



(12) 发明专利申请

(10) 申请公布号 CN 102546292 A

(43) 申请公布日 2012. 07. 04

(21) 申请号 201110422899. 2

(22) 申请日 2011. 12. 16

(71) 申请人 深信服网络科技(深圳) 有限公司
地址 518000 广东省深圳市南山区麒麟路 1 号南山科技创业服务中心 418、419

(72) 发明人 张洲亭

(74) 专利代理机构 深圳市世纪恒程知识产权代
理事务所 44287
代理人 胡海国

(51) Int. Cl.
H04L 12/26(2006. 01)
H04L 12/24(2006. 01)

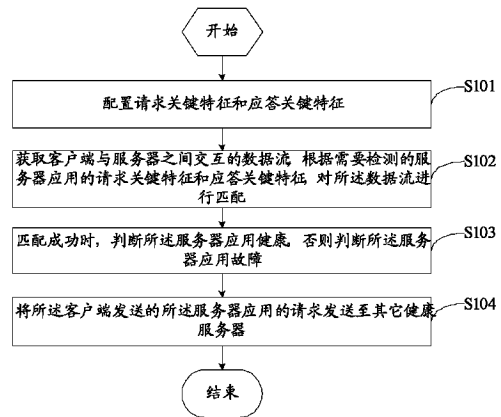
权利要求书 1 页 说明书 5 页 附图 3 页

(54) 发明名称

检测服务器应用健康状态的方法和装置

(57) 摘要

本发明揭示了一种检测服务器应用健康状态的方法和装置,该方法包括:获取客户端与服务器之间交互的数据流;根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配;匹配成功时,判定所述服务器应用健康,否则判定所述服务器应用故障。本发明提供的一种检测服务器应用健康状态的方法和装置,降低了检测设备的负担。



1. 一种检测服务器应用健康状态的方法,其特征在于,包括:
获取客户端与服务器之间交互的数据流;
根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配;
匹配成功时,判定所述服务器应用健康,否则判定所述服务器应用故障。
2. 如权利要求 1 所述的方法,其特征在于,所述根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配包括:
根据所述请求关键特征,对客户端发送的请求数据流执行请求特征匹配;
匹配成功时,根据所述应答关键特征,对服务器的应答数据流执行应答特征匹配。
3. 如权利要求 1 或 2 所述的方法,其特征在于,在执行所述匹配成功时,判定所述服务器应用健康,否则判定所述服务器应用故障之后,包括:
当判定所述服务器故障时,将所述客户端发送的所述服务器应用的请求发送至其它健康服务器。
4. 如权利要求 1 或 2 所述的方法,其特征在于,在执行所述获取客户端与服务器之间交互的数据流之前,还包括:
当交互协议已知时,指定所述请求关键特征;
当交互协议未知时,指定所述请求关键特征和应答关键特征。
5. 如权利要求 4 所述的方法,其特征在于,在执行当交互协议已知时,指定请求关键特征之后,还包括:
根据所述请求关键特征,在已知的交互协议查找对应的应答关键特征。
6. 一种检测服务器应用健康状态的装置,其特征在于,包括:
获取模块,用于获取客户端与服务器之间交互的数据流;
匹配模块,用于根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配;
判定模块,用于匹配成功时,判定所述服务器应用健康,否则判定所述服务器应用故障。
7. 如权利要求 6 所述的装置,其特征在于,所述匹配模块包括:
请求特征匹配单元,用于根据所述请求关键特征,对客户端发送的请求数据流执行请求特征匹配;
应答特征匹配单元,用于匹配成功时,根据所述应答关键特征,对服务器的应答数据流执行应答特征匹配。
8. 如权利要求 6 或 7 所述的装置,其特征在于,还包括:
故障处理模块,用于当判定所述服务器故障时,将所述客户端发送的所述服务器应用的请求发送至其它健康服务器。
9. 如权利要求 6 或 7 所述的装置,其特征在于,还包括:
设置模块,用于当交互协议已知时,指定所述请求关键特征;以及当交互协议未知时,指定所述请求关键特征和应答关键特征。
10. 如权利要求 9 所述的装置,其特征在于,所述设置模块还用于:
根据所述请求关键特征,在已知的交互协议查找对应的应答关键特征。

检测服务器应用健康状态的方法和装置

技术领域

[0001] 本发明涉及到通信领域,特别涉及到一种检测服务器应用健康状态的方法和装置。

背景技术

[0002] 目前通常利用负载均衡设备检测服务器应用健康状态,主要有如下几种方法:一是用 PING 命令检测服务器的通讯是否正常,如果 PING 命令失败,说明服务器不可用,则所有运行在该服务器上的应用也不可用,但是 PING 命令成功却并不能保证服务器上应用是正常的;另一种方法是针对服务器上使用 TCP 协议的应用,向服务器上开放的 TCP 服务端口建立连接,与 PING 命令一样,这种方法也不能完全反映服务器上应用的健康状态。

[0003] 为了解决上述两种方法不能完全真实反映服务器健康状态的问题,可向服务器应用发起一个业务请求并根据服务器的应答判断应用状态。以 WEB 应用为例,如图 1 所示:

[0004] 负载均衡设备等检测设备通过三次握手与待测的服务器建立 TCP 连接后,发送一个 HTTP GET 请求获取一 gif 图片,如果服务器返回 200OK,则认为服务器上 WEB 应用是正常的,否则就认为服务器应用失败。该检测方法一般是获取服务器一个静态的内容(比如一个文本文件)以判断应用的好坏,实际上,对于一个复杂的业务,需要检测设备与服务器多次交互才能完成,下面描述一个典型的 B/S 应用的数据录入过程:

[0005] 如图 2 所示,共需要 6 个步骤:

[0006] 1、检测设备与服务器建立 TCP 连接。

[0007] 2、浏览器获取登录界面。由于登录界面一般由多个元素组成,需要浏览器与服务器多次交互才能完成

[0008] 3、使用帐号、密码登录。

[0009] 4、浏览器向服务器提交业务数据。

[0010] 5、服务器执行业务处理。

[0011] 6、服务器向浏览器返回处理结果。

[0012] 步骤 4~步骤 6 是服务器应用最容易出故障的地方,在服务器应用出故障后,客户端依然可以成功执行步骤 1~步骤 3。此刻通过判断获取服务器上静态页面的方法不能准确的判断服务器应用状态的好坏。而检测设备又不能直接模拟步骤 4 的操作以监控应用,因为需要用户先执行登录,并且需要准备很多配置数据。

[0013] 为了解决上述客户端与待测服务器之间多次数据交互才能判断应用好坏的问题,负载均衡设备等检测设备上一般提供脚本语言的运行环境,允许客户编写脚本程序来实现复杂的服务器应用检测,常用的脚本编程语言有 perl、python 等,由于客户的应用系统差别很大,检测设备厂商要将所有的应用检测脚本内置到检测设备里,大大增加了检测设备的负担。且脚本编写人员也必须具备相当专业的知识,对检测人员的要求高。

发明内容

[0014] 本发明的主要目的为提供一种检测服务器应用健康状态的方法和装置,降低了检测设备的负担。

[0015] 本发明提出一种检测服务器应用健康状态的方法,包括:

[0016] 获取客户端与服务器之间交互的数据流;

[0017] 根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配;

[0018] 匹配成功时,判定所述服务器应用健康,否则判定所述服务器应用故障。

[0019] 优选地,所述根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配包括:

[0020] 根据所述请求关键特征,对客户端发送的请求数据流执行请求特征匹配;

[0021] 匹配成功时,根据所述应答关键特征,对服务器的应答数据流执行应答特征匹配。

[0022] 优选地,在执行所述匹配成功时,判定所述服务器应用健康,否则判定所述服务器应用故障之后,包括:

[0023] 当判定所述服务器故障时,将所述客户端发送的所述服务器应用的请求发送至其它健康服务器。

[0024] 优选地,在执行所述获取客户端与服务器之间交互的数据流之前,还包括:

[0025] 当交互协议已知时,指定所述请求关键特征;

[0026] 当交互协议未知时,指定所述请求关键特征和应答关键特征。

[0027] 优选地,在执行当交互协议已知时,指定请求关键特征之后,还包括:

[0028] 根据所述请求关键特征,在已知的交互协议查找对应的应答关键特征。

[0029] 本发明还提出一种检测服务器应用健康状态的装置,包括:

[0030] 获取模块,用于获取客户端与服务器之间交互的数据流;

[0031] 匹配模块,用于根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配;

[0032] 判定模块,用于匹配成功时,判定所述服务器应用健康,否则判定所述服务器应用故障。

[0033] 优选地,所述匹配模块包括:

[0034] 请求特征匹配单元,用于根据所述请求关键特征,对客户端发送的请求数据流执行请求特征匹配;

[0035] 应答特征匹配单元,用于匹配成功时,根据所述应答关键特征,对服务器的应答数据流执行应答特征匹配。

[0036] 优选地,所述装置还包括:

[0037] 故障处理模块,用于当判定所述服务器故障时,将所述客户端发送的所述服务器应用的请求发送至其它健康服务器。

[0038] 优选地,所述装置还包括:

[0039] 设置模块,用于当交互协议已知时,指定所述请求关键特征;以及当交互协议未知时,指定所述请求关键特征和应答关键特征。

[0040] 优选地,所述设置模块还用于:

[0041] 根据所述请求关键特征,在已知的交互协议查找对应的应答关键特征。

[0042] 本发明提出的一种检测服务器应用健康状态的方法和装置,针对现有的服务器应用健康状态检测方法都是由检测设备主动发起请求,然后检测设备检测请求的应答来判定服务器的健康状态,对于复杂的应用的健康检查,必须由用户在检测设备上编程实现。而在本发明中,不需要检测设备主动发送业务请求,检测设备只是依据检测配置,以及客户端与服务器之间的数据流,被动分析服务器对实际应用请求的应答情况来判定服务器上应用的健康状态,大大简化了检测设备的负担。同时也无需检测设备的管理员具备相当专业的编程知识。

附图说明

- [0043] 图 1 为现有技术检测服务器应用健康状态的数据交互示意图；
[0044] 图 2 为本发明检测服务器应用健康状态的方法一实施例的流程示意图；
[0045] 图 3 为本发明检测服务器应用健康状态的方法一实施例之实例的流程示意图；
[0046] 图 4 为本发明检测服务器应用健康状态的装置一实施例的结构示意图；
[0047] 图 5 为本发明检测服务器应用健康状态的装置一实施例中匹配模块的结构示意图。
[0048] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

- [0049] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。
[0050] 参照图 2,提出本发明检测服务器应用健康状态的方法一实施例,包括：
[0051] 步骤 S101、配置请求关键特征和应答关键特征；
[0052] 本实施例的应用环境包括：一或多个服务器,如各自运行一或多个应用程序的 WEB 服务器；一或多个客户端,向所述一或多个服务器提出应用请求；以及一可获取上述服务器和客户端往来数据流的检测设备,如随时监测服务器负载均衡的负载均衡器。当客户端通过网络发送应用请求时,应用请求可由负载均衡器接收并根据负载平稳技术路由至服务器,本实施例即以负载均衡设备为例进行说明,但在实际应用中,不仅限于负载均衡器,凡有监测数据流功能的检测设置均可适用于本实施例。
[0053] 负载均衡器首先在本端指定需要检测的请求关键特征以及服务器成功应答时的应答关键特征,因此对于已知交互协议的服务器应用,例如 HTTP 协议的应用,负载均衡器能从客户端和服务器的数据流中解析出完整的 HTTP 请求和应答,对于指定的应用请求关键特征,负载均衡器能够根据 HTTP 协议标准,得知指定检测请求的应答关键特征,因此对于该类服务器应用,负载均衡器只需指定需要检测的请求关键特征即可。可对于交互协议未知的服务器应用,由于负载均衡设备无法在客户端与服务器交互的数据流中识别完整的请求和应答,因此需要指定检测的请求关键特征,以及成功应答的应答关键特征,以便负载均衡器可在数据流中执行模式匹配或正则匹配以发现这些特征,从而进行判定。
[0054] 步骤 S102、获取客户端与服务器之间交互的数据流；根据需要检测的服务器应用的请求关键特征和应答关键特征,对所述数据流进行匹配；
[0055] 当客户端向服务器发送请求时,首先在请求数据流中执行请求关键特征匹配,匹配成功则将请求数据流转发给服务器。当服务器反馈应答的数据流时,负载均衡器在应答

的数据流中根据指定的应答关键特征（对于已知协议，可从协议中获知，对于未知协议，需要另行指定）对应答数据流进行特征匹配。

[0056] 步骤 S103、匹配成功时，判定所述服务器应用健康，否则判定所述服务器应用故障。当请求关键特征和应答关键特征均匹配成功则该服务器对客户端的该应用请求为健康的，否则判定为故障并进行步骤 S104。

[0057] 步骤 S104，将所述客户端发送的所述服务器应用的请求发送至其它健康服务器。负载均衡器检测到服务器故障时，当客户端再次发出该应用请求时，负载均衡器不再将该请求发送至故障的服务器，而改为路由至其它健康的服务器。

[0058] 如图 3 所示，图 3 所示为检测服务器的 HTTP POST 请求是否健康之示例，该应用请求的协议已知。当客户端与服务器通过 TCP 三次握手建立链接，并成功登录服务器后，客户端向服务器发送 POST 请求：请求访问的地址为 /abc.cgi，当且仅当负载均衡器获取到的服务器的应答为 2000k 时，才认为该服务器对该地址的 HTTP POST 请求应用是健康的，否则认为服务器故障。当负载均衡器检测到服务器故障时，当客户端再次发出对于该地址的 HTTPPOST 请求时，负载均衡器不同将该请求发送至故障的服务器，而改为路由至其它健康的服务器。

[0059] 本实施例中，针对现有的服务器应用健康状态检测方法都是由检测设备主动发起请求，然后检测设备检测请求的应答来判定服务器的健康状态，对于复杂的应用的健康检查，必须由用户在检测设备上编程实现。而在本发明中，不需要检测设备主动发送业务请求，检测设备只是依据检测配置，以及客户端与服务器之间的数据流，被动分析服务器对实际应用请求的应答情况来判定服务器上应用的健康状态，大大简化了检测设备的负担。同时也无需检测设备的管理员具备相当专业的编程知识。

[0060] 参照图 4，提出本发明一种检测服务器应用健康状态的装置一实施例，包括：

[0061] 设置模块 10，用于当交互协议已知时，指定请求关键特征；以及根据所述请求关键特征，在已知的交互协议查找对应的应答关键特征；当交互协议未知时，指定所述请求关键特征和应答关键特征。

[0062] 获取模块 20，用于获取客户端与服务器之间交互的数据流；

[0063] 匹配模块 30，用于根据需要检测的服务器应用的请求关键特征和应答关键特征，对所述数据流进行匹配；

[0064] 判定模块 40，用于匹配成功时，判定所述服务器应用健康，否则判定所述服务器应用故障。

[0065] 故障处理模块 50，用于当判定所述服务器故障时，将所述客户端发送的所述服务器应用的请求发送至其它健康服务器。

[0066] 参照图 5，所述匹配模块 30 包括：

[0067] 请求特征匹配单元 31，用于根据所述请求关键特征，对客户端发送的请求数据流执行请求特征匹配；

[0068] 应答特征匹配单元 32，用于匹配成功时，根据所述应答关键特征，对服务器的应答数据流执行应答特征匹配。

[0069] 本实施例的装置可以是负载均衡器等能够监测客户端与服务器交互数据流的检测设备，也可以是内置或外置于负载均衡器的一装置。其应用环境还包括：一或多个服务

器,如各自运行一或多个应用程序的 WEB 服务器;一或多个客户端,向所述一或多个服务器提出应用请求。当客户端通过网络发送应用请求时,应用请求可由本装置接收并路由至服务器,本实施例即以负载均衡设备为例进行说明,但在实际应用中,不仅限于负载均衡器。

[0070] 负载均衡器的设置模块 10 首先在负载均衡器上指定需要检测的请求关键特征以及服务器成功应答时的应答关键特征,对于已知交互协议的服务器应用,例如 HTTP 协议的应用,负载均衡器能从客户端和服务器的数据流中解析出完整的 HTTP 请求和应答,因此对于指定的应用请求关键特征,设置模块 10 能够根据 HTTP 协议标准,得知指定检测请求的应答关键特征,因此对于该类服务器应用,设置模块 10 只需指定需要检测的请求关键特征即可。可对于交互协议未知的服务器应用,由于负载均衡设备无法在客户端与服务器交互的数据流中识别完整的请求和应答,因此需要设置模块 10 指定检测的请求关键特征,以及成功应答的应答关键特征,以便匹配模块 30 可在数据流中执行模式匹配或正则匹配以发现这些特征,从而将由判定模块 40 进行判定。

[0071] 当客户端向服务器发送请求时,匹配模块 30 的请求特征匹配单元 31 首先在请求数据流中执行请求关键特征匹配,匹配成功则负载均衡器将请求数据流转发给服务器。当服务器反馈应答的数据流时,匹配模块 30 的应答特征匹配单元 32 在应答的数据流中根据指定的应答关键特征对应答数据流进行特征匹配,匹配成功则判定模块 40 判定该服务器对客户端的该应用请求为健康的,否则故障。此时,负载均衡器根据服务器的故障进行异常处理,避免将客户端的请求再次路由至故障服务器。

[0072] 如图 3 所示,图 3 所示为检测服务器的 HTTP POST 请求是否健康之示例,客户端与服务器通过 TCP 三次握手建立链接,并成功登录服务器后,客户端向服务器发送 POST 请求:请求访问的地址为 /abc.cgi,当且仅当获取模块 20 获取到的服务器的应答为 2000k 时,判定模块 40 才认为该服务器对该地址的 HTTP POST 请求应用是健康的,否则认为服务器故障。当负载均衡器检测到服务器故障时,故障处理模块 50 将该故障记录,并将客户端的该应用请求改为路由至其它健康的服务器。

[0073] 本实施例中,针对现有的服务器应用健康状态检测方法都是由检测设备主动发起请求,然后检测设备检测请求的应答来判定服务器的健康状态,对于复杂的应用的健康检查,必须由用户在检测设备上编程实现。而在本发明中,不需要检测设备主动发送业务请求,检测设备只是依据检测配置,以及客户端与服务器之间的数据流,被动分析服务器对实际应用请求的应答情况来判定服务器上应用的健康状态,大大简化了检测设备的负担。同时也无需检测设备的管理员具备相当专业的编程知识。

[0074] 以上所述仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

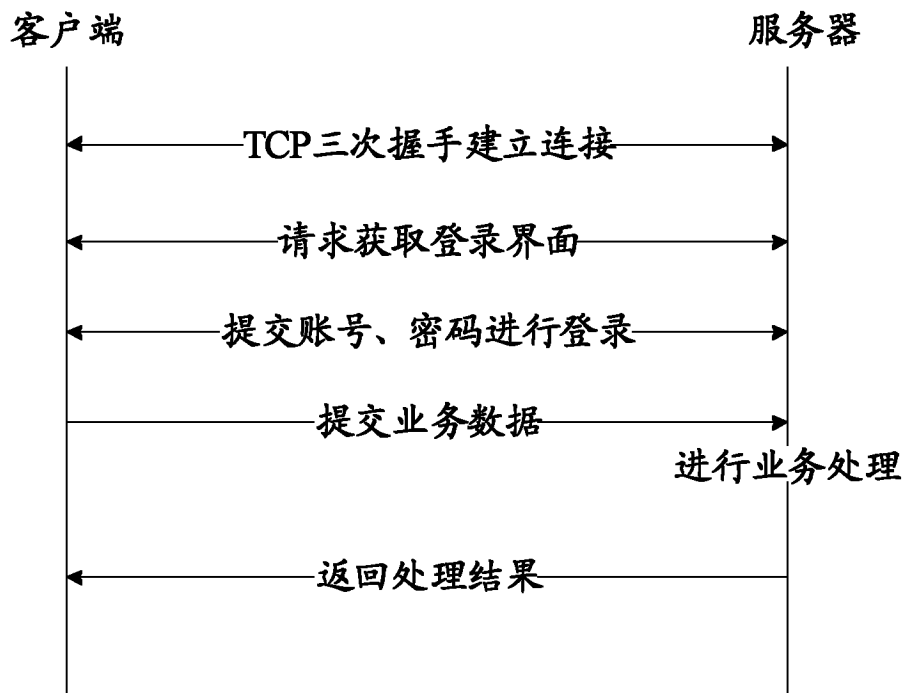


图 1

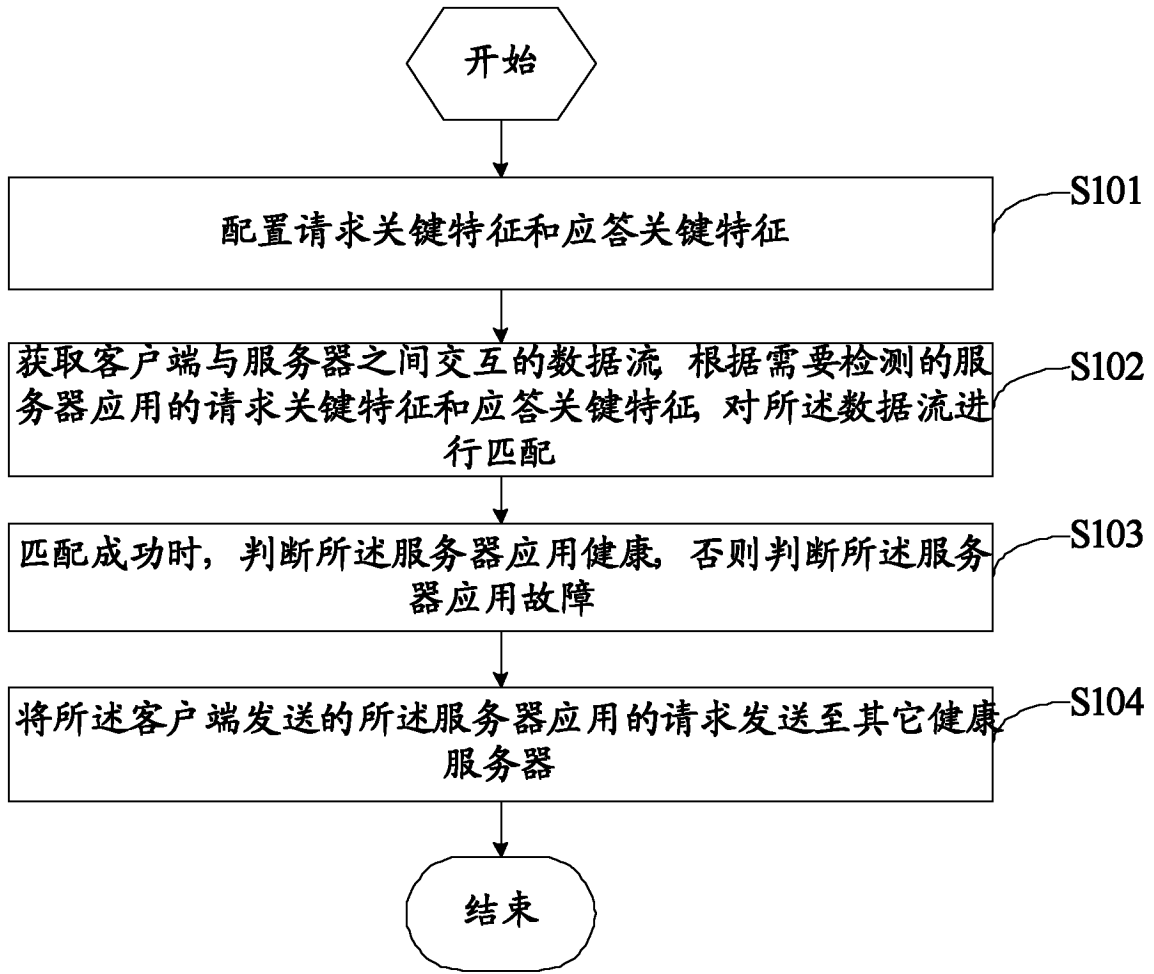


图 2

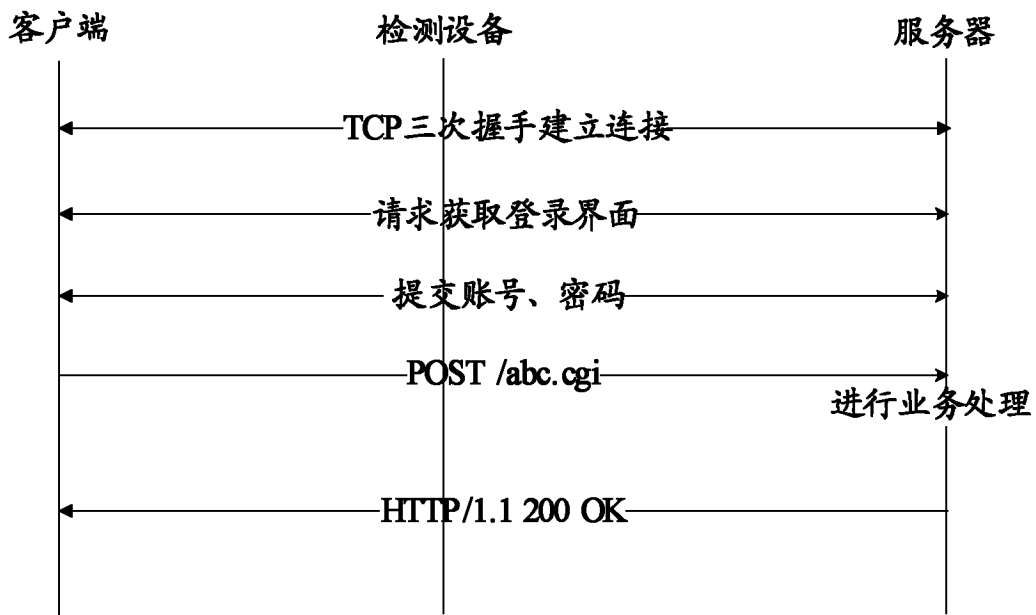


图 3

