



(12)发明专利申请

(10)申请公布号 CN 110445746 A

(43)申请公布日 2019. 11. 12

(21)申请号 201810417310.1

(22)申请日 2018.05.04

(71)申请人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区
科技中一路腾讯大厦35层

(72)发明人 黄宝权

(74)专利代理机构 深圳翼盛智成知识产权事务
所(普通合伙) 44300

代理人 黄威

(51) Int. Cl.

H04L 29/06(2006.01)

H04L 9/06(2006.01)

H04L 29/08(2006.01)

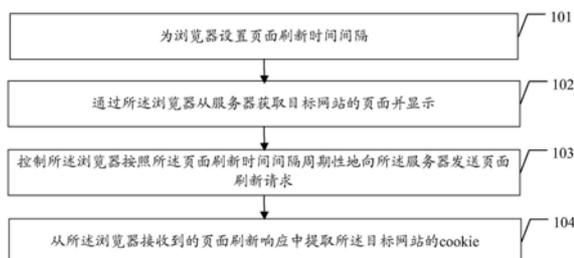
权利要求书2页 说明书14页 附图4页

(54)发明名称

cookie获取方法、装置及存储设备

(57)摘要

本发明实施例提供了一种cookie获取方法、装置及存储设备,其中,cookie获取方法包括:为浏览器设置页面刷新时间间隔;通过所述浏览器从服务器获取目标网站的页面并显示;控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求;从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。本发明实施例能够避免模拟请求获取cookie不完备的问题,且不需要花费很大的代价就能实现cookie的获取。



1. 一种cookie获取方法,其特征在于,包括:
为浏览器设置页面刷新时间间隔;
通过所述浏览器从服务器获取目标网站的页面并显示;
控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求;
从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。
2. 根据权利要求1所述的cookie获取方法,其特征在于,所述为浏览器设置页面刷新时间间隔,包括:
按照cookie的有效期为所述浏览器设置所述页面刷新时间间隔。
3. 根据权利要求1所述的cookie获取方法,其特征在于,所述通过所述浏览器从服务器获取目标网站的页面并显示,包括:
通过所述浏览器向所述服务器发送携带用户登陆信息的页面访问请求;
从所述浏览器接收到的页面访问响应中解析出所述目标网站的页面并显示。
4. 根据权利要求1所述的cookie获取方法,其特征在于,所述方法还包括:
通过所述浏览器向所述服务器发送携带所述目标网站的cookie的页面访问请求,以实现自动登录并访问所述目标网站。
5. 根据权利要求1所述的cookie获取方法,其特征在于,所述方法还包括:
将所述目标网站的cookie发送给网络爬虫,以使得所述网络爬虫携带所述目标网站的cookie自动登陆所述目标网站并爬取所述目标网站的网页内容。
6. 根据权利要求1至5任意一项所述的cookie获取方法,其特征在于,所述从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie,包括:
调用浏览器获取cookie的应用程序编程接口API,从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。
7. 根据权利要求1至5任意一项所述的cookie获取方法,其特征在于,所述方法还包括:
将所述目标网站的cookie加密并保存。
8. 一种cookie获取装置,其特征在于,包括:
设置单元,用于为浏览器设置页面刷新时间间隔;
获取单元,用于通过所述浏览器从服务器获取目标网站的页面并显示;
控制单元,用于控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求;
提取单元,用于从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。
9. 根据权利要求8所述的cookie获取装置,其特征在于,
所述设置单元具体用于,按照cookie的有效期为所述浏览器设置所述页面刷新时间间隔。
10. 根据权利要求8所述的cookie获取装置,其特征在于,所述获取单元包括:
发送子单元,用于通过所述浏览器向所述服务器发送携带用户登陆信息的页面访问请求;
解析子单元,用于从所述浏览器接收到的页面访问响应中解析出所述目标网站的页面并显示。

11. 根据权利要求8所述的cookie获取装置,其特征在于,所述装置还包括:

第一发送单元,用于通过所述浏览器向所述服务器发送携带所述目标网站的cookie的页面访问请求,以实现自动登录并访问所述目标网站。

12. 根据权利要求8所述的cookie获取装置,其特征在于,所述装置还包括:

第二发送单元,用于将所述目标网站的cookie发送给网络爬虫,以使得所述网络爬虫携带所述目标网站的cookie自动登陆所述目标网站并爬取所述目标网站的网页内容。

13. 根据权利要求8至12任意一项所述的cookie获取装置,其特征在于,

所述提取单元具体用于,调用浏览器获取cookie的应用程序编程接口API,从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

14. 根据权利要求8至12任意一项所述的cookie获取装置,其特征在于,所述装置还包括:

存储单元,用于将所述目标网站的cookie加密并保存。

15. 一种存储设备,其特征在于,所述存储设备用于储存多条指令,所述指令适于由处理器加载并执行如权利要求1至7任一项所述的cookie获取方法。

cookie获取方法、装置及存储设备

技术领域

[0001] 本发明实施例涉及互联网技术领域,具体涉及一种cookie获取方法、装置及存储设备。

背景技术

[0002] cookie,是指网站为了辨别用户身份、进行会话跟踪而储存在用户本地终端上的数据,cookie具有有效期。

[0003] 当用户第一次访问网站时,网站的服务器会为用户创建cookie,并将创建的cookie保存在用户终端侧,cookie中通常记载有一些关键的通信信息,例如:用户登陆信息、个人信息、个性化设置、历史访问记录等。用户在cookie的有效期内再次访问相同的网站时,终端会将cookie携带在操作请求中一起发送给网站的服务器,服务器根据所述cookie中携带的信息响应终端的操作请求;比如,根据所述cookie中携带的用户登陆信息对用户进行身份验证,如果验证通过,则允许用户登陆该网站,从而实现了自动登陆。

[0004] 由于cookie具有有效期,有效期内的cookie才可以使用,为了能够一直持有有效的cookie,现有技术往往通过模拟请求实现,例如,通过编程模拟向服务器发送一个携带cookie的超文本传输协议(Hyper Text Transfer Protocol,HTTP)请求,以请求获取新的cookie。模拟请求获取cookie的方式,容易出现模拟不完备的问题,且随着对方服务器防范策略的更新,需要不断对模拟方法进行改进,需要花费较大的代价。

发明内容

[0005] 本发明实施例提供了一种cookie获取方法、装置及存储设备,能够避免模拟不完备的问题,且不需要花费很大的代价就实现了cookie的获取。

[0006] 本发明实施例提供的cookie获取方法,包括:

[0007] 为浏览器设置页面刷新时间间隔;

[0008] 通过所述浏览器从服务器获取目标网站的页面并显示;

[0009] 控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求;

[0010] 从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0011] 本发明实施例提供的cookie获取装置,包括:

[0012] 设置单元,用于为浏览器设置页面刷新时间间隔;

[0013] 获取单元,用于通过所述浏览器从服务器获取目标网站的页面并显示;

[0014] 控制单元,用于控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求;

[0015] 提取单元,用于从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0016] 本发明实施例还提供了一种存储设备,所述存储设备用于储存多条指令,所述指

令适于由处理器加载并执行如本发明实施例提供的cookie获取方法。

[0017] 本发明实施例中,会为浏览器设置页面刷新时间间隔,在通过所述浏览器从服务器获取目标网站的页面并显示之后,可以控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求,然后从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie;即本发明实施例中,通过页面刷新来获取目标网站的cookie,而页面刷新是浏览器的真实请求行为,避免了模拟请求获取cookie不完备的问题,浏览器在进行页面刷新时,相当于真实用户在使用该网站,对方服务器很难对此技术方案进行防范,从而达到了以不变应万变的目的,继而简单又成功的获取到了所需要的cookie,另外,由于页面刷新周期性地进行,因而可以实现一直持有目标网站的有效cookie。

附图说明

[0018] 图1是本发明实施例所提供的cookie获取系统的场景示意图。

[0019] 图2是本发明实施例所提供的cookie获取方法的流程示意图。

[0020] 图3a是本发明实施例所提供的cookie获取方法的另一流程示意图。

[0021] 图3b是本发明实施例所提供的终端与服务器之间的信令交互示意图。

[0022] 图4是本发明实施例所提供的cookie获取方法的又一流程示意图。

[0023] 图5是本发明实施例所提供的cookie获取装置的结构示意图。

[0024] 图6是本发明实施例所提供的cookie获取装置的另一结构示意图。

[0025] 图7是本发明实施例所提供的终端的结构示意图。

具体实施方式

[0026] 请参照图示,其中相同的组件符号代表相同的组件,本申请的原理是以实施在一适当的运算环境中来举例说明。以下的说明是基于所例示的本申请具体实施例,其不应被视为限制本申请未在此详述的其它具体实施例。

[0027] 在以下的说明中,本申请的具体实施例将参考由一部或多部计算机所执行的步骤及符号来说明,除非另有述明。因此,这些步骤及操作将有数次提到由计算机执行,本文所指的计算机执行包括了由代表了以一结构化型式中的数据的电子信号的计算机处理单元的操作。此操作转换该数据或将其维持在该计算机的内存系统中的位置处,其可重新配置或另外以本领域测试人员所熟知的方式来改变该计算机的运作。该数据所维持的数据结构为该内存的实体位置,其具有由该数据格式所定义的特定特性。但是,本申请原理以上述文字来说明,其并不代表为一种限制,本领域测试人员将可了解到以下所述的多种步骤及操作亦可实施在硬件当中。

[0028] 本文所使用的术语“模块”可看做为在该运算系统上执行的软件对象。本文所述的不同组件、模块、引擎及服务可看做为在该运算系统上的实施对象。而本文所述的装置及方法可以以软件的方式进行实施,当然也可在硬件上进行实施,均在本申请保护范围之内。

[0029] 本申请中的术语“第一”、“第二”和“第三”等是用于区别不同对象,而不是用于描述特定顺序。此外,术语“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或模块的过程、方法、系统、产品或设备没有限定于已列出的步骤或模块,而是某些实施例还包括没有列出的步骤或模块,或某些实施例还包括对于这些过程、

方法、产品或设备固有的其它步骤或模块。

[0030] 在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

[0031] 请参阅图1，图1是本发明实施例的cookie获取系统的场景示意图，该cookie获取系统中包括cookie获取装置，cookie获取装置具体可以集成在终端中，终端例如手机、平板电脑、个人计算机等安装有浏览器的终端。另外，本发明实施例的cookie获取系统中还包括服务器，服务器可以是网页(web)服务器，即目标网站的网页服务器。

[0032] 具体在本发明实施例中，终端可以为其内安装的浏览器设置页面刷新时间间隔(例如5秒、1分钟等)，然后通过该浏览器从目标网站的服务器获取目标网站的页面并显示，此后，终端可以控制该浏览器按照所设置的页面刷新时间间隔周期性地向目标网站的服务器发送页面刷新请求，每发送一次页面刷新请求，终端都会通过浏览器接收到服务器反馈的对应的页面刷新响应，终端可以从浏览器每次接收到的页面刷新响应中提取所述目标网站的cookie，如此一来，由于页面刷新的周期性进行，终端将一直持有目标网站的有效cookie。

[0033] 需要说明的是，图1所示的cookie获取系统的场景示意图仅仅是一个示例，本发明实施例描述的cookie获取系统及场景是为了更加清楚的说明本发明实施例的技术方案，并不构成对于本发明实施例提供的技术方案的限定，本领域普通技术人员可知，随着cookie获取系统的演变和新业务平台的出现，本发明实施例的技术方案对于类似的技术问题，同样适用。

[0034] 以下将分别进行详细说明。

[0035] 本实施例将从cookie获取装置的角度描述本发明实施例提供的cookie获取方法，该cookie获取装置可以集成在终端中，终端内安装有浏览器，该浏览器可以包括但不限于谷歌Google浏览器、QQ浏览器、360浏览器、火狐浏览器等，如图2所示，本实施例的cookie获取方法包括以下步骤：

[0036] 步骤101、为浏览器设置页面刷新时间间隔。

[0037] 具体地，终端可以根据实际需求为浏览器设置页面刷新时间间隔。例如，终端可以根据cookie的有效期为浏览器设置页面刷新时间间隔，该页面刷新时间间隔小于cookie的有效期限，cookie的有效期限越长，该时间间隔相对来说，可以设置的越长。例如，当cookie的有效期限为10秒的时候，可以将页面刷新时间间隔设置为3秒，则后续终端可以控制浏览器每3秒向服务器发送一次页面刷新请求。再例如，当cookie的有效期限为30秒的时候，可以将页面刷新时间间隔设置为10秒，则后续终端可以控制浏览器每10秒向服务器发送一次页面刷新请求，这样保证了终端侧始终持有目标网站的有效cookie，且节省了信令开销。

[0038] 步骤102、通过所述浏览器从服务器获取目标网站的页面并显示。

[0039] 当终端的用户需要通过浏览器访问目标网站的页面时，终端可以通过所述浏览器从目标网站的服务器获取目标网站的页面并显示。

[0040] 例如，终端的用户需要访问某个需要进行身份认证的目标网站，则在首次访问时，终端可以通过浏览器向目标网站的服务器发送携带用户登陆信息的页面访问请求，该用户

登录信息可由用户手动在终端浏览器的对应输入栏输入,该用户登陆信息中可以包括终端的用户在注册目标网站时设置的用户名、账号、密码等信息。

[0041] 服务器在接收到浏览器发送的页面访问请求之后,会提取页面访问请求中的用户登陆信息,根据该用户登陆信息对终端的用户进行身份认证,认证成功时,会生成目标网站的cookie(即原始cookie),将目标网站的原始cookie以及用户想要访问的页面携带在页面访问响应中发送给浏览器。

[0042] 该原始cookie中可以包括上述用户登陆信息,当然还可以包括用户的个人信息、个性化设置、历史访问记录等信息,此处不做具体限定,该原始cookie具有有效期,该有效期由服务器进行设置。

[0043] cookie中存储数据通常采用“名—值”对儿的方式进行,将数据存入cookie可以采用setcookie(name,value)形式,例如将用户名放入到cookie中,即可采用setcookie(“userName”,userName)方式;另外,将数据从cookie中取出可以采用getcookie(name,value)形式,例如将用户名从cookie中取出,即可采用userName=getcookie(“userName”)的方式。

[0044] 原始cookie的发送可以通过页面访问响应报文的头部实现,服务器可以在页面访问响应报文头部设置该原始cookie,一个具体的cookie可如下:

[0045] Set-cookie:“name=value;domain=.domain.com;path=/;expires=Sat,11Jun2016 11:29:42”;

[0046] 其中,在本实施例中,name=value,name、value、expires是需要填入的项,其它项是可选项。

[0047] cookie的主要构成如下:

[0048] 名称(name):一个唯一确定的cookie名称。

[0049] 值(value):存储在cookie中的字符串值。

[0050] 域(domain):指明cookie对于哪个域是有效的。

[0051] 路径(path):表示这个cookie影响到的路径,终端的浏览器跟会根据这项配置,像指定域中匹配的路径发送cookie。

[0052] 有效期(expires):cookie的失效时间。

[0053] 终端可以从浏览器接收到的页面访问响应中提取目标网站的原始cookie,以及从该页面访问响应中解析出目标网站的页面并显示。

[0054] 步骤103、控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求。

[0055] 由于目标网站的原始cookie具有有效期,超过有效期的cookie即为无效cookie,因此为了一直持有目标网站的有效cookie,在浏览器显示目标网站的页面之后,终端可以控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求(即可以在用户没有操作页面的情况下,周期性地模拟用户刷新)。

[0056] 具体地,在控制浏览器第一次向服务器发送页面刷新请求时,终端可以将目标网站的原始cookie携带在该页面刷新请求中,服务器在接收到该页面刷新请求之后,会利用该原始cookie对用户进行身份认证,认证成功之后,根据用户的访问情况生成该目标网站的最新cookie,最新cookie中同样包含上述用户登陆信息,最新cookie同样具有有效期,服

务器可以将最新cookie携带在页面刷新响应的报文头部发送给终端。

[0057] 此后,当控制浏览器再次向服务器发送页面刷新请求时,再次发送的页面刷新请求中可以携带上一次得到的cookie。

[0058] 步骤104、从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0059] 在浏览器每次接收到服务器反馈的页面刷新响应之后,终端可以调用浏览器获取cookie的应用程序编程接口(Application Programming Interface,API),从浏览器每次接收到的页面刷新响应的头部提取所述目标网站的最新cookie。

[0060] 终端可以将提取的目标网站的cookie保存在终端的内存中,目标网站可以包括多个,终端可以将每个目标网站的标识(例如:域名)及该目标网站的cookie对应保存,后续在访问某个目标网站的时候,从内存中提取对应的cookie即可。

[0061] 当目标网站有多个时,多个目标网站的cookie的有效期可以不同,则在为浏览器设置页面刷新时间间隔时,可以针对不同的目标网站为浏览器设置不同的页面刷新时间间隔。

[0062] 由于目标网站的cookie中包含用户登陆信息,如果以明文的方式保存该cookie,存在用户登陆信息泄露的风险。因而,本实施例中,可以将目标网站的cookie加密后保存,具体的加密方法例如:可以对cookie进行256位的高级加密标准(Advanced Encryption Standard,AES)加密,当然还可以采用其他方式进行加密,此处不作具体限定。将cookie加密之后,攻击者即使扫描到了该cookie,也无法获知该cookie内的具体信息。

[0063] 另外,上述页面刷新请求中携带的目标网站的cookie可以是明文cookie,也可以是加密后的cookie,如果是加密后的cookie,则服务器在获取该cookie之后,可以先对该cookie解密,以进一步避免cookie内信息泄露。

[0064] 另外,因为每个cookie都有一个有效期,过了该有效期,该cookie就会失效,具体实现中,在得到目标网站的最新cookie之后,终端可以更改上次得到的目标网站的cookie的有效期,例如将上次得到的目标网站的cookie的有效期设置为当前时间减去1秒,以使得上次得到的cookie过期,则后续访问可以直接使用该最新cookie。

[0065] 本实施例中,会为浏览器设置页面刷新时间间隔,在通过所述浏览器从服务器获取目标网站的页面并显示之后,可以控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求,然后从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie;即本发明实施例中,通过页面刷新来获取目标网站的cookie,而页面刷新是浏览器的真实请求行为,避免了模拟请求获取cookie不完备的问题,浏览器在进行页面刷新时,相当于真实用户在使用该网站,对方服务器很难对此技术方案进行防范,从而达到了以不变应万变的目的,继而简单又成功的获取到了所需要的cookie,另外,由于页面刷新周期性的进行,因而可以实现一直持有目标网站的有效cookie。

[0066] 上述实施例描述的方法,下面两个实施例将举例做进一步的详细说明,如图3a所示,本实施例的cookie获取方法包括如下步骤:

[0067] 步骤201、为浏览器设置页面刷新时间间隔。

[0068] 具体地,终端可以根据目标网站的cookie的有效期为其内安装的浏览器设置页面刷新时间间隔,该页面刷新时间间隔小于cookie的有效时长,cookie的有效期越长,该时间间隔相对来说,可以设置的越长,以减少信令开销。例如,当cookie的有效期为30秒的时候,

可以将页面刷新时间间隔设置为10秒,则后续终端的浏览器可以每10秒向服务器发送一次页面刷新请求;再例如,当cookie的有效期为5分钟的时候,可以将页面刷新时间间隔设置为1分钟,则后续终端的浏览器可以每1分钟向服务器发送一次页面刷新请求。

[0069] 本实施例中,目标网站可以有多个,多个目标网站的cookie的有效期限可以不同,在设置页面刷新时间间隔时,可以针对不同的目标网站为浏览器设置不同的页面刷新时间间隔。例如,目标网站1的cookie的有效期限为30秒,则针对目标网站1,可以为浏览器设置10秒的页面刷新时间间隔;再例如,目标网站2的cookie的有效期限为5分钟,则可以为浏览器设置1分钟的页面刷新时间间隔,在实际应用中,可以根据用户访问的目标网站,选取对应的页面刷新时间间隔控制浏览器进行自动页面刷新。

[0070] 步骤202、通过所述浏览器向所述服务器发送携带用户登陆信息的页面访问请求。

[0071] 例如,终端的用户需要访问某个需要进行身份认证的目标网站,则在首次访问时,终端可以通过浏览器向目标网站的服务器发送携带用户登陆信息的页面访问请求,该用户登陆信息可由用户手动在终端浏览器的对应输入栏输入,该用户登陆信息中可以包括终端的用户在注册目标网站时设置的用户名、账号、密码等信息。

[0072] 服务器在接收到浏览器发送的页面访问请求之后,会提取页面访问请求中的用户登陆信息,根据该用户登陆信息对终端的用户进行身份认证,认证成功时,会生成目标网站的cookie(此处将通过手动登录得到的cookie记为第一cookie),将目标网站的第一cookie及用户想要访问的页面携带在页面访问响应中发送给终端的浏览器。

[0073] 该第一cookie中可以包括上述用户登陆信息,当然还可以包括用户的个人信息、个性化设置、历史访问记录等信息,此处不做具体限定,该第一cookie具有有效期,该有效期由服务器进行设置。

[0074] 步骤203、从所述浏览器接收到的页面访问响应中解析出所述目标网站的页面并显示。

[0075] 在浏览器接收到服务器反馈的页面登陆响应之后,终端可以解析该页面登陆响应,以获取目标网站的页面并显示,另外,终端还可以调用浏览器获取cookie的API,从所述页面登陆响应的头部提取所述目标网站的第一cookie。

[0076] 步骤204、判断是否到刷新时间,如果到了刷新时间,则执行步骤205,否则执行步骤209,继续等待刷新。

[0077] 步骤205、控制所述浏览器按照向所述服务器发送页面刷新请求。

[0078] 即终端控制浏览器按照针对用户当前访问的目标网站为浏览器设置的页面刷新时间间隔,周期性地向所述服务器发送页面刷新请求,以请求刷新浏览器上当前显示的网页,每次发送的页面刷新请求中可以携带上次得到的cookie。

[0079] 服务器在接收到该页面刷新请求之后,会利用页面刷新请求中携带cookie对用户进行身份认证,认证成功之后,根据用户的访问情况生成该目标网站的最新cookie(此处将通过自动页面刷新得到的cookie记为第二cookie),第二cookie中同样包含上述用户登陆信息,第二cookie同样具有有效期,服务器可以将第二cookie携带在页面刷新响应的报文头部发送给终端的浏览器。

[0080] 步骤206、从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0081] 在浏览器接收到服务器反馈的页面刷新响应之后,终端同样可以调用浏览器获取

cookie的应用程序编程接口(Application Programming Interface,API),从所述页面刷新响应的头部提取所述目标网站的第二cookie。

[0082] 需要说明的是,当浏览器第一次向服务器发送页面刷新请求时,携带的cookie可以是上述第一cookie,后续再次向服务器发送页面刷新请求时,携带的cookie可以是上一次得到的第二cookie。

[0083] 步骤207、将所述目标网站的cookie加密并保存。

[0084] 终端可以将提取的目标网站的cookie保存在终端的内存中,当目标网站有多个,终端可以将每个目标网站的标识(例如:域名)及该目标网站的cookie对应保存,后续在访问某个目标网站的时候,从内存中提取对应的cookie即可。

[0085] 由于目标网站的cookie中包含用户登陆信息,如果以明文的方式保存该cookie,存在用户登陆信息泄露的风险。因而,本实施例中,可以将目标网站的cookie加密后保存,具体的加密方法例如:可以对cookie进行256位的高级加密标准(Advanced Encryption Standard,AES)加密,当然还可以采用其他方式进行加密,此处不作具体限定。将cookie加密之后,攻击者即使扫描到了该cookie,也无法获知该cookie内的具体信息。

[0086] 步骤208、通过所述浏览器向所述服务器发送页面访问请求,所述页面访问请求中包括所述第二cookie,以实现自动登录并访问所述目标网站。

[0087] 在得到第二cookie之后,当终端的用户需要再次访问目标网站的页面时,终端可以从内存中提取该目标网站的第二cookie,通过浏览器直接向目标网站的服务器发送携带第二cookie的页面访问请求,服务器在接收到该页面访问请求之后,会提取第二cookie对用户进行身份认证,认证成功之后,向终端返回所需的页面,在这个过程中,服务器根据第二cookie实现了登陆目标网站的自动认证,避免了要求用户再次输入用户登陆信息进行身份认证,简化了操作。

[0088] 上面从终端侧描述了本发明实施例提供的cookie获取方法,下面将结合服务器对本发明实施例提供的cookie获取方法进行进一步地描述,请参阅图3b,在cookie获取过程中,终端与服务器之间的信令交互可如下:

[0089] 首先,终端为浏览器设置页面刷新时间间隔,在用户第一次访问目标网站时,终端可以通过浏览器向目标网站的服务器发送携带用户登录信息的页面访问请求;服务器在接收到该页面访问请求之后,根据用户登录信息对终端的用户进行身份认证,认证成功时,生成第一cookie,将第一cookie及需要访问的页面携带在页面访问响应中发送给终端;终端从页面访问响应中提取第一cookie并保存;此后,为了避免第一cookie失效后无法再自动登陆目标网站,终端可以在第一cookie失效前,控制浏览器按照所设置的页面刷新时间间隔周期性地向服务器发送页面刷新请求;服务器接收到页面刷新请求之后,会生成第二cookie,将第二cookie携带在页面刷新响应中反馈给终端的浏览器;在浏览器接收到服务器反馈的页面刷新响应之后,终端从该页面刷新响应中提取第二cookie,然后将保存的所述第一cookie更新为所述第二cookie;此后,当终端需要再次访问目标网站时,可以通过浏览器向目标网站的服务器发送携带第二cookie的页面访问请求,服务器在接收到该页面访问请求之后,会提取第二cookie对用户进行身份验证,如果第二cookie是过期的cookie或者是错误的cookie,那么将认证失败,服务器将给终端返回要求进行身份认证的页面;如果第二cookie是有效且正确的cookie,则认证成功,服务器向终端返回所需的页面,在这个过

程中,服务器根据第二cookie实现了登陆目标网站的自动认证,避免了要求用户再次输入用户登陆信息进行身份认证。

[0090] 本实施例中,通过页面刷新来获取目标网站的cookie,而页面刷新是浏览器的真实请求行为,避免了模拟请求获取cookie不完备的问题,浏览器在进行页面刷新时,相当于真实用户在使用该网站,对方服务器很难对此技术方案进行防范,从而达到了以不变应万变的目的,继而简单又成功的获取到了所需要的cookie,另外,由于页面刷新周期性地进行,因而可以实现一直持有目标网站的有效cookie。

[0091] 请参阅图4,本实施例的cookie获取方法包括:

[0092] 步骤301、为浏览器设置页面刷新时间间隔。

[0093] 具体地,终端可以根据目标网站的cookie的有效期为其内安装的浏览器设置页面刷新时间间隔,该页面刷新时间间隔小于cookie的有效时长,cookie的有效期越长,该时间间隔相对来说,可以设置的越长,以减少信令开销。例如,当cookie的有效期为30秒的时候,可以将页面刷新时间间隔设置为10秒,则后续终端的浏览器可以每10秒向服务器发送一次页面刷新请求;再例如,当cookie的有效期为5分钟的时候,可以将页面刷新时间间隔设置为1分钟,则后续终端的浏览器可以每1分钟向服务器发送一次页面刷新请求。

[0094] 本实施例中,目标网站可以有多个,多个目标网站的cookie的有效期可以不同,在设置页面刷新时间间隔时,可以针对不同的目标网站为浏览器设置不同的页面刷新时间间隔。例如,目标网站1的cookie的有效期为30秒,则针对目标网站1,可以为浏览器设置10秒的页面刷新时间间隔;再例如,目标网站2的cookie的有效期为5分钟,则可以为浏览器设置1分钟的页面刷新时间间隔,在实际应用中,可以根据用户访问的目标网站,选取对应的页面刷新时间间隔控制浏览器进行自动页面刷新。

[0095] 步骤302、通过所述浏览器向所述服务器发送携带用户登陆信息的页面访问请求。

[0096] 例如,终端的用户需要访问某个需要进行身份认证的目标网站,则在首次访问时,终端可以通过浏览器向目标网站的服务器发送携带用户登陆信息的页面访问请求,该用户登录信息可由用户手动在终端浏览器的对应输入栏输入,该用户登陆信息中可以包括终端的用户在注册目标网站时设置的用户名、账号、密码等信息。

[0097] 服务器在接收到浏览器发送的页面访问请求之后,会提取页面访问请求中的用户登陆信息,根据该用户登陆信息对终端的用户进行身份认证,认证成功时,会生成目标网站的cookie(此处将通过手动登录得到的cookie记为第一cookie),将目标网站的第一cookie及用户想要访问的页面携带在页面访问响应中发送给终端的浏览器。

[0098] 该第一cookie中可以包括上述用户登陆信息,当然还可以包括用户的个人信息、个性化设置、历史访问记录等信息,此处不做具体限定,该第一cookie具有有效期,该有效期由服务器进行设置。

[0099] 步骤303、从所述浏览器接收到的页面访问响应中解析出所述目标网站的页面并显示。

[0100] 在浏览器接收到服务器反馈的页面登陆响应之后,终端可以解析该页面登陆响应,以获取目标网站的页面并显示,另外,终端还可以调用浏览器获取cookie的应用程序编程接口(Application Programming Interface,API),从所述页面登陆响应的头部提取所述目标网站的第一cookie。

[0101] 步骤304、判断是否到刷新时间,如果到了刷新时间,则执行步骤305,否则执行步骤308,继续等待刷新。

[0102] 步骤305、控制所述浏览器按照向所述服务器发送页面刷新请求;

[0103] 即终端控制浏览器按照针对用户当前访问的目标网站为浏览器设置的页面刷新时间间隔,周期性地向所述服务器发送页面刷新请求,以请求刷新浏览器上当前显示的网页,每次发送的页面刷新请求中可以携带上次得到的cookie。

[0104] 服务器在接收到该页面刷新请求之后,会利用页面刷新请求中携带cookie对用户进行身份认证,认证成功之后,根据用户的访问情况生成该目标网站的最新cookie(此处将通过自动页面刷新得到的cookie记为第二cookie),第二cookie中同样包含上述用户登陆信息,第二cookie同样具有有效期,服务器可以将第二cookie携带在页面刷新响应的报文头部发送给终端的浏览器。

[0105] 步骤306、从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0106] 在浏览器接收到服务器反馈的页面刷新响应之后,终端同样可以调用浏览器获取cookie的应用程序编程接口(Application Programming Interface,API),从所述页面刷新响应的头部提取所述目标网站的第二cookie。

[0107] 需要说明的是,当浏览器第一次向服务器发送页面刷新请求时,携带的cookie可以是上述第一cookie,后续再次向服务器发送页面刷新请求时,携带的cookie可以是上一次得到的第二cookie。

[0108] 步骤307、将所述目标网站的cookie发送给网络爬虫,以使得所述网络爬虫携带所述目标网站的cookie自动登陆所述目标网站并爬取所述目标网站的网页内容。

[0109] 网络爬虫是一个自动提取网页的程序,它从万维网上下载网页,是搜索引擎的重要组成部分,其一般是从一个或若干个初始的网页的统一资源定位符(Uniform Resource Locator,URL)开始,获取初始网页上的URL,在爬取网页的过程中,不断地从当前网页上抽取新的URL放入队列,直至满足一定的停止条件。

[0110] 网络爬虫爬取网站的网页时,大部分网站要求进行身份认证,认证通过之后才可以访问网站的网页,网络爬虫一般会携带网站的cookie实现身份认证,以爬取网页,这就需要网络爬虫一直持有网站的有效cookie。

[0111] 具体在本发明实施例中,可以将网络爬虫需要爬取网页的网站作为目标网站,按照上述方法不断获取该目标网站的cookie,并将获取的cookie提供给网络爬虫,以使得网络爬虫始终持有目标网站的有效cookie,并携带有效cookie自动登陆所述目标网站,以爬取所述目标网站的网页内容。

[0112] 本实施例中,通过页面刷新来获取目标网站的cookie,而页面刷新是浏览器的真实请求行为,避免了模拟请求获取cookie不完备的问题,浏览器在进行页面刷新时,相当于真实用户在使用该网站,对方服务器很难对此技术方案进行防范,从而达到了以不变应万变的目的,继而简单又成功的获取到了所需要的cookie,另外,由于页面刷新周期性地进行,因而可以实现一直持有目标网站的有效cookie。

[0113] 进一步地,终端可以将持有的目标网站的有效cookie提供给网络爬虫,以使得网络爬虫也始终持有目标网站的有效cookie,并携带有效cookie自动登陆所述目标网站,以爬取所述目标网站的网页内容,这种方法可有效应对服务器的各种反爬虫策略,并解决了

网络爬虫的登录认证问题。

[0114] 为了更好地实施以上方法,本发明实施例还提供一种cookie获取装置,如图5所示,本实施例的装置包括设置单元401,获取单元402,控制单元403以及提取单元404,如下:

[0115] 设置单元401,用于为浏览器设置页面刷新时间间隔;

[0116] 获取单元402,用于通过所述浏览器从服务器获取目标网站的页面并显示;

[0117] 控制单元403,用于控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求;

[0118] 提取单元404,用于从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0119] 一实施例中,所述设置单元401具体用于:

[0120] 按照cookie的有效期为所述浏览器设置所述页面刷新时间间隔。

[0121] 一实施例中,如图6所示,所述获取单元402包括:

[0122] 发送子单元4021,用于通过所述浏览器向所述服务器发送携带用户登陆信息的页面访问请求;

[0123] 解析子单元4022,用于从所述浏览器接收到的页面访问响应中解析出所述目标网站的页面并显示。

[0124] 一实施例中,如图6所示,所述装置还包括:

[0125] 第一发送单元405,用于通过所述浏览器向所述服务器发送携带所述目标网站的cookie的页面访问请求,以实现自动登录并访问所述目标网站。

[0126] 一实施例中,如图6所示,所述装置还包括:

[0127] 第二发送单元406,用于将所述目标网站的cookie发送给网络爬虫,以使得所述网络爬虫携带所述目标网站的cookie自动登陆所述目标网站并爬取所述目标网站的网页内容。

[0128] 一实施例中,所述提取单元404具体用于:

[0129] 调用浏览器获取cookie的应用程序编程接口API,从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0130] 一实施例中,如图6所示,所述装置还包括:

[0131] 存储单元407,用于将所述目标网站的cookie加密并保存。

[0132] 需要说明的是,上述实施例提供的cookie获取装置在实现cookie获取时,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将设备的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。另外,上述实施例提供的cookie获取装置与cookie获取方法属于同一构思,其具体实现过程详见方法实施例,此处不再赘述。

[0133] 本实施例中,设置单元会为浏览器设置页面刷新时间间隔,在获取单元通过所述浏览器从服务器获取目标网站的页面并显示之后,控制单元可以控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求,然后提取单元从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie;即本实施例的装置,通过页面刷新来获取目标网站的cookie,而页面刷新是浏览器的真实请求行为,避免了模拟请求获取cookie不完备的问题,浏览器在进行页面刷新时,相当于真实用户在使用该网站,对方服务

器很难对此技术方案进行防范,从而达到了以不变应万变的目的,继而简单又成功的获取到了所需要的cookie,另外,由于页面刷新周期性的进行,因而可以实现一直持有目标网站的有效cookie。

[0134] 相应的,本发明实施例还提供了一种终端,该终端中集成有本发明实施例描述的cookie获取装置,如图7所示,该终端可以包括射频(RF, Radio Frequency)电路501、包括有一个或一个以上计算机可读存储介质的存储器502、输入单元503、显示单元504、传感器505、音频电路506、无线保真(WiFi, Wireless Fidelity)模块507、包括有一个或者一个以上处理核心的处理器508、以及电源509等部件。本领域技术人员可以理解,图7中示出的终端结构并不构成对终端的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。其中:

[0135] RF电路501可用于收发信息或通话过程中,信号的接收和发送,特别地,将基站的下行信息接收后,交由一个或者一个以上处理器508处理;另外,将涉及上行的数据发送给基站。通常,RF电路501包括但不限于天线、至少一个放大器、调谐器、一个或多个振荡器、用户身份模块(SIM, Subscriber Identity Module)卡、收发信机、耦合器、低噪声放大器(LNA, Low Noise Amplifier)、双工器等。此外,RF电路501还可以通过无线通信与网络和其他设备通信。所述无线通信可以使用任一通信标准或协议,包括但不限于全球移动通讯系统(GSM, Global System of Mobile communication)、通用分组无线服务(GPRS, General Packet Radio Service)、码分多址(CDMA, Code Division Multiple Access)、宽带码分多址(WCDMA, Wideband Code Division Multiple Access)、长期演进(LTE, Long Term Evolution)、电子邮件、短消息服务(SMS, Short Messaging Service)等。

[0136] 存储器502可用于存储软件程序以及模块,处理器508通过运行存储在存储器502的软件程序以及模块,从而执行各种功能应用以及数据处理。存储器502可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据终端的使用所创建的数据(比如音频数据、电话本等)等。此外,存储器502可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。相应地,存储器502还可以包括存储器控制器,以提供处理器508和输入单元503对存储器502的访问。

[0137] 输入单元503可用于接收输入的数字或字符信息,以及产生与用户设置以及功能控制有关的键盘、鼠标、操作杆、光学或者轨迹球信号输入。具体地,在一个具体的实施例中,输入单元503可包括触敏表面以及其他输入设备。触敏表面,也称为触摸显示屏或者触控板,可收集用户在其上或附近的触摸操作(比如用户使用手指、触笔等任何适合的物体或附件在触敏表面上或在触敏表面附近的操作),并根据预先设定的程式驱动相应的连接装置。可选的,触敏表面可包括触摸检测装置和触摸控制器两个部分。其中,触摸检测装置检测用户的触摸方位,并检测触摸操作带来的信号,将信号传送给触摸控制器;触摸控制器从触摸检测装置上接收触摸信息,并将它转换成触点坐标,再送给处理器508,并能接收处理器508发来的命令并加以执行。此外,可以采用电阻式、电容式、红外线以及表面声波等多种类型实现触敏表面。除了触敏表面,输入单元503还可以包括其他输入设备。具体地,其他输入设备可以包括但不限于物理键盘、功能键(比如音量控制按键、开关按键等)、轨迹球、鼠

标、操作杆等中的一种或多种。

[0138] 显示单元504可用于显示由用户输入的信息或提供给用户的信息以及终端的各种图形用户接口,这些图形用户接口可以由图形、文本、图标、视频和其任意组合来构成。显示单元504可包括显示面板,可选的,可以采用液晶显示器(LCD,Liquid Crystal Display)、有机发光二极管(OLED,Organic Light-Emitting Diode)等形式来配置显示面板。进一步的,触敏表面可覆盖显示面板,当触敏表面检测到在其上或附近的触摸操作后,传送给处理器508以确定触摸事件的类型,随后处理器508根据触摸事件的类型在显示面板上提供相应的视觉输出。虽然在图7中,触敏表面与显示面板是作为两个独立的部件来实现输入和输入功能,但是在某些实施例中,可以将触敏表面与显示面板集成而实现输入和输出功能。

[0139] 终端还可包括至少一种传感器505,比如光传感器、运动传感器以及其他传感器。具体地,光传感器可包括环境光传感器及接近传感器,其中,环境光传感器可根据环境光线的明暗来调节显示面板的亮度,接近传感器可在终端移动到耳边时,关闭显示面板和/或背光。作为运动传感器的一种,重力加速度传感器可检测各个方向上(一般为三轴)加速度的大小,静止时可检测出重力的大小及方向,可用于识别手机姿态的应用(比如横竖屏切换、相关游戏、磁力计姿态校准)、振动识别相关功能(比如计步器、敲击)等;至于终端还可配置的陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器,在此不再赘述。

[0140] 音频电路506、扬声器、传声器可提供用户与终端之间的音频接口。音频电路506可将接收到的音频数据转换后的电信号,传输到扬声器,由扬声器转换为声音信号输出;另一方面,传声器将收集的声音信号转换为电信号,由音频电路506接收后转换为音频数据,再将音频数据输出处理器508处理后,经RF电路501以发送给比如另一终端,或者将音频数据输出至存储器502以便进一步处理。音频电路506还可能包括耳塞插孔,以提供外设耳机与终端的通信。

[0141] WiFi属于短距离无线传输技术,终端通过WiFi模块507可以帮助用户收发电子邮件、浏览网页和访问流式媒体等,它为用户提供了无线的宽带互联网访问。虽然图7示出了WiFi模块507,但是可以理解的是,其并不属于终端的必须构成,完全可以根据需要在不改变发明的本质的范围内而省略。

[0142] 处理器508是终端的控制中心,利用各种接口和线路连接整个终端的各个部分,通过运行或执行存储在存储器502内的软件程序和/或模块,以及调用存储在存储器502内的数据,执行终端的各种功能和处理数据,从而对终端进行整体监控。可选的,处理器508可包括一个或多个处理核心;优选的,处理器508可集成应用处理器和调制解调处理器,其中,应用处理器主要处理操作系统、用户界面和应用程序等,调制解调处理器主要处理无线通信。可以理解的是,上述调制解调处理器也可以不集成到处理器508中。

[0143] 终端还包括给各个部件供电的电源509(比如电池),优选的,电源可以通过电源管理系统与处理器508逻辑相连,从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。电源509还可以包括一个或一个以上的直流或交流电源、再充电系统、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。

[0144] 尽管未示出,终端还可以包括摄像头、蓝牙模块等,在此不再赘述。具体在本实施例中,终端中的处理器508会按照如下的指令,将一个或一个以上的应用程序的进程对应的可执行文件加载到存储器502中,并由处理器508来运行存储在存储器502中的应用程序,从

而实现各种功能：

[0145] 为浏览器设置页面刷新时间间隔；

[0146] 通过所述浏览器从服务器获取目标网站的页面并显示；

[0147] 控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求；

[0148] 从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0149] 一实施例中，在为浏览器设置页面刷新时间间隔时，处理器508具体用于执行以下步骤：

[0150] 按照cookie的有效期为所述浏览器设置所述页面刷新时间间隔。

[0151] 一实施例中，在通过所述浏览器从服务器获取目标网站的页面并显示时，处理器508具体用于执行以下步骤：

[0152] 通过所述浏览器向所述服务器发送携带用户登陆信息的页面访问请求；

[0153] 从所述浏览器接收到的页面访问响应中解析出所述目标网站的页面并显示。

[0154] 一实施例中，处理器508还用于执行以下步骤：

[0155] 通过所述浏览器向所述服务器发送携带所述目标网站的cookie的页面访问请求，以实现自动登录并访问所述目标网站。

[0156] 一实施例中，处理器508还用于执行以下步骤：

[0157] 将所述目标网站的cookie发送给网络爬虫，以使得所述网络爬虫携带所述目标网站的cookie自动登陆所述目标网站并爬取所述目标网站的网页内容。

[0158] 一实施例中，在从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie时，处理器508具体用于执行以下步骤：

[0159] 调用浏览器获取cookie的应用程序编程接口API，从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0160] 一实施例中，处理器508还用于执行以下步骤：

[0161] 将所述目标网站的cookie加密并保存。

[0162] 本实施例的终端，会为浏览器设置页面刷新时间间隔，在通过所述浏览器从服务器获取目标网站的页面并显示之后，可以控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求，然后从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie；即本发明实施例中，通过页面刷新来获取目标网站的cookie，而页面刷新是浏览器的真实请求行为，避免了模拟请求获取cookie不完备的问题，浏览器在进行页面刷新时，相当于真实用户在使用该网站，对方服务器很难对此技术方案进行防范，从而达到了以不变应万变的目的，继而简单又成功的获取到了所需要的cookie，另外，由于页面刷新周期性地进行，因而可以实现一直持有目标网站的有效cookie。

[0163] 本申请实施例还提供一种存储设备，所述存储设备存储有计算机程序，当所述计算机程序在计算机上运行时，使得所述计算机执行上述任一实施例中的网页图片查看方法，比如：为浏览器设置页面刷新时间间隔；通过所述浏览器从服务器获取目标网站的页面并显示；控制所述浏览器按照所述页面刷新时间间隔周期性地向所述服务器发送页面刷新请求；从所述浏览器接收到的页面刷新响应中提取所述目标网站的cookie。

[0164] 在本申请实施例中，存储设备可以是磁碟、光盘、只读存储器(Read Only Memory，

ROM)、随机存取记忆体(Random Access Memory, RAM)等。

[0165] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中沒有详述的部分,可以参见其他实施例的相关描述。

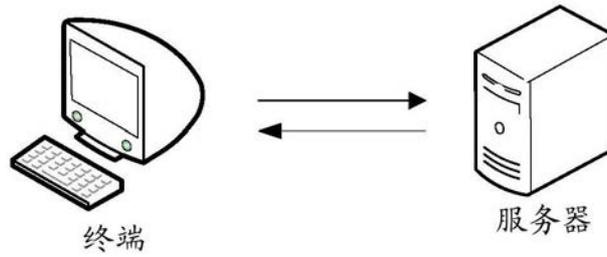


图1

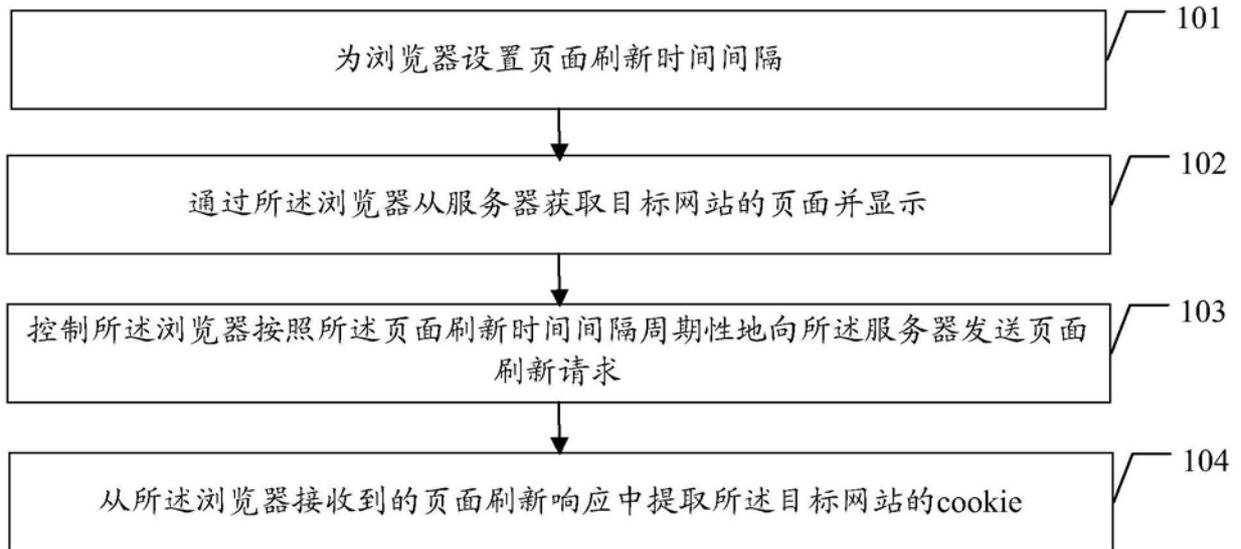


图2

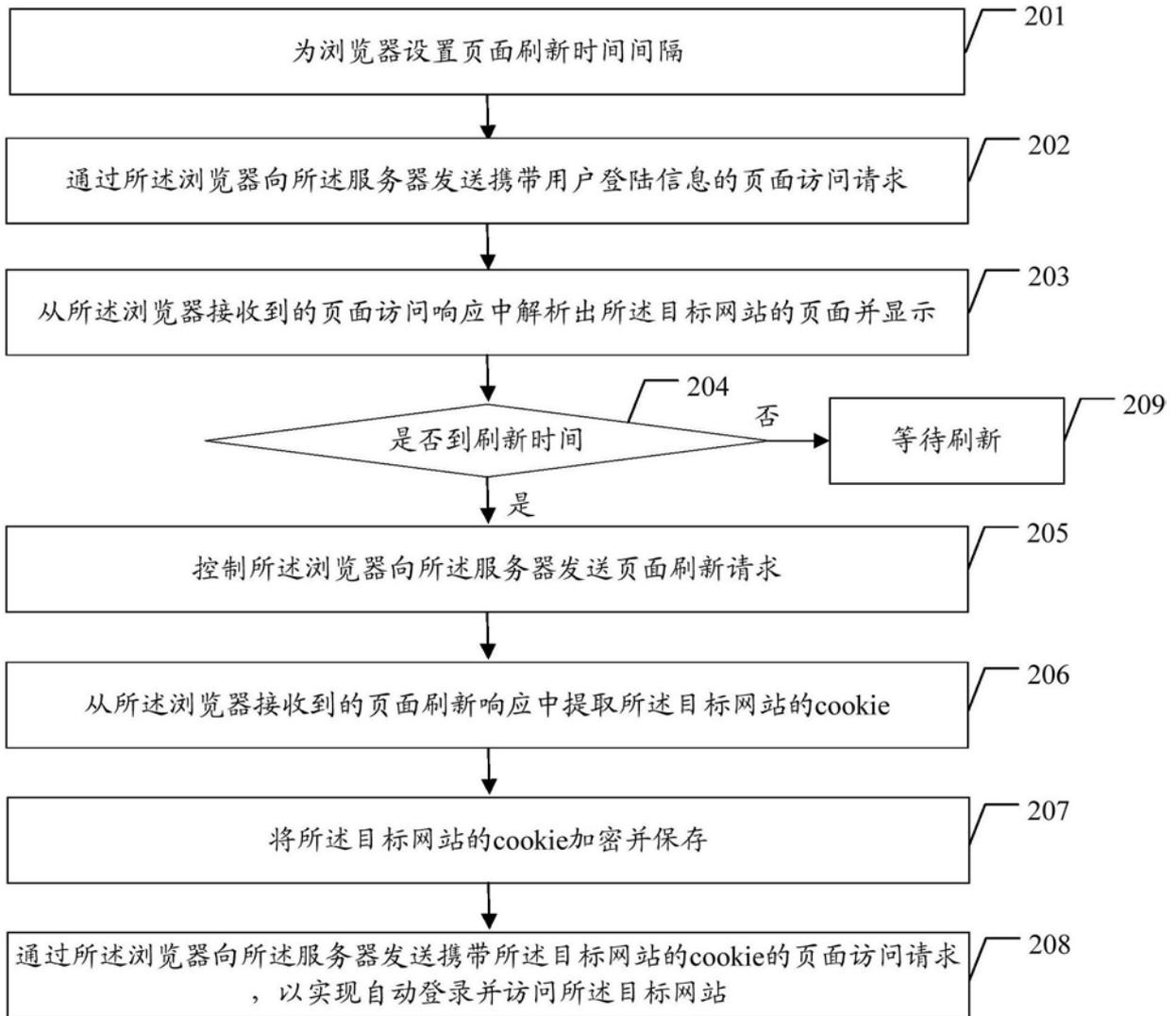


图3a

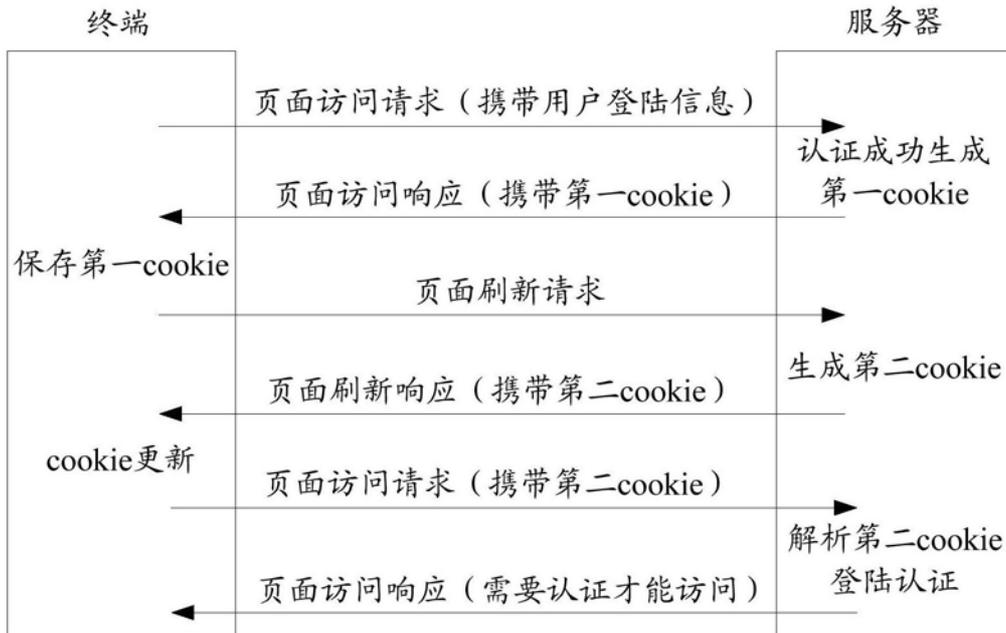


图3b

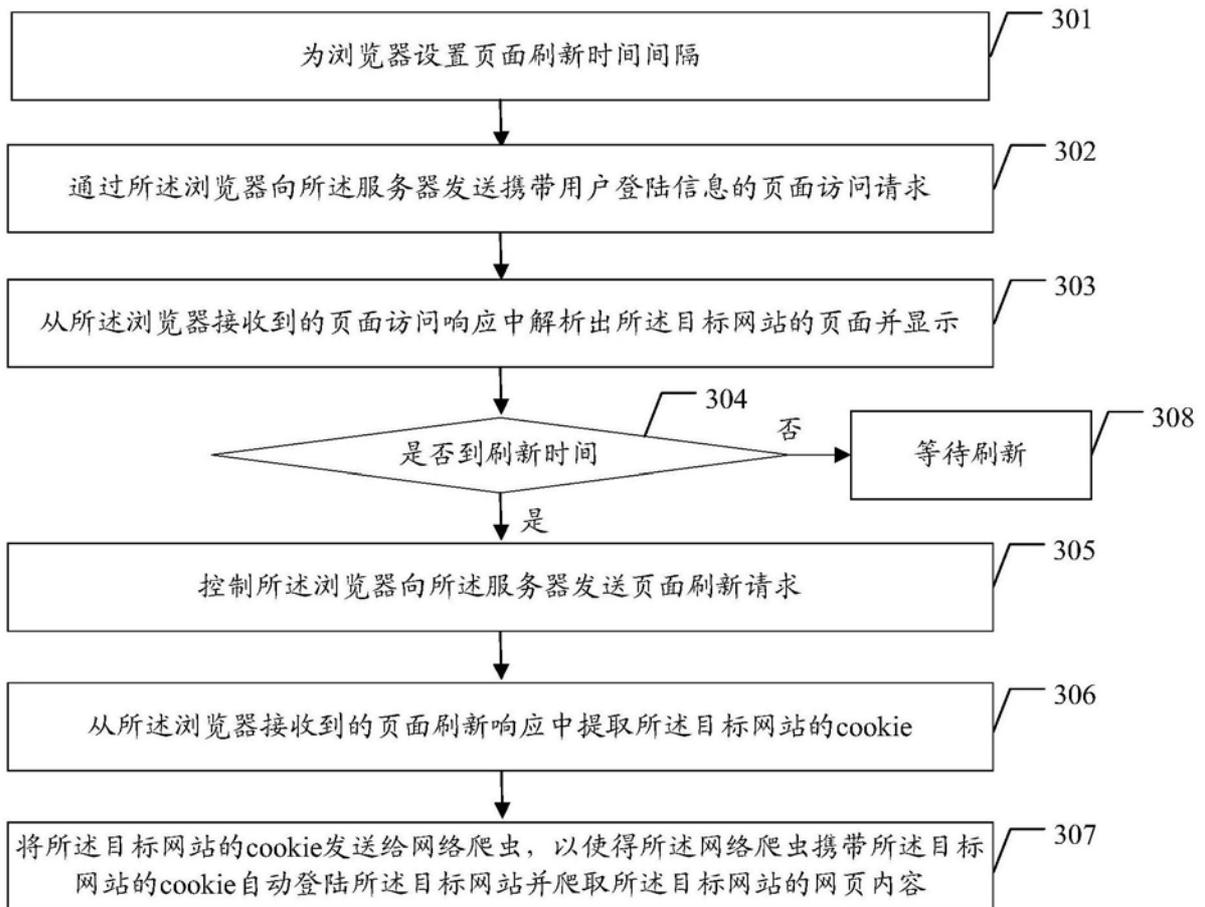


图4

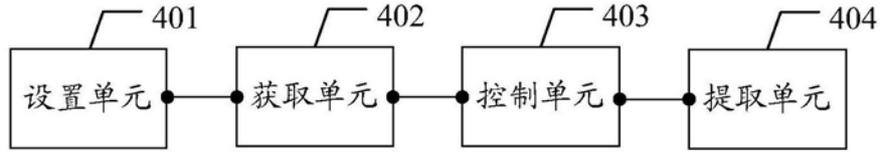


图5

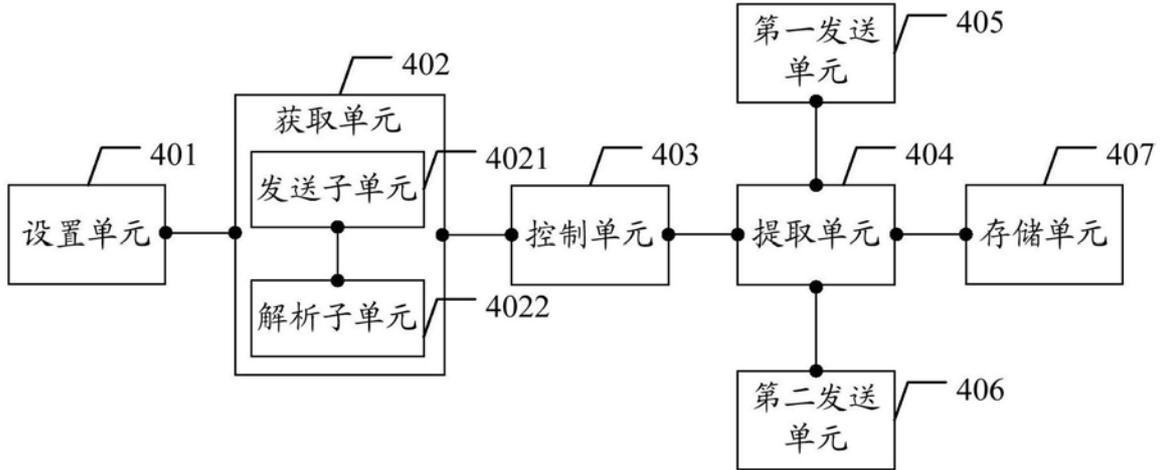


图6

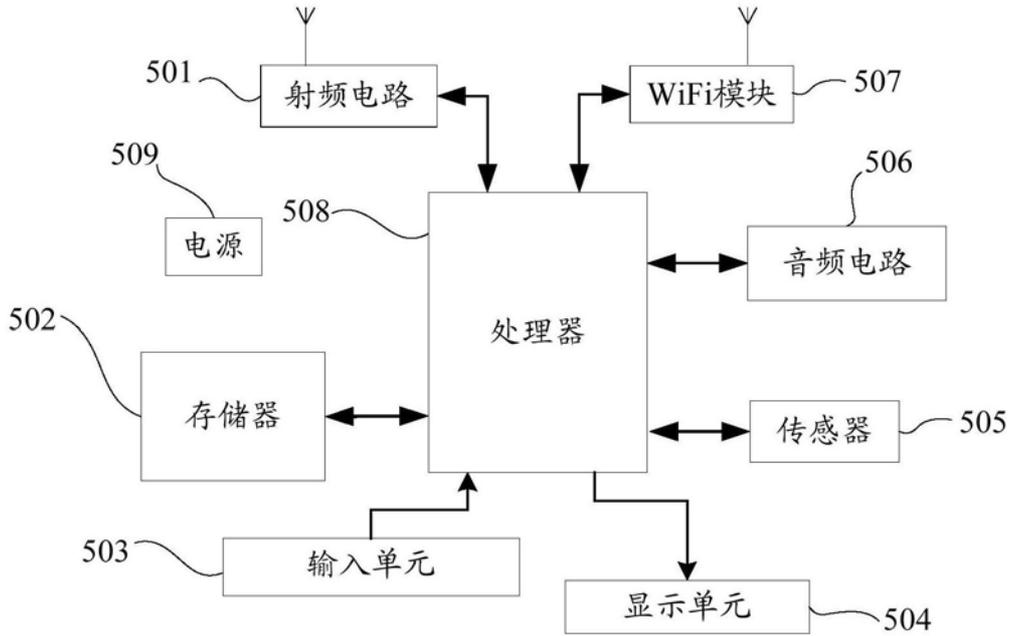


图7