(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
2018 年 6 月 21 日 (21.06.2018)

WIPO | PCT

(10) 国际公布号
WO 2018/107784 A1

(51) 国际专利分类号:
H04L 29/06 (2006.01)

(21) 国际申请号: PCT/CN2017/096502

(22) 国际申请日: 2017 年 8 月 8 日 (08.08.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201611167905.3 2016年12月16日 (16.12.2016) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 蒋武 (JIANG, Wu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, 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, ST, 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))。

(54) **Title:** METHOD AND DEVICE FOR DETECTING WEBSHELL

(54) **发明名称:** 检测网页后门的方法和装置

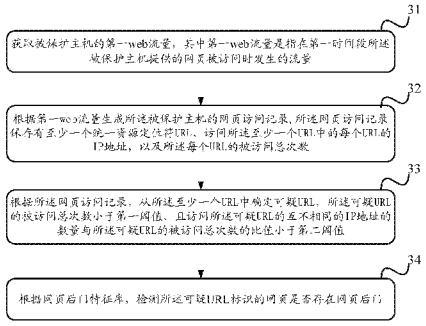

图 3

31 ACQUIRE FIRST WEB TRAFFIC OF A PROTECTED HOST, THE FIRST WEB TRAFFIC REFERRING TO THE TRAFFIC OCCURRING DURING A FIRST TIME PERIOD WHEN A WEBPAGE PROVIDED BY THE PROTECTED HOST HAS BEEN ACCESSED

32 GENERATE, ACCORDING TO THE FIRST WEB TRAFFIC, WEBPAGE ACCESS RECORDS OF THE PROTECTED HOST, THE WEBPAGE ACCESS RECORDS SAVING AT LEAST ONE UNIFORM RESOURCE LOCATOR (URL), AN IP ADDRESS HAVING ACCESSED EACH URL OF THE AT LEAST ONE URL, AND THE TOTAL NUMBER OF TIMES EACH URL HAS BEEN ACCESSED

33 DETERMINE, ACCORDING TO THE WEBPAGE ACCESS RECORDS, A SUSPICIOUS URL FROM THE AT LEAST ONE URL, THE TOTAL NUMBER OF TIMES THE SUSPICIOUS URL HAS BEEN ACCESSED BEING LESS THAN A FIRST THRESHOLD, AND THE RATIO OF THE NUMBER OF DIFFERENT IP ADDRESSES HAVING ACCESSED THE SUSPICIOUS URL TO THE TOTAL NUMBER OF TIMES THAT THE SUSPICIOUS URL HAS BEEN ACCESSED BEING LESS THAN A SECOND THRESHOLD

34 DETECT, ACCORDING TO A WEBSHELL FEATURE LIBRARY, WHETHER THE WEBPAGE OF THE SUSPICIOUS URL IDENTIFIER HAS A WEBSHELL

(57) Abstract: A method and a device for detecting a webshell, which are used for alleviating the problem in the prior art of low detection efficiency. The method comprises: acquiring first web traffic of a protected host; generating, according to the first web traffic, webpage access records of the protected host, the webpage access records being used for saving at least one uniform resource locator (URL), an IP address having accessed each URL of the at least one URL, and the total number of times each URL has been accessed, each URL identifying one webpage provided by the protected host; determining, according to the webpage access records, a suspicious URL from the at least one URL, the total number of times the suspicious URL has been accessed being less than a first threshold, and the ratio of the number of different IP addresses having accessed the suspicious URL to the total number of times that the suspicious URL has been accessed being less than a second threshold; and determining whether the webpage of the suspicious URL identifier comprises a shell feature, and detecting, according to the shell feature determination result, whether the webpage of the suspicious URL identifier has a webshell.

WO 2018/107784 A1

[见续页]

(57) 摘要：一种检测网页后门的方法和装置，用以缓解现有技术检测效率低的问题。该方法包括：获取被保护主机的第一web流量；根据第一web流量生成被保护主机的网页访问记录，所述网页访问记录用于保存至少一个统一资源定位符URL、访问所述至少一个URL中的每个URL的IP地址、以及所述每个URL的被访问总次数，其中每个URL标识所述被保护主机提供的一个网页；根据网页访问记录，从至少一个URL中确定可疑URL，所述可疑URL的被访问总次数小于第一阈值、且访问所述可疑URL的互不相同的IP地址的数量与所述可疑URL的被访问总次数的比值小于第二阈值；以及确定可疑URL标识的网页是否包含后门特征，根据后门特征确定结果检测所述可疑URL标识的网页是否存在网页后门。

检测网页后门的方法和装置

本申请要求于 2016 年 12 月 16 日提交中国专利局、申请号为 201611167905.3、申请名称为“检测网页后门的方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及网络安全技术领域，尤其涉及一种检测网页后门的方法及一种检测网页后门的装置。

背景技术

网页后门（webshell）是一种以网页文件形式存在的后门工具。通过 webshell 可以获得网站的操作权限，例如上传下载文件、查看数据库、执行脚本命令等。Webshell 文件可以是使用动态服务器页面（英文：Active Server Page，ASP）应用编写的网页文件，或使用超文本预处理器（英文：Hypertext Preprocessor，PHP）语言编写的网页文件，或通用网关界面（英文：Common Gateway Interface，CGI）程序文件。

网络中提供网页服务、开放网页服务相关端口的主机也被称为网站服务器、或者 web 服务器。网站服务器往往会成为 webshell 的攻击目标。攻击者利用开放端口等漏洞成功入侵网站服务器后，将 webshell 文件存放于该网站服务器的网页目录中，与正常网页文件混在一起。此后，攻击者可以通过浏览器访问存放于上述网页服务器的 webshell 文件以获得对于网站服务器的操作权限，从而达到控制网站服务器、盗取信息等非法目的。由于攻击者与被攻击网站服务器之间的数据通常是通过网页服务的默认端口 80 端口来传输的，而防火墙为了不影响网络用户的正常网页访问行为通常不会阻止访问 80 端口的超文本传输协议（英文：HyperText Transfer Protocol，HTTP）流量，因此简单的报文过滤方式并不能阻止上述攻击行为。

为了检测网页后门，现有技术通过人工分析 webshell 文件的代码、或者分析攻击者访问 webshell 文件时产生的流量获取 webshell 的特征，形成 webshell 特征库。安全设备获得 web 流量后，将 web 流量与 webshell 特征库中的特征进行匹配，来实现检测 webshell 的目的。然而由于现有网络中 web 流量的数据量巨大，导致耗费安全设备大量处理资源，检测效率较低。

发明内容

本申请实施例提供一种检测网页后门的方法，用以缓解现有技术检测效率低的问题。

本申请实施例提供的技术方案如下：

第一方面，提供了一种检测网页后门的方法，包括：获取被保护主机的第一 web 流量，所述第一 web 流量是指在第一时间段中所述被保护主机提供的网页被访问时发

生的流量；根据所述第一 web 流量生成所述被保护主机的网页访问记录，所述网页访问记录用于保存至少一个统一资源定位符（英文：Uniform Resource Locator，URL）、访问所述至少一个 URL 中的每个 URL 的 IP 地址、以及所述每个 URL 的被访问总次数，其中所述每个 URL 标识所述被保护主机提供的一个网页；根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，所述可疑 URL 的被访问总次数小于第一阈值、且访问所述可疑URL 的互不相同的 IP 地址的数量与所述可疑URL 的被访问总次数的比值小于第二阈值；以及确定所述可疑 URL 标识的网页是否包含网页后门特征库中的后门特征，根据后门特征确定结果检测所述可疑 URL 标识的网页是否存在网页后门。

本申请实施例基于已发生的被保护主机的 web 流量，构建能够反映被保护主机中各个网页被访问的次数、访问者 IP 分布等情况的网页访问记录。进一步根据该网页访问记录从被保护主机提供的多个网页 URL 中识别可疑程度较高的 URL，后续着重对可疑 URL 标识的网页进行检测，而无需对所有网页都进行网页后门检测。上述方法减少了需要进行网页后门检测的网页的数量，从而提高了 web 检测性能。

可选的，本申请还提供了网页访问记录的第一种具体结构，以及如何构建网页访问记录的详细步骤。通过这种结构的网页访问记录可以快捷地确定出可疑 URL。即，

在第一方面的第一种可能的实现方式中，所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述每个表项中保存有被访问总次数和 IP 地址列表；

所述第一 web 流量生成所述被保护主机的网页访问记录，包括：

从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址；

从所述至少一个访问请求报文中选择一个访问请求报文，对选择出的访问请求报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

解析选择出的访问请求报文，从而获得所述选择出的访问请求报文的源 IP 地址和携带的 URL；

在所述网页访问记录中查找所述选择出的访问请求报文携带的 URL 对应的表项；

如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1，在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；

如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页访问记录中创建所述选择出的访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

结合第一方面的第一种可能的实现方式，在第一方面的第二种可能的实现方式中所述根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，包括：

从所述网页访问记录中选择出一个表项；

确定选择出的表项的 IP 地址列表中互不相同的 IP 地址的数量；

如果所述选择出的表项的被访问总次数少于所述第一阈值、且确定出的互不相同的 IP 地址的数量与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL。

可选的，本申请还提供了网页访问记录的第二种具体结构，以及如何构建网页访

问记录的详细步骤。第二种具体结构在第一种具体结构的表项的基础上增加了 IP 地址计数值这一信息，通过这种结构的网页访问记录可以快捷地确定出可疑 URL。即，

在第一方面的第三种可能的实现方式中，所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述表项中保存有被访问总次数、IP 地址计数值和 IP 地址列表；

所述第一 web 流量生成所述被保护主机的网页访问记录，包括：

从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址；

从所述至少一个访问请求报文中选择出一个访问请求报文，对选择出的访问请求报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

获取所述选择出的访问请求报文的源 IP 地址和携带的 URL；

在所述网页访问记录查找所述选择出的访问请求报文携带的 URL 对应的表项；

如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1；确定所述查找到表项的 IP 地址列表中是否已保存所述源 IP 地址，如果所述查找到表项的 IP 地址列表中已保存所述源 IP 地址，则对所述选择出的访问请求报文处理结束；如果所述查找到的表项的 IP 地址列表中未保存所述源 IP 地址，则将所述查找到的表项的 IP 地址计数值加 1，并在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；

如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页访问记录中创建所述访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，将所述创建的表项的 IP 地址计数值设置为 1，并在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

结合第一方面的第三种可能的实现方式，在第一方面的第四种可能的实现方式中，所述根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，包括：

从所述网页访问记录中选择出一个表项；

如果选择出的表项的被访问总次数少于所述第一阈值、且所述选择出的表项的 IP 地址计数值与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL。

终端通过浏览器访问网页时，这一访问过程有可能并未成功。记录访问失败页面对应的表项将占用存储空间，后续对访问失败页面进行检测也会浪费处理资源。为了节约存储资源和处理资源，一种可能的实现方式是仅记录访问成功页面对应的表项，具体如下。

结合第一方面的第一种或第三种可能的实现方式，在第一方面的第五种实现方式中，从所述第一 web 流量中获得至少一个访问请求报文包括：

从所述第一 web 流量中选择至少一个访问应答报文，所述至少一个访问应答报文中的每个访问应答报文携带的状态码指示访问成功，所述每个访问应答报文的源地址为所述被保护主机的 IP 地址；

从所述第一 web 流量中获取所述每个网页访问应答报文分别对应的访问请求报文，

作为获得的所述至少一个访问请求报文。

终端通过安装的浏览器访问被保护主机提供的网页是，由于浏览器提供商、浏览器版本的差异，有可能造成不同浏览器访问网站服务器提供的同一网页时，产生的多个访问请求报文中携带不同的 URL。如果安全设备据此生成不同 URL 对应的表项，一方面与这些访问请求报文实际上访问的是同一网页这一实际情况不符，造成后续可疑 URL 识别时的偏差，另一方面会造成网页访问记录数据量过大。为了提高可疑 URL 识别的准确性，节约网页访问记录在存储器中占有的存储空间，安全设备在生成网页访问记录中的表项时，可以先对访问请求报文中的 URL 进行正规化处理，根据正规化处理后的 URL 生成表项。具体如下，

结合第一方面的第一种或第三种可能的实现方式，在第一方面的第六种实现方式中，在所述网页访问记录中查找所述选择出的访问请求报文携带的 URL 对应的表项，包括：

对所述选择出的访问请求报文携带的 URL 执行至少一种正规化处理，得到正规化处理后的 URL，所述正规化处理包括以下（1）～（3）中的一种或多种：（1）将所述选择出的访问请求报文携带的 URL 转换为预定编码格式，（2）将所述选择出的访问请求报文携带的 URL 中的字符转换为预定大小写类型，和（3）去除所述选择出的访问请求报文携带的 URL 中参数；

在所述网页访问记录中查找正规化处理后的 URL 对应的表项；

相应地，在所述网页访问记录中创建所述访问请求报文携带的 URL 对应的表项，具体为：

在所述网页访问记录中创建所述正规化处理后的 URL 对应的表项。

为了进一步降低网页访问列表占用的存储资源，可以对网页访问列表中记录的信息进行进一步精简，删除一些对识别可疑 URL 所用不大的信息。例如可以识别正常 URL 后删除并不再维护正常 URL 对应的表项中的被访问总次数和访问正常 URL 的 IP 地址，从而节省存储资源和后续更新表项耗费的处理资源。即，在第一方面的第七种可能的实现方式中，所述方法还包括：

根据所述网页访问记录，从所述至少一个 URL 中确定正常 URL，所述正常 URL 是所述至少一个 URL 中的被访问总次数大于所述第一阈值的 URL，或者网页后门检测结果指示所标识的网页不存在网页后门的可疑 URL；

删除所述网页访问记录中保存的访问所述正常 URL 的 IP 地址和所述正常 URL 的被访问总次数。

结合第一方面的第七种可能的实现方式，在第一方面的第八种可能的实现方式中，所述方式还包括：

获取所述被保护主机的第二 web 流量，所述第二 web 流量是指在所述第一时间段之后的第二时间段中所述被保护主机提供的网页被访问时发生的流量；

从所述第二 web 流量中获得第一访问请求报文、第二访问请求报文和第三访问请求报文；

解析所述第一访问请求报文，从而获得所述第一访问请求报文的源 IP 地址和携带的 URL；如果所述第一访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访

问记录中已保存所述第一访问请求报文携带的 URL，则将已保存的所述第一访问请求报文携带的URL 的被访问总次数加 1，在访问所述第一访问请求报文携带的URL 的 IP 地址中增加所述第一访问请求报文的源 IP 地址；

解析所述第二访问请求报文，从而获得所述第二访问请求报文的源 IP 地址和携带的 URL；如果所述第二访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访问记录中未保存所述第二访问请求报文携带的 URL，则在所述访问记录中保存所述第二访问请求报文携带的 URL，设置所述第二访问请求报文携带的 URL 的被访问总次数为 1，设置访问所述第二访问请求报文携带的URL 的 IP 地址为所述第二访问请求报文的源 IP 地址；

解析所述第三访问请求报文，从而获得所述第三访问请求报文携带的 URL；如果所述第三访问请求报文携带的URL 与所述正常 URL 相同，对所述第三访问请求的处理结束。

第二方面，提供了一种检测网页后门的装置，该装置具有实现上述第一方面所述方法或上述方面的任意一种可能的实现方式的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。所述硬件或软件包括一个或多个与上述功能相对应的模块。

第三方面，本申请实施例提供了一种计算机存储介质，用于储存为上述报文转发设备所用的计算机软件指令，其包含用于执行上述第一方面或上述方面的任意一种可能的实现方式所设计的程序。

附图说明

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一简单地介绍，显而易见地，下面描述中的附图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的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请实施例提供的检测网页后门的方法的应用场景示意图；

图 2 为本申请实施例提供的安全设备的结构示意图；

图 3 为本申请实施例提供的检测网页后门的方法的流程图；

图 4 为本申请实施例提供的哈希表的结构示意图；

图 5 为本申请实施例提供的根据第一 web 流量构建网页访问记录的方法的流程图；

图 6 为本申请实施例提供的一个表项的实例图；

图 7 为本申请实施例提供的另一种哈希表的结构示意图；

图 8 为本申请实施例提供的检测网页后门的方法的另一流程图；

图 9 为本申请实施例提供的安全设备处理三个访问请求报文之前网页访问记录的示意图；

图 10 为本申请实施例提供的安全设备处理三个访问请求报文之后网页访问记录的示意图；

图 11 为本申请实施例提供的检测网页后门的装置的结构示意图。

具体实施方式

下面将结合各个附图对本发明技术方案的实现原理、具体实施方式及其对应能够达到的有益效果进行详细的阐述。

终端使用浏览器访问网页这一行为产生的浏览器和网站服务器之间一系列的交互报文被称为 web 流量。随着网络中信息的爆炸性增长，一方面网站服务器往往会存储着以千百万计的网页文件，另一方面终端用户频繁进行网页访问活动，导致 web 流量急速增长。现有以防火墙、深度报文检测（英文：Deep Packet Inspection，DPI）等为例的安全设备受性能的制约，难以对接收到的 web 流量所承载的所有网页数据进行逐一检测，这也成为现有 web 安全技术的难点之一。

现有 web 检测性能不高的主要原因之一是由于待检测的网页数量巨大，对此本申请实施例提供了一种检测网页后门的方法。该方法基于已发生的被保护主机的 web 流量，构建能够反映被保护主机中各个网页被访问的次数、访问者 IP 分布等情况的网页访问记录。进一步根据该网页访问记录从被保护主机提供的所有网页的统一资源定位符（英文：Uniform Resource Locator，URL）中识别可疑程度较高的 URL，后续着重对可疑 URL 标识的网页进行检测，而无需对所有网页都进行网页后门检测。上述方法减少了待检测网页的数量，从而提高了 web 检测性能。

下面结合各个附图对本申请实施例技术方案的主要实现原理、具体实施方式及其对应能够达到的有益效果进行详细的阐述。

附图 1 为本申请实施例应用场景示意图。网络系统中包括网站服务器 11、安全设备 12、和多个终端 13。其中网站服务器 11 是被保护主机的一个示例。在本发明实施例中，被保护主机是指能提供网页服务的主机。在主机中安装 Apache 或微软公司的互联网信息服务（英文：Internet Information Services，IIS)应用软件后，主机可以作为网站服务器向网络中的其他用户提供网页服务。

终端 13 在本申请实施例中是指具有网页访问功能的终端设备，例如安装有浏览器的个人计算机、智能手机或者便携手计算机等等。浏览器是一种用于检索并展示互联网信息资源的应用程序。当前常用的浏览器包括 Internet Explorer、Mozilla Firefox、谷歌公司的 Chrome 等等。终端 13 可以位于局域网中，通过网络地址转换（英文：Network Address Translation，NAT）设备访问互联网中的网站服务器 11。终端 13 也可以直接通过公有 IP 地址直接访问互联网中的网站服务器 11。

安全设备 12 获取终端 13 访问网站服务器 11 时产生的 web 流量。如图 1 所示，安全设备 12 设置于终端 13 与网站服务器 11 之间的通信路径上，访问网站服务器 11 的流量都经由安全设备 12 转发给网站服务器。例如，安全设备 12 是设置于网站服务器 11 之前的防火墙，网站服务器 11 通过防火墙接入网络。在这种部署方式下，安全设备 12 保存流经安全设备 12 访问网站服务器 11 的 web 流量。安全设备 12 也可以以旁路方式部署，图 1 中未示出，例如网站服务器 11 通过网关设备 14 接入网络，安全设备 12 是与网关设备 14 相连的 DPI 设备。网关设备 14 对终端 13 访问网站服务器 11 的流量进行镜像处理，再将镜像处理得到的镜像流量发送给 DPI 设备。本申请实施例对安全设备 12 的具体部署方式不做限定，只要安全设备 12 能够获得终端 13 访问网站服务器 11 的 web 流量即可。

由于真实网络环境往往比较复杂，安全设备 12 可以参与其他网络设备的流量转发过程。在这种情况下，可以在安全设备 12 中预先存储一个或多个被保护主机的 IP 地址。安全设备 12 根据预先存储的被保护主机的 IP 地址结合 web 访问相关的协议类型，例如 HTTP，从获得的所有流量中筛选出被保护主机提供的网页被访问时发生的流量。

采用本申请实施例提供的方法对多个被保护主机提供的网页进行检测。为了描述简明，本申请实施例主要仅以被保护主机为一个网站服务器为例进行说明，对于多个被保护主机情况可以执行相类似的处理。

附图 2 是本申请实施例提供的安全设备的结构示意图。安全设备可以是附图 1 中的安全设备 12。安全设备包括处理器 210、存储器 220、网络接口 230、输入设备 240、显示器 250 和总线 260。其中处理器 210、存储器 220 以及网络接口 230、输入设备 240 和显示器 250 通过总线 304 相互连接。

处理器 210 可以是一个或多个中央处理器(英文:Central Processing Unit，CPU)，在处理器 210 是一个 CPU 的情况下，该 CPU 可以是单核 CPU，也可以是多核 CPU。

存储器 220 包括但不限于是随机存取存储器 (RAM)、只读存储器 (ROM)、可擦除可编程只读存储器(EPROM 或者快闪存储器)、或便携式只读存储器(CD-ROM)。

所述网络接口 230 用于可以是有线接口，例如光纤分布式数据接口（英文：Fiber Distributed Data Interface， FDDI）、千兆以太网（英文：Gigabit Ethernet， GE）接口；网络接口 230 也可以是无线接口。

处理器 210 用于读取存储器 220 中存储的程序代码 222，运行后执行以下操作。

具体地，处理器 210 通过网络接口 230 获取被保护主机的第一 web 流量，其中被保护主机的第一 web 流量是指在第一时间段所述被保护主机提供的网页被访问时发生的流量。为了区分不同阶段获取的 web 流量，本申请实施例将生成网页访问记录时所依据的 web 流量称为第一 web 流量。将生成网页访问记录后，接收到的 web 流量称为第二 web 流量。第二 web 流量可以用于更新网页访问记录。

处理器 210 通过所述第一 web 流量生成所述被保护主机的网页访问记录 221，其中网页访问记录保存至少一个 URL、访问所述至少一个 URL 中的每个 URL 的 IP 地址，以及所述每个 URL 的被访问总次数，其中所述每个 URL 标识所述被保护主机提供的一个网页。处理器 210 将生成的网页访问记录 221 存储于存储器 220 中。

处理器 210 根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，所述可疑 URL 的被访问总次数小于第一阈值、且访问所述可疑 URL 的互不相同的 IP 地址的数量与所述可疑 URL 的被访问总次数的比值小于第二阈值。处理器 210 根据存储器 220 中的网页后门特征库，检测所述可疑 URL 标识的网页是否存在网页后门。

由于仅有攻击者知晓 webshell 文件在网站服务器的网页目录中的存放位置，正常用户并不知晓 webshell 文件的存放位置，因此往往只有攻击者访问 webshell 文件，正常用户通常不会访问 webshell 文件。与此相比，网站服务器向公众提供的正常网页文件被大量正常用户频繁访问。因此 webshell 文件的访问分布情况与正常网页文件的访问分布情况有很大差异。正常网页文件具有被访问频率高、访问者 IP 分布广的特点，而 webshall 文件具有访问频率低、访问者 IP 较为单一的特点。当然，攻击者可以通过设置代理服务器、伪造 IP 地址等方式在一定程度上逃避监测。因此，本申请根据访

问行为的差异识别出可疑 URL，再进一步对可疑 URL 标识的网页进行检测。

本申请实施例中安全设备构建能够反映被保护主机中各个网页被访问的次数、访问者 IP 分布等情况的网页访问记录，从被保护主机提供的所有网页的 URL 中识别可疑程度较高的 URL，后续着重对可疑 URL 标识的网页进行检测，而不用对所有网页都进行检测。由于减少了待检测网页的数量，从而提高了 web 检测性能。

下面结合各个流程图，对本申请提供的检测网页后门的方法进行详细描述。

附图 3 是本申请实施例提供的检测网页后门的方法的原理流程图。该方法可以由附图 1 中的安全设备 12 执行。

步骤 31，获取被保护主机的第一 web 流量，其中第一 web 流量是指在第一时间段被保护主机提供的网页被访问时发生的流量。

安全设备中预先存储有被保护主机的 IP 地址。采用直路部署的情况下，安全设备接入网络后，将流经所述安全设备的报文的源地址或目的地址与被保护主机的 IP 地址进行比较，如果报文的源地址或目的地址与被保护主机的 IP 地址相同、且协议类型为 HTTP，则保存报文，从而获得被保护主机的第一 web 流量。采用旁路部署的情况下，安全设备将网关设备发来的镜像流量中的报文的源地址或目的地址与被保护主机的 IP 地址进行比较。如果报文的源地址或目的地址与被保护主机的 IP 地址相同、且协议类型为 HTTP，则保存报文；如果报文的源地址或目的地址与被保护主机的 IP 地址不同，或者协议类型与 web 访问无关，则删除报文，从而节省存储空间。

步骤 32，根据第一 web 流量生成所述被保护主机的网页访问记录。网页访问记录用于保存以下信息：至少一个 URL、访问所述至少一个 URL 中的每个 URL 的 IP 地址，以及所述每个 URL 的被访问总次数。其中所述每个 URL 标识所述被保护主机提供的一个网页。

具体地，网页访问记录中包含多个表项，每个表项与所述至少一个 URL 中的一个 URL 对应。每个表项不仅保存对应的 URL，还保存该表项对应的 URL 被访问的总次数，以及访问该表项对应的 URL 的 IP 地址。

安全设备可以采用多种不同的数据结构，例如多维数组、哈希表等来组织网页访问记录中的多个表项。

为了便于查找和更新存储的信息，本申请实施例提供了一种哈希表来保存上述网页访问记录。如图 4 所示，具体采用哈希桶来实现哈希表。每个被保护主机的 IP 地址对应一个哈希桶（Bucket）表。例如本实施例中每个被保护主机的 IP 地址用 41 表示，哈希桶表用 42 表示，每个地址 41 分别对应的哈希桶表 42 包括 256 个哈希桶。

哈希桶表 42 中的每个哈希桶是哈希表内表项的虚拟子群组。每个哈希桶对应一个由表项组成的长度不等的链表。在图 4 中链表用 43 表示，表项用 44 表示。链表 43 中存储有 0 个，1 个或多个表项 44。每个表项包括索引键和值。每个表项的索引键是对 URL 进行哈希运算得到的结果，值为 URL 本身，还保存有用于记录访问该 URL 的总次数的访问总次数 CountVisit，以及用于记录访问该 URL 的 IP 地址列表 IP List 等等信息。哈希算法包括信息摘要算法 5（Message-Digest Algorithm 5，MD5）。

在后续其他实施例中，将结合附图 5 至附图 7 介绍构建附图 4 所示的哈希表的详细过程。

步骤 33，根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，所述可疑 URL 的被访问总次数小于第一阈值、且访问所述可疑 URL 的互不相同的 IP 地址的数量与所述可疑 URL 的被访问总次数的比值小于第二阈值。安全设备中预先保存有第一阈值和第二阈值，其中第一阈值和第二阈值可以是网络管理人员根据经验和实际网络环境设定并通过附图 2 中的输入设备 240 输入安全设备的，也可以是根据预先标定的 web 流量样本，通过机器学习的方式获得的，本实施例对此不进行限定。

可选地，安全设备定期根据第一阈值、第二阈值对附图 4 所示的哈希表中表项存储的信息进行判别，从而识别可疑 URL。第一阈值为自然数、取值范围可以根据经验、存储器的存储空间和判别周期设定。随着判别周期越长，存储空间越大，第一阈值的取值范围也可以适当增大，从而获得更准确的识别效果。具体取值可以根据实际情况灵活设定。例如判别周期为 10 天，第一阈值的取值为 1000。

第二阈值为 0 到 1 之间的百分数。第二阈值的取值也可以根据经验和实际网络环境设定。第二阈值的取值越小，识别出的可疑 URL 误报率越低，但是会有一定的漏报率。第二阈值的取值越大，识别出的可疑 URL 误报率越高，漏报率将会降低。例如，第二阈值可以取 50%。

步骤 34，确定所述可疑 URL 标识的网页是否包含网页后门特征库中的后门特征，根据确定结果检测所述可疑 URL 标识的网页是否存在网页后门。

通常在网页访问过程中，浏览器先通过基于传输控制协议（英文：Transmission Control Protocol ，TCP）与网站服务器建立连接。然后通过建立的连接向网站服务器发送访问请求报文，例如 HTTP request GET 报文、HTTP request Post 报文。访问请求报文携带待访问页面的 URL。

网站服务器接收到访问请求报文后，根据访问请求报文中携带的 URL，从网页目录中查找到对应的网页文件。网站服务器根据查找结果向浏览器发送访问应答报文，例如 HTTP request Response 报文。访问应答报文中携带状态码，例如 HTTP 1.1 版本中定义了 5 类状态码， 状态码由三位数字组成，第一个数字定义了响应的类别，具体地

1XX 提示信息 - 表示请求已被成功接收，继续处理；

2XX 成功 - 表示请求已被成功接收，理解，接受；

3XX 重定向 - 要完成请求必须进行更进一步的处理；

4XX 客户端错误 - 请求有语法错误或请求无法实现；

5XX 服务器端错误 - 服务器未能实现合法的请求。

如果状态码指示访问成功，网站服务器根据查找到的网页文件的数据量大小，将网页文件通过一个或多个响应报文发送给浏览器。

安全设备通过步骤 31~步骤 33 得到可疑 URL 后，可以进一步得到可疑 URL 所标识的网页被访问时浏览器与网站服务器交互的报文。然后安全设备可以通过基于报文的检测方式和基于数据流的检测方式，根据网页后门特征库，检测上述交互报文承载的网页是否存在网页后门。

具体地，安全设备可以通过以下方式获取可疑 URL 所标识的网页被访问时浏览器与网站服务器交互的报文。

方式一

安全设备从保存的被保护主机的第一 web 流量中查找到终端访问可疑 URL 所标识的网页时产生的交互报文。例如，安全设备根据 HTTP 协议的相关标准，对第一 web 流量中的一个访问请求报文进行解析，从而得到该访问请求报文中携带的信息为：

Internet Protocol Version 4, Src: 219.133.94.158, Dst: 10.1.1.34

Transmission Control Protocol, Src Port: 1272(1272), Dst Port: 80(80), Seq:1, Ack:1, Len:89

Hypertext Transfer Protocol

GET http://www.google.com.hk/videohp HTTP/1.1

Accpet-Language: en-us

UA-CPU: X86

Accept-Encoding: gzip, deflate

User-Agent: Mozilla/4.0

Host: www.google.com.hk

Connection: Keep-Alive

Cache-Control: no-cache

安全设备得到访问请求报文携带的 URL 是 GET 关键字后面的 www.google.com.hk/videohp。安全设备将得到的 URL 与可疑 URL 进行比较，若访问请求报文携带的 URL 与可疑 URL 一致，则根据该访问请求报文的源地址、目的地址、源端口、目的端口、协议类型、序列号、时间戳等信息，从第一 web 流量中获得该访问请求报文所属数据流的所有报文，得到的报文即为访问可疑 URL 所标识的网页时，浏览器与网站服务器交互的报文。

方式二

安全设备通过该安全设备上安装的浏览器访问可疑 URL 所标识的页面，保存该过程中与网站服务器交互产生的一系列报文，从而得到访问可疑 URL 所标识的网页时，浏览器与网站服务器交互的报文。

在采用基于报文的检测方式的情况下，安全设备将得到的访问可疑 URL 所标识的网页时，浏览器与网站服务器交互的每个报文与网页后门特征库中的特征进行匹配，如果匹配命中的特征满足预设规则，例如匹配命中的特征超过预定数量，则确认可疑 URL 所标识的网页存在网页后门。在实施过程中，可以预先根据网页后门特征库中的特征生成多模式匹配状态机，将单个报文的内容输入状态机，通过一次扫描即可找到该报文匹配的所有特征，从而提高了检测性能。

在采用基于数据流的检测方式的情况下，安全设备得到访问可疑 URL 所标识的网页时，浏览器与网站服务器交互的各个报文后，对报文进行流重组从而得到数据流的载荷内容，将载荷内容与网页后门特征库中的特征进行匹配。根据匹配命中结果以及预定的网页后门识别规则，检测所述可疑 URL 标识的网页是否存在网页后门。预定的网页后门识别规则包括如果匹配命中的特征中先后出现特征 A、B、C，则确认可疑 URL 所标识的网页存在网页后门；或者，如果匹配命中的特征超过 3 个，则确认可疑 URL 所标识的网页存在网页后门。

附图 5 是本申请实施例提供的根据第一 web 流量构建网页访问记录的方法的流程图。

步骤 51，安全设备对第一 web 流量进行协议解析，得到第一 web 流量中的至少一个访问请求报文。在本实施例中，访问请求报文是指浏览器向网站服务器发送的 HTTP request GET 报文。HTTP request GET 报文的目的 IP 地址为所述被保护主机的 IP 地址。安全设备对至少一个访问请求报文中的每个访问请求报文执行步骤 52~58，直到处理完所有访问请求报文为止。具体地安全设备可以按照预设的选择规则，从至少一个访问请求报文中逐个选取访问请求报文，例如按照时间先后顺序，根据访问请求报文携带的时间戳，依次选取访问请求报文。

步骤 52~510 以一个访问请求报文为例，对处理过程进行详细说明。

步骤 52，安全设备通过协议解析获得该访问请求报文的目的 IP 地址、源地址和携带的 URL。

步骤 53，安全设备根据目的 IP 地址在网页访问记录中查找该目的 IP 地址对应的记录。即判断在网页访问记录中是否已记录有该目的 IP 地址、以及该目的 IP 地址对应的哈希桶表。如果网页访问记录中未记录该目的地址，则执行步骤 54；如果网页访问记录中已记录该目的地址，则执行步骤 55。

步骤 54，安全设备记录该目的 IP 地址，并创建该目的 IP 对应的哈希桶表。进一步执行步骤 56。

具体地，安全设备在网页访问记录中记录目的 IP 地址，创建该目的 IP 地址对应的包含 256 个哈希桶的哈希桶表。初始时，哈希桶表中的每个哈希桶对应的链表为空。

步骤 56，安全设备根据预定的哈希桶散列算法，对该访问请求报文中携带的 URL 进行计算，确定该访问请求报文中携带的 URL 所属的哈希桶。进一步执行步骤 57。

步骤 57，安全设备在确定出的哈希桶中创建一个表项。所创建的表项的索引键是对该访问请求报文中携带的 URL 进行哈希运算得到的结果，将该 URL 记录在创建的表项中。并且设置该创建的表项中保存的访问总次数为 1，在该表项的 IP 地址列表中记录步骤 52 解析得到的源地址。

步骤 55，安全设备根据预定的哈希桶散列算法，对该访问请求报文中携带的 URL 进行计算，确定该访问请求报文中携带的 URL 所属的哈希桶。进一步执行步骤 58。

步骤 58，安全设备在确定出的哈希桶对应链表中查找该 URL 对应的表项。

安全设备对该 URL 进行哈希运算，在查找到的哈希桶对应的链表中查找以哈希在运算结果为索引的表项。如果不存在以哈希在运算结果为索引的表项，则执行步骤 59。如果存在以哈希在运算结果为索引的表项，则执行步骤 510。

步骤 59，安全设备创建以哈希运算结果为索引的表项，在创建的表项中记录该 URL，在在该表项的 IP 地址列表中记录该访问请求报文中携带的源地址，设置创建的表项中的访问总次数为 1。

步骤 510，安全设备在以哈希运算结果为索引的表项的 IP 地址列表中记录该访问请求报文中携带的源地址，将该以哈希运算结果为索引的表项中保存的访问总次数加 1。

例如，安全设备通过协议解析获得第一 web 流量中的一个访问请求报文中携带的

目的IP地址为10.1.1.34，源地址为219.133.94.158，URL为www.google.com.hk/videohp。其中目的地址10.1.1.34与被保护主机的IP地址相同。

安全设备中预设的哈希算法为32位MD5算法，即输入为任意长度的URL，输出为32位16进制符号。本实例中对www.google.com.hk/videohp执行哈希运算的结果为a356bf63af5c8b348032bba8b44eceda。

哈希桶散列算法的目的是将任意一个哈希结果划归到256个哈希桶中的一个哈希桶中。在本实例中哈希桶散列算法具体是将哈希运算结果依次划分为16组，每组2位，依次执行相与运算，最终得到两个16进制符号；然后将两个16进制符号对256取余，将取余结果作为哈希桶的序号。

例如，a3|56|bf|63|af|5c|8b|34|80|32|bb|a8|b4|4e|ce|da=ab，ab%256=163，确认www.google.com.hk/videohp属于哈希桶163。

在哈希桶163中查找索引键为a356bf63af5c8b348032bba8b44eceda的表项。在本实例中假设哈希桶163中不存在索引键为

a356bf63af5c8b348032bba8b44eceda的表项，则安全设备在哈希桶163对应的链表的末尾新建索引键为a356bf63af5c8b348032bba8b44eceda的表项，或者按照预定规则插入链表的预定位置。在该表项中记录www.google.com.hk/videohp，在新建表项的IP地址列表中该访问请求报文中携带的源地址219.133.94.158，将创建表项中的访问总次数设置为1。经过上述处理创建的表项如图6所示。

相应地，采用附图5所示的方法构建出网页访问记录后，附图3的步骤33在确定每个表项对应的URL是否是可疑表项时，首先获取该表项中的IP地址列表IP List，从中确定出互不相同的IP地址，计算互不相同的IP地址的数量。然后取出被访问总次数CountVisit。如果被访问总次数CountVisit的值小于第一阈值、且计算出的互不相同的IP地址的数量与被访问总次数CountVisit的值的比值小于第二阈值，则确定该URL对应的URL是为可疑URL。

为了提高识别可疑URL的效率，还可以对附图4所示的表项44的数据结构进行改进，增加一项IP地址计数值Count IP，IP地址计数值用于记录访问该URL的互不相同的IP地址的数量。并且在IP地址列表IP Lisit中仅记录互不相同的IP地址，如附图7所示。

相应地，附图5所示的构建网页访问记录的方法也需要进行适应性调整。具体地，在步骤57或者步骤59中，如果未查找到访问请求报文携带的URL对应的表项，则在所述网页访问记录中创建所述访问请求报文携带的URL对应的表项，将创建的表项的被访问总次数设置为1，将所述创建的表项的IP地址计数值设置为1，并在所述创建的表项的IP地址列表中记录该访问请求报文的源IP地址。

在步骤510中，如果查找到访问请求报文携带的URL对应的表项，则将查找到的表项的被访问总次数加1。需要进一步确定所述查找到表项的IP地址列表中是否已保存该访问请求报文的源IP地址，如果所述查找到表项的IP地址列表中已保存该访问请求报文的源IP地址，则对所述访问请求报文处理结束。如果所述查找到的表项的IP地址列表中未保存该访问请求报文的源IP地址，则将所述查找到的表项的IP地址计数值加1，并在所述查找到的表项的IP地址列表中记录该访问请求报文的源IP地

址。

通过上述改进，在附图 3 的步骤 33 中，在确定每个表项对应的 URL 是否是可疑表项时，只需要取出被访问总次数 CountVisit 和 IP 地址计数值 CountIP，就可以简便地确认该 URL 对应的 URL 是否为可疑 URL。具体地，如果被访问总次数 CountVisit 的值小于第一阈值、且 IP 地址计数值 CountIP 的值与被访问总次数 CountVisit 的值的比值小于第二阈值，则确定该 URL 对应的 URL 是为可疑 URL。

可选地，终端通过浏览器访问网页时，这一访问过程有可能并未成功。对于攻击者来说，如果访问 webshell 文件失败，将无法攻击成功。如果安全设备对这些访问失败的页面进行检测将没有实际意义，因为在附图 3 的步骤 34 中无法得到浏览器与网站服务器交互的报文。为了避免后续对访问失败页面进行检测可能浪费处理资源、以及在网页访问记录中保存访问失败页面的 URL 对应表项浪费存储空间，在附图 5~附图 7 所示的方法构建网页访问记录的过程中，在步骤 51 从第一 web 流量中获取到的至少一个访问请求报文时可以进行如下改进。

安全设备首先从第一 web 流量中选择至少一个访问应答报文，其中选中的每个网页访问应答报文携带的状态码指示访问成功。访问应答报文是网站服务器接收到访问请求报文后，向浏览器返回的报文。本申请仅考虑源地址为所述被保护主机的 IP 地址的访问应答报文的。

例如，访问成功的访问应答报文解析后的内容如下

```
HTTP/1.1 200 OK
Date: Wed, 10 Jun 2009 11:22:58 GMT
Server: Microsoft-IIS/6.0
X-Powered-By: ASP.NET
Content-Length: 4218
Content-Type: text/html
Cache-control: private
```

其中状态码“200 OK” 指示访问成功。

此后，安全设备根据各个报文携带的源地址、源端口、目的地址、目的端口、协议类型、序列号、确认号等信息，确定第一 web 流量中各访问请求报文和各访问应答报文的对应关系，从而从第一 web 流量中获取所述每个指示访问成功的访问应答报文分别对应的访问请求报文，作为获得的所述至少一个访问请求报文。

此外，终端通过浏览器访问网站服务器时，由于终端可能安装不同厂商提供浏览器、或者不同版本的浏览器。不同的浏览器由于程序设计方面的差异，会导致不同浏览器访问网站服务器提供的同一网页时，产生的多个访问请求报文中携带不同的 URL。具体地，尽管这些多个访问请求报文访问同一网页，但是其中携带的 URL 采用不同的大小写方式、或者编码方式、或者携带不同的参数。安全设备会将这些访问请求报文按照携带不同 URL 进行处理，从而在网页访问记录中创建不同的表项。这样一方面，这种处理方式与这些访问请求报文实际上访问的是同一网页这一实际情况不符，造成后续可疑 URL 识别时的偏差，另一方面会造成网页访问记录数据量过大。为了提高可疑 URL 识别的准确性，节约网页访问记录在存储器中占有的存储空间，可选地，

在采用附图 5~附图 7 所示的方法构建网页访问记录的过程中，安全设备在步骤 58 在确定出的哈希桶对应链表中查找该 URL 对应的表项之前，先对解析得到的 URL 进行以下几种正规化处理中的至少一种正规化处理。

一、将解析得到的 URL 中的字符转换为预定大小写类型。例如将所有字符统一转换为小写。

二、将解析得到的 URL 转换为预定编码格式。URL 可能采用的编码方式有 GB2312、GBK、UTF8 等等。在本实例中将所有 URL 均转换为 GBK 编码。

三、去除解析得到的 URL 中参数。

例如解析得到的 URL 1 为 www.google.com.hk/videohp?hl=zh-cn&tab=wv，去除参数后的 URL 1 为 www.google.com.hk/videohp 。解析得到的 URL 2 为 www.google.com.hk/videohp?hl=zh-cn&tab=wv&aq=f，去除参数后的 URL 2 为 www.google.com.hk/videohp。

这样正规化处理后的 URL 1 和 URL 2 相同，在网页访问记录中对应同一个表项，从而有效控制网页访问记录的规模，节约存储资源。

在网站服务器提供的页面文件数目较多或者不断增长时，安全设备采用图 4 所示的数据结构分别存储访问所述至少一个 URL 中的每个 URL 的 IP 地址，以及所述每个 URL 的被访问总次数将占用较多存储资源。可选地，安全设备根据第一阈值、或者网页后门检测结果识别正常 URL，删除网页访问记录中保存的访问所述正常 URL 的 IP 地址和所述正常 URL 的被访问总次数，后续不再更新访问所述正常 URL 的 IP 地址和所述正常 URL 的被访问总次数，从而节省存储资源和处理资源。

基于上述考虑，对附图 3 所示的检测网页后门的方法进行改进，改进后的流程图请参照附图 8。附图 8 中的步骤 31~步骤 34 与附图 3 相同，在步骤 32 之后，还包括：

步骤 35，安全设备确定正常 URL，其中正常 URL 是指所述至少一个 URL 中的被访问总次数大于第一阈值的 URL。

在步骤 34 之后，还包括：

步骤 36，安全设备确定正常 URL，其中正常 URL 是指网页后门检测结果指示所标识的网页不存在网页后门的可疑 URL。

在步骤 35、36 之后，安全设备执行步骤 37，删除所述网页访问记录中保存的访问所述正常 URL 的 IP 地址和所述正常 URL 的被访问总次数。需要说明的是，步骤 35 和步骤 36 可以择一执行或同时执行。

由于当前信息的增长速度很快，网站服务器提供的正常网页数量也不断增长，需要适时更新网页访问记录。本申请实施例为了适应这种现状，在步骤 37 之后还包括：

步骤 38，安全设备获取所述被保护主机的第二 web 流量。所述第二 web 流量是指在所述第一时间段之后的第二时间段中所述被保护主机提供的网页被访问时发生的流量。

步骤 39，安全设备从所述第二 web 流量中获得访问请求报文，解析所述访问请求报文，从而获得所述访问请求报文的源地址和携带的 URL。

步骤 310，安全设备判断步骤 39 得到的访问请求报文携带的 URL 与正常 URL 是否相同，如果相同，对所述访问请求的处理结束。如果第二 web 流量中还有未处理的访

问请求报文，则继续处理另一个未处理的访问请求报文。如果不同，执行步骤311。

步骤311，安全设备判断网页访问记录中是否保存有所述访问请求报文携带的URL，如果保存有所述访问请求报文携带的URL，执行步骤312。如果未保存有所述访问请求报文携带的URL，执行步骤313。

步骤312，安全设备将已保存的所述访问请求报文携带的URL的被访问总次数加1，在访问所述访问请求报文携带的URL的IP地址中增加所述访问请求报文的源IP地址。如果第二web流量中还有未处理的访问请求报文，则继续处理另一个未处理的访问请求报文。

步骤313，安全设备在网页访问记录中保存所述访问请求报文携带的URL，设置所述访问请求报文携带的URL的被访问总次数为1，设置访问所述访问请求报文携带的URL的IP地址为该访问请求的源IP地址。如果第二web流量中还有未处理的访问请求报文，则继续处理另一个未处理的访问请求报文。

以第二web流量中的三个不同访问请求报文HTTP request 1、HTTP request 2和HTTP request 3为例，对附图8所示的方法进行举例说明。这里为了简明起见，仅以“IP+标识”的方式代替具体的32位2进制地址，用“URL+标识”的方式代替具体URL字符串。在本实例中安全设备处理三个访问请求报文之前，采用图7所示的数据结构构建出的网页访问记录如图9所示。其中，URL 3为正常URL，不保存URL 2对应的被访问总次数和IP地址列表。安全设备暂时无法识别URL 1是否为可疑URL或是正常URL，因此保存URL 3对应的被访问总次数和IP地址列表。

安全设备解析HTTP request 1、HTTP request 2和HTTP request 3得到这三个访问请求的目的地址均为IP 0，为被保护主机的IP地址。获得HTTP request 1携带的URL为URL 1、源IP地址为IP 1。HTTP request 2携带的URL为URL 2、源IP地址为IP 2。HTTP request 3携带的URL为URL 3、源IP地址为IP 3。

对于HTTP request 1，在附图4所示的哈希表中查找IP 0对应的哈希桶表，依次比较各表项保存URL与URL1是否相同。在本实例中URL 1与作为正常URL的URL3不同、且所述网页访问记录中已记录URL 1，则将已记录的URL 1的被访问总次数加1，在访问URL 1的IP地址中增加HTTP request 1的源地址IP 1，将IP地址计数值加1。

在本实例中HTTP request 2携带的URL 2与作为正常URL的URL3不同、且所述网页访问记录中未记录所述URL 2，则在所述访问记录中新建URL 2对应的表项，在新建表项中记录URL 2，设置URL 2的被访问总次数为1，设置IP地址计数值为1，在新建表项的IP地址列表中记录HTTP request 3的源地址IP 2。

在本实例中HTTP request 3携带的URL 3与正常URL相同，对HTTP request 3的处理结束。对上述三个访问请求处理后的网页访问记录如图10所示。

通过上述处理，安全设备在网页访问记录中对于正常URL只需要保存URL即可。对于新增的网页对应的URL、或者尚不能确认是正常URL还是可疑URL的待确认URL，保存待确认URL的IP地址，以及所述待确认URL的被访问总次数。以便后续根据记录的待确认URL的IP地址以及所述待确认URL的被访问总次数，确认待确认URL是正常URL还是可疑URL。一方面保证随着正常网页数目的快速增长，网页访问记录的数据量不至于急速增长，节约存储空间；另一方面能够识别出新出现的webshell文件，保证

了识别效果。

相应地，本申请实施例还提供了一种检测网页后门的装置，如图 11 所示，该装置包括获取单元 111，记录生成单元 112 和确定单元 113，具体如下。

获取单元 111，用于获取被保护主机的第一 web 流量，所述第一 web 流量是指在第一时间段中所述被保护主机提供的网页被访问时发生的流量。

记录生成单元 112，用于根据获取单元 111 获得的第一 web 流量生成所述被保护主机的网页访问记录，所述网页访问记录用于保存至少一个统一资源定位符 URL、访问所述至少一个 URL 中的每个 URL 的 IP 地址、以及所述每个 URL 的被访问总次数，其中所述每个 URL 标识所述被保护主机提供的一个网页。

确定单元 113，用于根据记录生成单元 112 生成的所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，所述可疑 URL 的被访问总次数小于第一阈值、且访问所述可疑 URL 的互不相同的 IP 地址的数量与所述可疑 URL 的被访问总次数的比值小于第二阈值；以及确定所述可疑 URL 标识的网页是否包含网页后门特征库中的后门特征，根据后门特征确定结果检测所述可疑 URL 标识的网页是否存在网页后门。

可选地，本申请实施例中所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述每个表项中保存有被访问总次数和 IP 地址列表。该表项的结构如图 4 所示。

所述记录生成单元，具体用于从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址；从所述至少一个访问请求报文中选择一个访问请求报文，对选择出的访问请求报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

解析选择出的访问请求报文，从而获得所述选择出的访问请求报文的源 IP 地址和携带的 URL；在所述网页访问记录中查找所述选择出的访问请求报文携带的 URL 对应的表项；如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1，在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页访问记录中创建所述选择出的访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

相应地，所述确定单元 113，具体用于从所述网页访问记录中选择出一个表项；确定选择出的表项的 IP 地址列表中互不相同的 IP 地址的数量；如果所述选择出的表项的被访问总次数少于所述第一阈值、且确定出的互不相同的 IP 地址的数量与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL

可选地，所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述表项中保存有被访问总次数、IP 地址计数值和 IP 地址列表。表项的结构如图 7 所示。

所述记录生成单元 112，具体用于从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址。

从所述至少一个访问请求报文中选择出一个访问请求报文，对选择出的访问请求

报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

获取所述选择出的访问请求报文的源 IP 地址和携带的 URL；在所述网页访问记录查找所述选择出的访问请求报文携带的 URL 对应的表项；如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1；确定所述查找到表项的 IP 地址列表中是否已保存所述源 IP 地址，如果所述查找到表项的 IP 地址列表中已保存所述源 IP 地址，则对所述选择出的访问请求报文处理结束；如果所述查找到的表项的 IP 地址列表中未保存所述源 IP 地址，则将所述查找到的表项的 IP 地址计数值加 1，并在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页访问记录中创建所述访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，将所述创建的表项的 IP 地址计数值设置为 1，并在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

相应地，确定单元 113，具体用于从所述网页访问记录中选择出一个表项；如果选择出的表项的被访问总次数少于所述第一阈值、且所述选择出的表项的 IP 地址计数值与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL。

可选地，记录生成单元 112 从所述第一 web 流量中选择至少一个访问应答报文，所述至少一个访问应答报文中的每个访问应答报文携带的状态码指示访问成功，所述每个访问应答报文的源地址为所述被保护主机的 IP 地址；从所述第一 web 流量中获取所述每个网页访问应答报文分别对应的访问请求报文，作为获得的所述至少一个访问请求报文。

可选地，记录生成单元 112 在所述网页访问记录查找所述选择出的访问请求报文携带的 URL 对应的表项，包括：对所述选择出的访问请求报文携带的 URL 执行至少一种正规化处理，得到正规化处理后的 URL，所述正规化处理包括以下（1）～（3）中的一种或多种：（1）将所述选择出的访问请求报文携带的 URL 转换为预定编码格式，（2）将所述选择出的访问请求报文携带的 URL 中的字符转换为预定大小写类型，和（3）去除所述选择出的访问请求报文携带的 URL 中参数；在所述网页访问记录中查找正规化处理后的 URL 对应的表项。

所述记录生成单元 112 在所述网页访问记录中创建所述访问请求报文携带的 URL 对应的表项，具体为：在所述网页访问记录中创建所述正规化处理后的 URL 对应的表项。

可选地，所述确定单元 113，还用于根据所述网页访问记录，从所述至少一个 URL 中确定正常 URL，所述正常 URL 是所述至少一个 URL 中的被访问总次数大于所述第一阈值的 URL，或者网页后门检测结果指示所标识的网页不存在网页后门的可疑 URL；删除所述网页访问记录中保存的访问所述正常 URL 的 IP 地址和所述正常 URL 的被访问总次数。

可选地，获取单元 111，还用于获取所述被保护主机的第二 web 流量，所述第二 web 流量是指在所述第一时间段之后的第二时间段中所述被保护主机提供的网页被访

问时发生的流量。

相应地，记录生成单元 112，还用于从所述第二 web 流量中获得第一访问请求报文、第二访问请求报文和第三访问请求报文；

解析所述第一访问请求报文，从而获得所述第一访问请求报文的源 IP 地址和携带的 URL；如果所述第一访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访问记录中已保存所述第一访问请求报文携带的 URL，则将已保存的所述第一访问请求报文携带的 URL 的被访问总次数加 1，在访问所述第一访问请求报文携带的 URL 的 IP 地址中增加所述第一访问请求报文的源 IP 地址。

解析所述第二访问请求报文，从而获得所述第二访问请求报文的源 IP 地址和携带的 URL；如果所述第二访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访问记录中未保存所述第二访问请求报文携带的 URL，则在所述网页访问记录中保存所述第二访问请求报文携带的 URL，设置所述第二访问请求报文携带的 URL 的被访问总次数为 1，设置访问所述第二访问请求报文携带的 URL 的 IP 地址为所述第二访问请求报文的源 IP 地址。

解析所述第三访问请求报文，从而获得所述第三访问请求报文携带的 URL；如果所述第三访问请求报文携带的 URL 与所述正常 URL 相同，对所述第三访问请求的处理结束。

本装置实施例中提供的检测网页后门的装置，可以集成在安全设备中，应用于方法实施例一附图 1 所示的场景中，实现其中安全设备的功能。检测网页后门的装置可以实现的其他附加功能、以及与其他网元设备的交互过程，请参照方法实施例中对安全设备的描述，在这里不再赘述。

本说明书中的各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

权 利 要 求 书

1、一种检测网页后门的方法，其特征在于，包括：

获取被保护主机的第一 web 流量，所述第一 web 流量是指在第一时间段中所述被保护主机提供的网页被访问时发生的流量；

根据所述第一 web 流量生成所述被保护主机的网页访问记录，所述网页访问记录用于保存至少一个统一资源定位符 URL、访问所述至少一个 URL 中的每个 URL 的 IP 地址、以及所述每个 URL 的被访问总次数，其中所述每个 URL 标识所述被保护主机提供的一个网页；

根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，所述可疑 URL 的被访问总次数小于第一阈值、且访问所述可疑 URL 的互不相同的 IP 地址的数量与所述可疑 URL 的被访问总次数的比值小于第二阈值；以及

确定所述可疑 URL 标识的网页是否包含网页后门特征库中的后门特征，根据后门特征确定结果检测所述可疑 URL 标识的网页是否存在网页后门。

2、根据权利要求 1 所述的方法，其特征在于，所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述每个表项中保存有被访问总次数和 IP 地址列表；

所述第一 web 流量生成所述被保护主机的网页访问记录，包括：

从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址；

从所述至少一个访问请求报文中选择一个访问请求报文，对选择出的访问请求报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

解析选择出的访问请求报文，从而获得所述选择出的访问请求报文的源 IP 地址和携带的 URL；

在所述网页访问记录中查找所述选择出的访问请求报文携带的 URL 对应的表项；

如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1，在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；

如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页访问记录中创建所述选择出的访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

3、根据权利要求 2 所述的方法，其特征在于，所述根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，包括：

从所述网页访问记录中选择出一个表项；

确定选择出的表项的 IP 地址列表中互不相同的 IP 地址的数量；

如果所述选择出的表项的被访问总次数少于所述第一阈值、且确定出的互不相同的 IP 地址的数量与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL。

4、根据权利要求 1 所述的方法，其特征在于，所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述表项中保存有被访问总次数、IP 地址计数值和 IP 地址列表；

所述第一 web 流量生成所述被保护主机的网页访问记录，包括：

从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址；

从所述至少一个访问请求报文中选择出一个访问请求报文，对选择出的访问请求报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

获取所述选择出的访问请求报文的源 IP 地址和携带的 URL；

在所述网页访问记录查找所述选择出的访问请求报文携带的 URL 对应的表项；

如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1；确定所述查找到表项的 IP 地址列表中是否已保存所述源 IP 地址，如果所述查找到表项的 IP 地址列表中已保存所述源 IP 地址，则对所述选择出的访问请求报文处理结束；如果所述查找到的表项的 IP 地址列表中未保存所述源 IP 地址，则将所述查找到的表项的 IP 地址计数值加 1，并在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；

如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页访问记录中创建所述访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，将所述创建的表项的 IP 地址计数值设置为 1，并在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

5、根据权利要求 4 所述的方法，其特征在于，所述根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，包括：

从所述网页访问记录中选择出一个表项；

如果选择出的表项的被访问总次数少于所述第一阈值、且所述选择出的表项的 IP 地址计数值与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL。

6、根据权利要求 2 或 4 所述的方法，其特征在于，从所述第一 web 流量中获得至少一个访问请求报文包括：

从所述第一 web 流量中选择至少一个访问应答报文，所述至少一个访问应答报文中的每个访问应答报文携带的状态码指示访问成功，所述每个访问应答报文的源地址为所述被保护主机的 IP 地址；

从所述第一 web 流量中获取所述每个网页访问应答报文分别对应的访问请求报文，作为获得的所述至少一个访问请求报文。

7、根据权利要求 2 或 4 所述的方法，其特征在于，在所述网页访问记录中查找所述选择出的访问请求报文携带的 URL 对应的表项，包括：

对所述选择出的访问请求报文携带的 URL 执行至少一种正规化处理，得到正规化处理后的 URL，所述正规化处理包括以下（1）～（3）中的一种或多种：（1）将所述选择出的访问请求报文携带的 URL 转换为预定编码格式，（2）将所述选择出的访问请求报文携带的 URL 中的字符转换为预定大小写类型，和（3）去除所述选择出的访问请求报文携带的 URL 中参数；

在所述网页访问记录中查找正规化处理后的 URL 对应的表项；

相应地，在所述网页访问记录中创建所述访问请求报文携带的 URL 对应的表项，具体为：

在所述网页访问记录中创建所述正规化处理后的 URL 对应的表项。

8、根据权利要求 1 所述的方法，其特征在于，还包括：

根据所述网页访问记录，从所述至少一个 URL 中确定正常 URL，所述正常 URL 是所述至少一个 URL 中的被访问总次数大于所述第一阈值的 URL，或者网页后门检测结果指示所标识的网页不存在网页后门的可疑 URL；

删除所述网页访问记录中保存的访问所述正常 URL 的 IP 地址和所述正常 URL 的被访问总次数。

9、根据权利要求 8 所述的方法，其特征在于，还包括：

获取所述被保护主机的第二 web 流量，所述第二 web 流量是指在所述第一时间段之后的第二时间段中所述被保护主机提供的网页被访问时发生的流量；

从所述第二 web 流量中获得第一访问请求报文、第二访问请求报文和第三访问请求报文；

解析所述第一访问请求报文，从而获得所述第一访问请求报文的源 IP 地址和携带的 URL；如果所述第一访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访问记录中已保存所述第一访问请求报文携带的 URL，则将已保存的所述第一访问请求报文携带的 URL 的被访问总次数加 1，在访问所述第一访问请求报文携带的 URL 的 IP 地址中增加所述第一访问请求报文的源 IP 地址；

解析所述第二访问请求报文，从而获得所述第二访问请求报文的源 IP 地址和携带的 URL；如果所述第二访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访问记录中未保存所述第二访问请求报文携带的 URL，则在所述网页访问记录中保存所述第二访问请求报文携带的 URL，设置所述第二访问请求报文携带的 URL 的被访问总次数为 1，设置访问所述第二访问请求报文携带的 URL 的 IP 地址为所述第二访问请求报文的源 IP 地址；

解析所述第三访问请求报文，从而获得所述第三访问请求报文携带的 URL；如果所述第三访问请求报文携带的 URL 与所述正常 URL 相同，对所述第三访问请求的处理结束。

10、一种检测网页后门的装置，其特征在于，包括：

获取单元，用于获取被保护主机的第一 web 流量，所述第一 web 流量是指在第一时间段中所述被保护主机提供的网页被访问时发生的流量；

记录生成单元，用于根据所述第一 web 流量生成所述被保护主机的网页访问记录，所述网页访问记录用于保存至少一个统一资源定位符 URL、访问所述至少一个 URL 中的每个 URL 的 IP 地址、以及所述每个 URL 的被访问总次数，其中所述每个 URL 标识所述被保护主机提供的一个网页；

确定单元，用于根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，所述可疑 URL 的被访问总次数小于第一阈值、且访问所述可疑 URL 的互不相同的 IP 地址的数量与所述可疑 URL 的被访问总次数的比值小于第二阈值；以及确定所述可疑 URL 标识的网页是否包含网页后门特征库中的后门特征，根据后门特征确定结果检测

所述可疑 URL 标识的网页是否存在网页后门。

11、根据权利要求 10 所述的装置，其特征在于，所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述每个表项中保存有被访问总次数和 IP 地址列表，

所述记录生成单元，具体用于从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址；

从所述至少一个访问请求报文中选择一个访问请求报文，对选择出的访问请求报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

解析选择出的访问请求报文，从而获得所述选择出的访问请求报文的源 IP 地址和携带的 URL；

在所述网页访问记录中查找所述选择出的访问请求报文携带的 URL 对应的表项；

如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1，在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；

如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页访问记录中创建所述选择出的访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

12、根据权利要求 11 所述的装置，其特征在于，

所述确定单元，具体用于从所述网页访问记录中选择出一个表项；确定选择出的表项的 IP 地址列表中互不相同的 IP 地址的数量；如果所述选择出的表项的被访问总次数少于所述第一阈值、且确定出的互不相同的 IP 地址的数量与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL。

13、根据权利要求 10 所述的方法，其特征在于，所述网页访问记录包括至少一个表项，所述至少一个表项中的每个表项分别与所述至少一个 URL 中的一个 URL 相对应，所述表项中保存有被访问总次数、IP 地址计数值和 IP 地址列表；

所述记录生成单元，具体用于从所述第一 web 流量中获得至少一个访问请求报文，所述访问请求报文的目的 IP 地址为所述被保护主机的 IP 地址；

从所述至少一个访问请求报文中选择出一个访问请求报文，对选择出的访问请求报文进行以下处理，直到处理完所述至少一个访问请求报文中的每个访问请求报文为止：

获取所述选择出的访问请求报文的源 IP 地址和携带的 URL；

在所述网页访问记录查找所述选择出的访问请求报文携带的 URL 对应的表项；

如果查找到所述选择出的访问请求报文携带的 URL 对应的表项，则将查找到的表项的被访问总次数加 1；确定所述查找到表项的 IP 地址列表中是否已保存所述源 IP 地址，如果所述查找到表项的 IP 地址列表中已保存所述源 IP 地址，则对所述选择出的访问请求报文处理结束；如果所述查找到的表项的 IP 地址列表中未保存所述源 IP 地址，则将所述查找到的表项的 IP 地址计数值加 1，并在所述查找到的表项的 IP 地址列表中记录所述源 IP 地址；

如果未查找到所述选择出的访问请求报文携带的 URL 对应的表项，则在所述网页

访问记录中创建所述访问请求报文携带的 URL 对应的表项，将创建的表项的被访问总次数设置为 1，将所述创建的表项的 IP 地址计数值设置为 1，并在所述创建的表项的所述 IP 地址列表中记录所述源 IP 地址。

14、根据权利要求 13 所述的装置，其特征在于，

所述确定单元，具体用于从所述网页访问记录中选择出一个表项；如果选择出的表项的被访问总次数少于所述第一阈值、且所述选择出的表项的 IP 地址计数值与所述选择出的表项的被访问总次数的比值小于所述第二阈值，则确定所述选择出的表项对应的 URL 为可疑 URL。

15、根据权利要求 12 或 14 所述的装置，其特征在于，

所述记录生成单元从所述第一 web 流量中选择至少一个访问应答报文，所述至少一个访问应答报文中的每个访问应答报文携带的状态码指示访问成功，所述每个访问应答报文的源地址为所述被保护主机的 IP 地址；

从所述第一 web 流量中获取所述每个网页访问应答报文分别对应的访问请求报文，作为获得的所述至少一个访问请求报文。

16、根据权利要求 12 或 14 所述的方法，其特征在于，

所述记录生成单元在所述网页访问记录查找所述选择出的访问请求报文携带的 URL 对应的表项，包括：

对所述选择出的访问请求报文携带的 URL 执行至少一种正规化处理，得到正规化处理后的 URL，所述正规化处理包括以下（1）～（3）中的一种或多种：（1）将所述选择出的访问请求报文携带的 URL 转换为预定编码格式，（2）将所述选择出的访问请求报文携带的 URL 中的字符转换为预定大小写类型，和（3）去除所述选择出的访问请求报文携带的 URL 中参数；

在所述网页访问记录中查找正规化处理后的 URL 对应的表项；

所述记录生成单元在所述网页访问记录中创建所述访问请求报文携带的 URL 对应的表项，具体为：

在所述网页访问记录中创建所述正规化处理后的 URL 对应的表项。

17、根据权利要求 10 所述的装置，其特征在于，

所述确定单元，还用于根据所述网页访问记录，从所述至少一个 URL 中确定正常 URL，所述正常 URL 是所述至少一个 URL 中的被访问总次数大于所述第一阈值的 URL，或者网页后门检测结果指示所标识的网页不存在网页后门的可疑 URL；删除所述网页访问记录中保存的访问所述正常 URL 的 IP 地址和所述正常 URL 的被访问总次数。

18、根据权利要求 17 所述的装置，其特征在于，

所述获取单元，还用于获取所述被保护主机的第二 web 流量，所述第二 web 流量是指在所述第一时间段之后的第二时间段中所述被保护主机提供的网页被访问时发生的流量；

所述记录生成单元，还用于从所述第二 web 流量中获得第一访问请求报文、第二访问请求报文和第三访问请求报文；

解析所述第一访问请求报文，从而获得所述第一访问请求报文的源 IP 地址和携带的 URL；如果所述第一访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访

问记录中已保存所述第一访问请求报文携带的 URL，则将已保存的所述第一访问请求报文携带的 URL 的被访问总次数加 1，在访问所述第一访问请求报文携带的 URL 的 IP 地址中增加所述第一访问请求报文的源 IP 地址；

解析所述第二访问请求报文，从而获得所述第二访问请求报文的源 IP 地址和携带的 URL；如果所述第二访问请求报文携带的 URL 与所述正常 URL 不同、且所述网页访问记录中未保存所述第二访问请求报文携带的 URL，则在所述网页访问记录中保存所述第二访问请求报文携带的 URL，设置所述第二访问请求报文携带的 URL 的被访问总次数为 1，设置访问所述第二访问请求报文携带的 URL 的 IP 地址为所述第二访问请求报文的源 IP 地址；

解析所述第三访问请求报文，从而获得所述第三访问请求报文携带的 URL；如果所述第三访问请求报文携带的 URL 与所述正常 URL 相同，对所述第三访问请求的处理结束。

19、一种安全设备，其特征在于，包括存储器，处理器，网络接口和总线，所述存储器、所述处理器和所述网络接口通过所述总线相互连接，其特征在于，

所述网络接口，用于获取被保护主机的第一 web 流量，所述第一 web 流量是指在第一时间段中所述被保护主机提供的网页被访问时发生的流量；

所述处理器读取所述存储器中存储的程序代码后，执行以下操作：

根据所述第一 web 流量生成所述被保护主机的网页访问记录，所述网页访问记录用于保存至少一个统一资源定位符 URL、访问所述至少一个 URL 中的每个 URL 的 IP 地址、以及所述每个 URL 的被访问总次数，其中所述每个 URL 标识所述被保护主机提供的一个网页；根据所述网页访问记录，从所述至少一个 URL 中确定可疑 URL，所述可疑 URL 的被访问总次数小于第一阈值、且访问所述可疑 URL 的互不相同的 IP 地址的数量与所述可疑 URL 的被访问总次数的比值小于第二阈值；以及确定所述可疑 URL 标识的网页是否包含网页后门特征库中的后门特征，根据后门特征确定结果检测所述可疑 URL 标识的网页是否存在网页后门。

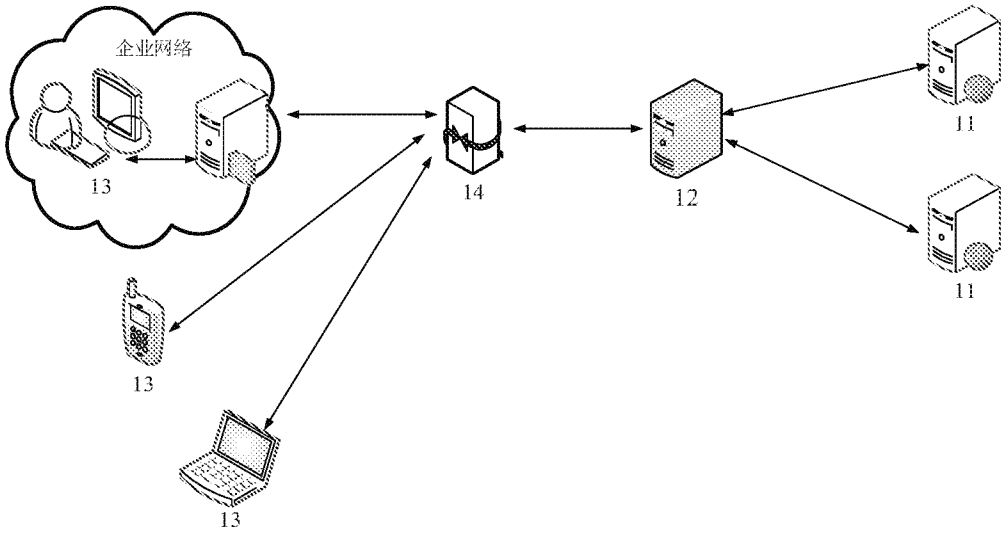
企业网络
13
14
12
11
11
13
13

图 1

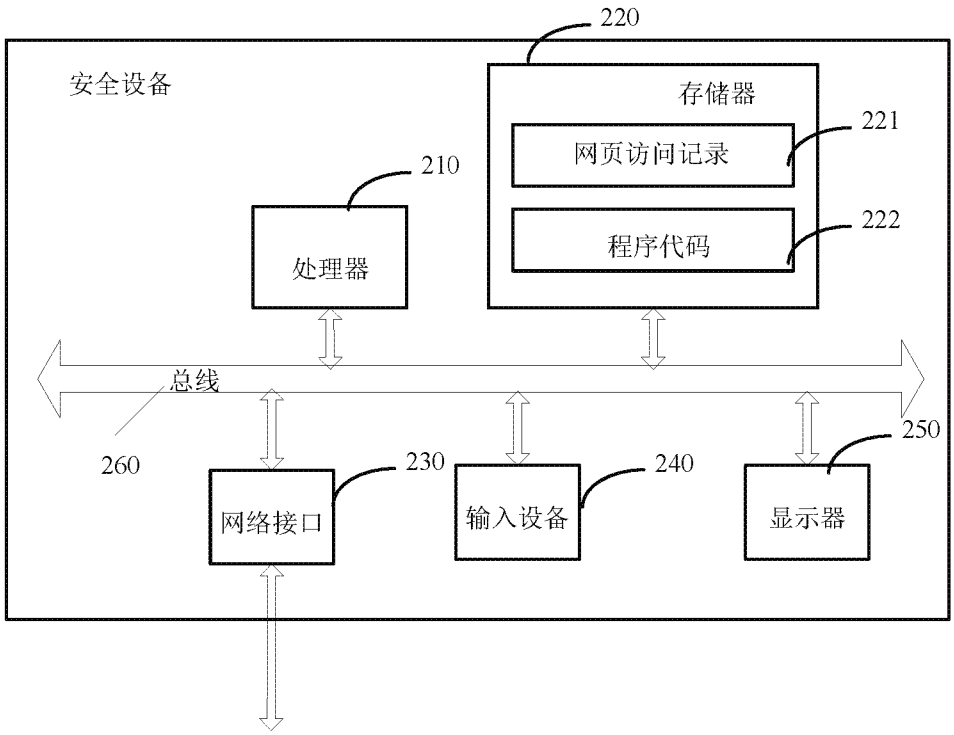
220
安全设备
存储器
221
网页访问记录
210
222
程序代码
处理器
总线
260
230
240
250
网络接口
输入设备
显示器

图 2

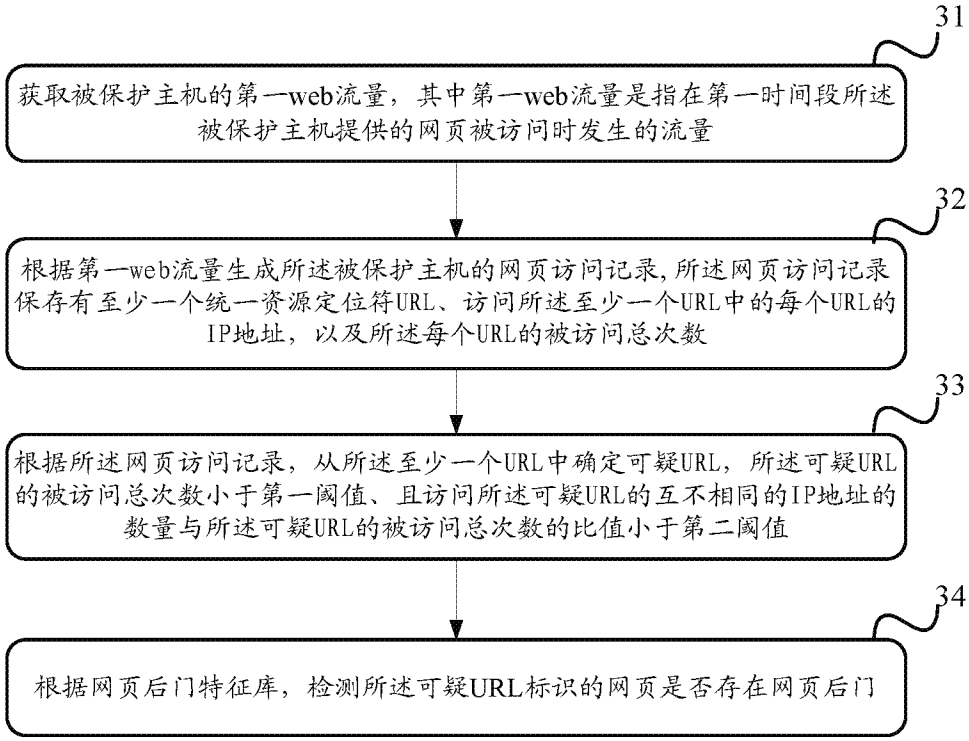
31
获取被保护主机的第一web流量，其中第一web流量是指在第一时间段所述被保护主机提供的网页被访问时发生的流量
32
根据第一web流量生成所述被保护主机的网页访问记录，所述网页访问记录保存有至少一个统一资源定位符URL、访问所述至少一个URL中的每个URL的IP地址，以及所述每个URL的被访问总次数
33
根据所述网页访问记录，从所述至少一个URL中确定可疑URL，所述可疑URL的被访问总次数小于第一阈值、且访问所述可疑URL的互不相同的IP地址的数量与所述可疑URL的被访问总次数的比值小于第二阈值
34
根据网页后门特征库，检测所述可疑URL标识的网页是否存在网页后门

图 3

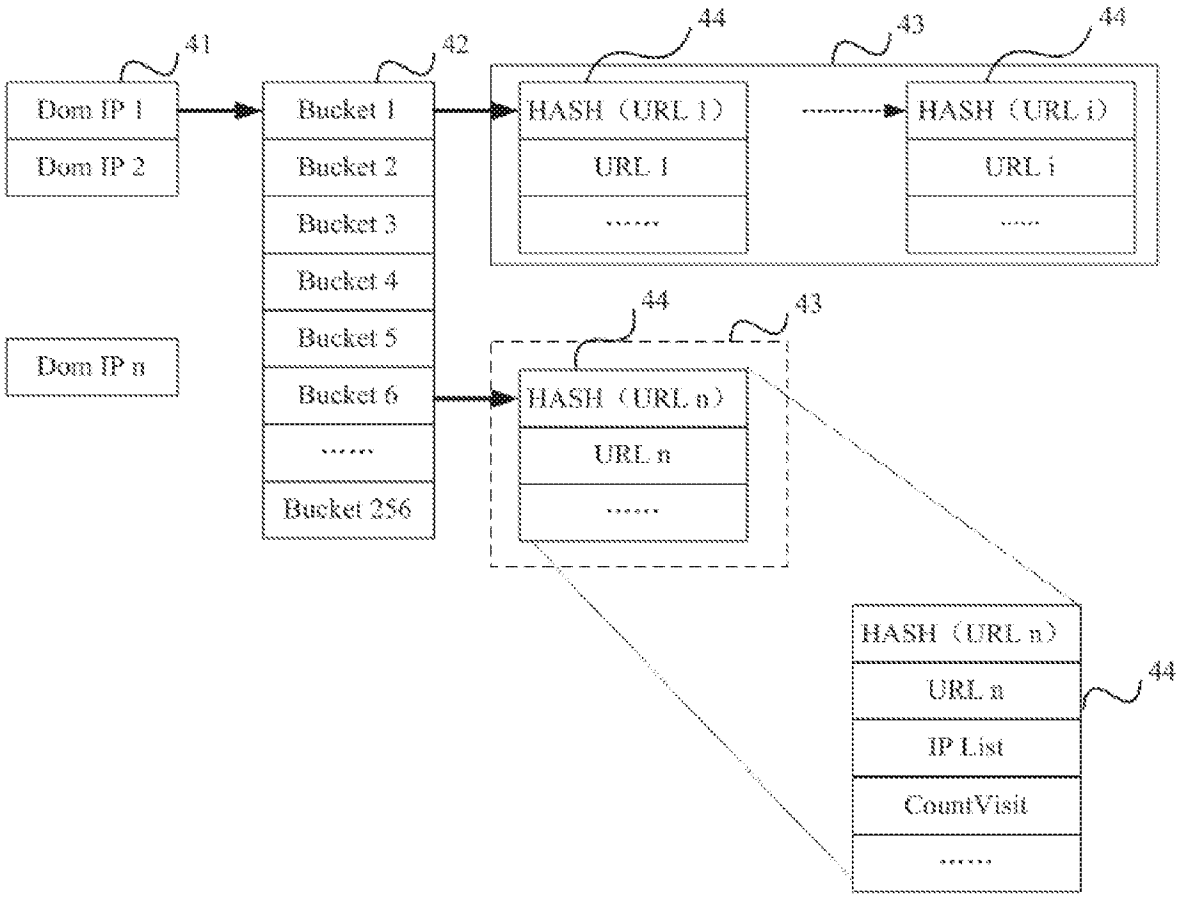
41
42
44
43
44
Dom IP 1
Dom IP 2
Dom IP n
Bucket 1
Bucket 2
Bucket 3
Bucket 4
Bucket 5
Bucket 6
Bucket 256
HASH（URL 1）
URL 1
HASH（URL i）
URL i
44
43
HASH（URL n）
URL n
HASH（URL n）
URL n
IP List
CountVisit
44

图 4

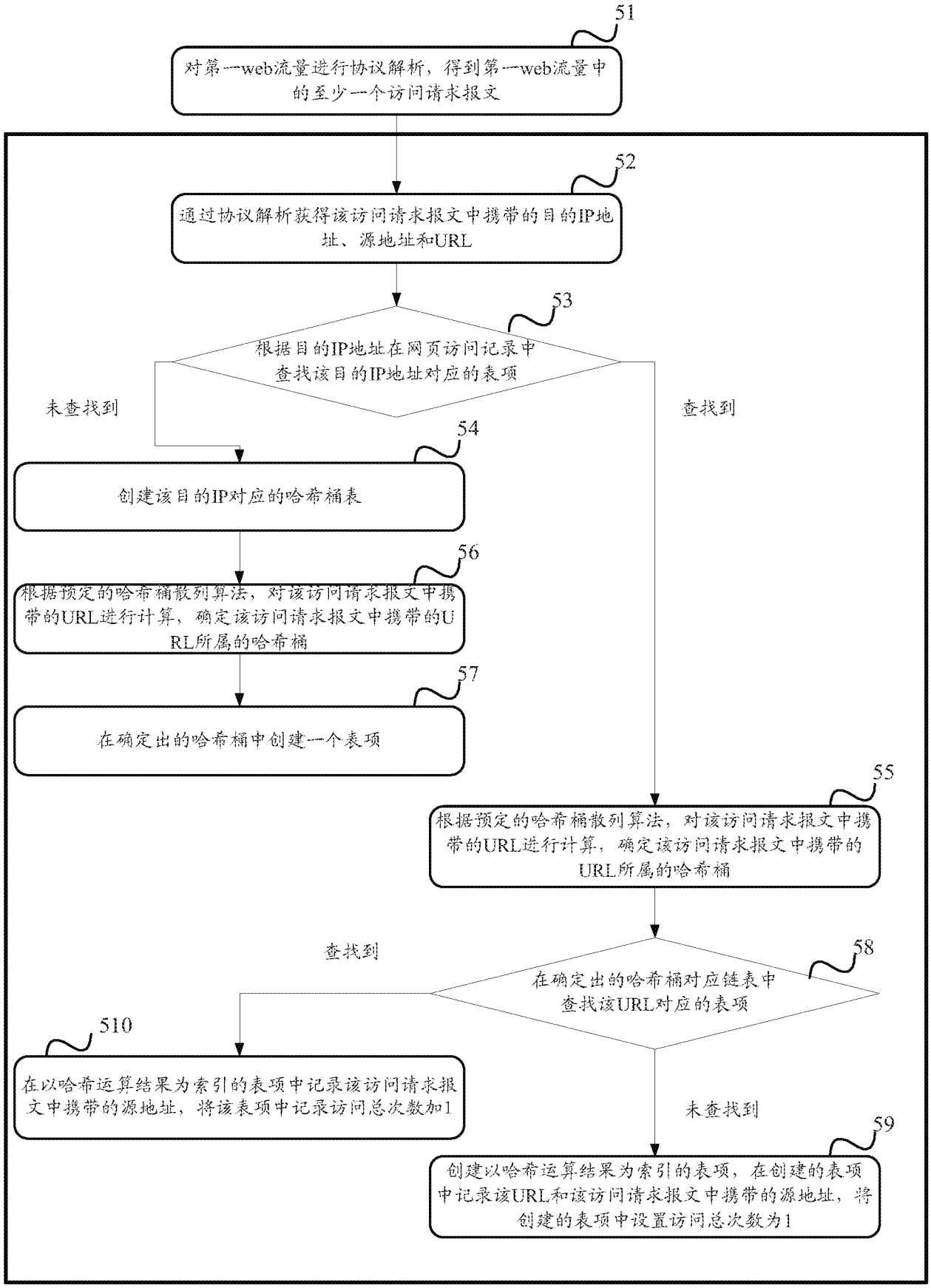
51
对第一web流量进行协议解析，得到第一web流量中的至少一个访问请求报文
52
通过协议解析获得该访问请求报文中携带的目的IP地址、源地址和URL
53
根据目的IP地址在网页访问记录中查找该目的IP地址对应的表项
未查找到
查找到
54
创建该目的IP对应的哈希桶表
56
根据预定的哈希桶散列算法，对该访问请求报文中携带的URL进行计算，确定该访问请求报文中携带的URL所属的哈希桶
57
在确定出的哈希桶中创建一个表项
55
根据预定的哈希桶散列算法，对该访问请求报文中携带的URL进行计算，确定该访问请求报文中携带的URL所属的哈希桶
查找到
58
在确定出的哈希桶对应链表中查找该URL对应的表项
510
在以哈希运算结果为索引的表项中记录该访问请求报文中携带的源地址，将该表项中记录访问总次数加1
未查找到
59
创建以哈希运算结果为索引的表项，在创建的表项中记录该URL和该访问请求报文中携带的源地址，将创建的表项中设置访问总次数为1

图 5

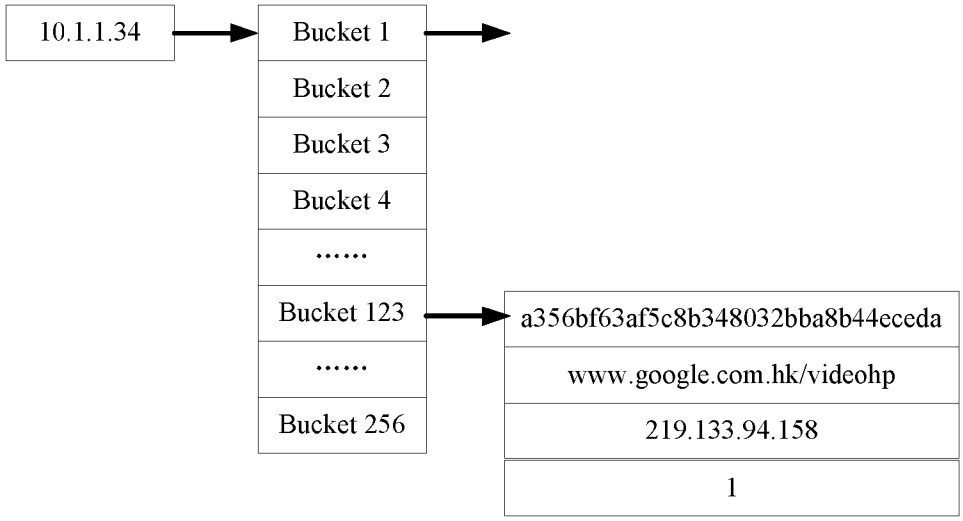
10.1.1.34
Bucket 1
Bucket 2
Bucket 3
Bucket 4
.....
Bucket 123
.....
Bucket 256
a356bf63af5c8b348032bba8b44eceda
www.google.com.hk/videohp
219.133.94.158
1

图 6

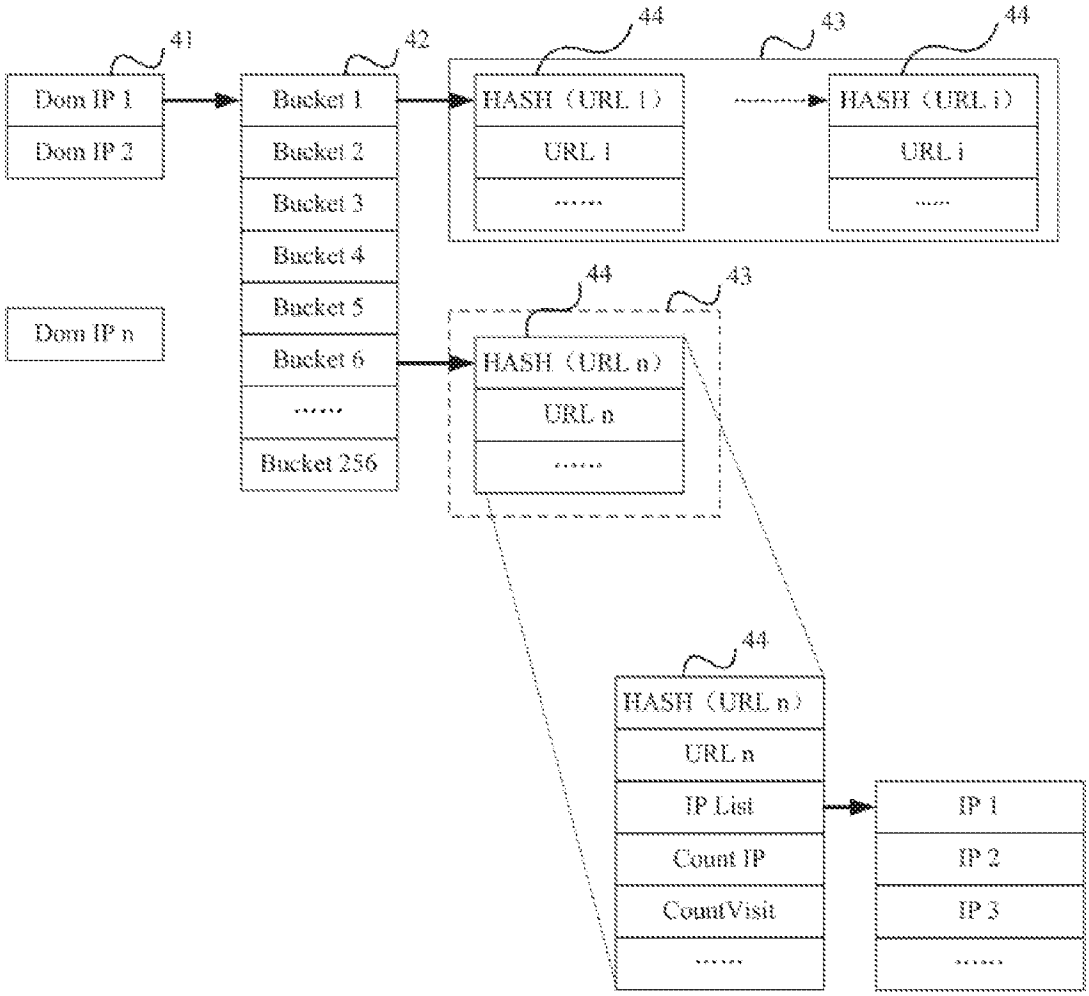
41
42
44
43
44
Dom IP 1
Dom IP 2
Dom IP n
Bucket 1
Bucket 2
Bucket 3
Bucket 4
Bucket 5
Bucket 6
.....
Bucket 256
HASH（URL 1）
URL 1
.....
HASH（URL i）
URL i
.....
44
43
HASH（URL n）
URL n
.....
44
HASH（URL n）
URL n
IP List
Count IP
CountVisit
.....
IP 1
IP 2
IP 3
.....

图 7

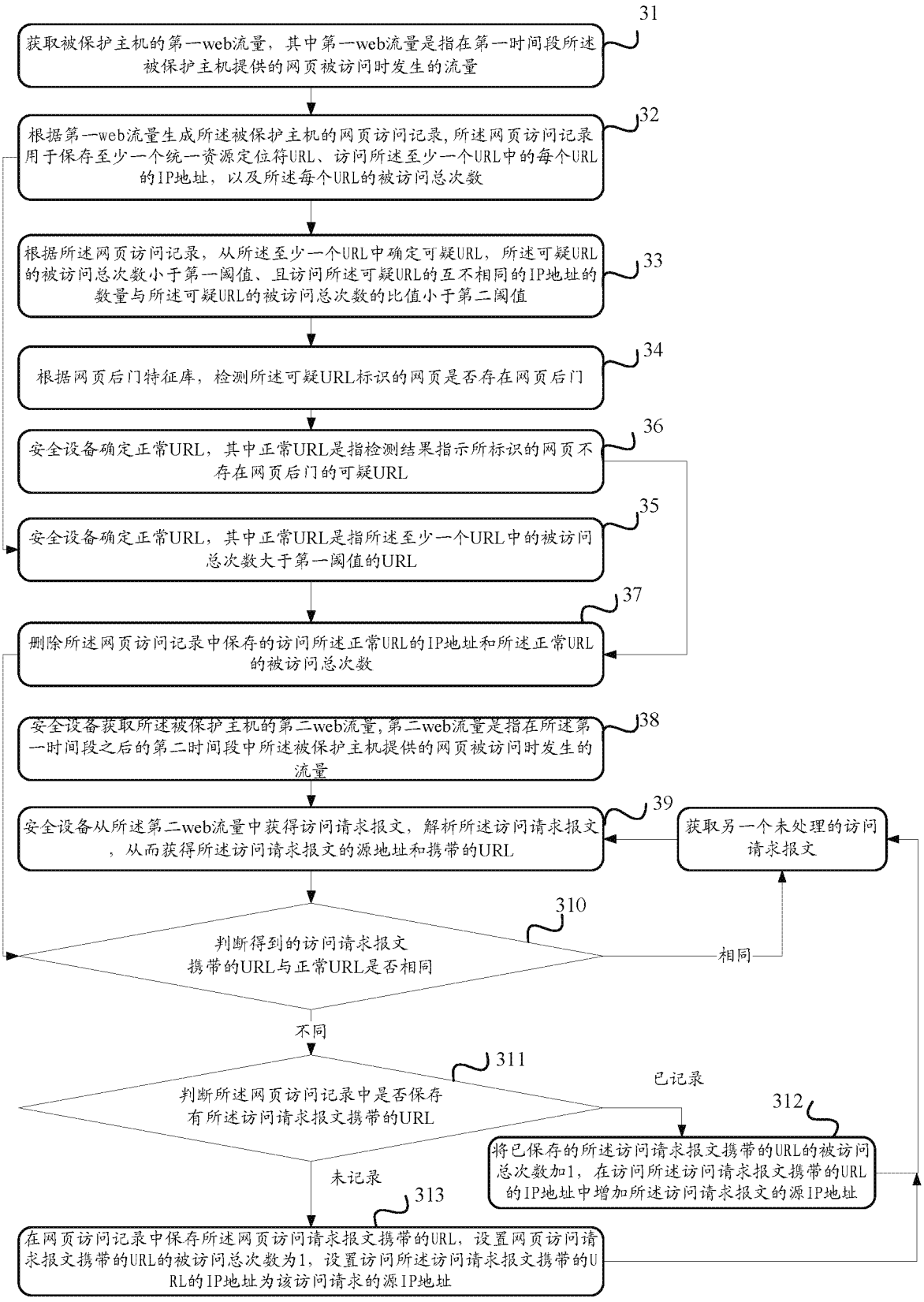
31
获取被保护主机的第一web流量，其中第一web流量是指在第一时间段所述被保护主机提供的网页被访问时发生的流量
32
根据第一web流量生成所述被保护主机的网页访问记录，所述网页访问记录用于保存至少一个统一资源定位符URL、访问所述至少一个URL中的每个URL的IP地址，以及所述每个URL的被访问总次数
33
根据所述网页访问记录，从所述至少一个URL中确定可疑URL，所述可疑URL的被访问总次数小于第一阈值、且访问所述可疑URL的互不相同的IP地址的数量与所述可疑URL的被访问总次数的比值小于第二阈值
34
根据网页后门特征库，检测所述可疑URL标识的网页是否存在网页后门
36
安全设备确定正常URL，其中正常URL是指检测结果指示所标识的网页不存在网页后门的可疑URL
35
安全设备确定正常URL，其中正常URL是指所述至少一个URL中的被访问总次数大于第一阈值的URL
37
删除所述网页访问记录中保存的访问所述正常URL的IP地址和所述正常URL的被访问总次数
38
安全设备获取所述被保护主机的第二web流量，第二web流量是指在所述第一时间段之后的第二时间段中所述被保护主机提供的网页被访问时发生的流量
39
安全设备从所述第二web流量中获得访问请求报文，解析所述访问请求报文，从而获得所述访问请求报文的源地址和携带的URL
获取另一个未处理的访问请求报文
310
判断得到的访问请求报文携带的URL与正常URL是否相同
相同
不同
311
判断所述网页访问记录中是否保存有所述访问请求报文携带的URL
已记录
312
将已保存的所述访问请求报文携带的URL的被访问总次数加1，在访问所述访问请求报文携带的URL的IP地址中增加所述访问请求报文的源IP地址
未记录
313
在网页访问记录中保存所述网页访问请求报文携带的URL，设置网页访问请求报文携带的URL的被访问总次数为1，设置访问所述访问请求报文携带的URL的IP地址为该访问请求的源IP地址

图 8

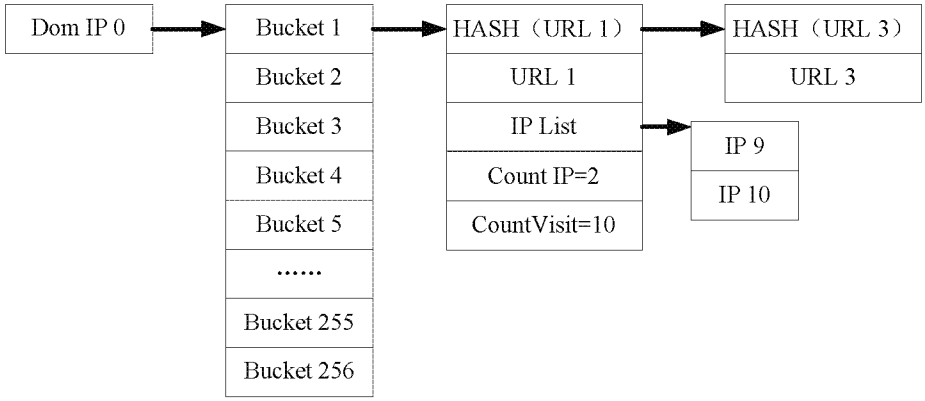
Dom IP 0
Bucket 1
Bucket 2
Bucket 3
Bucket 4
Bucket 5
.....
Bucket 255
Bucket 256
HASH（URL 1）
URL 1
IP List
Count IP=2
CountVisit=10
IP 9
IP 10
HASH（URL 3）
URL 3

图 9

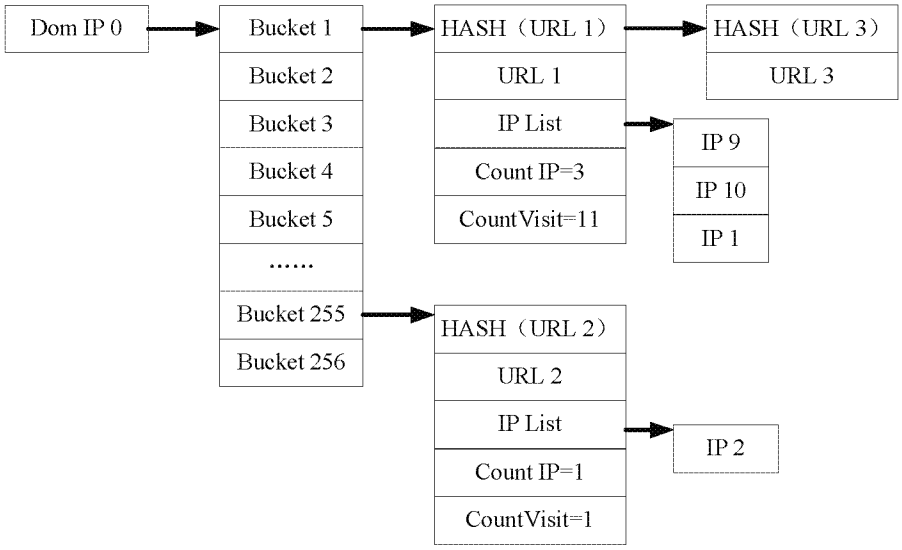
Dom IP 0
Bucket 1
Bucket 2
Bucket 3
Bucket 4
Bucket 5
.....
Bucket 255
Bucket 256
HASH（URL 1）
URL 1
IP List
Count IP=3
CountVisit=11
IP 9
IP 10
IP 1
HASH（URL 3）
URL 3
HASH（URL 2）
URL 2
IP List
Count IP=1
CountVisit=1
IP 2

图 10

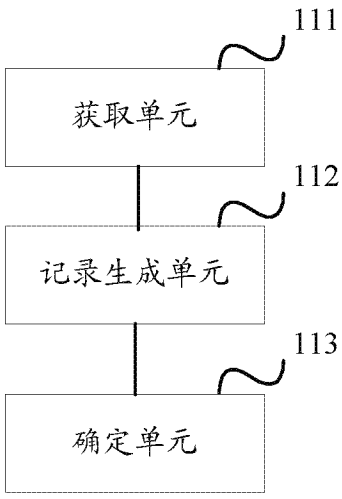
111
获取单元
112
记录生成单元
113
确定单元

图 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/096502

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06 (2006.01) i

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)

H04L

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)

WPI, EPODOC, CNKI, CNPAT, GOOGLE: 后门, 病毒, 木马, 蠕虫, 恶意, 访问, 日志, 记录, 网页, 次数, IP, URL, WEB, webshell, virus, trojan, worm, log?, access, record, time?, number

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104468477 A (HANGZHOU DPTECH TECHNOLOGIES CO., LTD.), 25 March 2015 (25.03.2015), description, paragraphs [0036]-[0047]	1-19
A	CN 105760379 A (CHINA MOBILE COMMUNICATIONS CORPORATION), 13 July 2016 (13.07.2016), entire document	1-19
A	CN 105187396 A (XIAOMI TECHNOLOGY CO., LTD.), 23 December 2015 (23.12.2015), entire document	1-19
A	CN 103701793 A (BEIJING QIHOO TECHNOLOGY CO., LTD.; QIZHI SOFTWARE (BEIJING) CO., LTD.), 02 April 2014 (02.04.2014), entire document	1-19
A	US 2015256551 A1 (KANG, M.H.), 10 September 2015 (10.09.2015), entire document	1-19

☐ 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
07 September 2017	11 October 2017
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 SU, Ning Telephone No. (86-10) 61648524

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/096502

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104468477 A	25 March 2015	None	
CN 105760379 A	13 July 2016	None	
CN 105187396 A	23 December 2015	None	
CN 103701793 A	02 April 2014	None	
US 2015256551 A1	10 September 2015	KR 101239401 B1	06 March 2013
		WO 2014054854 A1	10 April 2014

国际检索报告

国际申请号 PCT/CN2017/096502

A. 主题的分类

H04L 29/06(2006.01)i

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)

H04L

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）)

WPI, EPODOC, CNKI, CNPAT, GOOGLE: 后门，病毒，木马，蠕虫，恶意，访问，日志，记录，网页，次数，IP, URL, WEB, webshell, virus, trojan, worm, log?, access, record, time?, number

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
A	CN 104468477 A (杭州迪普科技有限公司) 2015年 3月 25日 (2015-03-25) 说明书第[0036]-[0047]段	1-19
A	CN 105760379 A (中国移动通信集团公司) 2016年 7月 13日 (2016-07-13) 全文	1-19
A	CN 105187396 A (小米科技有限责任公司) 2015年 12月 23日 (2015-12-23) 全文	1-19
A	CN 103701793 A (北京奇虎科技有限公司 奇智软件北京有限公司) 2014年 4月 2日 (2014-04-02) 全文	1-19
A	US 2015256551 A1 (KANG, MYOUNG HUN) 2015年 9月 10日 (2015-09-10) 全文	1-19

☐ 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
2017年 9月 7日	2017年 10月 11日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 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 苏宁 电话号码 (86-10)61648524

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/096502

检索报告引用的专利文件			公布日（年/月/日）	同族专利			公布日（年/月/日）
CN	104468477	A	2015年 3月 25日	无			
CN	105760379	A	2016年 7月 13日	无			
CN	105187396	A	2015年 12月 23日	无			
CN	103701793	A	2014年 4月 2日	无			
US	2015256551	A1	2015年 9月 10日	KR	101239401	B1	2013年 3月 6日
				WO	2014054854	A1	2014年 4月 10日