(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104766014 B
(45)授权公告日 2017.12.01

(21)申请号 201510219801.1

(22)申请日 2015.04.30

(65)同一申请的已公布的文献号
申请公布号 CN 104766014 A

(43)申请公布日 2015.07.08

(73)专利权人 安一恒通(北京)科技有限公司
地址 100091 北京市海淀区东北旺西路8号,中关村软件园4号楼C座1-03

(72)发明人 唐呈光 杨念 耿志峰

(74)专利代理机构 北京英赛嘉华知识产权代理有限责任公司 11204
代理人 王达佐 马晓亚

(51)Int.Cl.
*G06F 21/56*(2013.01)
*H04L 29/08*(2006.01)

(56)对比文件
CN 103902889 A,2014.07.02,
CN 103701779 A,2014.04.02,
US 2013/0305140 A1,2013.11.14,

审查员 唐剑

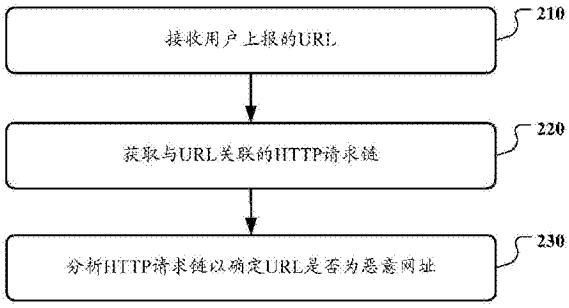
权利要求书3页 说明书12页 附图10页

(54)发明名称

用于检测恶意网址的方法和系统

(57)摘要

本申请公开了一种恶意网址检测方法和系统。该方法包括:接收用户上报的统一资源定位符URL;获取与URL关联的超文本传输协议HTTP请求链,HTTP请求链是包含访问URL的多次HTTP请求-响应交互信息的时序链表;以及分析HTTP请求链以确定URL是否为恶意网址。按照本申请的技术方案,对恶意网址的检测结果准确,能够检测各种新出现的恶意网址,而且用户友好,用户只需要上传URL,无需提供更多信息。




CN 104766014 B

1.一种检测恶意网址的方法,包括:

接收用户上报的统一资源定位符URL;

获取与所述URL关联的超文本传输协议HTTP请求链,所述HTTP请求链是包含访问所述URL的多次HTTP请求-响应交互信息的时序链表;以及

从所述HTTP请求链提取特征以确定所述URL是否为恶意网址。

2.根据权利要求1所述的方法,其中,获取HTTP请求链包括:

利用基于用户地理位置的分布式动态爬虫子系统来获取HTTP请求链。

3.根据权利要求2所述的方法,其中,利用基于用户地理位置的分布式动态爬虫子系统来获取HTTP请求链包括:

确定所述用户所在的地理位置和网络环境信息;

将所述URL调度至地理位置和网络环境信息与所述用户接近的动态爬虫服务器;以及

在所述动态爬虫服务器处下载与所述URL关联的网页内容以获得HTTP请求链。

4.根据权利要求3所述的方法,其中,确定所述用户所在的地理位置和网络环境信息包括:

基于所述用户上报URL的互联网协议IP地址确定所述用户的地理位置以及所使用的网络运营商信息;以及

基于所述网络运营商信息确定所述用户的网络环境信息,其中所述网络环境信息至少包括网络带宽。

5.根据权利要求3所述的方法,其中,下载与所述URL关联的网页内容以获得HTTP请求链包括:

抓取经过跳转的网页内容并保存中间结果。

6.根据权利要求5所述的方法,其中,抓取经过跳转的网页内容包括以下至少一项:

利用浏览器的排版引擎对超文本标记语言文档对象模型HTML DOM树进行渲染,以抓取通过HTML文档中的内联框架iframe标签进行跳转的网页内容;

通过JavaScript引擎执行JavaScript代码,以抓取通过JavaScript代码进行跳转的网页内容;以及

通过Flash播放器插件执行Flash以抓取通过Flash进行跳转的网页内容。

7.根据权利要求1-6任一所述的方法,其中,从所述HTTP请求链提取特征以确定所述URL是否为恶意网址包括:

从所述HTTP请求链中提取以下至少一个维度的特征:上下游信息,服务器维度,网页编程语言维度,时间维度,网页自身描述信息;以及

利用建立的、经过机器学习的分类模型,基于所提取的特征确定所述URL是正常网址还是可疑恶意网址。

8.根据权利要求7所述的方法,其中,

所述上下游信息包括以下至少一项信息:302跳转次数,404页面占比,子URL是否包含广告联盟链接,子URL是否包含恶意子链接,子URL是否包含小型网站统计工具;

所述服务器维度包括以下至少一项信息:是否为境外互联网协议IP地址,是否是Windows IIS,是否采用内容分发网络CDN技术,是否是kangle服务器,是否是netbox服务器,是否是nginx服务器,是否是apache服务器,是否是多媒体视频;