



(12) 发明专利申请

(10) 申请公布号 CN 102356388 A

(43) 申请公布日 2012. 02. 15

(21) 申请号 201080012902. 7

(51) Int. Cl.

(22) 申请日 2010. 03. 09

G06F 15/16(2006. 01)

(30) 优先权数据

G06F 9/305(2006. 01)

12/406, 699 2009. 03. 18 US

G06F 11/30(2006. 01)

(85) PCT申请进入国家阶段日

2011. 09. 16

(86) PCT申请的申请数据

PCT/US2010/026713 2010. 03. 09

(87) PCT申请的公布数据

W02010/107628 EN 2010. 09. 23

(71) 申请人 微软公司

地址 美国华盛顿州

(72) 发明人 J·张 L·李 C·A·小克拉克

I·D·G·夸菲耶 R·萨迪欧

D·A·赫奇佩思 S·A·萨努西

(74) 专利代理机构 上海专利商标事务所有限公司 31100

代理人 钱孟清

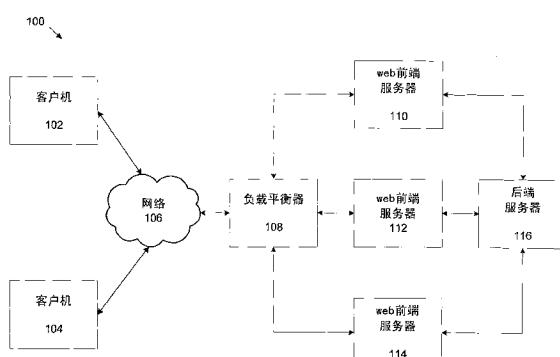
权利要求书 4 页 说明书 8 页 附图 8 页

(54) 发明名称

web 前端节流

(57) 摘要

服务器计算机包括性能监视器模块和节流逻辑模块。性能监视器模块包括监视服务器计算机的系统参数的性能监视器。节流逻辑模块确定性能监视器所监视的系统参数是否超过预定阈值。当系统参数超过预定阈值时，节流逻辑模块置位节流标志。当为预定数量的定时快照中的每一个置位至少一个节流标志时，节流逻辑模块在服务器计算机处激活节流。激活节流限制对服务器计算机接收到的请求消息的处理。



1. 一种服务器计算机 (110), 所述服务器计算机 (110) 包括：

性能监视器模块 (202), 所述性能监视器模块 (202) 包括一个或多个性能监视器, 所述一个或多个性能监视器中的每一个监视所述服务器计算机 (110) 的系统参数；

节流逻辑模块 (206), 所述节流逻辑模块 (206) 确定性能监视器所监视的系统参数是否超过预定阈值, 当系统参数超过预定阈值时所述节流逻辑模块 (206) 置位节流标志; 以及

节流配置模块 (204), 所述节流配置模块 (204) 存储针对所述一个或多个性能监视器所监视的每一个系统参数的所述预定阈值, 所述节流配置模块 (204) 还存储一个或多个节流逻辑参数；

其中, 当为预定数量的定时快照中的每一个置位至少一个节流标志时, 所述节流逻辑模块 (206) 在所述服务器计算机 (110) 处激活节流, 激活节流限制对所述服务器计算机 (110) 接收到的请求消息的处理。

2. 如权利要求 1 所述的服务器计算机 (110), 其特征在于, 限制对所述服务器计算机 (110) 接收到的请求消息的处理包括拒绝请求消息。

3. 如权利要求 1 所述的服务器计算机 (110), 其特征在于, 当与所述节流标志相关联的系统参数落在预定阈值以下时, 所述节流逻辑模块 (206) 停用节流标志。

4. 如权利要求 1 所述的服务器计算机 (110), 其特征在于, 当没有为所述预定数量的定时快照中的每一个置位至少一个节流标志时, 所述节流逻辑模块 (206) 停用节流。

5. 如权利要求 1 所述的服务器计算机 (110), 其特征在于, 还包括分类器模块 (208), 所述分类器模块 (208) 包括一个或多个分类器, 所述一个或多个分类器中的每一个标识一类或多类请求消息是可以被节流还是不可以被节流。

6. 如权利要求 1 所述的服务器计算机 (110), 其特征在于, 还包括性能分析模块 (210), 所述性能分析模块 (210) 分析所述服务器计算机处的性能数据以确定用于所述性能监视器的预定阈值。

7. 如权利要求 1 所述的服务器计算机 (110), 其特征在于, 还包括界面模块 (212), 所述界面模块 (212) 准许使用所述服务器计算机上的界面来配置所述预定阈值和所述节流逻辑参数。

8. 一种用于在服务器计算机 (110) 处对消息进行节流的方法, 所述方法包括：

在所述服务器计算机 (110) 处接收来自客户机计算机 (102) 的 HTTP 请求消息；

响应于在所述服务器计算机 (110) 处接收所述 HTTP 请求消息, 确定是否要在所述服务器计算机 (110) 处接受所述请求消息, 其中确定是否要在所述服务器计算机 (110) 处接受所述请求消息包括：

确定是否在所述服务器计算机 (110) 处激活节流；

确定所述请求消息是否可以被节流；

当在所述服务器计算机 (110) 处未激活节流时, 在所述服务器计算机 (110) 处接受所述请求消息；

当在所述服务器计算机 (110) 处激活节流并且所述请求消息不可以被节流时, 在所述服务器计算机 (110) 处接受所述请求消息; 以及

当在所述服务器计算机 (110) 处激活节流并且所述请求消息可以被节流时, 在所述服

务器计算机(110)处拒绝所述请求消息;以及

当确定在所述服务器计算机(110)处拒绝所述请求消息时,向所述客户机计算机(102)发送回复消息,所述回复消息通知所述客户机计算机(102、104)在所述服务器计算机(110)处已拒绝了所述请求消息。

9. 如权利要求8所述的方法,其特征在于,确定是否在所述服务器计算机(110)上激活节流包括:

在对应于第一定时快照的时刻,监视所述服务器计算机(110)上的一个或多个性能监视器的状态;

响应于在所述第一定时快照的时刻监视所述服务器计算机(110)上的一个或多个性能监视器的状态,当确定所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时,置位第一节流标志;

确定对应于一个或多个附加定时快照的时刻,所述一个或多个附加定时快照中的每一个在从所述第一定时快照起的一个或多个预定时间间隔处发生;

在对应于所述一个或多个附加定时快照中的每一个的时刻,监视所述服务器计算机(110)上的一个或多个性能监视器的状态;

响应于在所述一个或多个附加定时快照中的每一个的时刻监视所述服务器计算机(110)上的一个或多个性能监视器的状态,在所述一个或多个附加定时快照中的每一个确定所述一个或多个性能监视器中的至少一个是否确定了所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值;

响应于在所述一个或多个附加定时快照中的每一个的时刻监视所述服务器计算机(110)上的一个或多个性能监视器的状态,当在所述一个或多个附加定时快照中的每一个确定所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时,置位附加节流标志;

当确定在所述一个或多个附加定时快照的最后一个定时快照的时刻、所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时,确定是否置位了第一节流标志并且是否为所述一个或多个附加定时快照中的每一个置位了节流标志;以及

当确定在所述一个或多个附加定时快照的最后一个定时快照的时刻、所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时并且当确定置位了所述第一节流标志且为所述一个或多个附加定时快照中的每一个置位了节流标志时,在所述服务器计算机(110)上激活节流。

10. 如权利要求8所述的方法,其特征在于,确定所述请求消息是否可以被节流包括:

标识所述请求消息的HTTP消息类型;

标识所述请求消息的功能;

确定请求消息的所标识HTTP消息类型是否是可以被节流的HTTP消息类型;

确定请求消息的所标识功能是否是不可以被节流的功能;

当请求消息的所标识HTTP消息类型是不可以被节流的HTTP消息类型时或者当请求消息的所标识功能是不可以被节流的功能时,确定所述请求消息不可以被节流;以及

当请求消息的所标识HTTP消息类型是可以被节流的HTTP消息类型时并且当请求消息的所标识功能是尚未被标识为不可以被节流的功能的功能时,确定所述请求消息可以被节

流。

11. 如权利要求 8 所述的方法,其特征在于,确定所述请求消息是否可以被节流还包括使用一个或多个分类器来标识被指定节流的 HTTP 消息类型。

12. 如权利要求 8 所述的方法,其特征在于,定时快照的数量是可配置的。

13. 如权利要求 8 所述的方法,其特征在于,所述一个或多个性能监视器中的每一个的预定阈值是可配置的。

14. 如权利要求 8 所述的方法,其特征在于,所述第一定时快照和第二定时快照之间的预定时间间隔是可配置的。

15. 一种包括指令的计算机可读数据存储介质,所述指令在由服务器计算机(110)的处理单元执行时使所述服务器计算机(110):

接收来自客户机计算机(102)的 HTTP 请求消息;

响应于接收到所述 HTTP 请求消息,确定是否要接受所述请求消息,其中确定是否要接受所述请求消息包括:

在对应于第一定时快照的时刻,监视所述服务器计算机(110)上的一个或多个性能监视器的状态;

响应于在所述第一定时快照的时刻监视所述服务器计算机(110)上的一个或多个性能监视器的状态,当确定所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时,置位第一节流标志;

确定对应于一个或多个附加定时快照的时刻,所述一个或多个附加定时快照中的每一个在从所述第一定时快照起的一个或多个预定时间间隔处发生;

在对应于所述一个或多个附加定时快照中的每一个的时刻,监视所述服务器计算机(110)上的一个或多个性能监视器的状态;

响应于在所述一个或多个附加定时快照中的每一个的时刻监视所述服务器计算机(110)上的一个或多个性能监视器的状态,在所述一个或多个附加定时快照中的每一个时确定所述一个或多个性能监视器中的至少一个是否确定了所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值;

响应于在所述一个或多个附加定时快照中的每一个的时刻监视所述服务器计算机(110)上的一个或多个性能监视器的状态,当在所述一个或多个附加定时快照中的每一个确定所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时,置位附加节流标志;

当确定在所述一个或多个附加定时快照的最后一个定时快照的时刻所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时,确定是否置位了所述第一节流标志并且是否为所述一个或多个附加定时快照中的每一个置位了节流标志;

当确定在所述一个或多个附加定时快照的最后一个定时快照的时刻所述一个或多个性能监视器中的至少一个所监视的系统参数超过预定阈值时并且当确定置位了所述第一节流标志且为所述一个或多个附加定时快照中的每一个置位了节流标志时,在所述服务器计算机(110)上激活节流;

标识所述请求消息的 HTTP 消息类型;

标识所述请求消息的功能;

确定请求消息的所标识 HTTP 消息类型是否是可以被节流的 HTTP 消息类型；

确定请求消息的所标识功能是否是不可以被节流的功能；

当请求消息的所标识 HTTP 消息类型是不可以被节流的 HTTP 消息类型时或者当请求消息的所标识功能是不可以被节流的功能时，确定所述请求消息不可以被节流；

当请求消息的所标识 HTTP 消息类型是可以被节流的 HTTP 消息类型时并且当请求消息的所标识功能是尚未被标识为不可以被节流的功能的功能时，确定所述请求消息可以被节流；

当在所述服务器计算机（110）处未激活节流时，在所述服务器计算机（110）处接受所述请求消息；

当在所述服务器计算机（110）处激活节流并且所述请求消息不可以被节流时，在所述服务器计算机（110）处接受所述请求消息；以及

当在所述服务器计算机（110）处激活节流并且所述请求消息可以被节流时，在所述服务器计算机（110）处拒绝所述请求消息；以及

当确定在所述服务器计算机（110）处拒绝所述请求消息时，向所述客户机计算机（102、104）发送回复消息，所述回复消息通知所述客户机计算机（102）在所述服务器计算机（110）处已拒绝了所述请求消息。

## web 前端节流

[0001] 背景

[0002] web 前端服务器向客户机系统提供界面,这些客户机系统试图访问来自因特网上的数据库服务器的信息。web 前端服务器通常处理客户机请求,将该请求转发给数据库服务器并且将所获得的数据返回到客户机系统。

[0003] 基于在线浏览器的文档和协作系统通常提供多个 web 前端服务器来处理客户机请求。客户机期望他们访问数据库服务器的请求被及时地处理。然而,当系统忙碌时,web 前端服务器有时会过载从而不能及时地对客户机请求做出响应。

[0004] 概述

[0005] 本发明的各个实施例涉及提供服务器计算机的节流能力。服务器计算机包括性能监视器模块、节流逻辑模块和节流配置模块。性能监视器模块包括一个或多个性能监视器。每个性能监视器监视服务器计算机的系统参数。节流逻辑模块确定性能监视器所监视的系统参数是否超过预定阈值。当系统参数超过预定阈值时,节流逻辑模块置位节流标志。

[0006] 节流配置模块存储针对一个或多个性能监视器所监视的每个系统参数的预定阈值。节流配置模块还存储一个或多个节流逻辑参数。当为预定数量的定时快照 (time snapshot) 中的每一个置位至少一个节流标志时,节流逻辑模块在服务器计算机处激活节流。激活节流限制对服务器计算机接收到的请求消息的处理。

[0007] 一种或多种技术的细节在附图和以下描述中阐明。通过描述、附图和权利要求,这些技术的其他特征、目的和优点将变得显而易见。

[0008] 附图描述

[0009] 图 1 示出了用于 web 前端节流的示例系统。

[0010] 图 2 示出了 web 前端服务器的示例模块。

[0011] 图 3 示出了 web 前端节流中所使用的示例节流机制的示例逻辑图。

[0012] 图 4 示出了在示例 web 前端服务器处执行的示例操作的流程图。

[0013] 图 5 示出了图 4 所示的示例操作之一的流程图。

[0014] 图 6 示出了图 5 所示的示例操作之一的流程图。

[0015] 图 7 示出了图 5 所示的示例操作中的另一个的流程图。

[0016] 图 8 示出了实现 web 前端节流的系统的操作环境。

[0017] 详细描述

[0018] 本公开涉及用于提供 web 前端服务器的节流能力的系统和方法。系统和方法使用性能监视器来监视 web 前端服务器的系统健康。当一个或多个性能监视器指示系统资源在指定时间段内落在预定阈值以外时,在 web 前端服务器上激活节流。当在 web 前端服务器上激活节流时,web 前端服务器不处理特定类型的客户机请求消息,通常是具有低优先级的请求消息。相反,通常指示系统忙碌状态的出错消息被返回到进行请求的客户机系统。

[0019] 图 1 示出了用于 web 前端 (WFE) 节流的示例系统 100。示例系统 100 包括客户机 102 和 104、网络 106、负载平衡器 108、WFE 服务器 110、112、114 以及后端服务器 116。可使用更多或更少的客户机、WFE、后端服务器、负载平衡器和网络。在本公开中,术语 WFE 和 WFE

服务器互换地使用。

[0020] 在示例实施例中，客户机 102 和 104 是诸如台式计算机、膝上型计算机、终端计算机、个人数字助理、或蜂窝电话设备之类的计算设备。客户机 102 和 104 可包括输入 / 输出设备、中央处理单元（“CPU”）、数据存储设备和网络设备。在本申请中，术语客户机和客户机计算机互换地使用。

[0021] WFE 110、112 和 114 可由客户机 102 和 104 经由负载平衡器 108 通过网络 106 访问。后端服务器 116 对 WFE 110、112 和 114 是可访问的。负载平衡器 108 是服务器计算机。负载平衡器 108、WFE 110、112 和 114 以及后端服务器 116 可包括输入 / 输出设备、中央处理单元（“CPU”）、数据存储设备和网络设备。在示例实施例中，网络 106 是因特网，并且客户机 102 和 104 可访问 WFE 110、112 和 114 以及远程连接到 WFE 110、112 和 114 的资源。在本申请中，术语服务器和服务器计算机互换地使用。

[0022] 在示例实施例中，示例系统 100 是在线的、基于浏览器的文档协作系统。在线的、基于浏览器的文档协作系统的一个示例是来自美国华盛顿州雷蒙德市的微软公司的 Microsoft Sharepoint。在示例系统 100 中，后端服务器 116 是 SQL 服务器，例如，来自美国华盛顿州雷蒙德市的微软公司的 SQL Server 2008。

[0023] 在示例系统 100 中，WFE 110、112 和 114 提供客户机 102 和 104 与后端服务器 116 之间的界面。负载平衡器 108 是将请求从客户机 102 和 104 导向 WFE 110、112 和 114 的服务器计算机。负载平衡器 108 使用诸如 WFE 的利用率、连接到 WFE 的连接数量和整体 WFE 性能之类的因素来确定哪个 WFE 服务器接收客户机请求。

[0024] 在示例系统 100 中，客户机请求的一个示例可以是访问存储在后端服务器 116 上的文档、编辑存储在后端服务器 116 上的文档或者在后端服务器 116 上存储文档。当负载平衡器 108 通过网络 106 接收客户机请求时，负载平衡器 108 确定 WFE 服务器 110、112 和 114 中的哪一个接收客户机请求。尽管试图在 WFE 服务器 110、112 和 114 之间平衡请求，但 WFE 服务器 110、112 和 114 中的一个或多个的系统资源可能被耗尽以致 WFE 不能接受客户机请求。

[0025] 示例系统 100 中的每个 WFE 都包括一节流机制。该节流机制确定 WFE 何时激活节流。当在 WFE 上激活节流时，在该 WFE 处拒绝低优先级的客户机请求消息，并且指示 WFE 忙碌状态的出错回复消息从该 WFE 发送到作出请求的客户机 102、104。

[0026] 图 2 更详细地示出了示例 WFE 110。如图 2 所示，示例 WFE 110 包括示例性能监视器模块 202、示例节流配置模块 204、示例节流逻辑模块 206、示例分类器模块 208、示例性能分析模块 210 和示例界面模块 212。

[0027] 示例性能监视器模块 202 包括一个或多个性能监视器，这些性能监视器跟踪 WFE 110 上的指示服务器健康的系统参数。通常，WFE 110 针对每个所监视系统参数包括一个性能监视器。性能监视器通常是计数器，该计数器存储正在监视的系统参数的值的计数。在本公开中，术语性能监视器和性能计数器互换地使用。

[0028] 性能计数器监视的系统参数的示例包括诸如 CPU 使用的百分比和以兆字节为单位的可用 WFE 存储器之类的资源使用参数，以及诸如由 WFE 进行排队的请求的数量和以毫秒为单位的请求在队列中的等待时间之类的性能参数。在示例系统 100 中，排队的示例请求可以是 ASP.NET 请求，并且请求等待时间可以是 ASP.NET 请求等待时间。ASP.NET 是由

微软开发和投入市场的 web 应用程序框架，该 web 应用程序框架允许程序员构建动态网站、web 应用程序和 web 服务。ASP.NET 被频繁地用于构建 Microsoft Sharepoint 应用程序。

[0029] 示例节流配置模块 204 针对每个性能计数器存储一阈值。每个阈值都是可配置的预定阈值。每个阈值可由系统管理员手动地配置或者基于 WFE 的性能分析通过程序来配置。每个阈值表示对性能计数器所监视的系统参数的限制。取决于所监视的系统参数，阈值表示系统参数的上限或下限。超过上限或者落在下限以下的性能计数指示性能相关的问题。

[0030] 示例节流配置模块 204 还存储一个或多个节流逻辑参数。节流逻辑参数指定节流机制的各个方面。示例节流逻辑参数包括确定是否激活节流所需要的定时快照的数量和定时快照之间的时间间隔。其他节流逻辑参数是可能的。每个节流逻辑参数是可配置的。

[0031] 示例节流逻辑模块 206 监视一个或多个性能计数器并且确定何时要激活 WFE 的节流。通常，确定何时要激活例如 WFE 110 的节流是一个 n 步骤过程。在示例 n 步骤过程中，在 n 个时间点监视一个或多个性能计数器。第一个时间点被指定为定时快照 1，而第 n 个时间点被指定为定时快照 n。

[0032] 在示例 n 步骤过程中，示例节流逻辑模块 206 确定性能计数器所监视的系统参数是否超过预定阈值。在示例实施例中，当系统参数超过预定阈值时，取决于所监视的系统参数，系统参数的值超过上限或者落在下限以下。例如，当 CPU 使用超过预定阈值时，CPU 使用的值超过上限。然而，当可用存储器超过预定阈值时，可用存储器的值落在下限以下。

[0033] 示例 n 步骤过程中的第一步骤是在定时快照 1 监视一个或多个性能计数器。当在定时快照 1 监视一个或多个性能计数器时，如果性能计数器所监视的系统参数超过上限阈值或落在下限阈值以下，则与性能计数器相关联的性能监视器置位性能监视器的节流标志。

[0034] 示例 n 步骤过程中的下一步骤是在定时快照 2 监视一个或多个性能计数器。定时快照 2 是从定时快照 1 起的预定时间间隔。通常使用的示例时间间隔是 5 秒。时间间隔是可配置的值，并且可使用其他时间间隔。

[0035] 当在定时快照 2 监视一个或多个性能计数器时，如果性能计数器所监视的系统参数超过上限阈值或落在下限阈值以下，如果还没有为与性能计数器相关联的性能监视器置位节流标志，则与性能计数器相关联的性能监视器置位性能监视器的节流标志。

[0036] 对 n 步骤过程中剩余步骤的每一个步骤执行该相同过程。例如，如果 n 等于 5，则另外在定时快照 3、4 和 5 分别监视一个或多个性能计数器。当在定时快照 3、4 和 5 分别监视一个或多个性能计数器时，如果性能计数器所监视的系统参数超过上限阈值或落在下限阈值以下，则如果还没有为与性能计数器相关联的性能监视器置位节流标志，与性能计数器相关联的性能监视器置位性能监视器的节流标志。

[0037] 在定时快照 n，示例节流逻辑模块 206 确定在定时快照 1 是否置位了至少一个节流标志并且在定时快照 2-n（指定时快照 2 到定时快照 n）中的每一个是否置位至少一个节流标志。在 n 等于 2 的示例实施例中，只使用两个定时快照。如果示例节流逻辑模块 206 确定在定时快照 1 置位了至少一个节流标志并且在定时快照 2-n 中的每一个置位了至少一个节流标志，则示例节流逻辑模块 206 激活示例 WFE 110 的节流。当在 WFE 110 处激活节流状态时，WFE 110 不处理在 WFE 110 处接收的具有低优先级的所有请求消息。相反，WFE

110 向发起该请求的客户机发送出错消息，通常是忙碌错误。

[0038] 一个或多个性能监视器持续监视与相应的性能计数器相关联的系统参数。当为性能计数器置位节流标志时，如果性能计数器所监视的系统参数继续超过上限或者保持在上限以下，则节流标志保持置位。然而，当性能计数器所监视的系统参数返回到可接受值时，即小于或等于上限、或者大于或大于下限，重置性能计数器的节流标志。

[0039] 当确定是否激活示例 WFE 110 的节流时，示例节流逻辑模块 206 不考虑哪一个性能计数器置位节流标志。确定因素是在定时快照 1 置位至少一个节流标志且在定时快照 2-n 中的每一个置位至少一个节流标志。在示例实施例中，如果不同的性能计数器在定时快照 1 且在定时快照 2-n 中的每一个置位节流标志、或者如果在定时快照 1 且在定时快照 2-n 中的每一个为相同的性能计数器置位节流标志，则可激活节流状态。

[0040] 定时快照的数量 n 是可配置的。在一些示例实施例中，示例节流逻辑模块 206 使用两个定时快照来确定是否要在示例 WFE 110 处激活节流状态。示例节流逻辑模块 206 使用两个定时快照来使设置假警报节流状态的性能尖峰的可能性最小化。例如，可存在 CPU 活动的猝发，该猝发使监视 CPU 活动的性能计数器超过阈值。CPU 活动的这个猝发可以是只持续短时间段的尖峰。因为活动中的短尖峰激活 WFE 的节流状态是不合需要的，所以在该示例实施例中使用两个定时快照。在其他示例实施例中，可使用两个以上的定时快照。

[0041] 示例分类器模块 208 向消息类型分配优先级，并且确定在激活节流时可处理哪些消息以及在激活节流时可拒绝哪些消息。一般而言，准许写操作被完成并且不被节流是合乎需要的，而读操作（例如对网页的请求）更有可能被节流。例如，如果客户机 102 上的用户打开文档进行编辑，则允许用户完成文档的编辑且在编辑文档时不会因为 WFE 忙碌而被延迟是合乎需要的。类似地，如果客户机 104 上的用户尝试访问网页进行查看并且接收忙碌错误，则用户可在稍后时间简单地再试一次。

[0042] 在示例实施例中，分类器模块 208 按照 HTTP 消息类型对一些消息进行分类。例如，通常 HTTP Post 和 HTTP Put 消息类型分配到高优先级、而 HTTP Get 消息类型通常分配到低优先级。在示例实施例中，如果在示例 WFE 110 上激活节流并且 WFE 110 接收到 HTTP Post 或 HTTP Put 消息请求，则 WFE 110 处理 HTTP Post 或 HTTP Put 消息。类似地，如果在 WFE 110 上激活节流并且 WFE 110 接收到 HTTP Get 消息请求，则 WFE 110 通常通过返回忙碌错误来拒绝 HTTP Get 消息。

[0043] 除了按照消息类型对消息进行分类以外，示例分类器模块 208 还可按照消息功能对消息进行分类。例如，当在示例 WFE 110 处激活节流时，对来自搜索引擎爬行器的消息进行节流以防止搜索引擎爬行器中断用户触发的 HTTP Post 请求可能是合乎需要的。

[0044] 示例 WFE 110 可包括一个以上的分类器模块。例如，一个分类器模块可按照 HTTP 功能类型对请求进行分类。另一个分类器模块可按照功能对请求进行分类。例如，WFE 110 可包括用于 Microsoft Sharepoint 消息的唯一分类器模块。如果一个分类器模块将请求消息指定为高优先级而另一个分类器模块将请求消息指定为低优先级，则该请求消息被给予低优先级。如果只有一个分类器模块与请求消息相匹配，则请求消息被给予与该分类器相关联的优先级。

[0045] 示例性能分析模块 210 准许 WFE 随着时间的流逝分析一个或多个性能监视器以使性能计数器的阈值水平和其他节流参数可以被更准确地配置。例如，性能分析模块 210 可

确定同一性能计数器正在引起过量的节流状态从而超过阈值。对于该示例,为了使假警报最小化,性能分析模块 210 可确定应当增大性能计数器的阈值限制。在其他示例中,性能分析模块 210 可确定应当修改快照之间的时间间隔或激活节流所需的快照数量。

[0046] 示例界面模块 212 准许系统管理员手动地配置系统参数。例如,系统管理员可手动地配置一个或多个性能计数器的系统参数阈值。系统管理员可手动地配置 n 的值,由此配置用于确定是否要激活节流的定时快照的数量。系统管理员还可手动地配置快照之间的时间间隔。还可配置其他系统参数。

[0047] 示例界面模块 212 通常提供命令行界面以准许系统管理员配置系统参数。在一些实施例中,可使用图形用户界面。其他用户界面是可能的。

[0048] 图 3 示出了示例节流逻辑模块 206 中所使用的示例逻辑 300。对于图 2 所示的示例节流逻辑模块 206 的实施例,n 的值等于 2,指示一两步骤节流机制。由此,示出了两个 OR 门和两个节流标志。在其他示例实施例中,OR 门的数量和节流标志的数量等于 n。例如,对应于三个定时快照,三步骤节流机制使用三个 OR 门。OR 门仅作为示例示出。在示例实施例中,OR 门和 AND 门所使用的逻辑可通过其他手段(例如,软件)来实现。

[0049] 示例 OR 门 302 在时刻 t 接收来自一个或多个性能监视器的输入。时刻 t 构成定时快照 1。在时刻 t,如果一个或多个性能监视器中的任一个确定所监视的系统参数超过上限或落在下限以下,则置位示例标志 t。示例标志 t 是示例 AND 门 306 的一个输入。

[0050] 示例 OR 门 304 在时刻 t+1 接收来自一个或多个性能监视器的输入。时刻 t+1 构成定时快照 2。在时刻 t+1,如果一个或多个性能监视器中的任一个确定所监视的系统参数超过上限或落在下限以下,则置位示例标志 t+1。示例标志 t+1 是示例 AND 门 306 的第二输入。

[0051] 在时刻 t+1,如果标志 t 和标志 t+1 都被置位,则 AND 门 306 置位示例节流开 / 关(ON/OFF)标志。当示例节流开 / 关标志被置位时,在示例 WFE 110 处激活节流。

[0052] 图 4 是示出在示例 WFE 110 处执行的示例操作 400 的流程图。在操作 402,示例 WFE 110 接收来自示例客户机 102 的请求消息。在操作 404,示例 WFE 110 确定是否要接受请求消息以进行附加处理。当确定是否要接受请求消息时,示例 WFE 110 确定是否在 WFE 110 处激活节流以及某类型和功能的请求消息是否被节流。当在操作 406 示例 WFE 110 确定接受请求消息时,在操作 408 处理请求消息。当在操作 406 示例 WFE 110 确定拒绝请求消息时,示例 WFE 110 向示例客户机 102 发送回复消息。回复消息是向示例客户机 102 指示在示例 WFE 110 处拒绝了请求消息的出错消息,通常是忙碌消息。

[0053] 图 5 是示出在示例 WFE 110 处执行的更为详细的示例操作 404 的流程图。在示例操作 404,示例 WFE 110 确定是否要接受来自示例客户机 102 的请求消息。

[0054] 在操作 502,检查以确定在示例 WFE 110 上是否激活节流。在操作 504,确定示例请求消息是否可以被节流。当在操作 506 确定在示例 WFE 110 上不激活节流时,在示例 WFE 110 处接受请求消息。当在操作 506 确定在示例 WFE 110 上激活节流时并且当在操作 510 确定请求消息不可以被节流时,在操作 512,示例 WFE 110 接受示例请求消息。通常不可以被节流的示例请求消息是具有 HTTP Post 或 HTTP Put 消息类型的请求消息。

[0055] 当在操作 506 确定在示例 WFE 110 上激活节流时并且当在操作 510 确定请求消息可以被节流时,在操作 514,示例 WFE 110 拒绝示例请求消息。通常可以被节流的示例请求

消息是具有 HTTP Get 消息类型的请求消息。

[0056] 图 6 是示出在示例 WFE 110 处执行的更为详细的示例操作 600 的流程图。在示例操作 600, 示例 WFE 110 确定是否在示例 WFE 110 处激活节流。关于是否在示例 WFE 110 处激活节流的确定以有规律的时间间隔作出。出于性能的原因, 不针对每一个接收到的请求消息作出关于是否在示例 WFE 110 处激活节流的确定。

[0057] 在操作 602, 示例 WFE 110 在定时快照 1 监视 WFE 110 上的性能监视器。如果在定时快照 1 性能监视器所监视的至少一个系统参数超过为系统参数设置的预定阈值, 则在操作 604 示例 WFE 110 置位第一节流标志。

[0058] 在操作 606, 示例 WFE 110 在定时快照 2-n 监视 WFE 110 上的性能监视器, 其中定时快照 2-n 指定时快照 2 到定时快照 n。如果在定时快照 2-n 的每一个性能监视器所监视的至少一个系统参数超过为系统参数设置的预定阈值, 则在操作 608 示例 WFE 110 置位节流标志。为每个定时快照 2-n 置位单独的节流标志, 因为在每个定时快照 2-n 性能监视器所监视的至少一个系统参数超过为系统参数设置的预定阈值。应当理解, 操作 606 和 608 各自是多个操作, 从而在每个定时快照 2-n 执行一个操作。

[0059] 如果在操作 610 确定置位了两个节流标志, 则在操作 612 WFE 110 激活 WFE 110 处的节流。如果在操作 610 确定没有置位两个节流标志, 则在操作 614 WFE 110 在已激活节流的情况下停用 WFE 110 处的节流。

[0060] 图 7 是示出在示例 WFE 110 处执行的更为详细的示例操作 504 的流程图。在示例操作 504 中, 示例 WFE 110 确定来自示例客户机 102 的请求消息在示例 WFE 110 处是否可以被节流。

[0061] 在操作 702, 示例 WFE 110 标识来自示例客户机 102 的请求消息中的 HTTP 消息类型。HTTP 消息类型的一些示例是 HTTP Post、HTTP Put 和 HTTP Get。在操作 704, 示例 WFE 110 标识请求消息的功能。例如, 请求消息可以是爬行器搜索请求的一部分。

[0062] 在操作 706, 示例 WFE 110 确定具有所标识 HTTP 消息类型的请求消息是否可以被节流。通常, 低优先级的 HTTP 消息类型 (例如, HTTP Get) 可以被节流, 而高优先级的 HTTP 消息类型 (例如, HTTP Post 和 HTTP Put) 不可以被节流。

[0063] 在操作 708, 示例 WFE 110 确定具有所标识功能的请求消息是否不可被节流。具有指示消息具有高优先级的功能类型的消息不可被节流。相反, 具有未被标识为具有高优先级的功能类型的消息通常被节流。例如, 因为爬行器搜索可能是资源密集的并且这些搜索中断高优先级的请求消息 (例如, HTTP Post 或 HTTP Put 请求) 是不合需要的, 所以指示爬行器功能的消息类型通常被节流。因此, 在不节流的情况下通常无法实现爬行器搜索。

[0064] 在操作 710, 如果确定所标识消息类型不可被节流, 则在操作 712, 确定该请求消息不可被节流。在操作 710, 如果确定所标识消息类型可被节流并且如果在操作 714 确定请求消息的功能使请求消息不可被节流, 则在操作 716 确定请求消息不可被节流。在操作 710, 如果确定所标识消息类型可被节流并且如果在操作 714 确定请求消息的功能不能防止请求消息被节流, 则在操作 718 确定请求消息可以被节流。

[0065] 在一些示例实施例中, 消息类型和消息功能可被认为是相同的而且可不被单独考虑。在这些示例实施例中, 可组合操作 702 和 704, 可组合操作 706 和 708, 并且可组合操作 710-716。

[0066] 图 8 是示出电子计算设备 800 的示例物理组件的框图。客户机系统 102 和 104 以及服务器系统 108、110、112、114 和 116 和 / 或客户机系统 102 和 104 以及服务器系统 108、110、112、114 和 116 内的电子计算设备可用电子计算设备 800 的方式实现。如图 8 的示例所示,电子计算系统 800 包括存储器单元 801。存储器单元 801 是能够存储数据和指令的计算机可读数据存储介质。存储器单元 801 可以是各种不同类型的计算机可读数据存储介质,包括但不限于,动态随机存取存储器 (DRAM)、双倍数据率同步动态随机存取存储器 (DDR SDRAM)、减少等待时间的 DRAM、DDR SDRAM、DDR3 SDRAM、Rambus RAM、或者其他类型的计算机可读数据存储介质。

[0067] 此外,电子计算设备 800 包括处理单元 802。在第一示例中,处理单元 802 可执行使处理单元 802 提供专用功能的软件指令。在该第一示例中,处理单元 802 可实现为一个或多个处理核和 / 或一个或多个单独的微处理器。例如,在该第一示例中,处理单元 802 可实现为一个或多个 Intel Core2 微处理器。处理单元 802 能够执行指令集中的指令,诸如 x86 指令集、POWER 指令集、RISC 指令集、SPARC 指令集、IA-64 指令集、MIPS 指令集,或其他指令集。在第二示例中,处理单元 802 可实现为提供专用功能的专用集成电路 (ASIC)。在第三示例中,处理单元 1002 可通过使用 ASIC 和通过执行软件指令来提供专用功能。

[0068] 电子计算设备 800 还包括视频接口 804,视频接口 804 使客户机系统 (102 和 104) 或服务器系统 (108、110、112、114、116 和 118) 能将视频信息输出到显示设备 806。显示设备 806 可以是各种不同类型的显示设备。例如,显示设备 806 可以是阴极射线管显示器、LCD 显示面板、等离子屏幕显示面板、触敏显示面板、LED 阵列,或其他类型的显示设备。

[0069] 另外,电子设备 802 包括非易失性存储设备 808。非易失性存储设备 808 是能够存储数据和 / 或指令的计算机可读数据存储介质。非易失性存储设备 808 可以是各种不同类型的不同非易失性存储设备。例如,非易失性存储设备 808 可以是一个或多个硬盘驱动器、磁带驱动器、CD-ROM 驱动器、DVD-ROM 驱动器、蓝光盘驱动器,或其他类型的非易失性存储设备。

[0070] 电子计算设备 800 还包括外部组件接口 810,该外部组件接口 810 使得客户机系统 102 和 104 以及服务器系统 108、110、112、114 和 116 能与外部组件进行通信。如图 8 的示例所示,外部组件接口 810 与输入设备 812 和外部存储设备 814 进行通信。在电子计算设备 800 的一个实现中,外部组件接口 810 是通用串行总线 (USB) 接口。在电子计算设备 800 的其他实现中,电子计算设备 800 可包括允许电子计算设备 800 与输入设备和 / 或输出设备进行通信的另一类型的接口。例如,电子计算设备 800 可包括 PS/2 接口。输入设备 812 可以是各种不同类型的设备,包括但不限于,键盘、鼠标、跟踪球、指示笔输入设备、触摸板、触敏显示设备、触敏显示屏、或其他类型的输入设备。外部存储设备 814 可以是各种不同类型的计算机可读数据存储介质,包括磁带、闪存模块、磁盘驱动器、光盘驱动器、和其他计算机可读数据存储介质。

[0071] 此外,电子计算设备 800 包括网络接口 816,网络接口 816 允许电子计算设备 800 向网络 106 发送数据并从网络 106 接收数据。网络接口 816 可以是各种不同类型的网络接口。例如,网络接口 816 可以是以太网接口、令牌环接口、光纤接口、无线网络接口 (例如, WiFi、WiMax 等)、或另一类型的网络接口。

[0072] 电子计算设备 800 还包括通信介质 818,通信介质 818 便于电子计算设备 800 中的

各个部件之间的通信。通信介质 818 可包括一个或多个不同类型的通信介质, 包括但不限于, PCI 总线、PCI Express 总线、加速图形端口 (AGP) 总线、Infiniband 互连、串行高级技术附件 (ATA) 互连、平行 ATA 互连、光通道互连、USB 总线、小型计算机系统接口 (SCSI) 接口、或其他类型的通信介质。

[0073] 在图 8 的示例中示出若干计算机可读数据存储介质 (即, 存储器单元 801、非易失性存储设备 808、和外部存储设备 814)。这些计算机可读数据存储介质可共同构成一个单独的逻辑计算机可读数据存储介质。该单独的逻辑计算机可读数据存储介质可存储处理单元 802 可执行的指令。以上描述所述的动作可源自存储在该单独的逻辑计算机可读数据存储介质上的指令的执行。由此, 当本描述表示特定的逻辑模块执行特定的动作时, 这种陈述可解释为意指软件模块的指令使处理单元, 诸如处理单元 802, 执行该动作。

[0074] 上述各个实施例仅作为说明提供, 并且不应被解释为限制。可对上述实施例作出各种修改和改变, 而不背离本公开的真实精神和范围。

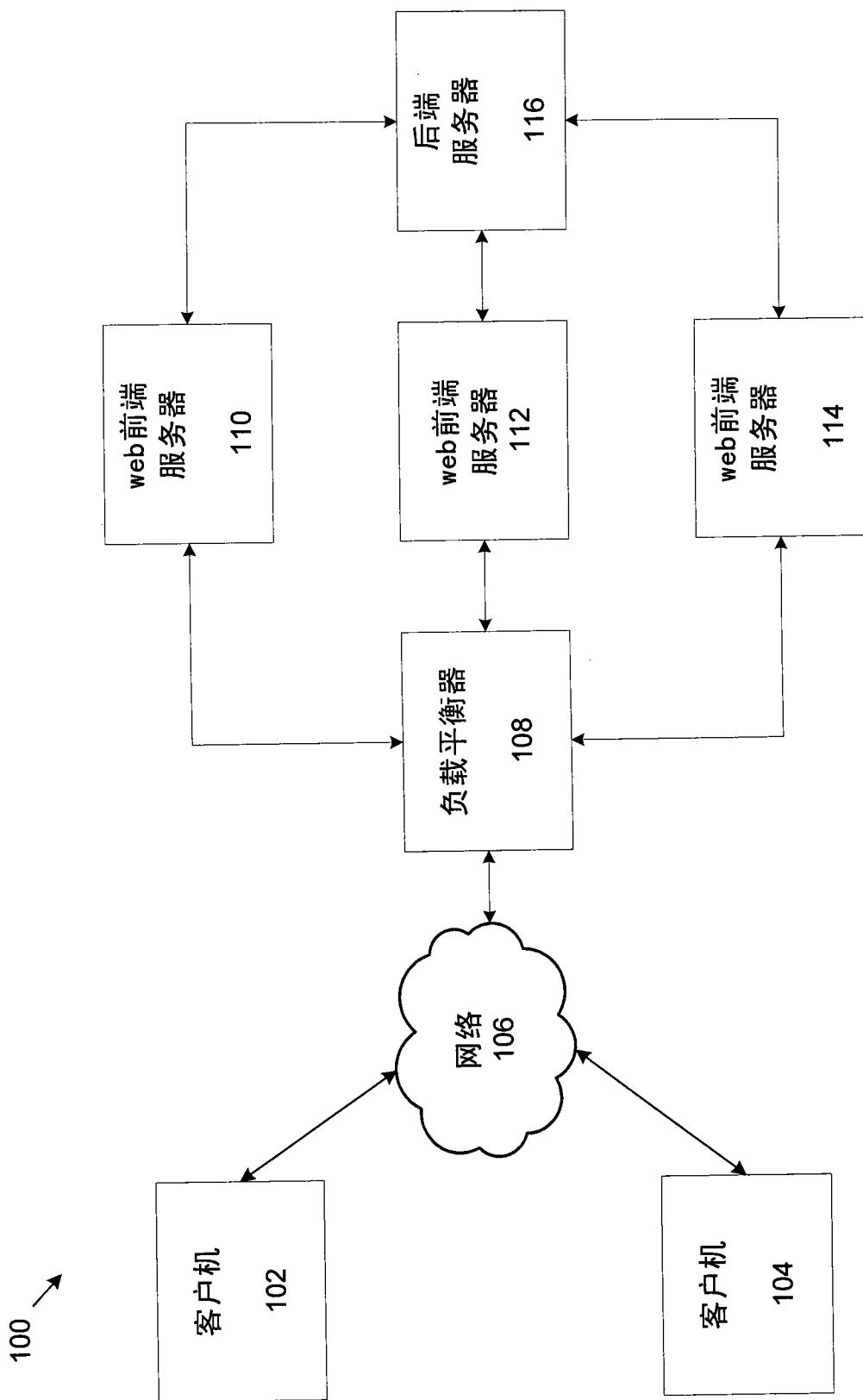


图 1

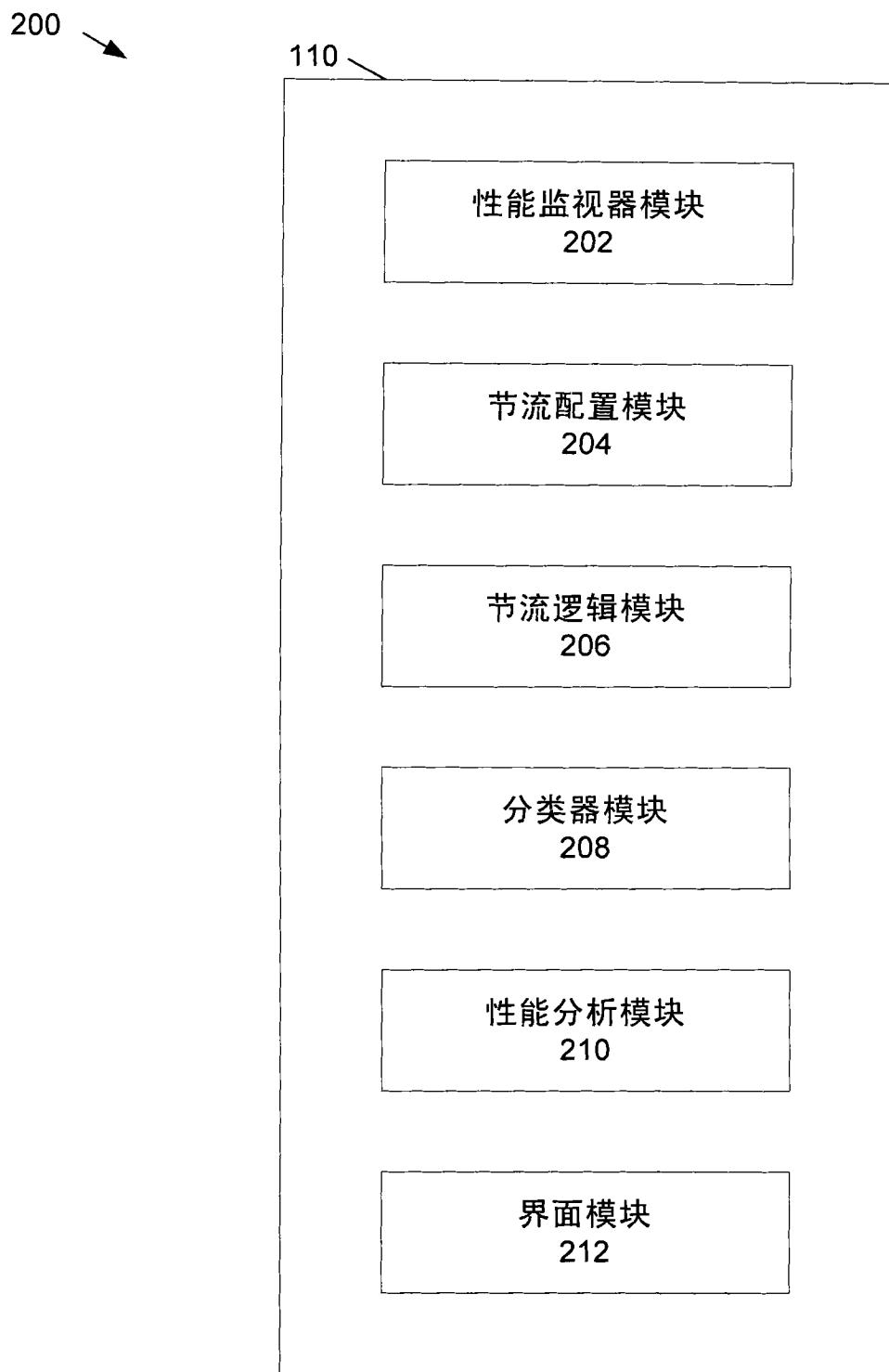


图 2

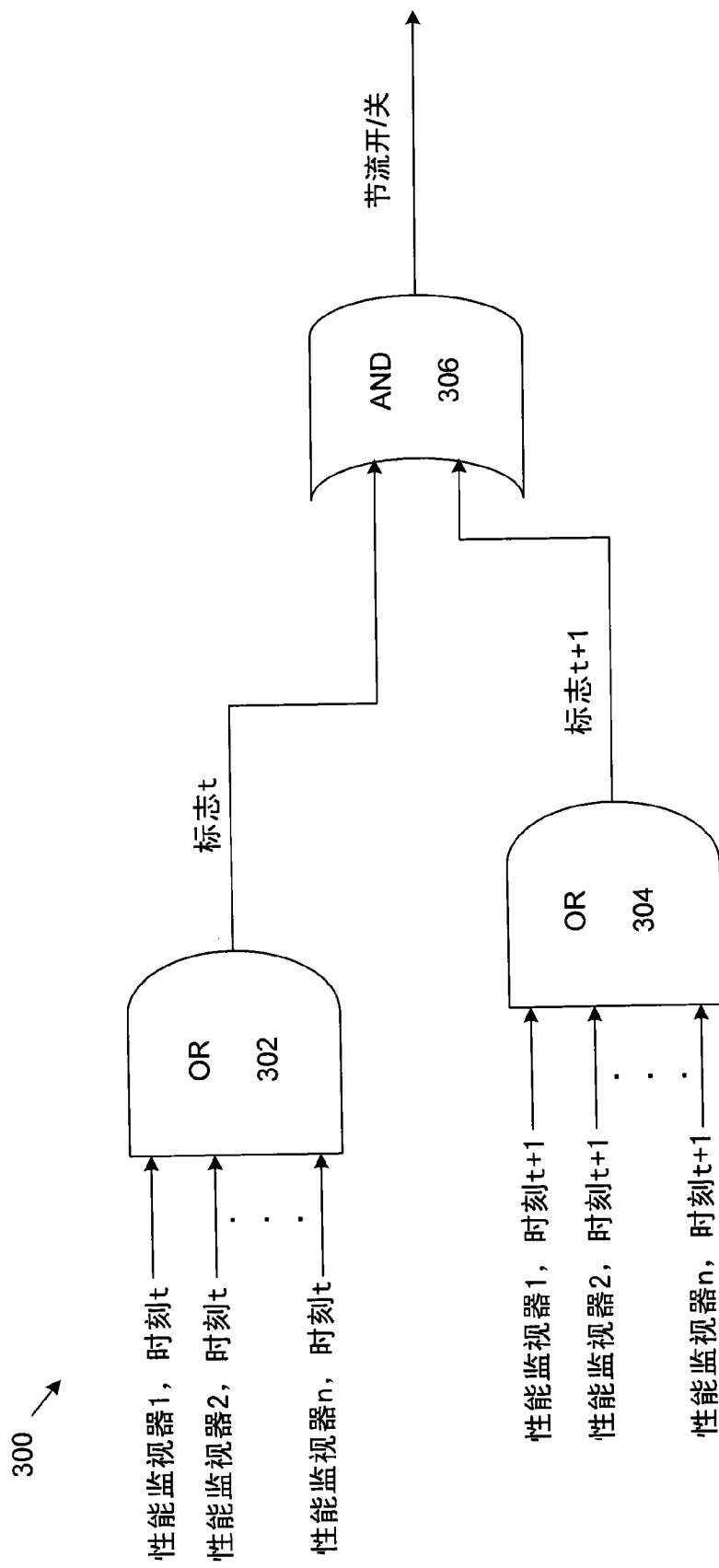


图 3

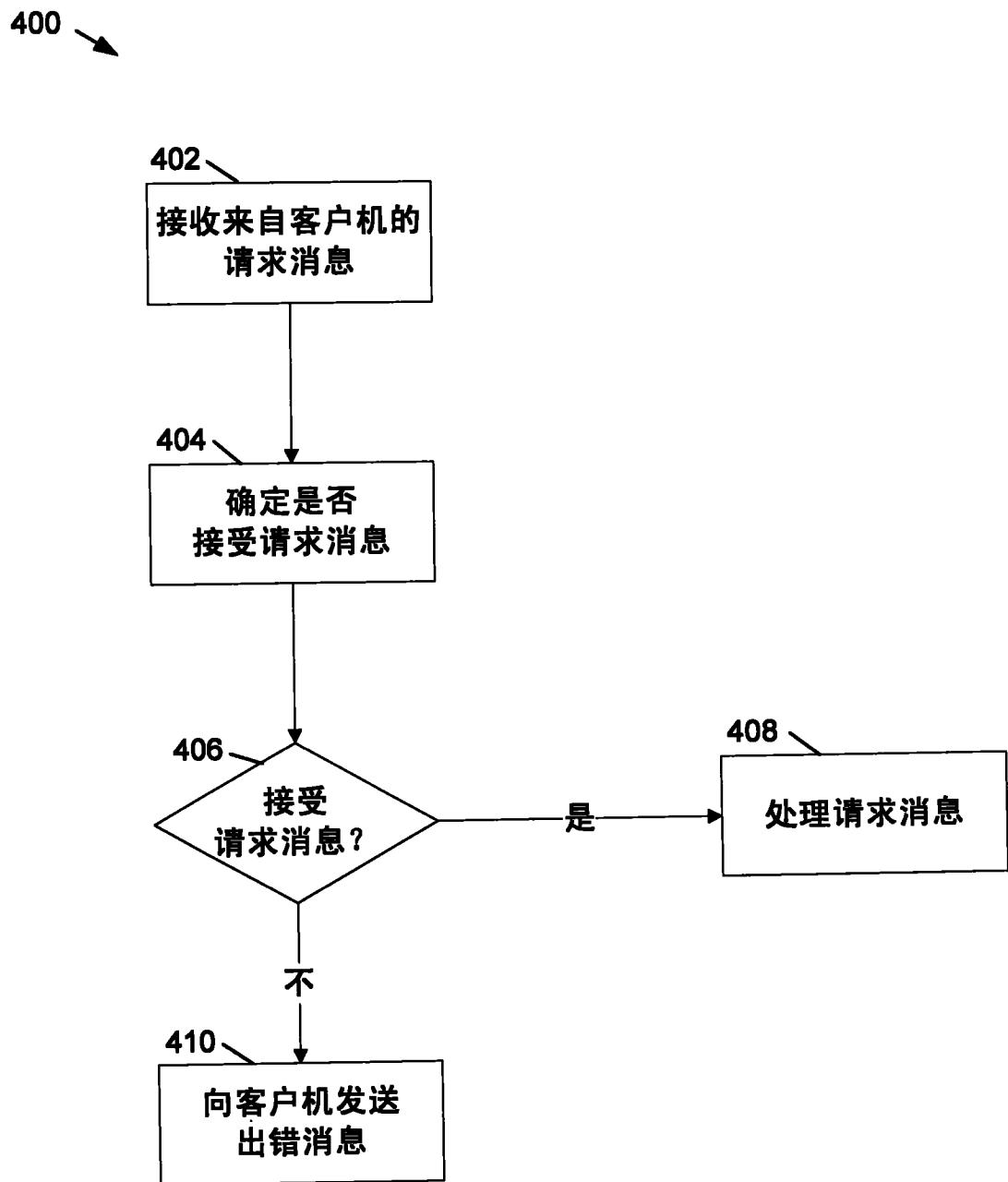


图 4

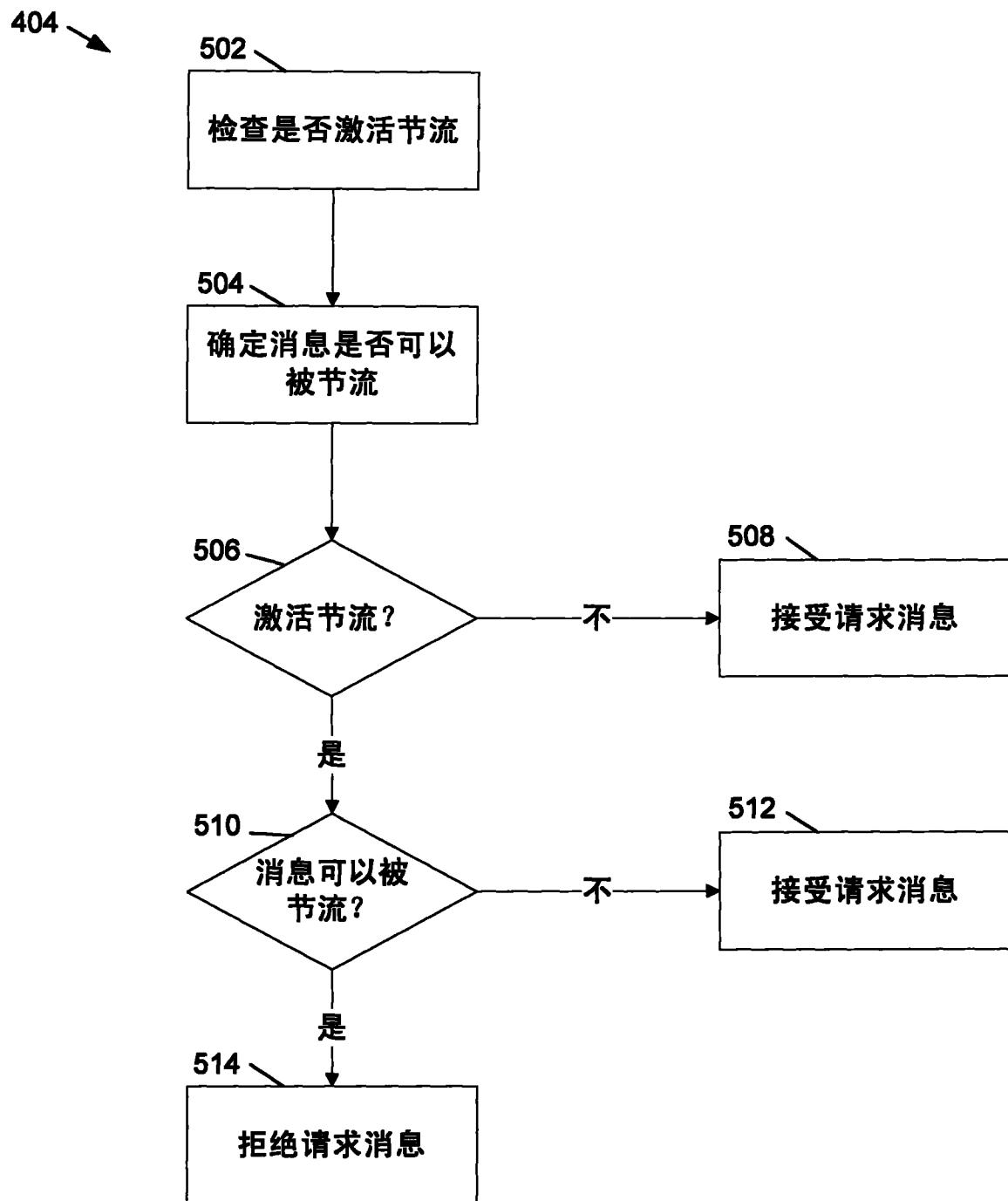


图 5

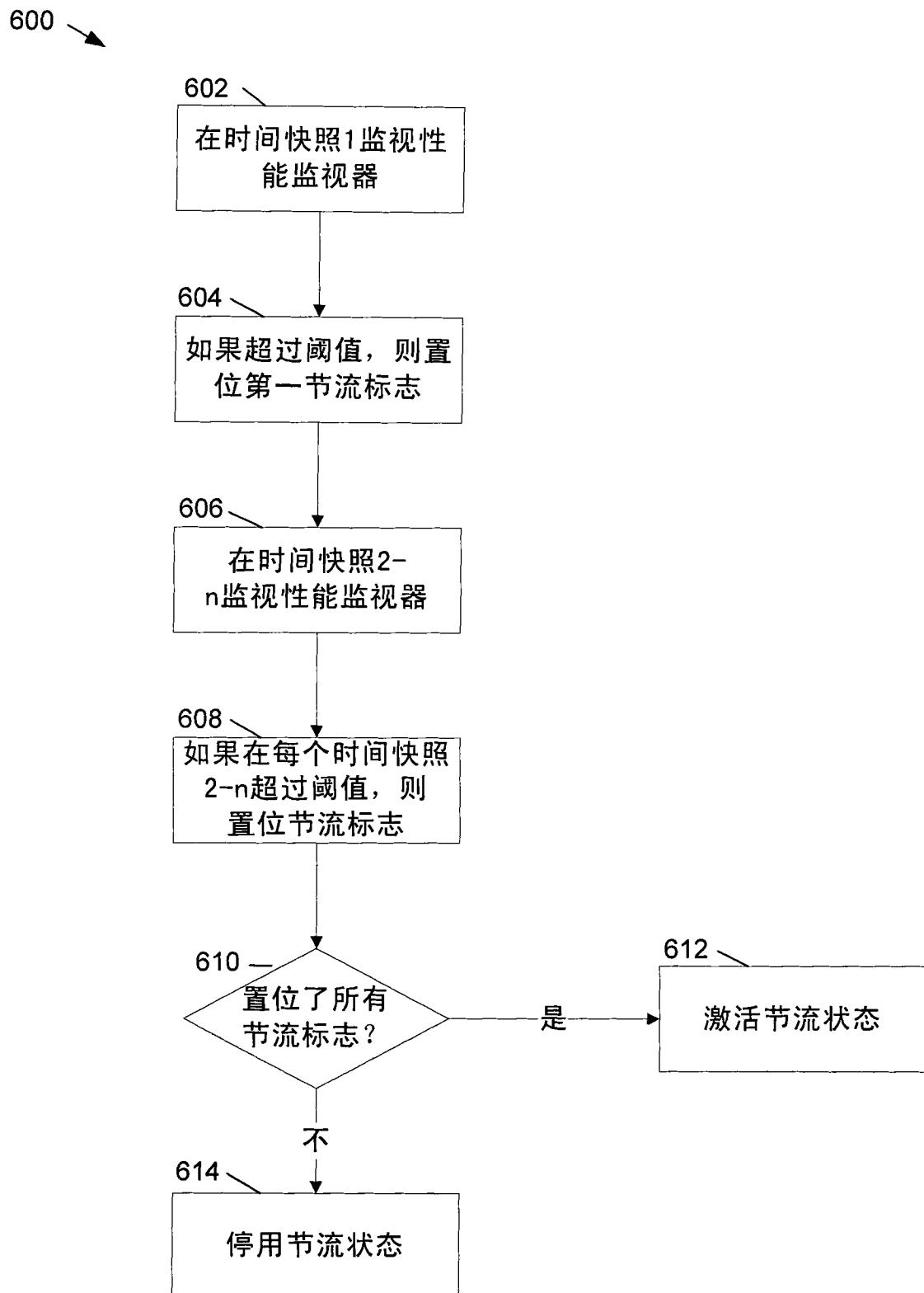


图 6

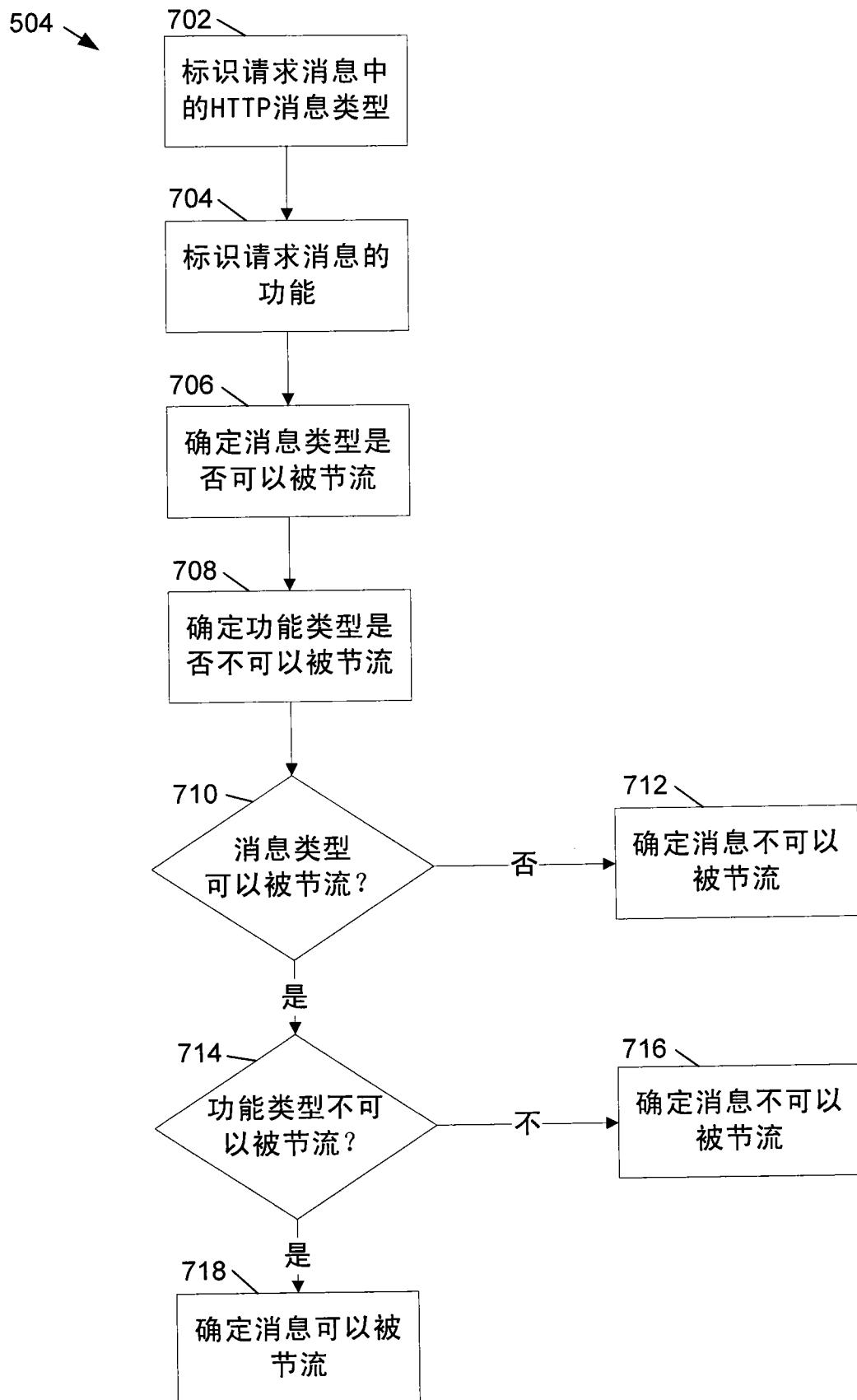


图 7

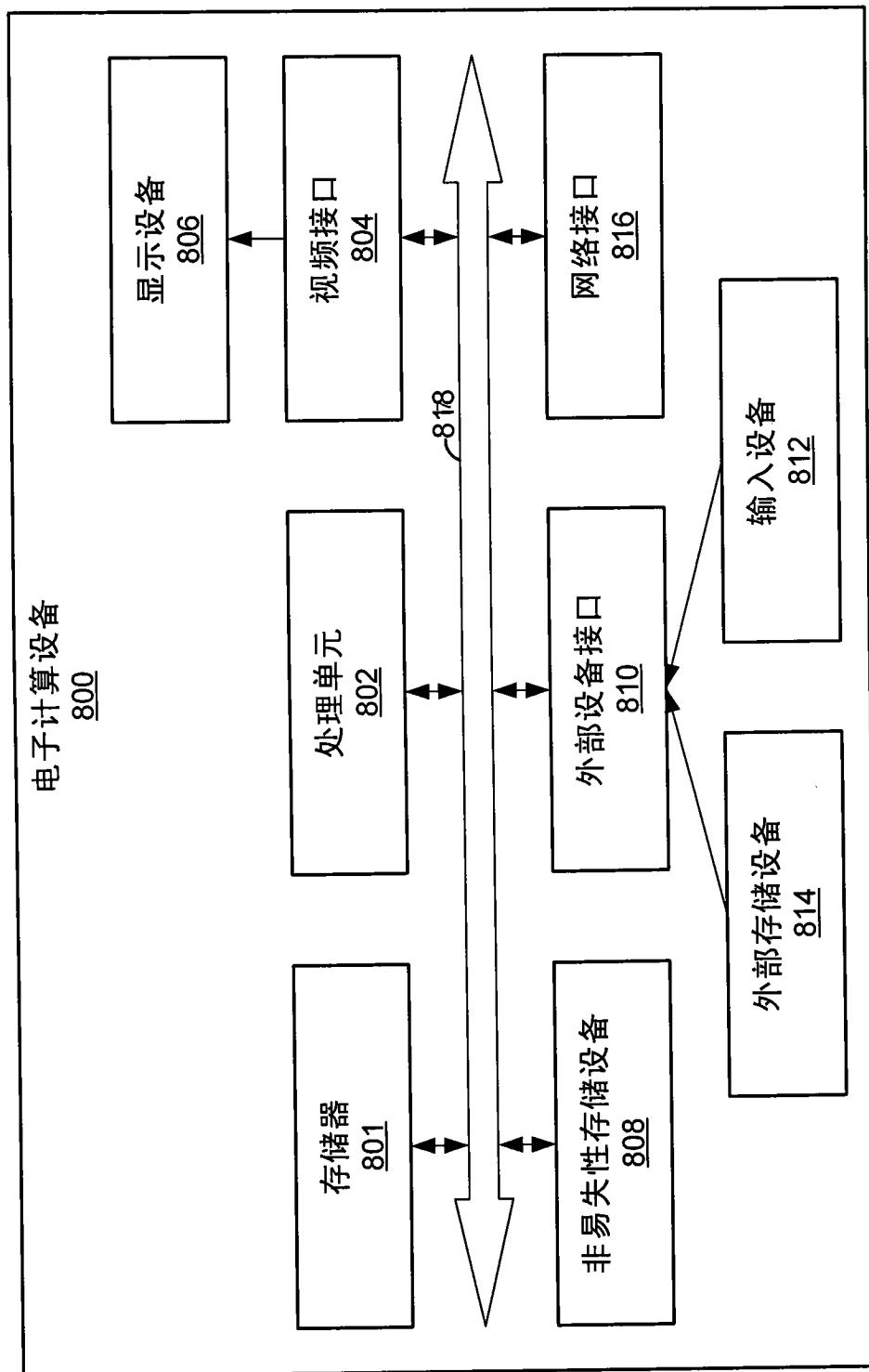


图 8