 地址;

[0114] 步骤 S330、确定所述 IP 地址不为合法 IP 地址。

[0115] 图 5 示出了本发明实施例提供的判断 IP 地址是否为合法 IP 地址的另一方法流程图,参照图 5,该方法可以包括:

[0116] 步骤 S400、确定所述 IP 地址对应的域名信息;

[0117] 可选的,本发明实施例可对 IP 地址进行反向查询,得到 IP 地址对应的域名信息。

[0118] 步骤 S410、判断所述域名信息是否与预设的搜索公司的域名信息对应,若是,执行步骤 S420,若否,执行步骤 S430;

[0119] 本发明实施例可预置搜索公司的域名集合,该域名集合记录有搜索公司的域名信息;在确定出携带蜘蛛标识的请求后,可根据所述请求的 IP 地址,反查出所述请求的域名信息;若反查得到的域名信息在域名集合中已记录,则可确定所述 IP 地址为合法 IP 地址;若反查得到的域名信息在域名集合中未记录,则可确定所述 IP 地址不为合法 IP 地址。

[0120] 步骤 S420、确定所述 IP 地址为合法 IP 地址;

[0121] 步骤 S430、确定所述 IP 地址不为合法 IP 地址。

[0122] 可选的,由于搜索公司的所在地址是可确定的,如百度公司的所在地址和谷歌公司的所在地址均是可确定的,而合法蜘蛛请求一般是由搜索公司的所在地址中的设备发出的,因此本发明实施例可通过携带有蜘蛛标识的请求的归属地址,识别该请求是合法蜘蛛请求还是伪造的蜘蛛请求;

[0123] 图 6 示出了本发明实施例提供的判断请求是否由搜索公司发出的另一方法流程图,参照图 6,该方法可以包括:

[0124] 步骤 S500、确定所述 IP 地址的归属地址;

[0125] 可选的,本发明实施例可在网络服务器中预置 IP 地址库,IP 地址库中记录有众多 IP 地址的归属地址;通过该 IP 地址库,网络服务器在确定携带有蜘蛛标识的请求后,可根据所述请求的 IP 地址,确定出所述 IP 地址的归属地址,从而得到所述请求的源发出地址。

[0126] 步骤 S510、判断所述归属地址与设定的搜索公司的所在地址是否对应,若是,执行步骤 S520,若否,执行步骤 S530;

[0127] 本发明实施例可统计各搜索公司的所在地址,从而得到合法蜘蛛请求的源发出地址,进而将所述归属地址(即携带有蜘蛛标识的请求的源发出地址)与合法蜘蛛请求的源发出地址进行比对,识别出所述请求是合法蜘蛛请求还是伪造的蜘蛛请求。

[0128] 步骤 S520、确定所述请求由搜索公司发出;

[0129] 对应的,所述请求为合法蜘蛛请求。

[0130] 步骤 S530、确定所述请求不是由搜索公司发出。

[0131] 对应的,所述请求为伪造的蜘蛛请求。

[0132] 优选的,本发明实施例可结合合法 IP 地址的判断逻辑,与搜索公司的所在地址的判断逻辑,判断请求是否由搜索公司发出,使得最后识别出的合法蜘蛛请求或伪造的蜘蛛请求的识别结果更为准确。

[0133] 图 7 示出了本发明实施例提供的判断请求是否由搜索公司发出的再一方法流程图,参照图 7,该方法可以包括:

[0134] 步骤 S600、判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址,若否,执行步骤 S610,若是,执行步骤 S620;

[0135] 可选的,可通过图 4 或图 5 所示方法,判断所述 IP 地址是否为合法 IP 地址。

[0136] 步骤 S610、确定所述请求不是由搜索公司发出;

[0137] 对应的,所述请求为伪造的蜘蛛请求。

[0138] 步骤 S620、确定所述 IP 地址的归属地址,判断所述归属地址与设定的搜索公司的所在地址是否对应,若是,执行步骤 S630,若否,执行步骤 S610;

[0139] 步骤 S630、确定所述请求由搜索公司发出。

[0140] 对应的,所述请求为合法蜘蛛请求。

[0141] 显然,本发明实施例也可先判断所述 IP 地址的归属地址,再判断所述 IP 地址是否为合法 IP 地址。

[0142] 图 8 示出了本发明实施例提供的判断请求是否由搜索公司发出的又一方法流程图,参照图 8,该方法可以包括:

[0143] 步骤 S700、确定所述 IP 地址的归属地址;

[0144] 步骤 S710、判断所述归属地址与设定的搜索公司的所在地址是否对应, 若否, 执行步骤 S720, 若是, 执行步骤 S730 ;

[0145] 步骤 S720、确定所述请求不是由搜索公司发出 ;

[0146] 对应的, 所述请求为伪造的蜘蛛请求。

[0147] 步骤 S730、判断所述 IP 地址是否为合法 IP 地址, 所述合法 IP 地址为搜索公司对应的 IP 地址, 若是, 执行步骤 S740, 若否, 执行步骤 S720 ;

[0148] 可选的, 可通过图 4 或图 5 所示方法, 判断所述 IP 地址是否为合法 IP 地址。

[0149] 步骤 S740、确定所述请求由搜索公司发出。

[0150] 对应的, 所述请求为合法蜘蛛请求。

[0151] 本发明实施例可从接入网站的请求中有效识别出伪造的蜘蛛请求, 为后续屏蔽伪造的蜘蛛请求, 保证网站信息的安全提供了可能。

[0152] 下面对本发明实施例提供的网站服务器进行介绍, 下文描述的网站服务器可与上文描述的识别请求的方法相互对应参照。

[0153] 图 9 为本发明实施例提供的网站服务器的结构框图, 参照图 9, 该网站服务器可以包括 :

[0154] 请求获取模块 100, 用于获取接入网站的请求 ;

[0155] 标识判断模块 200, 用于判断所述请求是否携带蜘蛛标识 ;

[0156] 地址确定模块 300, 用于若所述请求中携带蜘蛛标识, 确定所述请求的 IP 地址 ;

[0157] 判断模块 400, 用于根据所述 IP 地址, 判断所述请求是否由搜索公司发出 ;

[0158] 第一结果确定模块 500, 用于若所述请求由搜索公司发出, 则确定所述请求为合法蜘蛛请求 ;

[0159] 第二结果确定模块 600, 用于若所述请求不是由搜索公司发出, 则确定所述请求为伪造的蜘蛛请求。

[0160] 可选的, 图 10 示出了本发明实施例提供的判断模块 400 的一种可选结构, 参照图 10, 判断模块 400 可以包括 :

[0161] 第一地址合法判断单元 410, 用于判断所述 IP 地址是否为合法 IP 地址, 所述合法 IP 地址为搜索公司对应的 IP 地址 ;

[0162] 第一合法确定单元 411, 用于若第一地址合法判断单元的判断结果为是, 则确定所述请求由搜索公司发出 ;

[0163] 第二合法确定单元 412, 用于若第一地址合法判断单元的判断结果为否, 则确定所述请求不是由搜索公司发出。

[0164] 可选的, 图 11 示出了本发明实施例提供的第一地址合法判断单元 410 的一种可选结构, 参照图 11, 第一地址合法判断单元 410 可以包括 :

[0165] 第一判断执行子单元 4101, 用于调取预设的合法 IP 地址, 判断所述 IP 地址是否与预设的合法 IP 地址对应, 若是, 则确定所述 IP 地址为合法 IP 地址, 若否, 则确定所述 IP 地址不为合法 IP 地址。

[0166] 可选的, 图 12 示出了本发明实施例提供的第一地址合法判断单元 410 的另一种可选结构, 参照图 12, 第一地址合法判断单元 410 可以包括 :

[0167] 第二判断执行子单元 4102, 用于确定所述 IP 地址对应的域名信息, 判断所述域名

信息是否与预设的搜索公司的域名信息对应,若是,则确定所述 IP 地址为合法 IP 地址,若否,则确定所述 IP 地址不为合法 IP 地址。

[0168] 可选的,图 13 示出了本发明实施例提供的判断模块 400 的另一种可选结构,参照图 13,判断模块 400 可以包括:

[0169] 第一归属地址判断单元 420,用于确定所述 IP 地址的归属地址,判断所述归属地址与设定的搜索公司的所在地址是否对应;

[0170] 第一地址确定单元 421,用于若所述第一归属地址判断单元的判断结果为是,则确定所述请求由搜索公司发出;

[0171] 第二地址确定单元 422,用于若所述第一归属地址判断单元的判断结果为否,则确定所述请求不是由搜索公司发出。

[0172] 可选的,图 14 示出了本发明实施例提供的判断模块 400 的再一种可选结构,参照图 14,判断模块 400 可以包括:

[0173] 第二地址合法判断单元 430,用于判断所述 IP 地址是否为合法 IP 地址,所述合法 IP 地址为搜索公司对应的 IP 地址;

[0174] 可选的,第二地址合法判断单元 430 的结构可与第一地址合法判断单元类似,可参照图 11 和图 12 所示。

[0175] 第一确定单元 431,用于若所述第二地址合法判断单元的判断结果为否,则确定所述请求不是由搜索公司发出;

[0176] 第二归属地址判断单元 432,用于若第二地址合法判断单元的判断结果为是,确定所述 IP 地址的归属地址,并判断所述归属地址与设定的搜索公司的所在地址是否对应;

[0177] 第二确定单元 433,用于若所述第二归属地址判断单元的判断结果为是,则确定所述请求由搜索公司发出;

[0178] 第三确定单元 434,用于若所述第二归属地址判断单元的判断结果为否,则确定所述请求不是由搜索公司发出。

[0179] 本发明实施例中,网站服务器可从接入网站的请求中有效识别出伪造的蜘蛛请求,为后续屏蔽伪造的蜘蛛请求,保证网站信息的安全提供了可能。

[0180] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0181] 专业人员还可以进一步意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0182] 结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块,或者二者的结合来实施。软件模块可以置于随机存储器 (RAM)、内存、只读存储器 (ROM)、电可编程 ROM、电可擦除可编程 ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术

领域内所公知的任意其它形式的存储介质中。

[0183] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

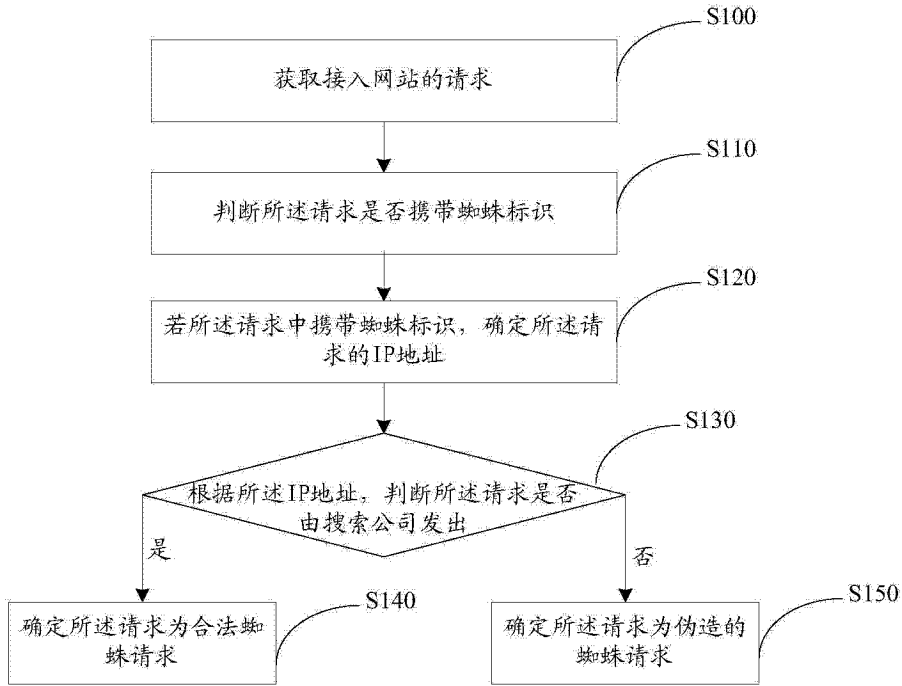


图 1

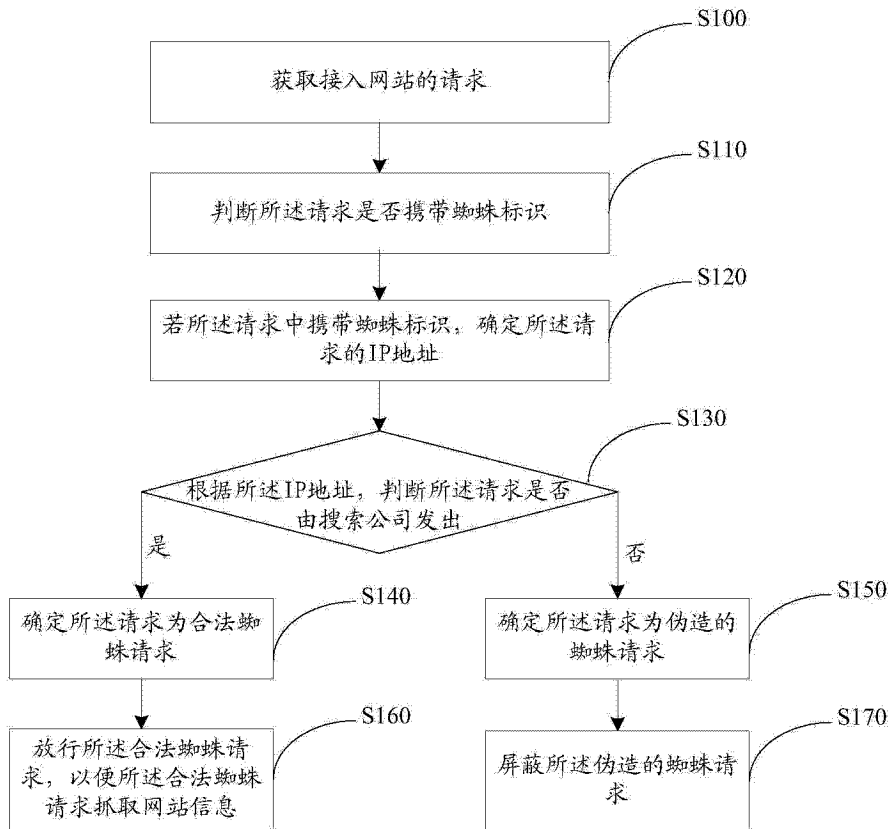


图 2

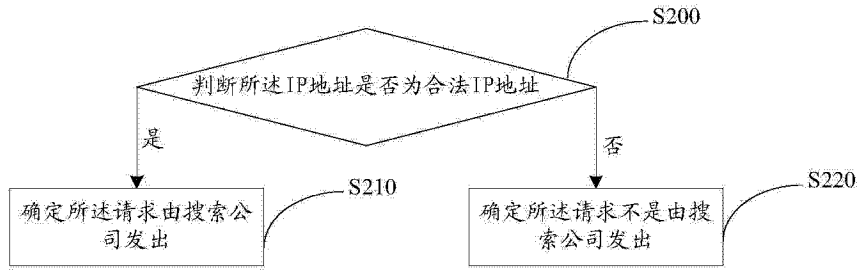


图 3

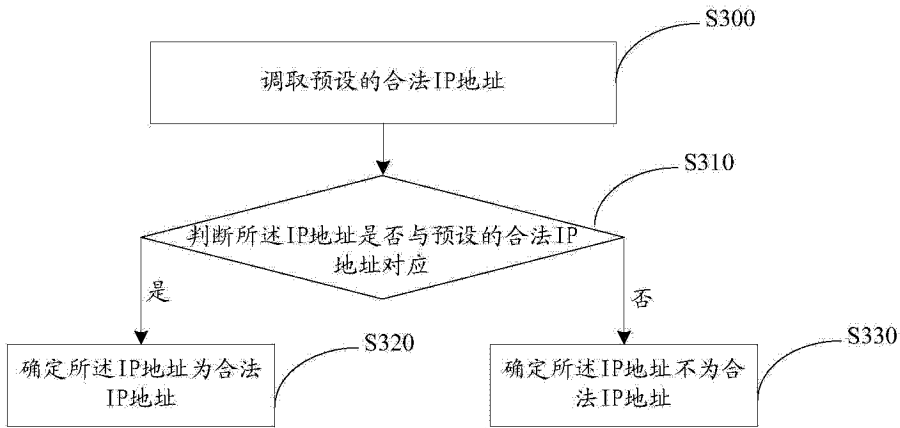


图 4

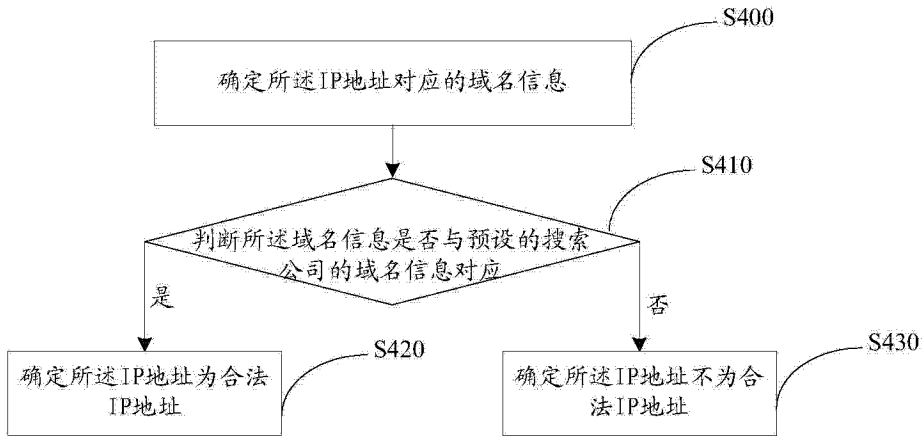


图 5

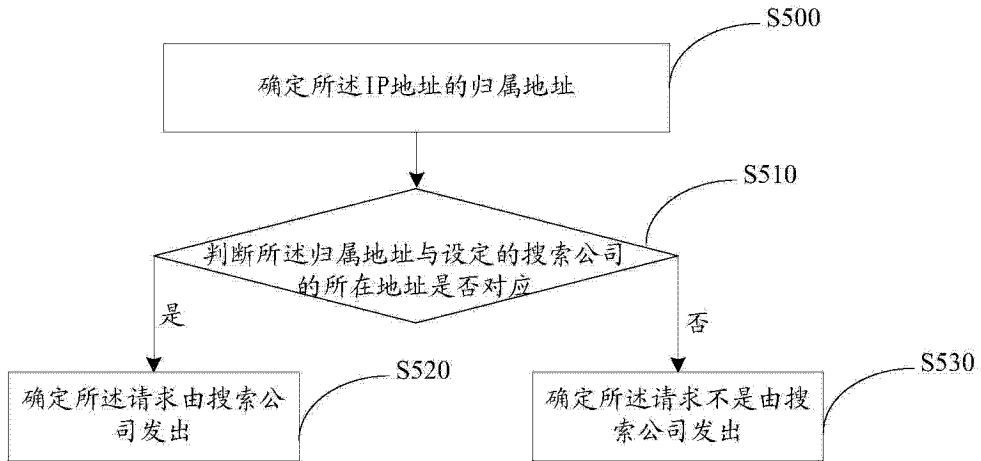


图 6

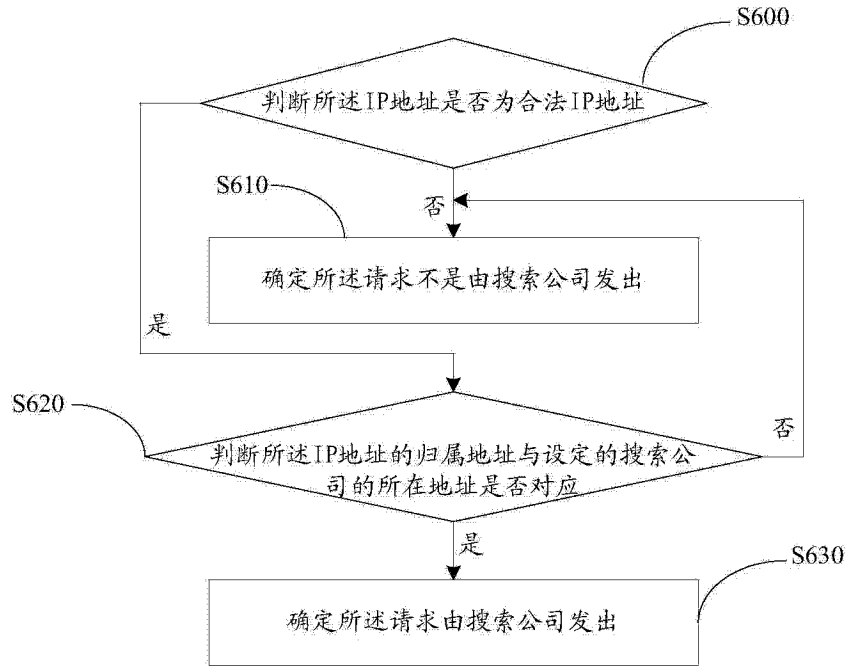


图 7

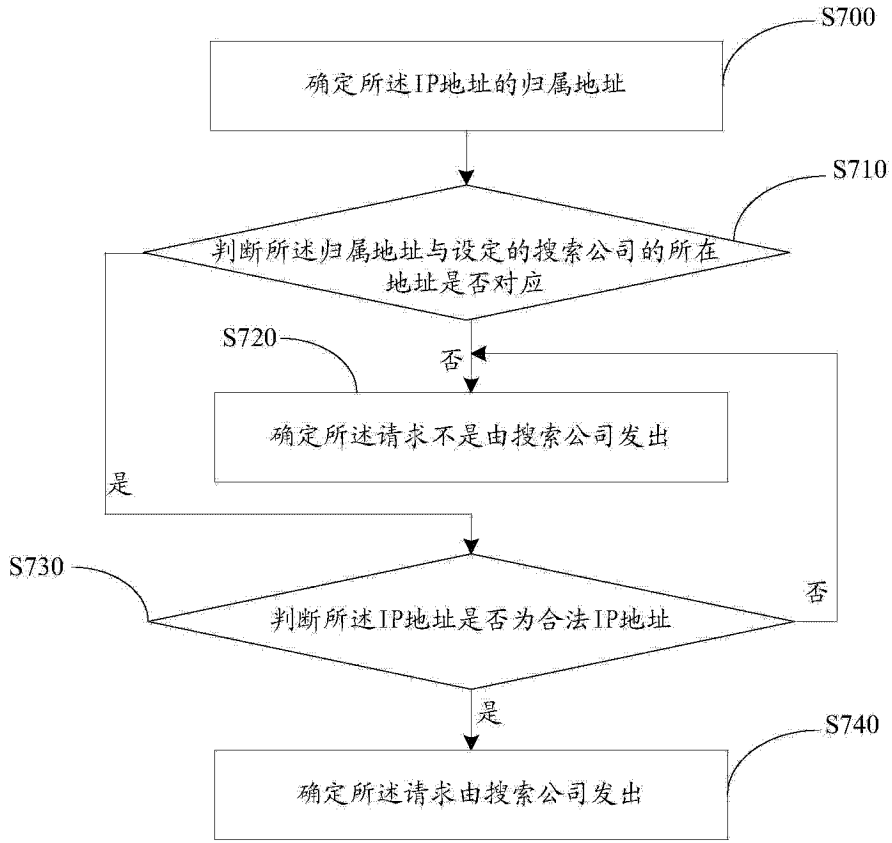


图 8

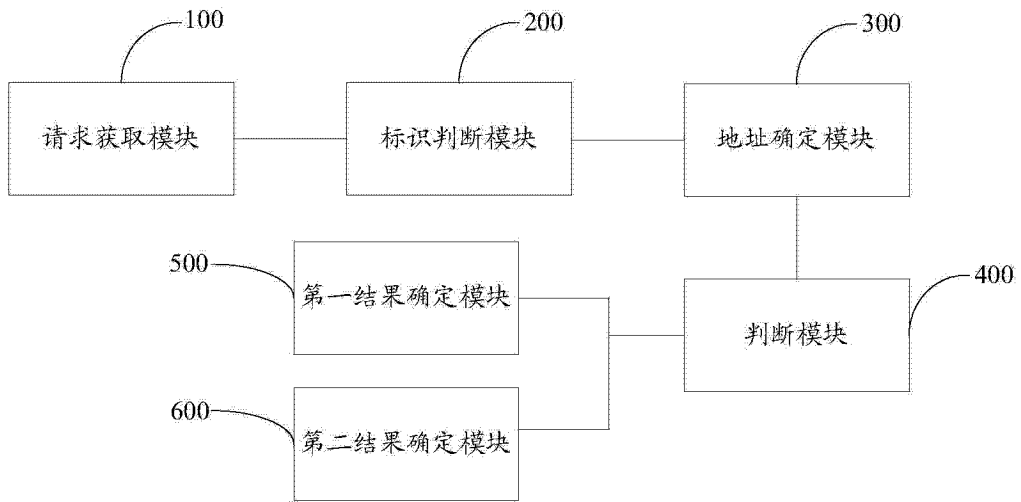


图 9

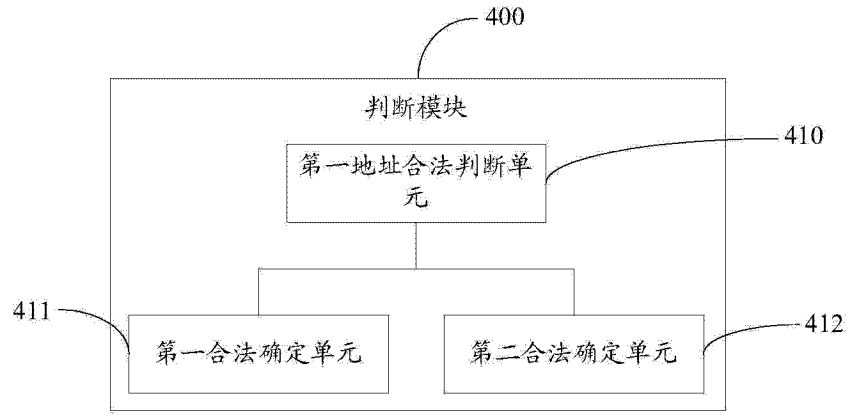


图 10

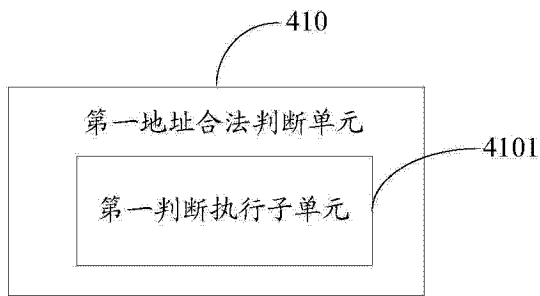


图 11

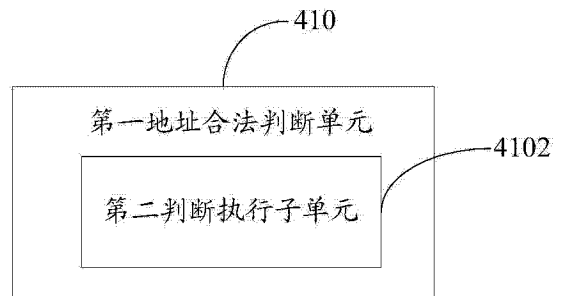


图 12

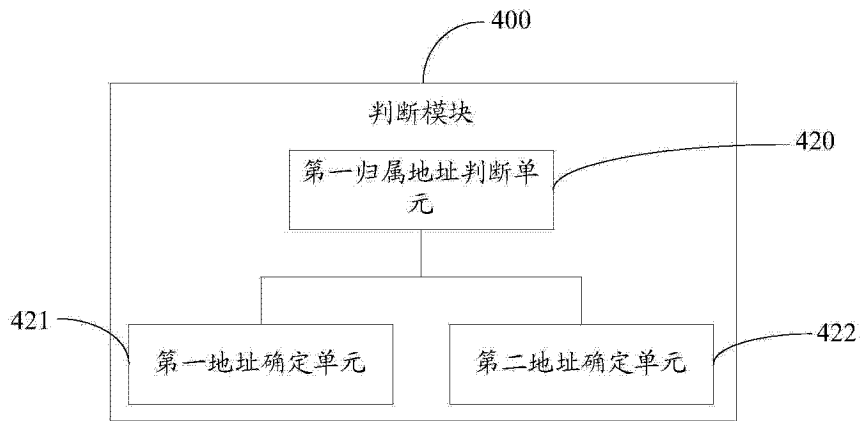


图 13

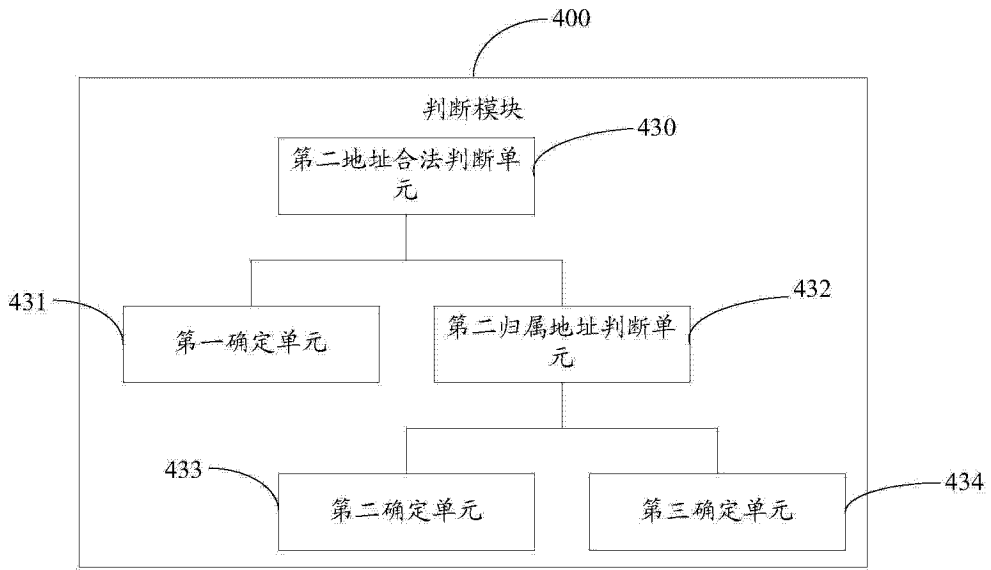


图 14