(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110430188 A
(43)申请公布日 2019.11.08

(21)申请号 201910709591.2

(22)申请日 2019.08.02

(71)申请人 武汉思普崚技术有限公司
地址 430070 湖北省武汉市东湖新技术开发区光谷大道77号金融港后台服务中心一期A4栋2层01号

(72)发明人 张晓东

(74)专利代理机构 北京弘权知识产权代理事务所(普通合伙) 11363
代理人 逯长明 许伟群

(51)Int.Cl.
H04L 29/06(2006.01)
G06F 16/953(2019.01)
G06F 16/955(2019.01)

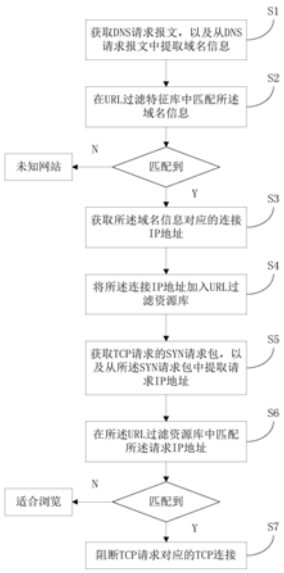
权利要求书2页 说明书9页 附图5页

(54)发明名称

一种快速URL过滤方法及装置

(57)摘要

本申请提供一种快速URL过滤方法及装置，所述方法先通过获取DNS请求报文并提取域名信息；再通过URL过滤特征库匹配域名信息，并且在匹配到所述域名信息后，获取连接IP地址，以及将连接IP地址加入URL过滤资源库中。当要建议TCP连接时，可以通过获取TCP请求的SYN请求包，提取请求IP地址，并在所述URL过滤资源库中匹配所述请求IP地址；如果在所述URL过滤资源库中匹配到所述请求IP地址，阻断所述TCP请求对应的TCP连接。所述方法可以利用DNS内容短，格式简单的特点，减小URL过滤的性能消耗。另外，本申请采用IP地址与TCP协议，能实现首包阻断，减少网络中无用流量的传输，提高网络传输的效率。



CN 110430188 A

1.一种快速URL过滤方法,其特征在于,包括:

获取DNS请求报文,以及从所述DNS请求报文中提取域名信息;

在URL过滤特征库中匹配所述域名信息;

如果在所述URL过滤特征库中匹配到所述域名信息,获取所述域名信息对应的连接IP地址;

将所述连接IP地址加入URL过滤资源库;所述URL过滤资源库包括多个域名信息,以及多个域名信息对应的IP地址;

获取TCP请求的SYN请求包,以及从所述SYN请求包中提取请求IP地址;

在所述URL过滤资源库中匹配所述请求IP地址;

如果在所述URL过滤资源库中匹配到所述请求IP地址,阻断所述TCP请求对应的TCP连接。

2.根据权利要求1所述的URL过滤方法,其特征在于,获取DNS请求报文,以及从所述DNS请求报文中提取域名信息的步骤,包括:

获取DNS请求报文以及当前应用领域下的域名模板;

根据所述域名模板,在所述DNS请求报文中匹配符合域名模板形式的文本片段;

提取所述文本片段作为所述域名信息。

3.根据权利要求1所述的URL过滤方法,其特征在于,所述URL过滤特征库包括多个预置域名信息,以及与每个所述预置域名信息对应的分类信息;在URL过滤特征库中匹配所述域名信息的步骤,包括:

逐一对比所述域名信息与预置域名信息;

如果所述域名信息与任一预置域名信息一致,提取匹配到的所述预置域名信息对应的分类信息;

如果所述域名信息与任一预置域名信息均不一致,确定当前域名信息为未知域名;

将所述未知域名发送至上位服务器。

4.根据权利要求1所述的URL过滤方法,其特征在于,如果在URL过滤特征库中匹配到所述域名信息,获取所述域名信息对应的连接IP地址的步骤,包括:

获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文;

从所述DNS响应报文中,提取所述域名信息对应的连接IP地址。

5.根据权利要求4所述的URL过滤方法,其特征在于,获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文的步骤前,还包括:

提取本地网络中的DNS缓存数据;

在所述DNS缓存数据中,匹配所述域名信息;

如果在所述DNS缓存数据匹配到所述域名信息,提取所述域名信息对应的连接IP地址;

如果在所述DNS缓存数据中未匹配到所述域名信息,获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文。

6.根据权利要求1所述的URL过滤方法,其特征在于,获取TCP请求的SYN请求包,以及从所述SYN请求包中提取请求IP地址的步骤,包括:

在接收到所述SYN请求包后,在目标栏提取请求IP地址;

暂停将所述SYN请求包转发至所述请求IP地址对应的服务器。

7.根据权利要求6所述的URL过滤方法,其特征在于,所述方法还包括:

如果在URL过滤资源库中未匹配到所述请求IP地址,将所述SYN请求包转发至所述请求IP地址对应的服务器,以建立TCP连接。

8.根据权利要求1所述的URL过滤方法,其特征在于,所述方法还包括:

获取客户端输入的访问信息;

根据所述访问信息判断访问信息类型,所述访问信息类型包括IP地址访问和非IP地址访问;

如果所述访问信息类型为非IP地址访问,从所述访问信息中提取请求IP地址;

如果所述访问信息类型为IP地址访问,将所述访问信息作为所述请求IP地址。

9.一种快速URL过滤装置,其特征在于,包括:

域名信息模块,用于获取DNS请求报文,以及从所述DNS请求报文中提取域名信息;

特征匹配模块,用于在URL过滤特征库中匹配所述域名信息;

连接IP地址模块,用于如果所述在URL过滤特征库中匹配到所述域名信息,获取所述域名信息对应的连接IP地址;

学习模块,用于将所述连接IP地址加入URL过滤资源库;所述URL过滤资源库包括多个域名信息,以及多个域名信息对应的IP地址;

请求IP地址模块,用于获取TCP请求的SYN请求包,以及从所述SYN请求包中提取请求IP地址;

资源匹配模块,用于在所述URL过滤资源库中匹配所述请求IP地址;

阻断模块,用于如果在所述URL过滤资源库中匹配到所述请求IP地址,阻断所述TCP请求对应的TCP连接。

10.根据权利要求9所述的URL过滤装置,其特征在于,所述连接IP地址模块包括:

DNS响应报文单元,用于获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文;

连接IP地址提取单元,用于从所述DNS响应报文中,提取所述域名信息对应的连接IP地址。

一种快速URL过滤方法及装置

技术领域

[0001] 本申请涉及URL过滤技术领域,尤其涉及一种快速URL过滤方法及装置。

背景技术

[0002] URL(Uniform Resource Locator,统一资源定位符)是对可以从互联网上得到的资源的位置和访问方法的一种简洁表示,是互联网上标准资源的地址。互联网上的每个文件都有一个唯一的URL,它包含的信息可以指出文件的位置以及浏览器的处理方式。URL过滤技术,应用于对互联网上的网站进行分类,并通过将所有Web流量与URL过滤库进行比较,以及通过引用已经分类的中央数据库或根据分类中包含的信息,来允许或阻止用户对Web进行访问。

[0003] 典型URL过滤方法包括:先识别HTTP流量,再通过解析http协议的报文,获取域名信息;再对域名做hash运算,查找到对应的hash位置;以及遍历hash链(根据URL过滤库生成),比较字符串是否一致,一致则匹配到对应的分类,否则未匹配,最后根据匹配结果选择性的进行阻断。

[0004] 但由于这种URL过滤方法需要对HTTP的协议数据进行识别,并且要从报文内容中解析获取域名,再进行URL过滤库的匹配,导致这种URL过滤方法的过滤性能较低。并且在实际应用中,要在获取匹配结果后才能进行阻断,导致有较多的数据被发送至用户处,造成网络资源浪费,以及可能存在的网络安全隐患。

发明内容

[0005] 本申请提供了一种快速URL过滤方法及装置,以解决URL过滤性能低的问题。

[0006] 一方面,本申请提供一种快速URL过滤方法,包括:

[0007] 获取DNS请求报文,以及从所述DNS请求报文中提取域名信息;

[0008] 在URL过滤特征库中匹配所述域名信息;

[0009] 如果在所述URL过滤特征库中匹配到所述域名信息,获取所述域名信息对应的连接IP地址;

[0010] 将所述连接IP地址加入URL过滤资源库;所述URL过滤资源库包括多个域名信息,以及多个域名信息对应的IP地址;

[0011] 获取TCP请求的SYN请求包,以及从所述SYN请求包中提取请求IP地址;

[0012] 在所述URL过滤资源库中匹配所述请求IP地址;

[0013] 如果在所述URL过滤资源库中匹配到所述请求IP地址,阻断所述TCP请求对应的TCP连接。

[0014] 可选的,获取DNS请求报文,以及从所述DNS请求报文中提取域名信息的步骤,包括:

[0015] 获取DNS请求报文以及当前应用领域下的域名模板;

[0016] 根据所述域名模板,在所述DNS请求报文中匹配符合域名模板形式的文本片段;

[0017] 提取所述文本片段作为所述域名信息。

[0018] 可选的,所述URL过滤特征库包括多个预置域名信息,以及与每个所述预置域名信息对应的分类信息;在URL过滤特征库中匹配所述域名信息的步骤,包括:

[0019] 逐一对比所述域名信息与预置域名信息;

[0020] 如果所述域名信息与任一预置域名信息一致,提取匹配到的所述预置域名信息对应的分类信息;

[0021] 如果所述域名信息与任一预置域名信息均不一致,确定当前域名信息为未知域名;

[0022] 将所述未知域名发送至上位服务器。

[0023] 可选的,如果在URL过滤特征库中匹配到所述域名信息,获取所述域名信息对应的连接IP地址的步骤,包括:

[0024] 获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文;

[0025] 从所述DNS响应报文中,提取所述域名信息对应的连接IP地址。

[0026] 可选的,获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文的步骤前,还包括:

[0027] 提取本地网络中的DNS缓存数据;

[0028] 在所述DNS缓存数据中,匹配所述域名信息;

[0029] 如果在所述DNS缓存数据匹配到所述域名信息,提取所述域名信息对应的连接IP地址;

[0030] 如果在所述DNS缓存数据中未匹配到所述域名信息,获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文。

[0031] 可选的,获取TCP请求的SYN请求包,以及从所述SYN请求包中提取请求IP地址的步骤,包括:

[0032] 在接收到所述SYN请求包后,在目标栏提取请求IP地址;

[0033] 暂停将所述SYN请求包转发至所述请求IP地址对应的服务器。

[0034] 可选的,所述方法还包括:

[0035] 如果在URL过滤资源库中未匹配到所述请求IP地址,将所述SYN请求包转发至所述请求IP地址对应的服务器,以建立TCP连接。

[0036] 可选的,所述方法还包括:

[0037] 获取客户端输入的访问信息;

[0038] 根据所述访问信息判断访问信息类型,所述访问信息类型包括IP地址访问和非IP地址访问;

[0039] 如果所述访问信息类型为非IP地址访问,从所述访问信息中提取请求IP地址;

[0040] 如果所述访问信息类型为IP地址访问,将所述访问信息作为所述请求IP地址。

[0041] 另一方面,本申请还提供一种快速URL过滤装置,包括:

[0042] 域名信息模块,用于获取DNS请求报文,以及从所述DNS请求报文中提取域名信息;

[0043] 特征匹配模块,用于在URL过滤特征库中匹配所述域名信息;

[0044] 连接IP地址模块,用于如果所述在URL过滤特征库中匹配到所述域名信息,获取所述域名信息对应的连接IP地址;

[0045] 学习模块，用于将所述连接IP地址加入URL过滤资源库；所述URL过滤资源库包括多个域名信息，以及多个域名信息对应的IP地址；

[0046] 请求IP地址模块，用于获取TCP请求的SYN请求包，以及从所述SYN请求包中提取请求IP地址；

[0047] 资源匹配模块，用于在所述URL过滤资源库中匹配所述请求IP地址；

[0048] 阻断模块，用于如果在所述URL过滤资源库中匹配到所述请求IP地址，阻断所述TCP请求对应的TCP连接。

[0049] 可选的，所述连接IP地址模块包括：

[0050] DNS响应报文单元，用于获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文；

[0051] 连接IP地址提取单元，用于从所述DNS响应报文中，提取所述域名信息对应的连接IP地址。

[0052] 由以上技术方案可知，本申请提供一种快速URL过滤方法及装置，所述方法先通过获取DNS请求报文并提取域名信息；再通过URL过滤特征库匹配域名信息，并且在匹配到所述域名信息后，获取连接IP地址，以及将连接IP地址加入URL过滤资源库中。当要建议TCP连接时，可以通过获取TCP请求的SYN请求包，提取请求IP地址，并在所述URL过滤资源库中匹配所述请求IP地址；如果在所述URL过滤资源库中匹配到所述请求IP地址，阻断所述TCP请求对应的TCP连接。所述方法可以利用DNS内容短，格式简单的特点，减小URL过滤的性能消耗。另外，本申请采用IP地址与TCP协议，能实现首包阻断，减少网络中无用流量的传输，提高网络传输的效率。

附图说明

[0053] 为了更清楚地说明本申请的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，对于本领域普通技术人员而言，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0054] 图1为本申请一种快速URL过滤方法的流程示意图；

[0055] 图2为本申请提取域名信息的流程示意图；

[0056] 图3为本申请在URL过滤特征库中匹配域名信息的流程示意图；

[0057] 图4为本申请获取连接IP地址的流程示意图；

[0058] 图5为本申请在DNS缓存数据中匹配域名信息的流程示意图；

[0059] 图6为本申请提取请求IP地址的流程示意图；

[0060] 图7为本申请从访问信息中提取请求IP地址的流程示意图；

[0061] 图8为本申请一种快速URL过滤装置的结构示意图。

具体实施方式

[0062] 下面将详细地对实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下实施例中描述的实施方式并不代表与本申请相一致的所有实施方式。仅是与权利要求书中所详述的、本申请的一些方面相一致的系统和方法的示例。

[0063] 本申请所述快速URL过滤方法及装置，可应用于上网行为管理设备，上网行为管理产品是指帮助互联网用户控制和管理互联网的使用情况。包括对访问网页过滤、网络应用控制、带宽流量管理、信息收发审计、用户行为分析等。

[0064] 参见图1，为本申请一种快速URL过滤方法的结构示意图。由图1可知，本申请提供的快速URL过滤方法，包括以下步骤：

[0065] S1：获取DNS请求报文，以及从所述DNS请求报文中提取域名信息。

[0066] 本申请提供的技术方案中，DNS(Domain Name System，域名系统)是互联网的一项服务。它作为将域名和IP地址相互映射的一个分布式数据库，能够实现更方便地访问互联网。实际应用中，客户端可以在浏览器中输入想要访问的域名信息，浏览器发送查询报文至DNS服务器，以触发DNS服务器查询对应的IP地址，再将IP地址返回至客户端，通过浏览器的后台进程，访问该IP地址。

[0067] 例如，客户端发送查询报文“query www.sohu.com”至DNS服务器，DNS服务器首先检查自身缓存，如果存在缓存记录则直接返回结果，如：“220.181.90.8”。如果记录老化或不存在，则DNS服务器向根域名服务器发送查询报文“query www.sohu.com”，根域名服务器返回顶级域.com的权威域名服务器地址。DNS服务器向.com域的权威域名服务器发送查询报文“query www.sohu.com”，得到二级域.sohu.com的权威域名服务器地址。DNS服务器向.sohu.com域的权威域名服务器发送查询报文“query www.sohu.com”，得到主机www的记录，存入自身缓存并返回给客户端IP地址“220.181.90.8”。在DNS服务中，一般使用TCP和UDP端口完成数据传输。

[0068] 同理，在本申请中，用户在客户端输入要访问的域名后，客户端浏览器会生成一组DNS请求报文，并且将该DNS请求报文发送给DNS服务器的过程中，由上网行为管理设备对请求报文进行抓取，获取到DNS请求报文。上网行为管理设备再对请求报文进行分析，提取其中的域名信息，例如，从“query www.sohu.com”中提取域名信息为“www.sohu.com”。

[0069] 在本申请的部分实施例中，如图2所示，可以按照如下方式获取域名信息，即获取DNS请求报文，以及从所述DNS请求报文中提取域名信息的步骤，包括：

[0070] S101：获取DNS请求报文以及当前应用领域下的域名模板；

[0071] S102：根据所述域名模板，在所述DNS请求报文中匹配符合域名模板形式的文本片段；

[0072] S103：提取所述文本片段作为所述域名信息。

[0073] 本实施例中，可以根据不同的应用场景定义不同的域名模板，例如，根据不同地区的政策要求，其可以访问的网站也不相同，相应的访问域名结构也存在着部分差异；因此，可以根据不同的应用领域，预定义一些域名模板，以实现更加准确的域名信息提取。

[0074] 在获取DNS请求报文后，可以根据域名模板在请求报文内容中匹配符合域名模板形式的文本片段。例如，域名模板为“www.××.com”，则可以在请求报文中逐一匹配这一格式的文本片段，则可以获取到请求报文中的“www.sohu.com”的文本片段，再将这部分文本片段从请求报文中提取出来，即可作为所述域名信息，以进行后续过滤。

[0075] S2：在URL过滤特征库中匹配所述域名信息。

[0076] 本申请提供的技术方案中，所述URL过滤特征库为内置在上网行为管理设备中的数据库，所述URL过滤特征库可以是随着设备出厂而内置其中的数据库，也可以是在上位服

务器中下载的数据库。URL过滤特征库可作为上网行为管理设备的网站分类库，其中存储有多个网站的域名信息和分类信息，分类信息可以用于判断对应域名信息是否为不合适客户端浏览的网站。

[0077] URL过滤特征库可以仅用来记载不适合客户端浏览的网站，而对于适合浏览的网站，可以不进行存储，从而减少URL过滤特征库所占用的存储空间。进一步地，如图3所示，所述URL过滤特征库包括多个预置域名信息，以及与每个所述预置域名信息对应的分类信息；在URL过滤特征库中匹配所述域名信息的步骤，还包括：

[0078] S201：逐一对比所述域名信息与预置域名信息；

[0079] S202：如果所述域名信息与任一预置域名信息一致，提取匹配到的所述预置域名信息对应的分类信息；

[0080] S203：如果所述域名信息与任一预置域名信息均不一致，确定当前域名信息为未知域名；

[0081] S204：将所述未知域名发送至上位服务器。

[0082] 本实施例中，URL过滤特征库中记载的预置域名信息可以为上位服务器已经确定为不适合浏览的多个网站域名。并且每一个预置域名信息都对应于该网站的分类信息，例如，病毒网站、钓鱼网站、非法网站等。实际应用时，可以通过逐一比对从DNS请求报文中提取的域名信息和预置域名信息，确定其是否一致。如果DNS请求报文中的域名信息和预置域名信息一致，确定当前客户端想要访问的页面为不适合浏览的页面，因此可以提取对应的分类信息，以便向客户端展示不适合访问的原因。

[0083] 在本实施例中，如果提取的域名信息与任一预置域名信息都不一致，即上述URL过滤特征库中没有存储当前DNS请求中的域名信息，这可能是由于两种原因，一种为URL过滤特征库中仅存储有不适合客户端浏览的域名，而用户输入的为适合浏览的域名；另一种为用户输入的域名是一个全新的域名信息，即是一个未知网站。其中，对于适合浏览的网站域名，可以直接放行，也可以通过上位服务器，再进行验证。

[0084] 由于不适合客户端浏览的网站是少数网站，因此在实际应用中，对于适合浏览的域名，也可以在URL过滤特征库中增加合法网站的分类信息，适合浏览的网站也能够匹配到预置域名信息，而未匹配到预置域名信息的，则均为未知网站，对于未知网站需要通过上位服务器进一步进行验证。

[0085] S3：如果在所述URL过滤特征库中匹配到所述域名信息，获取所述域名信息对应的连接IP地址。

[0086] 本申请中，如果在URL过滤特征库中匹配到所述域名信息，说明客户端想要访问的域名信息可能是不适合浏览的网站，因此可以通过获取域名信息对应的连接IP地址来更新该上网行为管理设备的URL过滤资源库，以便后续对该网站IP进行阻断，阻止客户端访问该网站。

[0087] 在本申请的部分实施例中，如图4所示，如果在URL过滤特征库中匹配到所述域名信息，获取所述域名信息对应的连接IP地址的步骤，还包括：

[0088] S301：获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文；

[0089] S302：从所述DNS响应报文中，提取所述域名信息对应的连接IP地址。

[0090] 由于在实际应用中，DNS服务器可以根据DNS请求报文反馈DNS响应报文，即根据域

名信息反馈IP地址。例如,针对请求报文“query www.sohu.com”,DNS服务器反馈的DNS响应报文中,包含内容“Address:220.181.90.8”。因此,可以从所述DNS响应报文中,提取所述域名信息对应的连接IP地址,即提取“220.181.90.8”。

[0091] 进一步地,如图5所示,获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文的步骤前,还包括:

[0092] S3011:提取本地网络中的DNS缓存数据;

[0093] S3012:在所述DNS缓存数据中,匹配所述域名信息;

[0094] S3013:如果在所述DNS缓存数据匹配到所述域名信息,提取所述域名信息对应的连接IP地址;

[0095] S3014:如果在所述DNS缓存数据中未匹配到所述域名信息,获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文。

[0096] 由于在实际应用中,DNS服务器返回的IP地址信息可以在本地网络中暂时保存一段时间,即在网络中缓存有DNS数据。因此,可以在向DNS服务器查询IP地址前,先在本地网络的DNS缓存数据中进行匹配,如果在所述DNS缓存数据匹配到所述域名信息,可以直接获取对应的IP地址作为连接IP地址,从而无需向DNS服务器进行查询,提高域名查询的效率。如果在所述DNS缓存数据中未匹配到所述域名信息,再执行获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文的步骤,以获取到DNS服务器返回的IP地址作为连接IP地址。

[0097] S4:将所述连接IP地址加入URL过滤资源库。

[0098] 本申请提供的技术方案中,所述URL过滤资源库包括多个域名信息,以及多个域名信息对应的IP地址。URL过滤资源库是根据URL过滤特征库建立的包含域名信息和IP地址的数据表,用于记载所有已经认证为不适合客户端浏览的网站信息。所述URL过滤数据库可以仅内置在上网行为管理设备中,其内容的更新是通过上述URL过滤特征库为基础的。

[0099] 可见,对于已经应用的上网行为管理设备,其中的URL过滤资源库可以仅仅存储有客户端已经试图访问过的不适合浏览网站,而对于未曾浏览过的其他网站,可以不进行保存,从而大大节省上网行为管理设备的存储空间,并且由于存储的网站数量较少,其匹配速度也得到提升。另外,由于URL过滤资源库可以依据URL过滤特征库,在客户端访问不适合网站的域名时,进行学习更新,可以使得上网行为管理设备在较小的存储数据的前提下,拥有更全面的过滤性能。

[0100] S5:获取TCP请求的SYN请求包,以及从所述SYN请求包中提取请求IP地址。

[0101] 本申请提供的技术方案中,当DNS服务器向客户端返回请求响应后,客户端浏览器会通过后台程序进程,自动访问连接IP地址对应的服务器。而想要访问IP地址对应的服务器,则需要通过TCP协议建立数据连接,即发送TCP请求给IP地址对应的服务器。TCP请求中,包括SYN(Synchronize Sequence Numbers,同步序列编号)请求包,即TCP请求的首个数据包。

[0102] 由于在实际应用中,客户端可能不仅仅采用DNS的方式访问网站,但都需要指定服务器的IP地址才能实现访问。因此,在本申请中,如果用户直接输入IP地址进行访问,可以直接获取该IP地址。并且,实际应用中,也可以通过其他协议,如HTTP等,仅需要直接从数据包中提取对应的IP地址即可。

[0103] 需要说明的是,本申请提供的技术方案中,所述连接IP地址和请求IP地址可以相

同，也可以不同。其中，通过DNS方式访问，并通过浏览器后台进程完成指定IP地址服务器访问时，所述连接IP地址和请求IP地址是相同的。而通过其他方式访问，或者访问与上述URL过滤资源库学习更新过程不在同一个时间时，其连接IP地址和请求IP地址是不相同的。

[0104] 在本申请的部分实施例中，如图6所示，获取TCP请求的SYN请求包，以及从所述SYN请求包中提取请求IP地址的步骤，还包括：

[0105] S501：在接收到所述SYN请求包后，在目标栏提取请求IP地址；

[0106] S502：暂停将所述SYN请求包转发至所述请求IP地址对应的服务器。

[0107] 即在实际应用时，上网行为管理设备在接收到SYN请求包后，可以先对请求IP地址进行提取，并将提取的请求IP地址进行进一步的判断，确定请求IP地址对应的网站是否为不是后客户端浏览的网站。与此同时，上网行为管理设备可以暂时停止将SYN请求包转发至所述请求IP地址对应的服务器，而等待请求IP地址的判断结果，从而选择阻断还是允许访问该网站。

[0108] 进一步地，如图7所示，所述方法还包括：

[0109] S511：获取客户端输入的访问信息；

[0110] S512：根据所述访问信息判断访问信息类型，所述访问信息类型包括IP地址访问和非IP地址访问；

[0111] S513：如果所述访问信息类型为非IP地址访问，从所述访问信息中提取请求IP地址；

[0112] S514：如果所述访问信息类型为IP地址访问，将所述访问信息作为所述请求IP地址。

[0113] 实际应用中，可以通过获取客户端中输入的访问信息，来进一步判断访问信息的类型，从而根据不同的访问信息类型确定请求IP地址。访问信息类型可以包括IP地址访问和非IP地址访问，客户端输入的访问信息是否直接是IP地址，如果是IP地址访问的方式，可以直接将访问信息作为请求IP地址。如果所述访问信息类型为非IP地址访问，即客户端没有直接输入IP地址的形式进行访问，则需要对访问信息进行进一步分析处理，以获得请求IP地址。例如，用户仍然以输入域名的方式进行网站的访问，则需要从DNS服务器反馈的请求响应中，获取该域名信息对应的IP地址。

[0114] 需要说明的是，在实际应用中，客户端对于部分网站的访问，可能是通过页面跳转的方式实现。而这种跳转的方式本质上，也是浏览器后台程序根据搜索引擎的检索结果，或者超链接中的域名信息所对应的IP地址发出的访问请求，因此这种方式与IP地址访问方式相同。

[0115] S6：在所述URL过滤资源库中匹配所述请求IP地址。

[0116] 在获取请求IP地址后，本申请可以根据获取的请求IP地址，在URL过滤资源库中进行匹配，具体的匹配方式可以与上述方式相同，都是逐一与数据库中每一个表项进行对比，确定URL过滤资源库中是否存在所述请求IP地址。

[0117] 在实际应用中，有些网站的域名不只一个，但其IP地址一般是不变的，因此在本申请提供的技术方案中，无论客户端输入的网站为何种形式，都可以通过IP地址进行快速过滤，从而验证其合法性。应用本申请URL过滤方法的上网行为管理设备，可以在客户端以任何方式访问服务器时，都能够对URL地址进行过滤。

[0118] S7:如果在所述URL过滤资源库中匹配到所述请求IP地址,阻断所述TCP请求对应的TCP连接。

[0119] 本申请提供的技术方案中,如果在URL过滤资源库中匹配到请求IP地址,说明当前SYN包中的请求IP地址对应的网站为不适合在客户端上浏览的网站,可以通过阻断该TCP请求,而使客户端不能访问该网站。由于本申请中请求IP地址是在SYN包中提取的,作为TCP连接的首包数据,因此,可以实现首包阻断,从而减少网络中无用流量的传输,提高网络传输有效率。

[0120] 实际应用中,当在所述URL过滤资源库中匹配到所述请求IP地址时,可以在阻断所述TCP请求对应的TCP连接后,向客户端推送一个页面,以显示当前网站不适合在客户端上进行浏览。进一步地,根据不同的分类信息,向客户端推送的页面也可以不同,例如,分类信息为非法网站时,可以推送的页面包含“根据××法律法规,您不能浏览该页面的内容”;分类信息为病毒网站时,可以在推送的页面中包含“该页面可能存在病毒,因此无法显示该页面内容”。

[0121] 进一步地,如果在URL过滤资源库中未匹配到所述请求IP地址,将所述SYN请求包转发至所述请求IP地址对应的服务器,以建立TCP连接。即针对URL过滤资源库中没有记载的IP地址信息,可以认定其为适合在客户端上访问的网站信息。对于适合在客户端上访问的信息,可以在判断合法后,对SYN请求包进行放行,并不再对该网站对应的数据进行监控,以使用户能够正常浏览该网页。

[0122] 可见,本申请提供的URL过滤方法,可以通过DNS获取请求的域名及域名对应的IP地址,减少HTTP协议识别及HTTP协议解析获取域名消耗的性能;利用DNS请求内容较短,且格式简单的优势,相对于HTTP报文携带的内容较大,且需要字符比较查找,较大的消耗的URL过滤的性能,提高过滤效率。同时,本申请中后续的阻断只需要比对IP地址是否在URL过滤资源库中,IP地址转化为数字比较,性能较高。并且,因阻断匹配采用的是IP地址与TCP协议,所以能实现首包阻断,减少网络中无用流量的传输,提高网络传输有效率。此外,因将学习到的IP地址放在URL过滤资源库中,网络中存在DNS缓存或者直接地址访问也可以进行匹配,极大的提高阻断率。

[0123] 基于上述URL过滤方法,本申请还提供一种快速URL过滤装置,所述快速URL过滤装置分别与客户端和上位服务器之间建立网络连接,且所述客户端通过所述快速URL过滤装置连接至互联网。如图8所示,所述快速URL过滤装置进一步包括:域名信息模块1、特征匹配模块2、连接IP地址模块3、学习模块4、请求IP地址模块5、资源匹配模块6以及阻断模块7,其中:

[0124] 域名信息模块1,用于获取DNS请求报文,以及从所述DNS请求报文中提取域名信息;

[0125] 特征匹配模块2,用于在URL过滤特征库中匹配所述域名信息;

[0126] 连接IP地址模块3,用于如果所述在URL过滤特征库中匹配到所述域名信息,获取所述域名信息对应的连接IP地址;

[0127] 学习模块4,用于将所述连接IP地址加入URL过滤资源库;所述URL过滤资源库包括多个域名信息,以及多个域名信息对应的IP地址;

[0128] 请求IP地址模块5,用于获取TCP请求的SYN请求包,以及从所述SYN请求包中提取

请求IP地址；

[0129] 资源匹配模块6，用于在所述URL过滤资源库中匹配所述请求IP地址；

[0130] 阻断模块7，用于如果在所述URL过滤资源库中匹配到所述请求IP地址，阻断所述TCP请求对应的TCP连接。

[0131] 在本申请的部分实施例中，所述连接IP地址模块3还包括：DNS响应报文单元和连接IP地址提取单元，其中：

[0132] DNS响应报文单元，用于获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文；

[0133] 连接IP地址提取单元，用于从所述DNS响应报文中，提取所述域名信息对应的连接IP地址。

[0134] 由以上技术方案可知，本申请提供一种快速URL过滤方法及装置，所述方法先通过获取DNS请求报文并提取域名信息；再通过URL过滤特征库匹配域名信息，并且在匹配到所述域名信息后，获取连接IP地址，以及将连接IP地址加入URL过滤资源库中。当要建议TCP连接时，可以通过获取TCP请求的SYN请求包，提取请求IP地址，并在所述URL过滤资源库中匹配所述请求IP地址；如果在所述URL过滤资源库中匹配到所述请求IP地址，阻断所述TCP请求对应的TCP连接。所述方法可以利用DNS内容短，格式简单的特点，减小URL过滤的性能消耗。另外，本申请采用IP地址与TCP协议，能实现首包阻断，减少网络中无用流量的传输，提高网络传输的效率。

[0135] 本申请提供的实施例之间的相似部分相互参见即可，以上提供的具体实施方式只是本申请总的构思下的几个示例，并不构成本申请保护范围的限定。对于本领域的技术人员而言，在不付出创造性劳动的前提下依据本申请方案所扩展出的任何其他实施方式都属于本申请的保护范围。

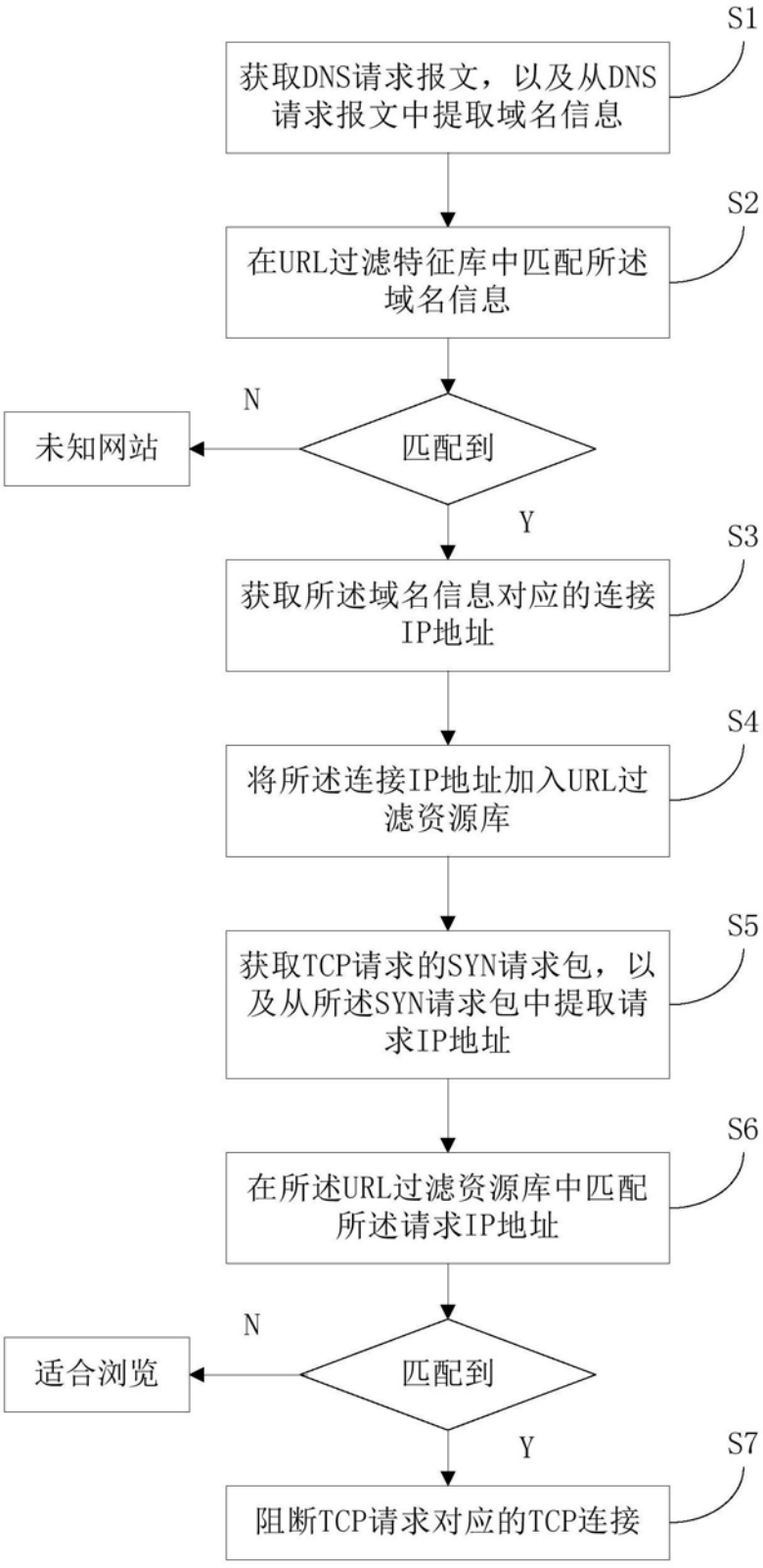

S1
获取DNS请求报文，以及从DNS请求报文中提取域名信息
S2
在URL过滤特征库中匹配所述域名信息
N
匹配到
未知网站
Y
S3
获取所述域名信息对应的连接IP地址
S4
将所述连接IP地址加入URL过滤资源库
S5
获取TCP请求的SYN请求包，以及从所述SYN请求包中提取请求IP地址
S6
在所述URL过滤资源库中匹配所述请求IP地址
N
匹配到
适合浏览
Y
S7
阻断TCP请求对应的TCP连接

图1

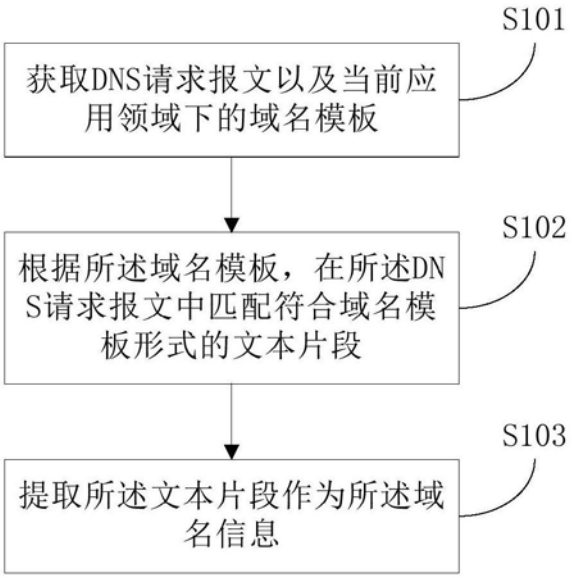

S101
获取DNS请求报文以及当前应用领域下的域名模板
S102
根据所述域名模板，在所述DNS请求报文中匹配符合域名模板形式的文本片段
S103
提取所述文本片段作为所述域名信息

图2

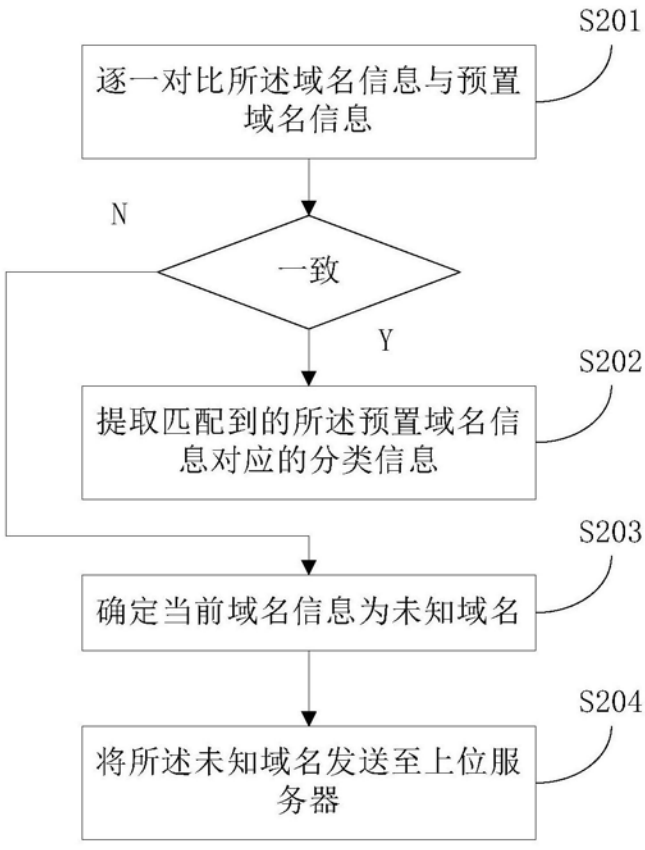

S201
逐一对比所述域名信息与预置域名信息
N
一致
Y
S202
提取匹配到的所述预置域名信息对应的分类信息
S203
确定当前域名信息为未知域名
S204
将所述未知域名发送至上位服务器

图3

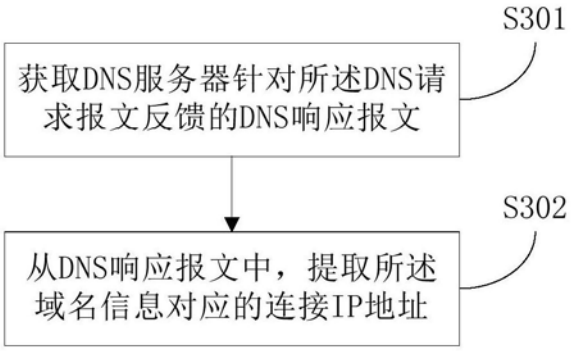

S301
获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文
S302
从DNS响应报文中，提取所述域名信息对应的连接IP地址

图4

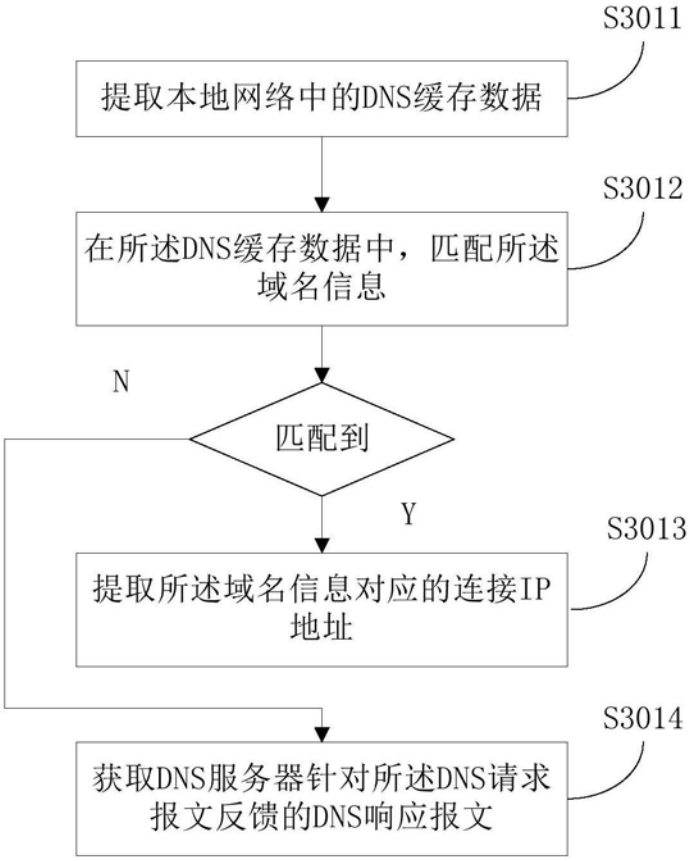

S3011
提取本地网络中的DNS缓存数据
S3012
在所述DNS缓存数据中，匹配所述域名信息
N
匹配到
Y
S3013
提取所述域名信息对应的连接IP地址
S3014
获取DNS服务器针对所述DNS请求报文反馈的DNS响应报文

图5

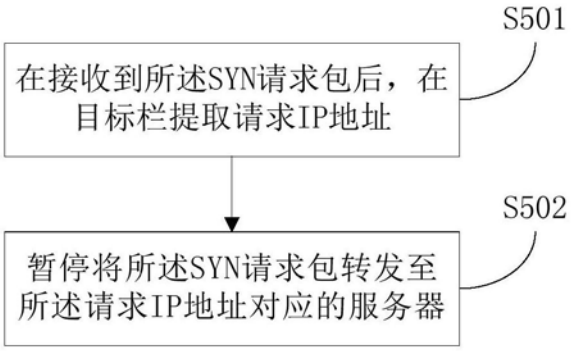

S501
在接收到所述SYN请求包后，在目标栏提取请求IP地址
S502
暂停将所述SYN请求包转发至所述请求IP地址对应的服务器

图6

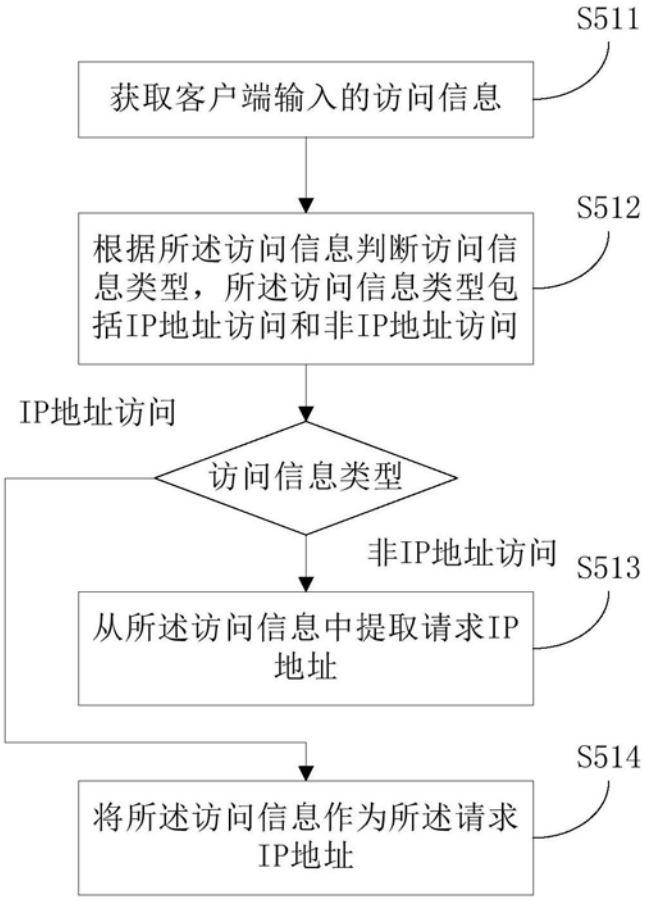

S511
获取客户端输入的访问信息
S512
根据所述访问信息判断访问信息类型，所述访问信息类型包括IP地址访问和非IP地址访问
IP地址访问
访问信息类型
非IP地址访问
S513
从所述访问信息中提取请求IP地址
S514
将所述访问信息作为所述请求IP地址

图7

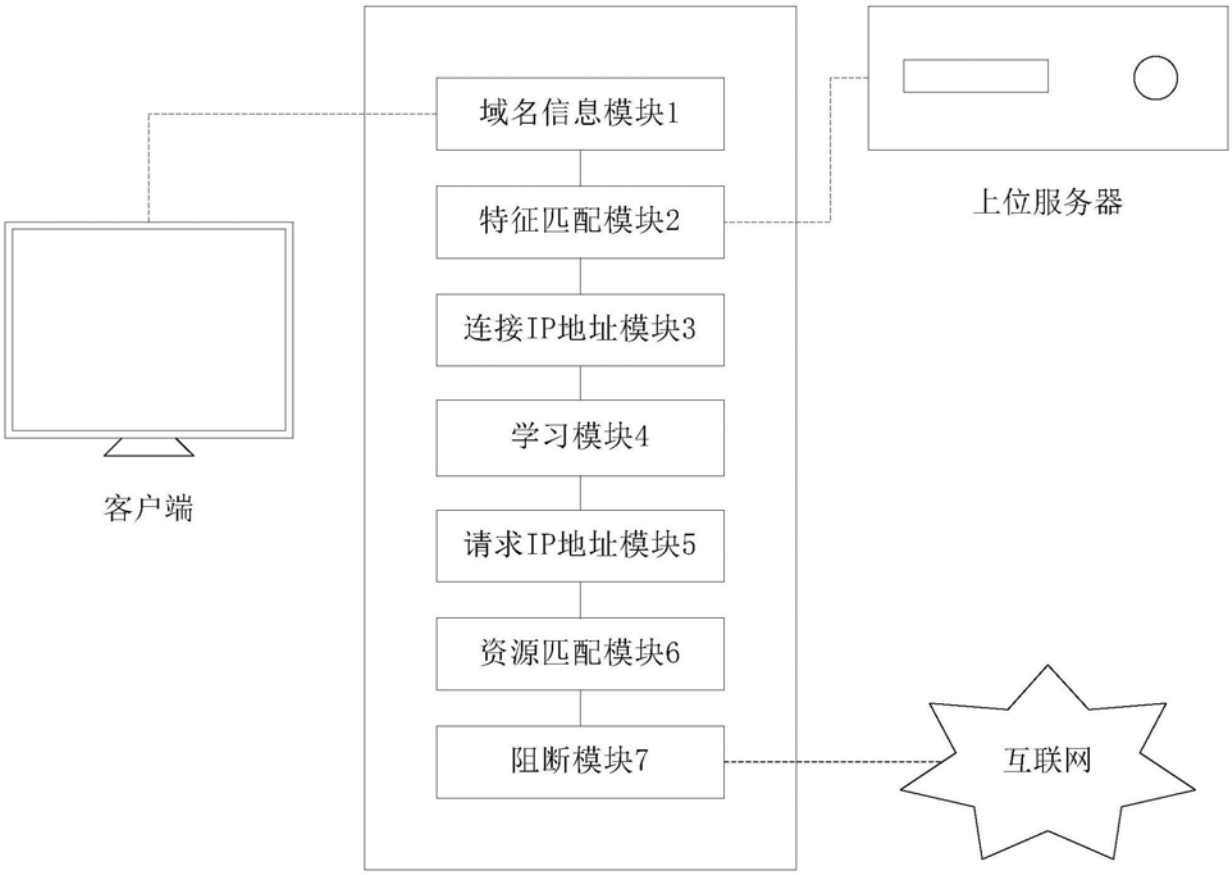

域名信息模块1
特征匹配模块2
连接IP地址模块3
学习模块4
请求IP地址模块5
资源匹配模块6
阻断模块7
客户端
上位服务器
互联网

图8