(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
2014 年 3 月 6 日 (06.03.2014)

WIPO | PCT

(10) 国际公布号
WO 2014/032619 A1

(51) 国际专利分类号:
H04L 29/06 (2006.01) *G06F 17/30* (2006.01)

(21) 国际申请号: PCT/CN2013/082729

(22) 国际申请日: 2013 年 8 月 30 日 (30.08.2013)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201210320707.1 2012 年 8 月 31 日 (31.08.2012) CN

(71) 申请人: **北京奇虎科技有限公司 (BEIJING QIHOO TECHNOLOGY COMPANY LIMITED)** [CN/CN]; 中国北京市西城区新街口外大街 28 号 D 座 112 室（德胜园区）, Beijing 100088 (CN)。 **奇智软件（北京）有限公司 (QIZHI SOFTWARE (BEIJING) COMPANY LIMITED)** [CN/CN]; 中国北京市朝阳区酒仙桥路 14 号兆维大厦 4 层东侧单元, Beijing 100016 (CN)。

(72) 发明人: **肖鹏 (XIAO, Peng)**; 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。 **李晓波 (LI, Xiaobo)**; 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。 **宋申雷 (SONG, Shenlei)**; 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。 **刘起 (LIU, Qi)**; 中国北京市朝阳区酒仙桥路 6 号院 2 号楼, Beijing 100015 (CN)。

(74) 代理人: **北京智汇东方知识产权代理事务所（普通合伙） (WISEAST INTELLECTUAL PROPERTY LAW FIRM)**; 中国北京市海淀区成府路 28 号优盛大厦 D-1111 室, Beijing 100083 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA,

[见续页]

(54) **Title**: WEB ADDRESS ACCESS METHOD AND SYSTEM

(54) **发明名称** : 网址访问方法及系统

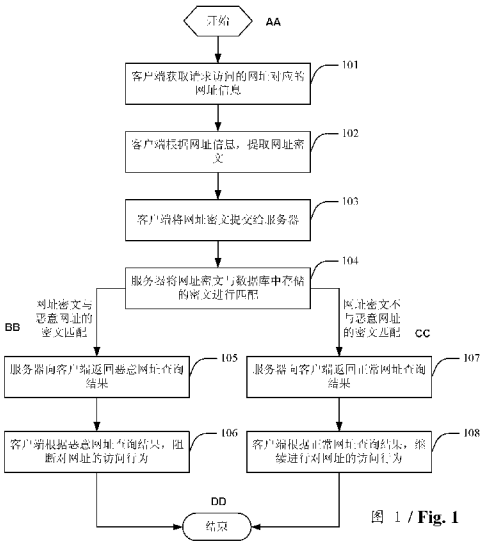

图 1 / **Fig. 1**

101 A CLIENT ACQUIRING WEB ADDRESS INFORMATION CORRESPONDING TO A WEB ADDRESS REQUESTED TO BE ACCESSED
102 THE CLIENT EXTRACTING A WEB ADDRESS CIPHERTEXT ACCORDING TO THE WEB ADDRESS INFORMATION
103 THE CLIENT SUBMITTING THE WEB ADDRESS CIPHERTEXT TO A SERVER
104 THE SERVER MATCHING THE WEB ADDRESS CIPHERTEXT WITH A CIPHERTEXT STORED IN A DATABASE
105 THE SERVER RETURNING TO THE CLIENT THE QUERY RESULT OF A MALICIOUS WEB ADDRESS
106 ACCORDING TO THE QUERY RESULT OF A MALICIOUS WEB ADDRESS, THE CLIENT BLOCKING THE BEHAVIOUR OF ACCESSING THE WEB ADDRESS
107 THE SERVER RETURNING TO THE CLIENT THE QUERY RESULT OF A NORMAL WEB ADDRESS
108 ACCORDING TO THE QUERY RESULT OF A NORMAL WEB ADDRESS, THE CLIENT CONTINUING THE BEHAVIOUR OF ACCESSING THE WEB ADDRESS
AA START
BB THE WEB ADDRESS CIPHERTEXT MATCHES WITH A CIPHERTEXT OF A MALICIOUS WEB ADDRESS
CC THE WEB ADDRESS CIPHERTEXT DOES NOT MATCH WITH A CIPHERTEXT OF A MALICIOUS WEB ADDRESS
DD END

(57) **Abstract**: Disclosed are a web address access method and system. The method comprises: a client acquiring web address information corresponding to a web address requested to be accessed; the client extracting a web address ciphertext according to the web address information; the client submitting the web address ciphertext to a server; the server matching the web address ciphertext with a ciphertext stored in a database; if the web address ciphertext matches with a ciphertext marked as a malicious web address in the database, returning to the client the query result of a malicious web address; according to the query result of a malicious web address, the client blocking the behaviour of accessing the web address; if the web address ciphertext does not match with the ciphertext marked as a malicious web address in the database, returning to the client the query result of a normal web address; and according to the query result of a normal web address, the client continuing the behaviour of accessing the web address. According to the web address access method and system provided in the present embodiments, a malicious website can be quickly and effectively blocked.

(57) **摘要**:

[见续页]

WO 2014/032619 A1

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, 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))。

本发明公开了一种网址访问方法及系统，其中方法包括：客户端获取请求访问的网址对应的网址信息；所述客户端根据所述网址信息，提取网址密文；所述客户端将所述网址密文提交给服务器；所述服务器将网址密文与数据库中存储的密文进行匹配；若网址密文与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；所述客户端根据所述恶意网址查询结果，阻断对所述网址的访问行为；若网址密文不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果；所述客户端根据所述正常网址查询结果，继续进行对所述网址的访问行为。根据本实施例提供的网址访问方法及系统，从而能够快速有效地拦截恶意网站。

网址访问方法及系统

技术领域

本发明涉及网络技术领域，具体涉及一种网址访问方法及系统。

背景技术

在所有木马、恶意软件传播的途径中，有70%以上的安全威胁来源于网络浏览，主要的方式包括网页挂马、网络钓鱼、恶意下载等，各式各样的恶意网站已经严重威胁到用户个人信息安全、国家信息安全和互联网健康发展，因此针对恶意网站的实时拦截是信息安全厂商必备的核心功能。

现有技术的恶意网址库都存在于本地客户端，而互联网上的恶意网站在不断更新变化，恶意网址库的生成也需要不断更新，现有技术需要依靠本地客户端不断升级新的本地恶意网址库才能保证恶意网址的拦截效果，然而本地恶意网址的升级时间周期过长，往往存在滞后性，无法及时更新互联网上层出不穷的各类恶意网址，导致无法快速有效地拦截恶意网站。

发明内容

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者至少部分地解决上述问题的网址访问方法和相应的网址访问系统。

根据本发明的一个方面，提供了一种网址访问方法，包括：

客户端获取请求访问的网址对应的网址信息；

所述客户端根据所述网址信息，提取网址密文；

所述客户端将所述网址密文提交给服务器；

所述服务器将网址密文与数据库中存储的密文进行匹配；

若网址密文与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；所述客户端根据所述恶意网址查询结果，阻断对所述网址的访问行为；

若网址密文不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果；所述客户端根据所述正常网址查询结果，继续进行对所述网址的访问行为。

可选地，所述网址信息具体为至少一个第一 URL。

可选地，所述数据库中标记为恶意网址的密文包括以下信息的一种或多种：恶意 URL 的特征值、恶意 URL 的主机名的特征值和恶意 URL 的子域名的特征值。

可选地，所述至少一个第一 URL 包括：所述请求访问的网址对应的网页的 URL 或所述请求访问的网址对应的网页内容中链接的 URL 或下载文件的 URL 或以上信息的任一组合。

可选地，所述客户端获取请求访问的网址对应的网址信息包括：

通过指定响应事件接口，获取所述客户端请求访问的网址对应的网页的URL。

可选地，所述客户端获取请求访问的网址对应的网址信息包括：

获得客户端的浏览器内部的页面对象；

通过调用所述页面对象的方法，获取所述客户端请求访问的网址对应的网页内容中链接的URL。

可选地，所述客户端获取请求访问的网址对应的网址信息包括：

监控所述客户端的浏览器内部与下载有关的函数；

当所述浏览器发生下载行为时，获取所述下载文件的URL。

可选地，在所述客户端根据所述网址信息，提取网址密文之前还包括：所述客户端对所述至少一个第一URL进行规范化处理。

可选地，所述客户端对所述至少一个第一URL进行规范化处理包括：

将所述第一URL中的字母大小写进行统一；

去除所述第一URL中重复多余的路径符和参数。

可选地，所述客户端根据所述网址信息，提取网址密文包括：

获取所述第一URL的主机名和所述第一URL的第一域名段；

分别计算所述第一URL的特征值、所述第一URL的主机名的特征值和所述第一URL的第一域名段的特征值；

所述第一URL的特征值、所述第一URL的主机名的特征值和所述第一URL的第一域名段的特征值即为所述网址密文。

可选地，若所述第一URL的主机名从右至左的第一级根域名为国际顶级域名，则所述第一URL的第一域名段为所述第一URL的主机名的第一级子域名；若所述第一URL的主机名从右至左的第一级根域名为国家地区顶级域名，第一级子域名包括国际顶级域名，则所述第一URL的第一域名段为所述第一URL的主机名的第二级子域名；若所述第一URL的主机名使用了动态域名，则所述第一URL的第一域名段为第一URL的主机名从动态域名开始，向右提取的下一级子域名。

可选地，所述若网址密文与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果具体为：若所述至少一个第一URL中任一第一URL的特征值、所述至少一个第一URL中任一第一URL的主机名的特征值和所述至少一个第一URL中任一第一URL的第一域名段的特征值中的任一个与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果。

可选地，所述服务器将网址密文与数据库中存储的密文进行匹配包括：

将所述至少一个第一URL中任一第一URL的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一URL中任一第一URL的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一URL中任一第一URL的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一URL中任一第一URL的主机名的特征值与数

据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果。

可选地，还包括构建所述数据库的步骤；

所述构建数据库的步骤包括：

获取已知为恶意网址且第一域名段相同的至少一个第二 URL；

获取所述至少一个第二 URL 中包含子域名级数最高的第三 URL，从右至左逐级追溯第三 URL 所包含的子域名，提取至少一级子域名；

若所述第二 URL 的第一域名段属于预设的可信名单，将所述每个第二 URL 的特征值和每个第二 URL 的主机名的特征值标记为恶意网址的密文，存储在数据库中；

若所述第二 URL 的第一域名段属于预设的不可信名单，获取至少一个第二 URL 中包含子域名级数最低的第四 URL，将所述每个第二 URL 的特征值、每个第二 URL 的主机名的特征值以及除了各个第二 URL 的主机名以外的追溯提取的至少一级子域名中级数高于第四 URL 的子域名的特征值标记为恶意网址的密文，存储在数据库中。

可选地，所述追溯提取的至少一级子域名的级数为设定阈值。

根据本发明的另一方面，提供了一种网址访问系统，包括：客户端和服务器；

所述客户端包括：

监控模块，用于获取请求访问的网址对应的网址信息；

提取模块，用于根据所述网址信息，提取网址密文；

通信模块，用于将所述网址密文提交给服务器；

保护模块，用于根据服务器返回的恶意网址查询结果，阻断对所述网址的访问行为；

访问模块，用于根据服务器返回的正常网址查询结果，继续进行对所述网址的访问行为。

所述服务器包括：

数据库，用于存储密文；

查询模块，用于将网址密文与数据库中存储的密文进行匹配；若网址密文与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；若网址密文不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果。

可选地，所述监控模块具体用于获取请求访问的网址对应的至少一个第一 URL，所述至少一个第一 URL 包括：所述请求访问的网址对应的网页的 URL 或所述请求访问的网址对应的网页内容中链接的 URL 或下载文件的 URL 或以上信息的任一组合；

所述数据库中标记为恶意网址的密文包括以下信息的一种或多种：恶意 URL 的特征值、恶意 URL 的主机名的特征值和恶意 URL 的子域名的特征值。

可选地，所述监控模块包括：

第一监控单元，用于通过指定响应事件接口，获取所述客户端请求访问的网址对应的网页的 URL。

可选地，所述监控模块包括：

第二监控单元，用于获得客户端的浏览器内部的页面对象；通过调用所述页面对象的方法，获取所述客户端请求访问的网址对应的网页内容中链接的 URL。

可选地，所述监控模块包括：

第三监控单元，用于监控所述客户端的浏览器内部与下载有关的函数；当所述浏览器发生下载行为时，获取所述下载文件的 URL。

可选地，所述客户端还包括：处理模块，用于对所述至少一个第一 URL 进行规范化处理。

可选地，所述处理模块包括：

统一单元，用于将所述第一 URL 中的字母大小写进行统一；

去除单元，用于去除所述第一 URL 中重复多余的路径符和参数。

可选地，所述提取模块包括：

获取单元，用于获取所述第一 URL 的主机名和所述第一 URL 的第一域名段；

计算单元，用于分别计算所述第一 URL 的特征值、所述第一 URL 的主机名的特征值和所述第一 URL 的第一域名段的特征值；

所述第一 URL 的特征值、所述第一 URL 的主机名的特征值和所述第一 URL 的第一域名段的特征值即为所述网址密文。

可选地，若所述第一 URL 的主机名从右至左的第一级根域名为国际顶级域名，则所述获取单元具体用于获取所述第一URL的主机名的第一级子域名为所述第一URL的第一域名段；若所述第一 URL 的主机名从右至左的第一级根域名为国家地区顶级域名，第一级子域名包括国际顶级域名，则所述获取单元具体用于获取所述第一 URL 的主机名的第二级子域名为所述第一 URL 的第一域名段；若所述第一 URL 使用了动态域名，则所述获取单元具体用于获取从动态域名开始，向右提取的下一级子域名为所述第一 URL 的第一域名段。

可选地，所述查询模块具体用于将网址密文与数据库中存储的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的特征值、所述至少一个第一 URL 中任一第一 URL 的主机名的特征值和所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值中的任一个与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果。

可选地，所述查询模块具体用于：

将所述至少一个第一 URL 中任一第一 URL 的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一 URL 中任一第一 URL 的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果。

可选地，所述服务器还包括：构建模块，用于构建所述数据库；

所述构建模块包括：

第一获取单元，用于获取已知为恶意网址且第一域名段相同的至少一个第二 URL；

第二获取单元，用于获取所述至少一个第二 URL 中包含子域名级数最高的第三 URL，从右至左逐级追溯第三 URL 所包含的子域名，提取至少一级子域名；

第一标记单元，用于若所述第二 URL 的第一域名段属于预设的可信名单，将所述每个第二 URL 的特征值和每个第二 URL 的主机名的特征值标记为恶意网址的密文，存储在数据库中；

第二标记单元，用于若所述第二 URL 的第一域名段属于预设的不可信名单，获取至少一个第二 URL 中包含子域名级数最低的第四 URL，将所述每个第二 URL 的特征值、每个第二 URL 的主机名的特征值以及除了各个第二 URL 的主机名以外的追溯提取的至少一级子域名中级数高于第四 URL 的子域名的特征值标记为恶意网址的密文，存储在数据库中。

根据本实施例提供的网址访问方法及系统，当客户端请求访问网址时，从网址信息中提取网址密文，将网址密文提交给服务器，服务器将网址密文与数据库中存储的密文匹配，完成网址的安全查询和验证，客户端根据服务器的验证结果决定是否继续对网址的访问行为。该方法不依赖客户端本地的数据库，将网址的安全查询和验证放在服务器侧完成。由于服务器侧的数据库可以及时的更新互联网上的各类恶意网址，它的升级周期远远短于客户端本地的数据库，而且服务器侧的数据库中恶意网址的信息存储量很大，覆盖面很广，从而能够快速有效地拦截恶意网站。

上述说明仅是本发明技术方案的概述，为了能够更清楚了解本发明的技术手段，而可依照说明书的内容予以实施，并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明的具体实施方式。

附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

图 1 示出了根据本发明一个实施例的网址访问方法的流程图；

图 2 示出了根据本发明一个实施例的网址访问方法的流程图；

图 3 示出了本发明实施例中网址密文匹配过程的流程图；

图 4 示出了根据本发明一个实施例的网址访问系统的结构示意图。

图 5 示意性地示出了可以实现根据本发明的网址访问系统；以及

图 6 示意性地示出了用于保持或者携带实现根据本发明的方法的程序代码的存储单元。

具体实施方式

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人员。

图 1 示出了根据本发明一个实施例的网址访问方法的流程图。本实施例中，以客户端访问的网址为统一资源定位符（Universal Resource Locator，以下简称：URL）为例进行介绍。如图 1 所示，该方法包括如下步骤：

步骤 101、客户端获取请求访问的网址对应的网址信息。

监控客户端各种类型的浏览器的网页访问行为，所请求访问的网址信息称为第一 URL。该第一 URL 可以包括如下几种：

ⅰ.请求访问的网址对应的网页的 URL；

例如，客户端请求访问“新浪”主页，该网页的 URL 即为：http://www.sina.com.cn/。

ⅱ.请求访问的网址对应的网页内容中链接的 URL；

在客户端请求访问的网页的内容中有可能存在一些链接网址，这些链接网址的 URL 也属于监控的范围。

iii.下载文件的 URL。

客户端请求下载文件，该下载文件的 URL 也属于监控的范围。

客户端的某一网页访问行为可能涉及到以上三种 URL 中的一种或多种，即第一

URL 包括以上三种 URL 中的任一种或任意几种的组合。

步骤 102、客户端根据网址信息，提取网址密文。

客户端根据第一 URL 所包含的信息，提取第一 URL 对应的网址密文。

步骤 103、客户端将网址密文提交给服务器。

步骤 104、服务器将网址密文与数据库中存储的密文进行匹配，数据库中存储的密文包括被标记为恶意网址的密文；若网址密文与数据库中标记为恶意网址的密文匹配，执行步骤 105；否则，执行步骤 107。

本实施例在服务器侧预先构建了数据库，该数据库中至少存储了被标记为恶意网址的密文。这些密文都是根据大量已知为恶意网址的 URL 而获得的。

步骤 105、服务器向客户端返回恶意网址查询结果，执行步骤 106。

客户端提交的网址密文与数据库中标记为恶意网址的密文匹配表明客户端要访问的第一 URL 为恶意网址，在此种情况下，服务器向客户端返回恶意网址查询结果。

步骤 106、客户端根据恶意网址查询结果，阻断对网址的访问行为，结束。

步骤 107、服务器向客户端返回正常网址查询结果，执行步骤 108。

客户端提交的网址密文与数据库中标记为恶意网址的密文不匹配表明客户端要访问的第一 URL 为正常网址，在此种情况下，服务器向客户端返回正常网址查询结果。

步骤 108、客户端根据正常网址查询结果，继续进行对网址的访问行为，结束。

根据本实施例提供的网址访问方法，当客户端请求访问网址时，从网址信息中提取网址密文，将网址密文提交给服务器，服务器将网址密文与数据库中存储的密文匹配，完成网址的安全查询和验证，客户端根据服务器的验证结果决定是否继续对网址的访问行为。该方法不依赖客户端本地的数据库，将网址的安全查询和验证放在服务器侧完成。由于服务器侧的数据库可以及时的更新互联网上的各类恶意网址，它的升级周期远远短于客户端本地的数据库，而且服务器侧的数据库中恶意网址的信息存储量很大，覆盖面很广，从而能够快速有效地拦截恶意网站。

图 2 示出了根据本发明一个实施例的网址访问方法的流程图。本实施例提供了一种基于云安全的网址访问方法，不依赖客户端本地的网址数据库，将网址的安全查询和验证放在服务器侧完成。如图 2 所示，该方法包括如下步骤：

步骤 201、客户端获取请求访问的至少一个第一 URL。

本实施例的至少一个第一 URL 可以包括上述实施例所述的三种 URL 中的任一种或任意几种的组合。

上述三种 URL 的获取方法分别描述如下：

通过指定响应事件接口，例如通过实现标准插件机制的指定响应事件接口，获取客户端请求访问的网址对应的网页的 URL。例如，在 IE（Internet Explorer）浏览器中使用浏览器辅助对象（Browser Helper Object，简称：BHO）插件机制，通过响应“BeforeNavigate2”事件可以获取 IE 当前加载的 URL。在火狐（Firefox）浏览器中使用火狐扩展机制提供的指定响应事件接口，获取火狐浏览器当前加载的 URL。在谷歌（chrome）浏览器中使用网景插件应用程序编程接口（Netscape Plugin

Application Programming Interface，简称：NPAPI）插件机制，获取谷歌浏览器当前加载的URL。

从浏览器环境中获得浏览器访问的网页内容中的链接 URL，包括但不限于页面内的超级链接地址。具体方法是，获得浏览器内部的页面对象，再通过调用页面对象的方法，获取网页内容中的链接URL。其中，可以通过浏览器提供的标准插件机制获得浏览器内部的页面对象。

从浏览器环境中获得浏览器正在下载文件的 URL。具体方法是，监控浏览器内部与下载有关的函数，当发现浏览器发生下载行为时，能够分析获得下载文件的URL。其中，可以使用钩子（HOOK）机制监控浏览器内部与下载有关的函数。

步骤202、客户端对至少一个第一URL进行规范化处理。

该规范化处理过程可以包括：将第一URL中的字母大小写进行统一，包括协议、主机名、路径名、文件名和参数等信息；去除第一URL中重复多余的路径符和参数。

例如，第一URL为：HTTp://www.A.com//aBc/abc.Php?A=1。

将其中的大小写字母统一为小写：http://www.a.com//abc/abc.php?a=1；

去除其中重复多余的路径符：http://www.a.com/abc/abc.php?a=1。

步骤203、客户端根据第一URL，提取网址密文。

对于第一URL来说，第一URL本身（url）、第一URL的主机名（host）和第一URL的第一域名段（domain）是三段关键信息。在得到第一URL之后，获取第一URL的主机名和第一URL的第一域名段。其中，第一URL的主机名是去除掉第一URL中的路径符、协议头和端口号等信息之后的主机部分；第一URL的第一域名段是根据第一URL的主机名从右至左逐级追溯得到的。优选地，在获取第一URL的第一域名段时，从右至左最高追溯7级。

若第一URL的主机名从右至左的第一级根域名为国际顶级域名，则第一URL的第一域名段为第一URL的主机名的第一级子域名。其中国际顶级域名是指“com”、“net”、“org”、“edu”、“gov”等常见顶级域名。例如，第一URL的主机名为 www.a.com，它从右至左的第一级根域名为“com”，那么提取它的第一级子域名“a.com”为第一URL的第一域名段。

若第一URL的主机名从右至左的第一级根域名为国家地区顶级域名，第一级子域名包括国际顶级域名，则第一URL的第一域名段为第一URL的主机名的第二级子域名。其中国家地区顶级域名是指“cn”、“hk”等特殊顶级域名。例如，第一URL的主机名为 www.a.com.cn，它从右至左的第一级根域名为“cn”，第一级子域名为“com.cn”，那么提取它的第二级子域名“a.com.cn”为第一URL的第一域名段。

若第一URL的主机名使用了动态域名，则第一URL的第一域名段为第一URL的主机名从动态域名开始，向右提取的下一级子域名。其中动态域名是指一些二级或三级动态域名，如“3322.org”、“s.3322.org”、“s.3322.net”等动态域名。例如，第一URL的主机名为 www.a.3322.org，它使用了动态域名“3322.org”，那么从动态域名开始，向右提取下一级子域名“a.3322.org”为第一URL的第一域名段。

本实施例分别计算上述三段关键信息的特征值作为网址密文。所述特征值可以具体为哈希值，优选地，所述特征值可以为根据消息摘要算法第五版(Message Digest Algorithm，以下简称:md5)计算得到的哈希值，或SHA1码，或CRC(Cyclic Redundancy Check，循环冗余校验）码等可唯一标识原程序的特征码。在下面的例子中，以特征值为32位md5哈希值为例进行说明。

例如，第一URL为：http://www.a.com/abc/abc.php?a=1；根据上述方法，获取第一URL的主机名为：www.a.com；获取第一URL的第一域名段为：a.com。

计算第一URL本身的32位md5哈希值为：

md5(http://www.a.com/abc/abc.php?a=1,32)=e2a6b69ff15c6a8e276f089250ab3f7d

计算第一URL的主机名的32位md5哈希值为：

md5(www.a.com,32) = 30f4a7bbefe70d75616707c80921a7e8

计算第一URL的第一域名段的32位md5哈希值为：

md5(a.com,32) = b3655bd7aad56513fcdacbd4254ed6b7

对于具有一个第一URL的情况，上述计算得到的第一URL的32位md5哈希值、第一URL的主机名的32位md5哈希值和第一URL的第一域名段的32位md5哈希值即为第一URL的网址密文。对于多个第一URL的情况，分别计算每个第一URL的上述三段关键信息的32位md5哈希值，将每个第一URL的上述三段关键信息的32位md5哈希值形成一组，从而得到包括但不限于一组32位md5哈希值的网址密文。

在上述第一URL为：http://www.a.com/abc/abc.php?a=1的例子中，得到的第一URL的一组网址密文如下：

domain|host|url

a.com | www.a.com| http://www.a.com/abc/abc.php?a=1

b3655bd7aad56513fcdacbd4254ed6b7| 30f4a7bbefe70d75616707c80921a7e8 | e2a6b69ff15c6a8e276f089250ab3f7d

步骤204、客户端将网址密文提交给服务器。

步骤205、服务器将网址密文与数据库中存储的密文进行匹配，数据库中存储的密文至少包括被标记为恶意网址的密文；若网址密文与数据库中标记为恶意网址的密文匹配，执行步骤206；否则，执行步骤208。

本实施例在服务器侧预先构建了网址数据库，该网址数据库中至少存储了被标记为恶意网址的密文。具体地，网址数据库中的数据键值按照网址url、网址host和网址domain三种关键信息的特征值进行存储，三种关键信息的键值可以分别按照正常网址和恶意网址进行标记。具体地，标记为恶意网址的密文包括以下信息的一种或多种：恶意URL的特征值、恶意URL的主机名的特征值和恶意URL的子域名的特征值。

网址数据库中的密文都是根据大量已知为恶意网址的URL而获得的。

本实施例中，构建网址数据库可以包括以下步骤：

（a）获取已知为恶意网址且第一域名段相同的至少一个第二URL。

在获取到大量已知为恶意网址的URL之后，按照客户端提取主机名和第一域名段的方法，获取这些恶意网址的URL的主机名和第一域名段。在这些恶意网址的URL中，经常会出现第一域名段相同的URL。例如，对于以下恶意网址的URL：

http://a.b.c.d.e.f.g.com/abc/abc1.php?a=1

http://b.c.d.e.f.g.com/abc/abc.php?a=1

http://d.e.f.g.com/abc/abc.php?a=1

其第一域名段均为g.com。在这里，以上三个URL被称为第二URL。

（b）获取至少一个第二URL中包含子域名级数最高的第三URL，从右至左逐级追溯第三URL所包含的子域名，提取至少一级子域名；

在上述例子中，三个第二URL中包含子域名级数最高的第三URL是：http://a.b.c.d.e.f.g.com/abc/abc1.php?a=1，它共包含7级子域名。从右至左逐级追溯第三URL所包含的子域名，提取到如下7级子域名：

第一级子域名：g.com

第二级子域名：f.g.com

第三级子域名：e.f.g.com

第四级子域名：d.e.f.g.com

第五级子域名：c.d.e.f.g.com

第六级子域名：b.c.d.e.f.g.com

第七级子域名：a.b.c.d.e.f.g.com

优选地，本步骤追溯提取的至少一级子域名的级数为设定阈值N。由于一个域名中的多个子域名中既有恶意网址也有正常网址，一般6级以下都会出现这种情况，所以优选地，N大于或等于6。

（c）若第二URL的第一域名段属于预设的可信名单，例如，白名单，则将每个第二URL的特征值、每个第二URL的主机名的特征值标记为恶意网址的密文，存储在数据库中。

对于一些访问量较大的正常网站，例如：sina.com.cn，sohu.com等网站，可将它们写入预设的可信名单。如果第二URL的第一域名段属于这样的可信名单，那么将每个第二URL的特征值和每个第二URL的主机名的特征值标记为恶意网址的密文，存储在数据库中。

在上述例子中，如果g.com属于预设的可信名单，那么被标记为恶意网址的密文包括以下信息的特征值：

各个第二URL：

http://a.b.c.d.e.f.g.com/abc/abc1.php?a=1

http:// b.c.d.e.f.g.com/abc/abc.php?a=1

http:// d.e.f.g.com/abc/abc.php?a=1

各个第二URL的主机名：

a.b.c.d.e.f.g.com

b.c.d.e.f.g.com

d.e.f.g.com

将上述信息的特征值存储在云端网址数据库中，且被标记为恶意网址的密文。然而，未出现恶意网址的其他子域名的特征值，可以被标记为正常网址也存储在云端网址数据库中，包括：

g.com

f.g.com

e.f.g.com

c.d.e.f.g.com

（d）若第二URL的第一域名段属于预设的不可信名单，例如，黑名单，则获取至少一个第二URL中包含子域名级数最低的第四URL，将每个第二URL的特征值、每个第二URL的主机名的特征值以及除了各个第二URL的主机名以外的追溯提取的至少一级子域名中级数高于第四URL的子域名的特征值标记为恶意网址的密文，存储在数据库中。

对于一些访问量很小的网站，可以将它们写入不可信名单。如果第二URL的第一域名段属于这样的不可信名单，那么获取至少一个第二URL中包含子域名级数最低的第四URL，将每个第二URL的特征值、每个第二URL的主机名的特征值以及除了各个第二URL的主机名以外的追溯提取的至少一级子域名中级数高于第四URL的子域名的特征值标记为恶意网址的密文，存储在数据库中。

在上述例子中，如果g.com属于预设的不可信名单，获取其中包含子域名级数最低的第四URL为：http:// www.d.e.f.g.com/abc/abc.php?a=1，它共包含4级子域名，那么被标记为恶意网址的密文包括以下信息的特征值：

各个第二URL：

http://a.b.c.d.e.f.g.com/abc/abc1.php?a=1

http:// b.c.d.e.f.g.com/abc/abc.php?a=1

http:// d.e.f.g.com/abc/abc.php?a=1

各个第二URL的主机名：

a.b.c.d.e.f.g.com

b.c.d.e.f.g.com

d.e.f.g.com

所述追溯提取的至少一级子域名中级数高于第四URL的子域名包括：a.b.c.d.e.f.g.com，b.c.d.e.f.g.com，c.d.e.f.g.com，其中a.b.c.d.e.f.g.com和b.c.d.e.f.g.com是第二URL的主机名，那么除了各个第二URL的主机名以外的追溯提取的至少一级子域名中级数高于第四URL的子域名就是：

c.d.e.f.g.com

将上述信息的特征值存储在云端网址数据库中，且被标记为恶意网址的密文。然而，未出现恶意网址的其他子域名的特征值，可以被标记为正常网址也存储在云端网址数据库中，包括：

g.com

f.g.com

e.f.g.com

本步骤所提到的特征值应该是与客户端所提交的特征值是相同类型的。该特征值可以具体为哈希值，优选地，该特征值可以为根据 md5 算法计算得到的哈希值。

服务器将客户端提交的网址密文与云端网址数据库中标记为恶意网址的密文匹配，具体匹配过程如下：

若至少一个第一 URL 中任一第一 URL 的特征值、至少一个第一 URL 中任一第一 URL 的主机名的特征值和至少一个第一 URL 中任一第一 URL 的第一域名段的特征值中的任一个与云端网址数据库中标记为恶意网址的密文匹配，则执行步骤 206；否则，执行步骤 208。

图 3 示出了本发明实施例中网址密文匹配过程的流程图。图 3 所示的匹配过程是本发明实施例一种优选的实施方式，但本发明不仅限于此。如图 3 所示，服务器将客户端提交的网址密文与数据库中存储的密文进行匹配的过程还可以包括如下步骤：

步骤 301、将至少一个第一 URL 中任一第一 URL 的特征值与数据库中标记为恶意网址的密文进行匹配；若匹配，执行步骤 206；否则，执行步骤 302；

步骤 302、将至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文进行匹配；若匹配，执行步骤 206；否则，执行步骤 303；

步骤 303、将至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文进行匹配；若匹配，执行步骤 206；否则，执行步骤 208。

综上所述，上述匹配过程包括如下三种情况：

（1）至少一个第一 URL 的三段关键信息的特征值中的任一特征值与云端网址数据库中标记为恶意网址的密文匹配，执行步骤 206；

（2）至少一个第一 URL 的三段关键信息的特征值都不与云端网址数据库中标记为恶意网址的密文匹配，执行步骤 208；

（3）至少一个第一 URL 的三段关键信息的其中一特征值与云端网址数据库中标记为正常网址的密文匹配，且其他特征值都不与云端网址数据库中标记为恶意网址的密文匹配，执行步骤 208。

步骤 206、服务器向客户端返回恶意网址查询结果，执行步骤 207。

客户端提交的网址密文与数据库中标记为恶意网址的密文匹配表明客户端要访问的第一 URL 为恶意网址，在此种情况下，服务器向客户端返回恶意网址查询结果。

步骤 207、客户端根据恶意网址查询结果，阻断对网址的访问行为，结束。

客户端根据恶意网址查询结果，阻断对网址的访问行为，并提示用户。

步骤 208、服务器向客户端返回正常网址查询结果，执行步骤 209。

客户端提交的网址密文与数据库中标记为恶意网址的密文不匹配表明客户端要访问的第一 URL 为正常网址，在此种情况下，服务器向客户端返回正常网址查询结果。

步骤 209、客户端根据正常网址查询结果，继续进行对网址的访问行为，结束。

根据本实施例提供的网址访问方法，当客户端请求访问网址时，从网址信息中提取网址密文，将网址密文提交给服务器，服务器将网址密文与数据库中存储的密文匹配，完成网址的安全查询和验证，客户端根据服务器的验证结果决定是否继续对网址的访问行为。该方法不依赖客户端本地的数据库，将网址的安全查询和验证放在服务器侧完成。由于服务器侧的数据库可以及时的更新互联网上的各类恶意网址，它的升级周期远远短于客户端本地的数据库，而且服务器侧的数据库中恶意网址的信息存储量很大，覆盖面很广，从而能够快速有效地拦截恶意网站。

图 4 示出了根据本发明一个实施例的网址访问系统的结构示意图。如图 4 所示，该网址访问系统包括：客户端 1 和服务器 2。

客户端 1 包括：监控模块 10、提取模块 11、通信模块 12、保护模块 13 和访问模块 14。其中，监控模块 10 用于获取请求访问的网址对应的网址信息；提取模块 11 用于根据网址信息，提取网址密文；通信模块 12 用于将网址密文提交给服务器 2；保护模块 13 用于根据服务器 2 返回的恶意网址查询结果，阻断对网址的访问行为；访问模块 14 用于根据服务器 2 返回的正常网址查询结果，继续进行对网址的访问行为。

服务器 2 包括：数据库 20 和查询模块 21。其中，数据库 20 用于存储密文；查询模块 21 用于将网址密文与数据库 20 中存储的密文进行匹配；若网址密文与数据库 20 中标记为恶意网址的密文匹配，则向客户端 1 返回恶意网址查询结果；若网址密文不与数据库 20 中标记为恶意网址的密文匹配，则向客户端 1 返回正常网址查询结果。

进一步的，监控模块 10 具体用于获取请求访问的网址对应的至少一个第一 URL，所述至少一个第一 URL 包括：请求访问的网址对应的网页的 URL 或请求访问的网址对应的网页内容中链接的 URL 或下载文件的 URL 或以上信息的任一组合。所述数据库中标记为恶意网址的密文包括以下信息的一种或多种：恶意 URL 的特征值、恶意 URL 的主机名的特征值和恶意 URL 的子域名的特征值。

监控模块 10 可以包括：第一监控单元 10a，用于通过指定响应事件接口，获取客户端 1 请求访问的网址对应的网页的 URL。

监控模块 10 也可以包括：第二监控单元 10b，用于获得客户端 1 的浏览器内部的页面对象；通过调用页面对象的方法，获取客户端 1 请求访问的网址对应的网页内容中链接的 URL。

监控模块 10 还可以包括：第三监控单元 10c，用于监控客户端 1 的浏览器内部与下载有关的函数；当浏览器发生下载行为时，获取下载文件的 URL。

客户端 1 还可以包括：处理模块 15，用于对至少一个第一 URL 进行规范化处理。进一步的，处理模块 15 可以包括：统一单元 15a 和去除单元 15b，统一单元 15a 用于将第一 URL 中的字母大小写进行统一；去除单元 15b 用于去除第一 URL 中重复多余的路径符和参数。

提取模块 11 可以包括：获取单元 11a 和计算单元 11b。其中，获取单元 11a 用于获取第一 URL 的主机名和第一 URL 的第一域名段；计算单元 11b，用于分别计算第

一 URL 的特征值、第一 URL 的主机名的特征值和第一 URL 的第一域名段的特征值；所述第一 URL 的特征值、所述第一 URL 的主机名的特征值和所述第一 URL 的第一域名段的特征值即为所述网址密文。

若第一 URL 的主机名从右至左的第一级根域名为国际顶级域名，则获取单元 11a 具体用于获取第一 URL 的主机名的第一级子域名为第一 URL 的第一域名段；

若第一 URL 的主机名从右至左的第一级根域名为国家地区顶级域名，第一级子域名包括国际顶级域名，则获取单元 11a 具体用于获取第一 URL 的主机名的第二级子域名为第一 URL 的第一域名段；

若第一 URL 使用了动态域名，则获取单元 11a 具体用于获取从动态域名开始，向右提取的下一级子域名为第一 URL 的第一域名段。

查询模块 21 具体用于将网址密文与数据库 20 中存储的密文进行匹配；若至少一个第一 URL 中任一第一 URL 的特征值、至少一个第一 URL 中任一第一 URL 的主机名的特征值和至少一个第一 URL 中任一第一 URL 的第一域名段的特征值中的任一个与数据库 20 中标记为恶意网址的密文匹配，则向客户端 1 返回恶意网址查询结果。

作为一种优选的实施方式，该查询模块 21 可以具体用于：

将所述至少一个第一 URL 中任一第一 URL 的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一 URL 中任一第一 URL 的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果。

服务器 2 还包括构建模块 22，该构建模块 22 可以包括：第一获取单元 22a、第二获取单元 22b、第一标记单元 22c 和第二标记单元 22d。其中，第一获取单元 22a 用于获取已知为恶意网址且第一域名段相同的至少一个第二 URL；第二获取单元 22b 用于获取至少一个第二 URL 中包含子域名级数最高的第三 URL，从右至左逐级追溯第三 URL 所包含的子域名，提取至少一级子域名；第一标记单元 22c 用于若第二 URL 的第一域名段属于预设的可信名单，将每个第二 URL 的特征值和每个第二 URL 的主机名的特征值标记为恶意网址的密文，存储在数据库 20 中；第二标记单元 22d，用于若第二 URL 的第一域名段属于预设的不可信名单，获取至少一个第二 URL 中包含

子域名级数最低的第四URL，将每个第二URL的特征值、每个第二URL的主机名的特征值以及除了各个第二 URL 的主机名以外的追溯提取的至少一级子域名中级数高于第四URL的子域名的特征值标记为恶意网址的密文，存储在数据库20中。

根据本实施例提供的网址访问系统，当客户端请求访问网址时，从网址信息中提取网址密文，将网址密文提交给服务器，服务器将网址密文与数据库中存储的密文匹配，完成网址的安全查询和验证，客户端根据服务器的验证结果决定是否继续对网址的访问行为。该方法不依赖客户端本地的数据库，将网址的安全查询和验证放在服务器侧完成。由于服务器侧的数据库可以及时的更新互联网上的各类恶意网址，它的升级周期远远短于客户端本地的数据库，而且服务器侧的数据库中恶意网址的信息存储量很大，覆盖面很广，从而能够快速有效地拦截恶意网站。

图5示出了可以实现根据本发明的网址访问系统。该网址访问系统上包括处理器510和以存储器520形式的计算机程序产品或者计算机可读介质。存储器520可以是诸如闪存、EEPROM（电可擦除可编程只读存储器）、EPROM、硬盘或者ROM之类的电子存储器。存储器520具有用于执行上述方法中的任何方法步骤的程序代码1331的存储空间530。例如，用于程序代码的存储空间1330可以包括分别用于实现上面的方法中的各种步骤的各个程序代码531。这些程序代码可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。这些计算机程序产品包括诸如硬盘，紧致盘（CD）、存储卡或者软盘之类的程序代码载体。这样的计算机程序产品通常为如参考图 6 所述的便携式或者固定存储单元。该存储单元可以具有与图 5 的服务器中的存储器 520 类似布置的存储段、存储空间等。程序代码可以例如以适当形式进行压缩。通常，存储单元包括计算机可读代码531’，即可以由例如诸如 510 之类的处理器读取的代码，这些代码当由服务器运行时，导致该服务器执行上面所描述的方法中的各个步骤。

在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述，构造这类系统所要求的结构是显而易见的。此外，本发明也不针对任何特定编程语言。应当明白，可以利用各种编程语言实现在此描述的本发明的内容，并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

类似地，应当理解，为了精简本公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释成反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本发明的单独实施例。

本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它们分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

本发明的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本发明实施例的网址访问系统中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序（例如，计算机程序和计算机程序产品）。这样的实现本发明的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何其他形式提供。

应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

此外，还应当注意，本说明书中使用的语言主要是为了可读性和教导的目的而选择的，而不是为了解释或者限定本发明的主题而选择的。因此，在不偏离所附权利要求书的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。对于本发明的范围，对本发明所做的公开是说明性的，而非限制性的，本发明的范围由所附权利要求书限定。

权 利 要 求

1、一种网址访问方法，包括：

客户端获取请求访问的网址对应的网址信息；

所述客户端根据所述网址信息，提取网址密文；

所述客户端将所述网址密文提交给服务器；

所述服务器将网址密文与数据库中存储的密文进行匹配；

若网址密文与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；所述客户端根据所述恶意网址查询结果，阻断对所述网址的访问行为；

若网址密文不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果；所述客户端根据所述正常网址查询结果，继续进行对所述网址的访问行为。

2、根据权利要求 1 所述的方法，所述网址信息具体为至少一个第一 URL。

3、根据权利要求 2 所述的方法，所述数据库中标记为恶意网址的密文包括以下信息的一种或多种：恶意 URL 的特征值、恶意 URL 的主机名的特征值和恶意 URL 的子域名的特征值。

4、根据权利要求 2 所述的方法，所述至少一个第一 URL 包括：所述请求访问的网址对应的网页的 URL 或所述请求访问的网址对应的网页内容中链接的 URL 或下载文件的 URL 或以上信息的任一组合。

5、根据权利要求 4 所述的方法，所述客户端获取请求访问的网址对应的网址信息包括：

通过指定响应事件接口，获取所述客户端请求访问的网址对应的网页的 URL。

6、根据权利要求 4 所述的方法，所述客户端获取请求访问的网址对应的网址信息包括：

获得客户端的浏览器内部的页面对象；

通过调用所述页面对象的方法，获取所述客户端请求访问的网址对应的网页内容中链接的 URL。

7、根据权利要求 4 所述的方法，所述客户端获取请求访问的网址对应的网址信息包括：

监控所述客户端的浏览器内部与下载有关的函数；

当所述浏览器发生下载行为时，获取所述下载文件的 URL。

8、根据权利要求 2 至 7 任一项所述的方法，在所述客户端根据所述网址信息，提取网址密文之前还包括：所述客户端对所述至少一个第一 URL 进行规范化处理。

9、根据权利要求 8 所述的方法，所述客户端对所述至少一个第一 URL 进行规范化处理包括：

将所述第一 URL 中的字母大小写进行统一；

去除所述第一 URL 中重复多余的路径符和参数。

10、根据权利要求 2 至 7 任一项所述的方法，所述客户端根据所述网址信息，提取网址密文包括：

获取所述第一 URL 的主机名和所述第一 URL 的第一域名段；

分别计算所述第一 URL 的特征值、所述第一 URL 的主机名的特征值和所述第一 URL 的第一域名段的特征值；

所述第一 URL 的特征值、所述第一 URL 的主机名的特征值和所述第一 URL 的第一域名段的特征值即为所述网址密文。

11、根据权利要求 10 所述的方法，若所述第一 URL 的主机名从右至左的第一级根域名为国际顶级域名，则所述第一 URL 的第一域名段为所述第一 URL 的主机名的第一级子域名；

若所述第一 URL 的主机名从右至左的第一级根域名为国家地区顶级域名，第一级子域名包括国际顶级域名，则所述第一 URL 的第一域名段为所述第一 URL 的主机名的第二级子域名；

若所述第一 URL 的主机名使用了动态域名，则所述第一 URL 的第一域名段为第一 URL 的主机名从动态域名开始，向右提取的下一级子域名。

12、根据权利要求 10 所述的方法，所述若网址密文与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果具体为：若所述至少一个第一 URL 中任一第一 URL 的特征值、所述至少一个第一 URL 中任一第一 URL 的主机名的特征值和所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值中的任一个与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果。

13、根据权利要求 12 所述的方法，所述服务器将网址密文与数据库中存储的密文进行匹配包括：

将所述至少一个第一 URL 中任一第一 URL 的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一 URL 中任一第一 URL 的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；

若所述至少一个第一 URL 中任一第一 URL 的主机名的特征值不与数据库中标记为恶意网址的密文匹配，则将所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文进行匹配；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；若所述至少一个第一 URL 中任一第一 URL 的第一域名段的特征值不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回

正常网址查询结果。

14、根据权利要求 3 所述的方法，还包括构建所述数据库的步骤；

所述构建数据库的步骤包括：

获取已知为恶意网址且第一域名段相同的至少一个第二 URL；

获取所述至少一个第二 URL 中包含子域名级数最高的第三 URL，从右至左逐级追溯第三 URL 所包含的子域名，提取至少一级子域名；

若所述第二 URL 的第一域名段属于预设的可信名单，将所述每个第二 URL 的特征值和每个第二 URL 的主机名的特征值标记为恶意网址的密文，存储在数据库中；

若所述第二 URL 的第一域名段属于预设的不可信名单，获取至少一个第二 URL 中包含子域名级数最低的第四 URL，将所述每个第二 URL 的特征值、每个第二 URL 的主机名的特征值以及除了各个第二 URL 的主机名以外的追溯提取的至少一级子域名中级数高于第四 URL 的子域名的特征值标记为恶意网址的密文，存储在数据库中。

15、一种网址访问系统，包括：客户端和服务器；

所述客户端包括：

监控模块，用于获取请求访问的网址对应的网址信息；

提取模块，用于根据所述网址信息，提取网址密文；

通信模块，用于将所述网址密文提交给服务器；

保护模块，用于根据服务器返回的恶意网址查询结果，阻断对所述网址的访问行为；

访问模块，用于根据服务器返回的正常网址查询结果，继续进行对所述网址的访问行为。

所述服务器包括：

数据库，用于存储密文；

查询模块，用于将网址密文与数据库中存储的密文进行匹配；若网址密文与数据库中标记为恶意网址的密文匹配，则向所述客户端返回恶意网址查询结果；若网址密文不与数据库中标记为恶意网址的密文匹配，则向所述客户端返回正常网址查询结果。

16、一种计算机程序，包括计算机可读代码，当所述计算机可读代码在服务器上运行时，导致所述服务器执行根据权利要求 1-14 中的任一个所述的 Android 应用程序的安全检测方法。

17、一种计算机可读介质，其中存储了如权利要求 16 所述的网址访问方法。

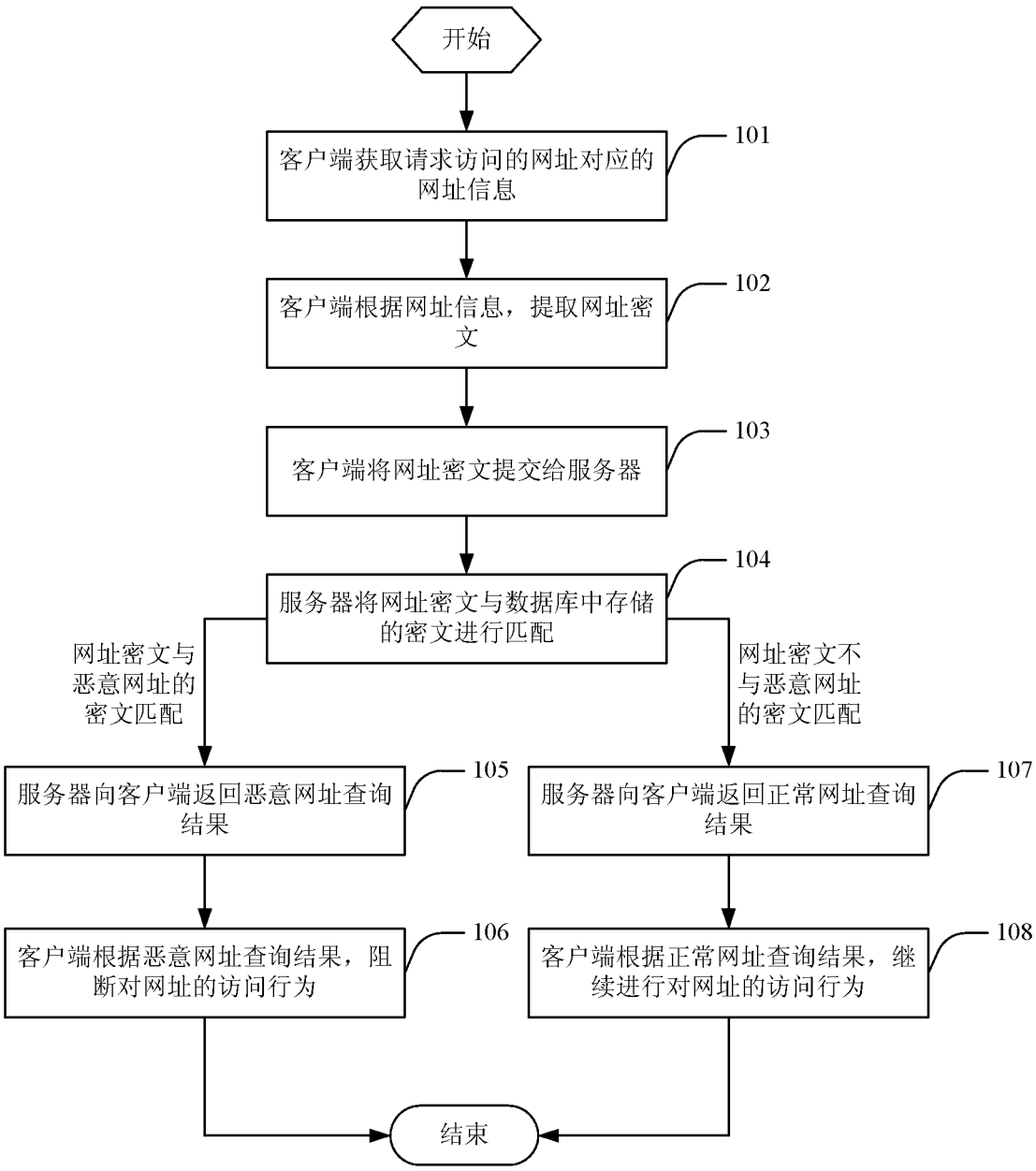
开始
客户端获取请求访问的网址对应的网址信息
101
客户端根据网址信息，提取网址密文
102
客户端将网址密文提交给服务器
103
服务器将网址密文与数据库中存储的密文进行匹配
104
网址密文与恶意网址的密文匹配
网址密文不与恶意网址的密文匹配
服务器向客户端返回恶意网址查询结果
105
服务器向客户端返回正常网址查询结果
107
客户端根据恶意网址查询结果，阻断对网址的访问行为
106
客户端根据正常网址查询结果，继续进行对网址的访问行为
108
结束

图 1

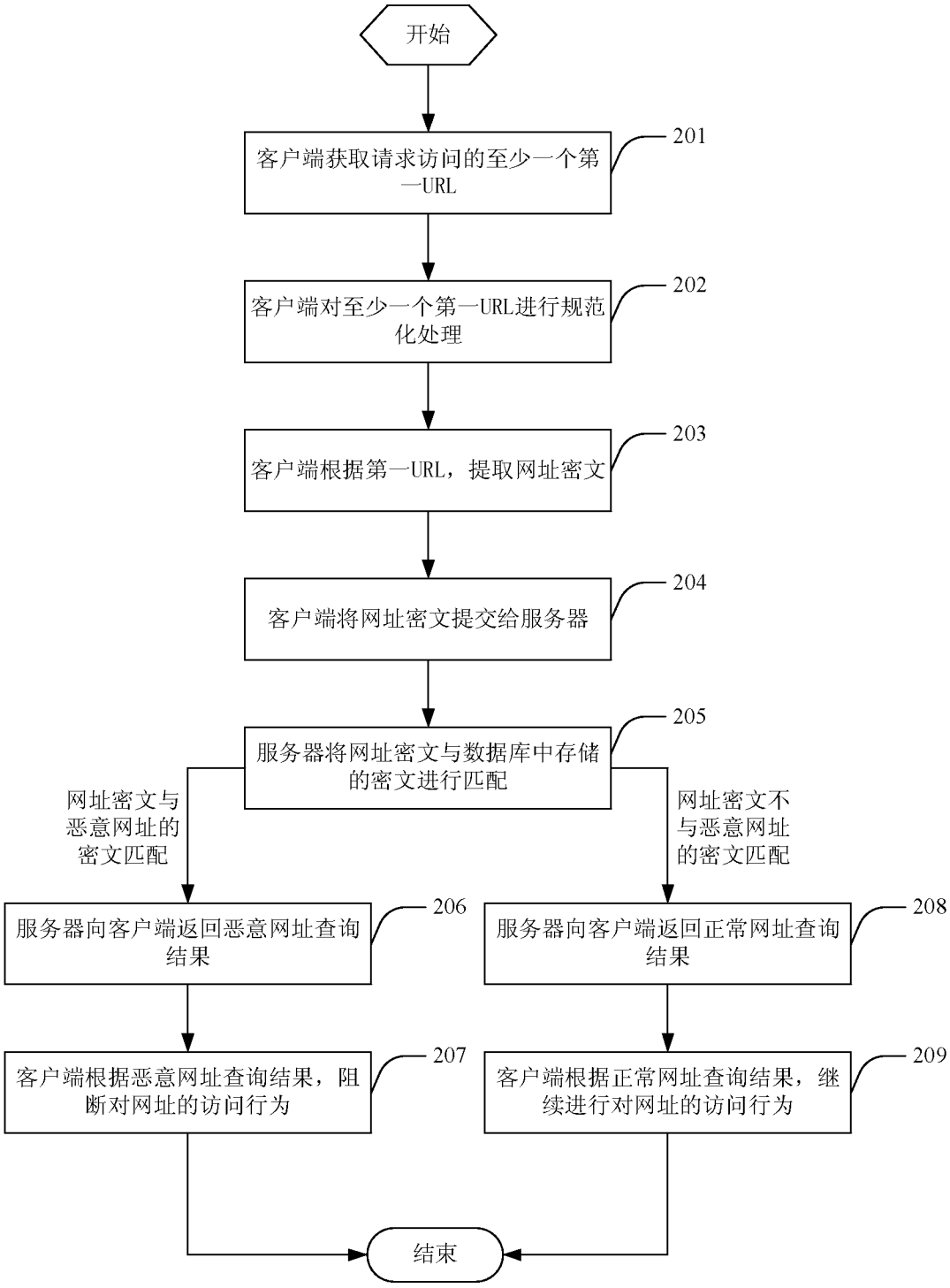
开始
客户端获取请求访问的至少一个第一URL
201
客户端对至少一个第一URL进行规范化处理
202
客户端根据第一URL，提取网址密文
203
客户端将网址密文提交给服务器
204
服务器将网址密文与数据库中存储的密文进行匹配
205
网址密文与恶意网址的密文匹配
网址密文不与恶意网址的密文匹配
服务器向客户端返回恶意网址查询结果
206
服务器向客户端返回正常网址查询结果
208
客户端根据恶意网址查询结果，阻断对网址的访问行为
207
客户端根据正常网址查询结果，继续进行对网址的访问行为
209
结束

图 2

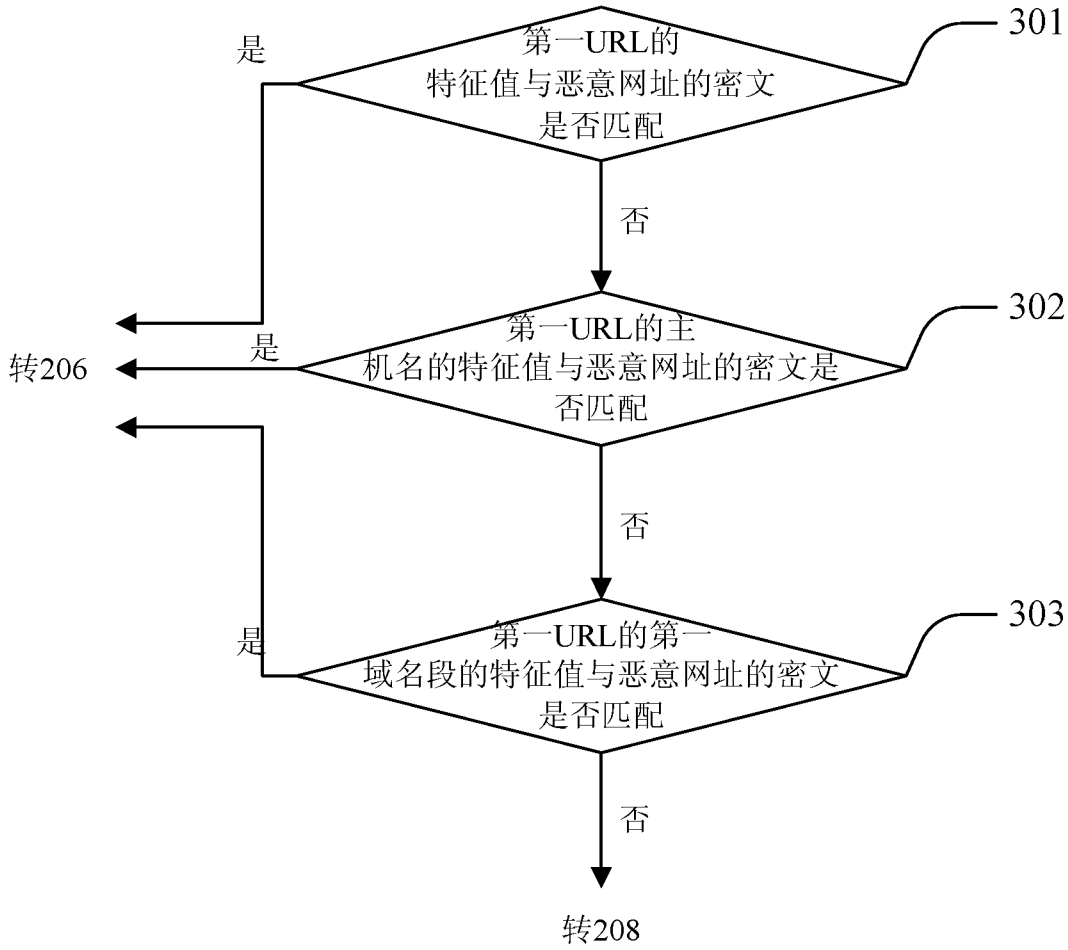
第一URL的
特征值与恶意网址的密文
是否匹配
301
是
否
第一URL的主
机名的特征值与恶意网址的密文是
否匹配
302
是
转206
否
第一URL的第一
域名段的特征值与恶意网址的密文
是否匹配
303
是
否
转208

图 3

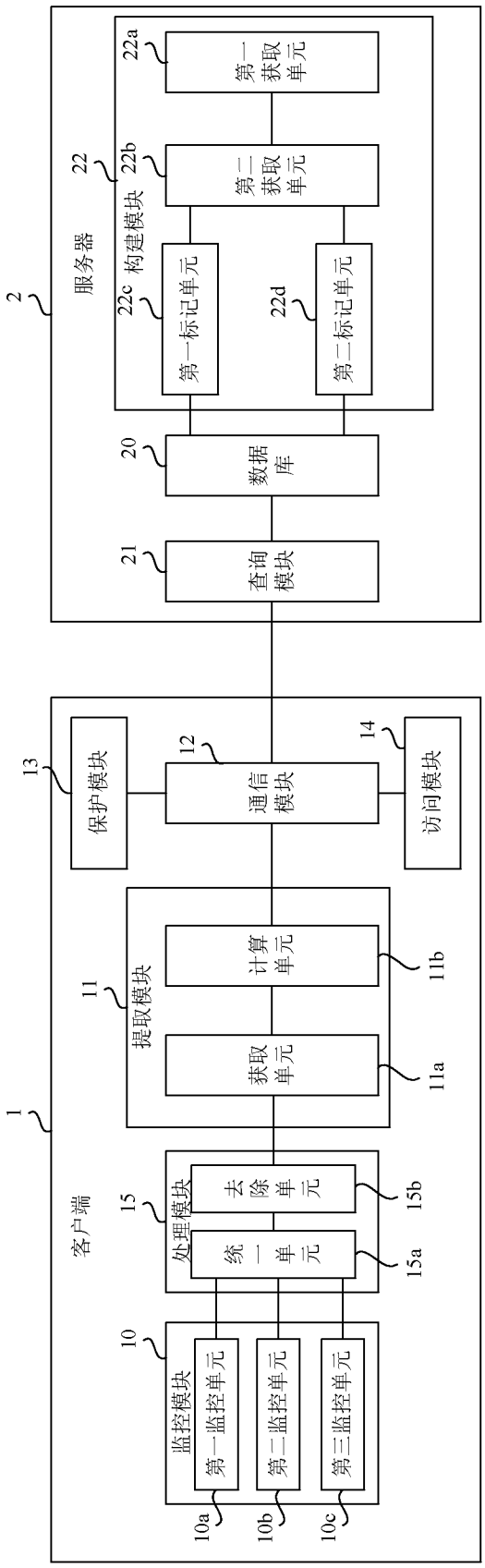
2
服务器
22
构建模块
22a
第一获取单元
22b
第二获取单元
22c
第一标记单元
22d
第二标记单元
20
数据库
21
查询模块
1
客户端
13
保护模块
12
通信模块
14
访问模块
11
提取模块
计算单元
11b
获取单元
11a
15
处理模块
去除单元
15b
统一单元
15a
10
监控模块
第一监控单元
第二监控单元
第三监控单元
10a
10b
10c

图 4

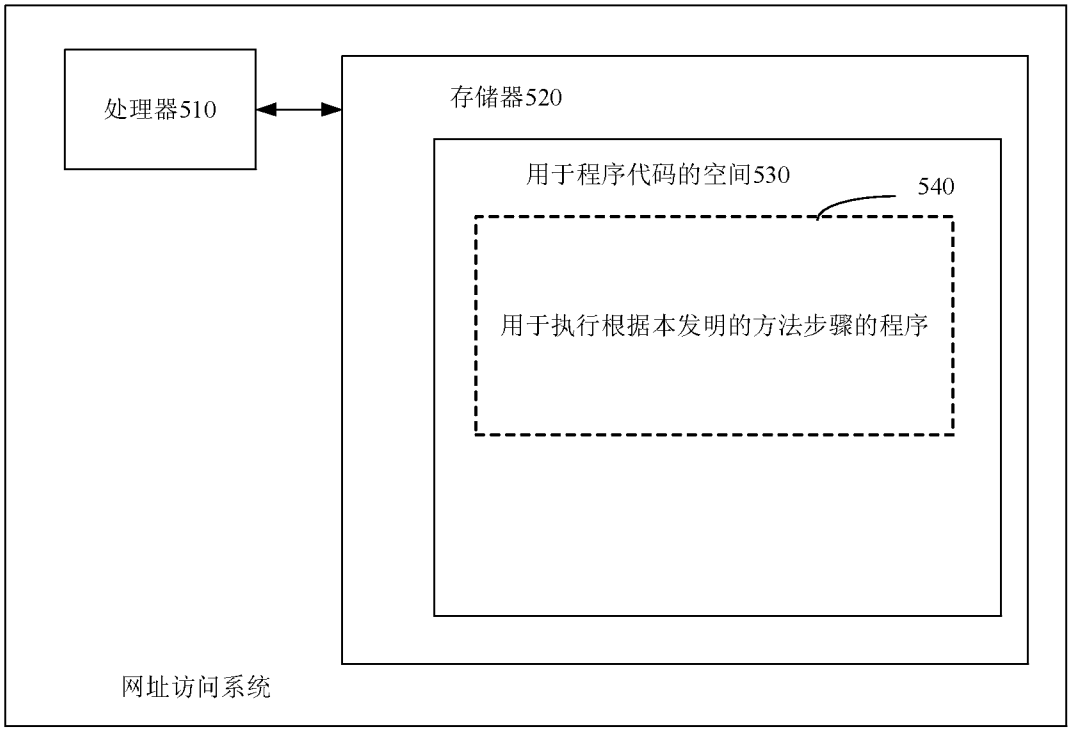
处理器510
存储器520
用于程序代码的空间530
540
用于执行根据本发明的方法步骤的程序
网址访问系统

图 5

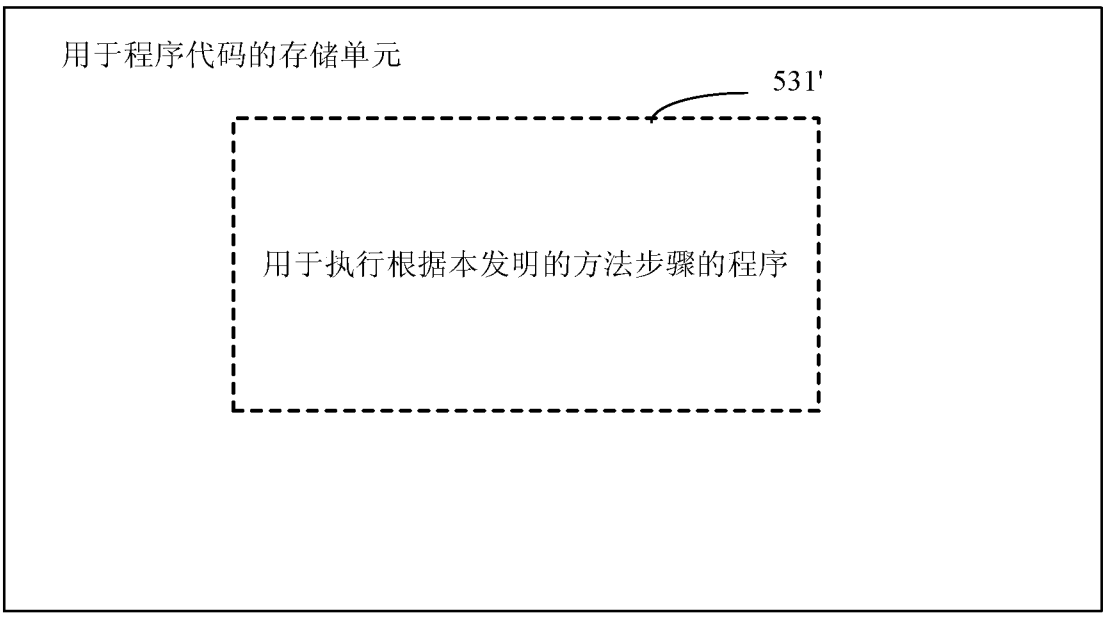
用于程序代码的存储单元
531'
用于执行根据本发明的方法步骤的程序

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/082729

A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

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)

IPC: H04L; G06F

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)

CNABS, CNTXT, CNKI, VEN: match, compare, website, access, hostile, malicious, Trojan, virus, phish, ciphertext

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 102833258 A (BEIJING QIHOO SCI&TECHNOLOGY CO., LTD. et al.) 19 December 2012 (19.12.2012) the whole document	1-17
X	CN 102054030 A (HUIZHOU TCL MOBILE COMMUNICATION CO., LTD.) 11 May 2011 (11.05.2011) see description, paragraphs [0038]-[0065] and figures 1 and 2	1, 15-17
Y	the same as above	2-14
X	CN 102594825 A (BEIJING BAIDU NETCOM SCI&TECHNOLOGY CO.) 18 July 2012 (18.07.2012) see description, paragraphs [0123]-[0159] and figure 4	1, 15-17

☒ 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
14 November 2013 (14.11.2013)	28 November 2013 (28.11.2013)
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer YU, Wenqing Telephone No. (86-10) 62411450

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/082729

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	the same as above	2-14
Y	CN 101958896 A (UNIV. ZHEJIANG) 26 January 2011 (26.01.2011) see description, paragraphs [0033]-[0058]	2-14
A	US 2005/0283835 A1 (LALONDE, C. et al.) 22 December 2005 (22.12.2005) the whole document	1-17
A	CN 102571770 A (BEIJING NSFOCUS INFORMATION SECURITY TEC.) 11 July 2012 (11.07.2012) the whole document	1-17
A	CN 102647417 A (QIZHI SOFTWARE BEIJING CO., LTD.) 22 August 2012 (22.08.2012) the whole document	1-17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/082729

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:16, 17
because they relate to subject matter not required to be searched by this Authority, namely:
The subject matter of claim 16 relates to a computer program including computer readable codes, the subject matter of claim 17 relates to a computer readable medium storing the program claimed in claim 16, said claims do not warrant an international search according to the criteria set out in Rule 39. 1(iv). The subject matters of claims 16 and 17 are amended to a network accessing method after being anticipated reasonably by the examiner, which are the same as the subject matter of claim 1.

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/082729

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102833258 A	19.12.2012	None	
CN 102054030 A	11.05.2011	None	
CN 102594825 A	18.07.2012	None	
CN 101958896 A	26.01.2011	None	
US 2005/0283835 A1	22.12.2005	US 7526810 B2	28.04.2009
CN 102571770 A	11.07.2012	None	
CN 102647417 A	22.08.2012	WO 2013143405 A1	03.10.2013

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/082729

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06 (2006.01) i

G06F 17/30 (2006.01) i

国际检索报告

国际申请号
PCT/CN2013/082729

A. 主题的分类

参见附加页

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)

IPC：H04L, G06F

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）)

CNABS, CNTXT, CNKI：网址，访问，恶意，钓鱼，挂马，木马，匹配，对照，密文

VEN：match, compare, website, access, hostile, malicious, Trojan, virus, phish

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
PX	CN 102833258 A (北京奇虎科技有限公司等) 19.12 月 2012 (19.12.2012) 全文	1-17
X	CN 102054030 A (惠州 TCL 移动通信有限公司) 11.5 月 2011 (11.05.2011) 说明书第[0038]-[0065]段及附图 1, 2	1, 15-17
Y	同上	2-14
X	CN 102594825 A (北京百度网讯科技有限公司) 18.7 月 2012 (18.07.2012) 说明书第[0123]-[0159]段及附图 4	1, 15-17
Y	同上	2-14
Y	CN 101958896 A (浙江大学) 26.1 月 2011 (26.01.2011)	2-14

☒ 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
14.11 月 2013 (14.11.2013)	**28.11 月 2013 (28.11.2013)**
ISA/CN 的名称和邮寄地址： 中华人民共和国国家知识产权局 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 喻文清 电话号码：(86-10) **62411450**

C(续). 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
	说明书第[0033]-[0058]段	
A	US 2005/0283835 A1 (LALONDE, C.等) 22.12 月 2005 (22.12.2005) 全文	1-17
A	CN 102571770 A (北京神州绿盟信息安全科技股份有限公司) 11.7 月 2012 (11.07.2012) 全文	1-17
A	CN 102647417 A (奇智软件(北京)有限公司) 22.8 月 2012 (22.08.2012) 全文	1-17

国际检索报告

国际申请号
PCT/CN2013/082729

第II栏　某些权利要求被认为是不能检索的意见(续第 1 页第 2 项)

根据条约第 17 条(2)(a)，对某些权利要求未做国际检索报告的理由如下：

1. ☒ 权利要求：16，17

因为它们涉及不要求本单位进行检索的主题，即：权利要求 16 所保护的主题是一种包括计算机可读代码的计算机程序，权利要求 17 所保护的主题是一种存储了权利要求 16 所述程序的计算机可读介质，上述主题属于细则 39.1 定义的不要求国际检索单位检索的主题之(6)。

（参见附加页）

2. ☐ 权利要求：

因为它们涉及国际申请中不符合规定的要求的部分，以致不能进行任何有意义的国际检索，

具体地说：

3. ☐ 权利要求：

因为它们是从属权利要求，并且没有按照细则 6.4(a)第 2 句和第 3 句的要求撰写。

第III栏　缺乏发明单一性的意见(续第 1 页第 3 项)

本国际检索单位在该国际申请中发现多项发明，即：

1. ☐ 由于申请人按时缴纳了被要求缴纳的全部附加检索费，本国际检索报告涉及全部可作检索的权利要求。

2. ☐ 由于无需付出有理由要求附加费的劳动即能对全部可检索的权利要求进行检索，本单位未通知缴纳任何附加费。

3. ☐ 由于申请人仅按时缴纳了部分被要求缴纳的附加检索费，本国际检索报告仅涉及已缴费的那些权利要求。具体地说，是权利要求：

4. ☐ 申请人未按时缴纳被要求缴纳的附加检索费。因此，本国际检索报告仅涉及权利要求书中首先提及的发明；包含该发明的权利要求是：

关于异议的说明：
☐ 申请人缴纳了附加检索费，同时提交了异议书，适用时，缴纳了异议费。
☐ 申请人缴纳了附加检索费，同时提交了异议书，但未在通知书规定的时间期限内缴纳异议费。
☐ 缴纳附加检索费时未提交异议书。

第II栏 某些权利要求被认为是不能检索的意见(续第 1 页第 2 项)

审查员对上述主题进行合理预期，将权利要求 16、17 的主题均修改为一种网址访问方法，上述合理预期后的主题与权利要求 1 的主题相同。

A. 主题的分类

H04L 29/06 (2006.01) i

G06F 17/30 (2006.01) i

国际检索报告

关于同族专利的信息

国际申请号
PCT/CN2013/082729

检索报告中引用的专利文件	公布日期	同族专利	公布日期
CN 102833258 A	19.12.2012	无	
CN 102054030 A	11.05.2011	无	
CN 102594825 A	18.07.2012	无	
CN 101958896 A	26.01.2011	无	
US 2005/0283835 A1	22.12.2005	US 7526810 B2	28.04.2009
CN 102571770 A	11.07.2012	无	
CN 102647417 A	22.08.2012	WO 2013143405 A1	03.10.2013