



[12] 发明专利说明书

[21] ZL 专利号 99123172.4

[45] 授权公告日 2004 年 5 月 26 日

[11] 授权公告号 CN 1151448C

[22] 申请日 1999.10.27 [21] 申请号 99123172.4

[30] 优先权

[32] 1998.11.20 [33] US [31] 09/196, 944

[71] 专利权人 国际商业机器公司

地址 美国纽约

[72] 发明人 丹尼尔·曼努埃尔·迪亚斯

阿朗·康吉尔·扬加尔

埃里克·莱维-阿贝格诺里 宋俊华

审查员 王京霞

[74] 专利代理机构 中国国际贸易促进委员会专利

商标事务所

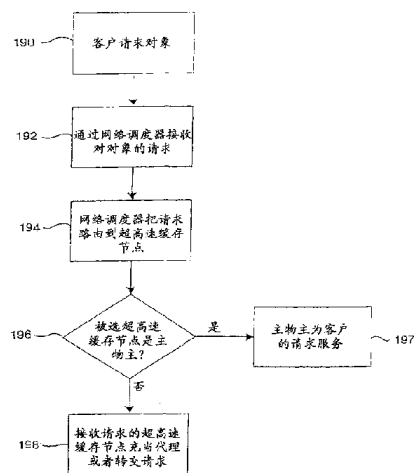
代理人 于 静

权利要求书 4 页 说明书 15 页 附图 9 页

[54] 发明名称 可扩缩的超高速缓存检索方法

[57] 摘要

从由多个节点组成的系统中的一个或多个客户接收对对象的请求。从一个客户发送一个请求。所述多个节点的第二节点接收来自所述多个节点的第一节点接收的请求。利用多种协议之一从多个节点的第二节点回送被请求对象。该方法可应用到可扩缩的和高利用的超高速缓存阵。该超高速缓存阵可增强和诸如因特网的网络连接的 Web 服务器的性能和吞吐量。网络调度器可把请求发送到超高速缓存阵的超高速缓存节点。网络调度器选择的超高速缓存节点或为请求服务、或把请求转交到超高速缓存阵的另一个超高速缓存节点、或经多种协议之一和超高速缓存的另一节点通信以共同为请求服务。网络调度器还可充当基于内容的路由器，从而可能利用基于内容的和不基于内容的路由选择的组合路由请求以进一步改进系统性能。



1. 在适用于从一个或多个客户接收诸对象请求的系统中，所述系统由多个节点组成，一种用于检索所述诸对象的被请求对象的方法，所述方法包括步骤：

(a) 从所述客户中的一个发送所述请求之一；

(b) 由所述多个节点中的另一个节点接收来自所述多个节点的一个节点的所述请求，其中所述一个节点包括多个超高速缓存中的至少一个，所述另一个节点包括所述多个超高速缓存中的至少一个；以及

(c) 利用多种协议中的一种从所述多个节点的所述另一个节点回送所述被请求对象。

2. 权利要求 1 的方法，其中所述一个节点从所述客户中的所述一个客户接收所述请求。

3. 权利要求 1 的方法，其中所述一个节点从所述多个节点中的另一个节点接收所述请求。

4. 权利要求 1 的方法，其中根据所述被请求对象的尺寸选择多种协议中的所述一种协议。

5. 权利要求 4 的方法，其中响应所述被请求对象的尺寸小于预定尺寸，把被请求对象从所述另一个节点回送到所述一个节点，并且继而从所述一个节点回送到所述客户中的所述一个客户。

6. 权利要求 4 的方法，其中响应所述被请求对象的尺寸大于预定尺寸，在不先通过所述一个节点下，把被请求对象从所述另一个节点回送到所述客户中的所述一个客户。

7. 权利要求 5 的方法，其中所述被请求对象在回送到所述客户中的所述一个客户之前从所述一个节点回送到所述多个节点中的至少一个别的节点上。

8. 权利要求 6 的方法，其中所述被请求对象在回送到所述客户中的所述一个客户之前从所述另一个节点回送到所述多个节点中的至少一个别的节点上。

9. 权利要求 1 的方法，其中响应所述一个节点不具有所述被请求对象的超高速缓存复制执行所述步骤 (b)。

10. 权利要求 1 的方法，其中所述多个节点中的一个节点充当后端服务器。

5 11. 权利要求 10 的方法，还包括响应所述另一个节点不具有所述被请求对象的超高速缓存复制把所述请求发送到后端服务器的步骤。

12. 权利要求 1 的方法，还包括在所述多个节点的一些上分配所述对象中的一些对象的物主权的步骤，并且其中所述多个节点中所述一些

10 13. 权利要求 12 的方法，其中所述另一个节点是所述被请求对象的物主，并且响应所述一个节点不是所述被请求对象的物主执行所述步骤 (b)。

14. 权利要求 12 的方法，其中所述对象中的任一个至多具有一个物主。

15 15. 权利要求 12 的方法，其中所述对象中的任一个具有多个物主。

16. 权利要求 1 的方法，其中多个协议中之一是 HTTP (超文本传输协议)、UDP (用户数据报协议) 和转交接口中之一。

17. 权利要求 1 的方法，其中所述多个节点中的一个充当后端服务器。

20 18. 权利要求 17 的方法，还包括响应所述另一个节点不具有所述被请求对象的超高速缓存复制把所述请求发送到后端服务器的步骤。

19. 一种检索多个对象中的一个目标对象的方法，该目标对象是由一个用户请求的，所述方法包括步骤：

(a) 把多个对象中的每个对象分配给多个节点的至少一个节点；

25 (b) 把多个超高速缓存中的至少一个分配给多个节点中的至少一个节点；

(c) 把对该目标对象的请求发送到该多个节点中的第一节点；

(d) 判定是否把该目标对象分配给了第一节点；以及

(e) 若未把该目标对象分配给第一节点；

30 (i) 把该请求发送到该多个节点中的第二节点，第二节点分配着

该目标对象，以及

(ii) 若该目标对象存储在该多个超高速缓存中和第二节点关联的超高速缓存中，利用多种协议中的一种从和第二超高速缓存关联的超高速缓存中检索该目标对象。

5 20. 依据权利要求 19 的方法，其中通过多种协议中的至少一种协议利用通信执行步骤 (ii)。

21. 依据权利要求 20 的方法，其中根据目标对象的尺寸选择多个协议中的所述至少一个协议。

22. 依据权利要求 21 的方法，还包括步骤：

10 若未把该目标对象分配给第一节点并且若该目标对象的尺寸小于预定目标尺寸，从第二节点把该目标对象传送到第一节点，并且从第一节点把该目标对象传送到客户。

23. 依据权利要求 21 的方法，还包括步骤：

15 若未把该目标对象分配给第一节点并且若该目标对象的尺寸大于预定目标尺寸，把该目标对象从第二节点传送到客户。

24. 依据权利要求 20 的方法，其中在第一节点和第二节点之间执行多种协议中的第一协议，并且在第一节点和客户之间执行多种协议中的第二协议。

20 25. 依据权利要求 20 的方法，其中在第二节点和客户之间执行多种协议中的第一协议。

26. 依据权利要求 19 的方法，还包括步骤：

 若该目标对象分配给第一节点，如果该目标对象存储在分配给第一节点的多个超高速缓存的某个超高速缓存中，从分配的该超高速缓存检索目标对象。

25 27. 依据权利要求 26 的方法，还包括步骤：

 若该目标对象未存储在分配给第一节点的该高速缓存中，从和第一节点对应的服务器中检索该目标对象。

30 28. 依据权利要求 19 的方法，其中若目标对象未存储在分配给第二节点的高速缓存中，步骤 (e) 从和第二节点对应的服务器检索目标对象。

29. 依据权利要求 19 的方法，其中多种协议中的一种是 HTTP(超文本传输协议)、UDP(用户数据报协议)和转交接口中的一种。

30. 根据权利要求 19 的方法，其中还包括步骤：

5 (m) 根据请求内容把请求之一选择性地发送到节点中的至少一个上；

(n) 和请求的内容无关地把请求之一选择性地发送到节点中的任何一个上；以及

(i) 为了改进系统性能，改变执行步骤 (m) 和 (n) 的频率。

10 31. 权利要求 30 的方法，其中为了减少对系统资源的使用和改进系统性能，增大执行步骤 (m) 的频率。

32. 权利要求 30 的方法，其中为了减少对系统资源的使用和改进系统性能，增大执行步骤 (n) 的频率。

33. 权利要求 30 的方法，其中步骤 (m) 还包括把连接从一个节点转交到另一个节点上。

可扩展的超高速缓存检索方法

5 技术领域

本发明一般地涉及超高速缓存系统。本发明尤其涉及计算机网络中的可扩展的和高利用的超高速缓存。此外，本发明涉及一种应用于客户/服务器环境，尤其例如万维网环境的超高速缓存阵列。

技术背景

10 为改进计算机系统的性能常常使用超高速缓存。一旦把一个对象存储到超高速缓存器中，对该超高速缓存的对象的随后请求可由该超高速缓存器满足。和从远程位置再生或检索该对象相比，从超高速缓存器满足对对象的请求可带来较少的开销。对 Web 服务日益增加的需求带来的慢性能可能会使 Web 服务器变成低效的或者不可用的。

15 超高速缓存提供一种应付不断增加 Web 服务和 Proxy 服务吞吐量的要求的方法。例如万维网上的客户和服务器系统可利用超高速缓存改进性能。在一些情况下，没有超高速缓存的帮助可能会很慢地和低效地执行 Web 服务器应用程序。不利用超高速缓存方法，Web 服务器可能成为系统瓶颈。在某 Web 服务器上运行的基础操作系统例如可能会有妨碍该 Web 服务器的吞吐量的性能问题。一种用于改进 Web 服务器的性能的技术是把经常请求的数据（例如，Web 页面）存储在超高速缓存器中。从超高速缓存器检索数据所需的开销可少于从 Web 服务器检索数据所需的开销。

发明内容

在适用于从一个或多个客户接收诸对象请求的系统中，所述系统由多个节点组成。一种用于检索所述诸对象的被请求对象的方法，所述方法包括步骤：（a）从所述
25 客户中的一个发送所述请求之一；（b）由所述多个节点中的另一个节点接收来自所述多个节点的一个节点的所述请求，其中所述一个节点包括多个超高速缓存中的至少一个，所述另一个节点包括所述多个超高速缓存中的至少一个；以及（c）利用多种协议中的一种从所述多个节点的所述另一个节点回送所述被请求对象。

一种检索多个对象中的一个目标对象的方法，该目标对象是由一个用户请求的，
30 所述方法包括步骤：（a）把多个对象中的每个对象分配给多个节点的至少一个节点；（b）把多个超高速缓存中的至少一个分配给多个节点中的至少一个节点；（c）把对该目标对象的请求发送到该多个节点中的第一节点；（d）判定是否把该目标对象分配给了第一节点；以及（e）若未把该目标对象分配给第一节点；（i）把该请求发送到该多个节点中的第二节点，第二节点分配着该目标对象，以及（ii）若该目标对象存储

在该多个超高速缓存中和第二节点关联的超高速缓存中，利用多种协议中的一种从和第二超高速缓存关联的超高速缓存中检索该目标对象。

根据各附图阅读下列详细说明将对本发明得到最好的理解。要强调的是，根据通常的习惯，附图的各细节不是按尺度的。反之，出于清晰，各细节的尺度任意扩大或缩小。其中：

附图说明

图1(a)是依据本发明的一种实施例的示例计算机网络系统的框图；

图1(b)是依据本发明的示范实施例检索被请求对象的方法的流程图；

图2是一个框图，说明依据本发明的示范实施例在命中某超高速缓存器成员和命中超高速缓存阵列的情况下检索被请求对象的方法；

图3是一个框图，说明依据本发明的示范实施例在命中某超高速缓存器成员但未命中某超高速缓存阵列的情况下检索被请求对象的方法；

图4至图7是框图，说明依据本发明的示范实施例在未命中超高速缓存器成员情况下检索被请求对象的方法；以及

图8是一个框图，说明依据本发明的示范实施例在基于内容进行路由选择情况下检索被请求对象的方法。

具体实施方式

术语表

尽管本文中各术语仍蕴含着词典含义，下述定义可能是有帮助的。

客户

客户计算机典型地向一个或多个执行与命令和/或请求关联的任务的服务器发出命令和/或请求。

服务器

服务器是在一个或多个客户计算机的命令和/或请求下执行某任务的计算机。

万维网 (Web)

通过提供从服务器到服务器超链接链接文档的因特网服务。用户通过点击亮度增强的感兴趣的词组或短语(超文本链接)、点击诸如小应用程序或图象图的图形、填表和输入URL可在文档间“跳跃”。无论文档存储在因特网上的何处，用户能在文档间“跳跃”。因特网Web服务器支持用户并且提供信息。用户利用客户程序“浏览Web”。可把Web当作带有用URL定址的资源、并且使用HTTP协议在计算机之间传送信息的因特网。可利用HTML(除别的格式之外)显示和URL相对应的信息并且提供对其它URL的点击接口。

统一资源定位符 (URL)

因特网上资源的地址。Web浏览器利用URL定位因特网资源。URL

规定访问资源时所采用的协议（例如 `http:` 用于万维网页面或者 `ftp:` 用于 FTP 网点），在其上驻留资源的服务器名（例如 `//www.whitehouse.gov`），以及任选的至某资源的通路（例如 HTML 文档或该服务器上的文件）。

5 超文本标记语言（HTML）

该标记语言用于万维网上的文档。HTML 是 SGML（标准广义标记语言）的一种应用，其利用标志标记文档中的元素，例如文本和图形，以指示 Web 浏览器应如何向用户显示这些元素以及应如何响应用户通过键击或鼠标点击发出的例如激励某键击的动作。HTML 是 Web 服务器上内容的与设备无关的表达。Web 服务器可在相信客户会选择适当的表达下以 HTML 向客户发送内容（包括超链接）。

超文本传输协议（HTTP）

该客户/服务器协议用于访问万维网上的信息。HTTP 是无状态协议的一种例子。换言之、独立地处理每个从客户到服务器的请求。利用该协议客户向服务器发送请求并且服务器用该协议应答。

用户数据报协议（UDP）

TCP/IP 内的一种无连接协议，其对应于 ISO/OSI 模型中的传送层。UDP 把应用程序生成的数据报文转换成经 IP 发送的分组，但可能不验证报文被正确地发送。从而，UDP 可能比 TCP 更为有效，因此它可用于各种用途，包括 SNMP（简单网络管理协议）；其可靠性可能取决于生成报文的应用程序。

路由器

通信网络上加快报文传送的中间设备。在通过各种可能连接的网格连接许多计算机的单网络中，路由器接收被发送的报文并且沿着可得到的最有效路由把报文发送到它们的正确目的地上。在采用相同通信协议的互连局域网（LAN）集中，路由器为有些困难的动作功能充当 LAN 之间的链接，实现报文从一个 LAN 发送到另一个 LAN。

Web 浏览器

一种能使用户查看万维网、其它网络或用户计算机上的 HTML（或其它）文档、追踪文档之间超链接并传送文件的客户应用程序。

传输控制协议 (TCP)

TCP/IP 内的协议, 其管理把数据报文分解成经 IP 发送的分组并且从通过 IP 接收的分组重组完整的报文并进行验证。TCP 对应于 ISO/OSI 模型中的传送层。

5 网际协议 (IP)

TCP/IP 内的协议, 其管理把数据报文分解成分组、对从发送方到目的地网络和网站进行路由选择并且在目的地把各分组重组原始数据报文。IP 对应于 ISO/OSI 模型中的网络层。

TCP/IP

10 由美国国防部开发的用于在计算机之间通信的一种协议。它建立在 UNIX 系统中并且事实上已成为网络 (包括因特网) 上数据传输的标准。

代理服务器

管理对和从局域网 (LAN) 的因特网业务的部件, 并可提供其它性能, 例如文档超高速缓存和存取控制。代理服务器可以通过提供频繁
15 请求的数据, 例如大众的 Web 网页, 改进性能, 并且可以滤掉和丢弃物主未适当考虑的请求, 例如未授权下对专有文件的访问请求。

超高速缓存器

专用存储器子系统, 其中复制常使用的数据以供快速访问。

对象

20 对象是可存储在超高速缓存器、服务器或客户机中的数据。

图 1 (a) 示出一个示例性的计算机网络系统, 包括: 客户 110、120 和 130, 网络 150, 网络调度器 160, 超高速缓存器组 170 以及服务器组 180。超高速缓存器组 170 包括超高速缓存器节点 172、174、176 和 178。服务器组 180 包括服务器 (“后端”) 服务器 182、184、186 和 188。

25 客户计算机 110、120 和 130 发出对例如 Web 页的对象的请求。

图 1 (b) 是依据本发明的示范实施例检索被请求对象的方法的流程图。在步骤 190 某客户, 例如客户 110, 请求可能由某服务器, 例如服务器 182, 存储 (或生成或取出) 的一个对象。在步骤 192, 经例如网络 150 的网络该对对象的请求由网络调度器, 例如网络调度器 160,
30 接收。在步骤 194, 网络调度器 160 把该对对象的请求路由到一个超高

速缓存器节点, 例如超高速缓存器 172。在步骤 196, 判定步骤 194 中从网络调度器 160 接收该对象请求的超高速缓存器 172 是否是该对象的主物主 (primary owner)。若在步骤 194 从网络调度器 160 接收该对象请求的超高速缓存器 172 是主物主, 则超高速缓存器 172 可在步骤 197 为该客户请求服务。反之, 在步骤 198, 超高速缓存器 172 和主物主可充当代理 (例如通过 HTTP 或 UDP 接口) 以为该请求服务并检索被请求的对象, 或者超高速缓存器 172 可把该请求转交给主物主。在步骤 197 和 198, 主物主从超高速缓存器或通过和服务器 182 的通信检索被请求的对象。请注意, 一般的业内人士可修改本发明以使多于一个的超高速缓存器可以是一个对象的主物主。

客户 110、120 和 130 可以是和因特网连接的工作站、个人计算机或别的计算机。例如, 一个在家里使用个人计算机的用户可通过用 Web 浏览器输入相应的 URL 请求检索和观看某 Web 网页。由该 URL 定址的被请求 Web 网页可能属于可通过因特网上的 Web 服务访问的服务器。

超高速缓存器组 170 可以是一个或多个网络节点。超高速缓存器组 170 中所包括的每个节点可以是一个或多个处理机。超高速缓存器组 170 的每个节点的每个处理机可包括一组一个或多个超高速缓存部件 (超高速缓存器节点组), 它们可形成单个超高速缓存空间和单个超高速缓存映象。换言之, 客户可把一个超高速缓存器阵看成是单个映象。例如, 客户可通过和网络调度器 160 相关的地址访问超高速缓存器组 170, 而超高速缓存器组 170 中的每个节点可具有一个独立的 IP 地址。超高速缓存器组可在内部把各个超高速缓存部件的资源组合起来。例如, 可以把超高速缓存部件的超高速缓存空间组合起来以改变可用于超高速缓存的存储器空间。此外, 还可以组合超高速缓存部件的各吞吐量以改变可使用的吞吐量。通过不同的地址 (例如 IP 地址) 可在超高速缓存器组 170 内部对超高速缓存部件 172、174、176 和 178 定址。例如, 可通过一个路由器实现超高速缓存节点 (件) 172、174、176 和 178。例如可采用 IBM 22XX 路由器系列中的路由器。

网络调度器 160 例如可在 TCP 路由器上实现。当网络调度器 160

从一个客户，例如客户 110，接收请求时，网络调度器可在不“查看”该请求下把该请求路由到超高速缓存器组 170 中的某个超高速缓存节点。网络调度器 160 例如可以是路由器节点的业务，以把对 Web 网页（或别的对象）的客户请求路由到超高速缓存器组 170。网络调度器 160 5 可得到有关超高速缓存器组 170 的各超高速缓存节点 172、174、176、178 的可使用性及负载信息。网络调度器 160 还可根据超高速缓存节点的可用性以及负载信息向选定的超高速缓存节点路由并发送对对象的请求（例如参加，G. Hunt 等的“网络调度器：用于可扩展因特网服务的连接路由器”，第七届国际万维网会议预告集，1998 年 4 月）。

10 在超高速缓存节点和被请求对象之间可形成一种对应。被请求对象可对应到为该被请求对象的主物主的某超高速缓存节点，或者对应到别的超高速缓存节点。例如，用 URL 表达的因特网地址可在若干 Web 超高速缓存节点之中分配。对于每个 URL，可把一个 Web 超高速缓存节点指定为主物主。除主物主之外的一个或多个 Web 超高速缓存节点也 15 可和单个 URL 对应。例如利用散列函数在若干 Web 超高速缓存节点中分配各 URL。换言之，可利用散列函数形成超高速缓存节点和被请求对象之间的对应。

当客户 110、120 或 130 发送检索某对象（即目标对象）的请求时，网络调度器 160 把该请求路由并发送到超高速缓存器组 170 中的第一超 20 高速缓存器。被网络调度器 160 选择的第一超高速缓存器可能对应于或不对应该被请求的对象。换言之，可能把或者未把该被请求对象分配给该第一超高速缓存器。若由网络调度器 160 选择的第一超高速缓存器对应于被请求的对象（即它是被请求对象的主物主）并且被请求对象存储在 25 第一超高速缓存器中，则第一超高速缓存器可服务该请求。若由网络调度器 160 选择的第一超高速缓存器和被请求对象对应（即它是被请求对象的主物主）但被请求对象未存储在第一超高速缓存器中，则第一超高速缓存器可从服务器群 180 的适当服务器 182、184、186 或 188 检索被请求对象。

例如，假定某用户利用客户机上的 Web 浏览器请求检索和观看某 30 Web 网页。该被请求的 Web 网页可用某 URL 定址并存储在与网络连接

的某服务器中。网络调度器 160 选择超高速缓存器组 170 的第一超高速缓存节点。网络调度器 160 把对该 Web 网页的请求路由到第一超高速缓存节点。若被请求的 Web 网页存储在第一超高速缓存中，则第一超高速缓存节点可从超高速缓存器检索该 Web 网页，并且向该客户机回
5 送该 Web 网页。若第一超高速缓存节点是被请求 Web 网页的主物主，但该 Web 网页未存储在第一超高速缓存节点中，则第一超高速缓存节点可访问通过 URL 定址的适当服务器，从该服务器检索该 Web 网页，并把该 Web 网页回送到客户机。

然而，若网络调度器 160 选择的第一超高速缓存器不对应于被请求
10 对象（即它不是被请求对象的主物主），则第一超高速缓存器可把该请求发送到和该被请求对象对应的第二超高速缓存器（即第二超高速缓存器是被请求对象的主物主）。第一和第二超高速缓存器可进行通信，以对请求服务并检索被请求对象。替代地，第一超高速缓存器可把该请求转交给第二超高速缓存器。在转交的情况下，可从第一超高速缓存器把
15 该请求以及和 TCP 连接有关的信息（例如，序号、IP 地址和 TCP 端口）发送到超高速缓存器组 170 的第二超高速缓存器。

例如可以根据被请求对象的大小做出是使第一和第二超高速缓存器通信来为请求服务还是使第一超高速缓存器把请求转交给第二超高速缓存器的决策。

20 例如，可遵循下述规则：

1. 若被请求对象的大小小于阈尺寸，则第二超高速缓存器把被请求对象发送给第一超高速缓存器。第二超高速缓存器从超高速缓存器或与适当的服务器通信检索被请求对象。接着第一超高速缓存器可通过向
25 请求客户发送被请求对象为该请求服务。

2. 若被请求对象的大小大于阈尺寸，则第一和第二超高速缓存器相协调以转交该请求：

(i) 从第一超高速缓存器向第二超高速缓存器转交 TCP 连接；

(ii) 通知网络调度器 160 将由第二超高速缓存器对该请求服务（请
30 注意还可通知网络调度器 160 把对某对象的以后请求发送到第二超高速缓存器）；

(iii) 第二超高速缓存器从超高速缓存器或者与适当服务器的通信检索被请求对象，并把被请求对象发送到请求客户。

请注意可动态地调整或改变阈尺寸。

在这种类型的示例协调下，第二超高速缓存器可存储用于执行 TCP
5 转交（接管）和/或 HTTP 转交（接管）的信息字段。最初从网络调度器 160 接收请求的第一超高速缓存器可等待，直至第二超高速缓存器（被请求对象的主物主）向第一超高速缓存器发送被请求对象或者通知第一超高速缓存器由第二超高速缓存器对该请求服务。从而，上面说明的这种类型的示例协调可由被请求对象的主物主驱动。

10 当网络调度器 160 向是被请求对象的主物主的第一超高速缓存器传送连接请求时，出现超高速缓存部件命中。若例如网络调度器 160 采用循环方案，则某特定超高速缓存部件被选择的概率是均匀分布的（即在 n 个超高速缓存部件的情况下概率为 $1/n$ ）。此外，网络调度器 160 可访问和被请求对象相关的寻址信息（例如 URL），但这是以明显的开销
15 为代价的。这样，网络调度器 160 可访问足够的信息以确定和服务器群集 180 相关的超高速缓存器组 170，接着随机地选择一个超高速缓存部件。若向网络调度器 160 提供负载及可用性信息，可对选择某特定超高速缓存器的可能加权。换言之，若一个超高速缓存器负载重，则较少地选择它，并可更经常地选择其余的超高速缓存部件。

20 当超高速缓存器组 170 可从一个超高速缓存部件 172、174、176、178 的超高速缓存空间为某对象的请求服务时，出现超高速缓存器阵击中。从而有可能同时出现超高速缓存部件命中和超高速缓存器阵未命中。例如，网络调度器 160 可能选择是被请求对象的主物主的一个超高速缓存部件，但该被请求对象的主物主并不在超高速缓存器中具有被请
25 求对象。这样，在超高速缓存部件命中及超高速缓存器阵未命中的情况下，主物主从服务器检索被请求对象。

另外，可同时出现超高速缓存部件未命中和超高速缓存器命中。例如，网络调度器 160 可能选择一个不是被请求对象的主物主的超高速缓存部件，但该被请求对象的主物主，即超高速缓存器组 170 的某超高速
30 缓存部件，在超高速缓存器中具有该被请求对象。

从而，可能出现下列四种情况：

1. 超高速缓存部件命中，超高速缓存器阵命中
2. 超高速缓存部件命中，超高速缓存器阵未命中
3. 超高速缓存部件未命中，超高速缓存器阵命中
- 5 4. 超高速缓存部件未命中，超高速缓存器阵未命中

另外，可能使用不同的通信协议检索被请求对象。可能取决于是否第一和第二超高速缓存器通信以服务某请求或者是否第一超高速缓存器把某请求转交给第二超高速缓存器，采用不同的协议。例如，可能采用 HTTP 接口（其中第一超高速缓存器可充当 HTTP 代理）、基于 UDP
10 的请求或转交。

可能出现下列八种情况：

1. 超高速缓存部件命中，超高速缓存器阵命中。
2. 超高速缓存部件命中，超高速缓存器阵列未命中。
3. 超高速缓存部件未命中，超高速缓存器阵命中，
- 15 3. 1. 利用 HTTP 检索对象，
3. 2. 利用 UDP 检索对象，或
3. 3. 通过转交请求检索对象。
4. 超高速缓存部件未命中，超高速缓存器阵未命中，
4. 1. 利用 HTTP 检索对象，
- 20 4. 2. 利用 UDP 检索对象，或
4. 3. 通过转交请求检索对象。

超高速缓存部件命中

图 2 和图 3 是框图，说明依据本发明的示范实施例在超高速缓存部件命中下检索被请求对象的方法。客户 110 发出经网络 150 检索某对象
25 的请求。该被请求对象存储在服务器群集 180 中的一个服务器 182、184 上或者可由一个服务器 182、184 生成或取得。经网络 150 把客户 110 发出的请求传送到网络调度器 160。接着网络调度器 160 选择第一超高速缓存器，例如超高速缓存器组 170 的超高速缓存部件 172。假定网络调度器 160 选择的第一超高速缓存器 172 是被请求对象的主物主，出现
30 超高速缓存部件命中。

图 2 表示第一超高速缓存器在超高速缓存中具有该被请求的对象（超高速缓存器命中）的情况。在该情况中，第一超高速缓存器 172 可从超高速缓存检索该被请求的对象，并且经网络 150 把被请求对象发送给客户 110。

5 图 3 表示第一超高速缓存器在超高速缓存中不具有该被请求的对象（超高速缓存器未命中）的情况。在该情况下，第一超高速缓存器 172 可能先从服务器 182 检索被请求的对象，并接着经网络 150 把被请求对象发送给客户 110。和超高速缓存部件命中及超高速缓存器命中情况相比，超高速缓存部件命中但超高速缓存器未命中情况的系统性能
10 （例如开销和吞吐量）可能下降。

超高速缓存部件未命中 - HTTP 接口

图 4 和图 5 是框图，说明依据本发明的示范实施例在未命中超高速缓存部件情况下检索被请求对象的方法。客户 110 发出经网络检索某对象的请求。该被请求对象存储在服务器群集 180 中的一个服务器 182、184
15 上或者可由一个服务器 182、184 生成或取得。经网络 150 把客户 110 发出的请求传送到网络调度器 160。接着网络调度器 160 选择第一超高速缓存器，例如超高速缓存器组 170 的超高速缓存部件 172。假定网络调度器选择的第一超高速缓存器 172 不是被请求对象的主物主，出现未命中超高速缓存部件。请注意在网络调度器 160 随机地（以均匀分布）
20 选择超高速缓存器组 170（带有 n 个超高速缓存部件）中的第一部件的情况下，未命中超高速缓存部件的概率是 $(n-1) \div n$ 。

图 4 表示第二超高速缓存，例如超高速缓存部件 176，是被请求对象的主物主并且超高速缓存部件 176 在超高速缓存器中具有被请求对象（超高速缓存器命中）的情况。在该情况下，第一超高速缓存器 172
25 可接受和客户 110 的连接。然后第一超高速缓存器 172 可确定第二超高速缓存器 176 并且建立 HTTP 接口。第一超高速缓存器 172 和第二超高速缓存器 176 可通过 HTTP 接口进行通信以检索被请求对象。第二超高速缓存器 176 可从超高速缓存器检索被请求对象并把被请求对象发送到
30 第一超高速缓存器 172。第一超高速缓存器 172 可经过网络 150 把被请求对象发送给客户 110。

图 5 表示第二超高速缓存器 176 不在超高速缓存器中具有被请求对象（超高速缓存器阵未击中）的情况。在这种情况下，第二超高速缓存器 176 可先从服务器 184 检索被请求对象，并接着把被请求对象发送到第一超高速缓存器 172。然后第一超高速缓存器 172 可经网络 150 把被请求对象发送到客户 110。

超高速缓存部件未命中 - UDP 接口

图 4 表示第二超高速缓存器，例如超高速缓存部件 176 是被请求对象的主物主并且超高速缓存部件 176 在超高速缓存器中具有被请求对象（超高速缓存器阵命中）的情况。在这种情况下，第一超高速缓存器 172 可接受和客户 110 的连接。然后第一超高速缓存器 172 可确定第二超高速缓存器 176 并且建立 UDP 接口。第一超高速缓存器 172 和第二超高速缓存器 176 可通过 UDP 接口通信，以检索被请求对象。第二超高速缓存器 176 可从超高速缓存器检索被请求对象并把被请求对象发送到第一超高速缓存器 172。第一超高速缓存器 172 可经过网络 150 把被请求对象发送给客户 110。

图 5 表示第二超高速缓存器 176 不在超高速缓存器中具有被请求对象（超高速缓存器阵未击中）的情况。在这种情况下，第二超高速缓存器 176 可先从服务器 184 检索被请求对象，并接着把被请求对象发送到第一超高速缓存器 172。然后第一超高速缓存器 172 可经网络 150 把被请求对象发送给客户 110。

UDP 接口的性能可好于 UTTP 接口，因为 UDP 接口可避免带有伴生开销的 TCP 连接。

超高速缓存部件未命中 - 转交接口

图 6 和图 7 是框图，说明依据本发明的示范实施例在未命中超高速缓存部件情况下检索被请求对象的方法。客户 110 发出经网络检索某对象的请求。该被请求对象存储在服务器群集 180 中的一个服务器 182、184 上或者可由一个服务器 182、184 生成或取得。经网络 150 把客户 110 发出的请求传送到网络调度器 160。接着网络调度器 160 选择第一超高速缓存器，例如超高速缓存器组 170 的超高速缓存部件 172。假定网络调度器选择的第一超高速缓存器 172 不是被请求对象的主物主，出现未

命中超高速缓存部件。

图 6 表示第二超高速缓存，例如超高速缓存部件 176，是被请求对象的主物主并且超高速缓存部件 176 在超高速缓存器中具有被请求对象（超高速缓存器命中）的情况。在该情况下，第一超高速缓存 172 进行把该请求转交给第二超高速缓存 176（例如和 TCP 连接一起）。接着第二超高速缓存 176 可从超高速缓存器检索被请求对象并把被请求对象发送到客户 110。

图 7 表示第二超高速缓存器 176 不在超高速缓存器中具有被请求对象（超高速缓存器未命中）的情况。在该情况下，第一超高速缓存 172 进行把该请求转交给第二超高速缓存 176（例如和 TCP 连接一起）。第二超高速缓存器 176 可先从服务器 184 检索被请求对象，并且接着把被请求对象发送到客户 110。

超高速缓存器组 170 可备有几种支持转交接口的性能。首先，超高速缓存器组 170 的所有超高速缓存部件可用一个群集地址（例如一个 IP 地址）定址。这样，第一超高速缓存 172 和第二超高速缓存 174 都可接收和该群集地址对应的请求，并且对客户 110 应答。第二，例如可设置一种诸如 TCP 核心扩展的机制，以实现从第一超高速缓存 172（利用一个 TCP 栈）到第二超高速缓存 176（利用第二 TCP 栈）的转交。在转交期间，可利用第一超高速缓存 172 和第二超高速缓存 176 之间的 UDP 接口传送连接信息。利用 UDP 接口可帮助改进性能。第三，和 TCP 接口能力一起，可实施一种用于从第一超高速缓存 172 向第二超高速缓存 176 转交 HTTP 请求的附加机制。第四，可利用网络调度器 160 支持 TCP 转交。当从第一超高速缓存 172 向第二超高速缓存 176 转交 TCP/HTTP 连接时，可能需要确保该连接流（从客户到服务器）通过第二超高速缓存 176（连接的新物主）。第一超高速缓存 172 或第二超高速缓存 176 可指示网络调度器 160 更新其信息，从而把后继的分组发送到第二超高速缓存 176，而不是发送到第一超高速缓存 172。

转交接口下的开销可能高于 UDP 接口情况下的开销。但是，在 UDP 或 HTTP 接口情况下，随着对象大小的增加性能可能明显下降。和 UDP 或 HTTP 接口情况不同，转交接口的性能对对象尺寸的增加不太敏感。

转交接口性能对对象尺寸增加的敏感性可能类似于超高速缓存部件命中性能对对象尺寸增加的敏感性。从而，通过对相对大的对象采用转交接口而对相对小的对象采用 UDP、HTTP 或其它接口，可改进性能。

超高速缓存部件未命中 - 混合式

5 可根据被请求对象的大小选择所使用的接口类型（例如，UDP 或转交）。可确定尺寸阈值，从而通过对小于该尺寸阈的对象采用一种类型的接口（例如 UDP）而对大于该尺寸阈的对象采用另一种类型的接口（例如转交），可改进性能。例如通过对某给定计算机网络和系统测量吞吐量，确定尺寸阈值。

10 例如，可以对小于 2K 字节的被请求对象选择 UDP 接口，而对大于 2K 字节的被请求对象选择转交接口。例如参照图 1(a)，假定客户请求一个由服务器 186 存储或者可由服务器 186 生成或取得的对象。还假定超高速缓存部件 178 是被请求对象的主物主。可应用下述示范性规则：

15 1. 由网络调度器 160 选择的第一超高速缓存部件，比如说超高速缓存部件 172，可向第二超高速缓存 178 发送请求以及例如 TCP 连接信息。

2. 第二超高速缓存 178 可确定被请求对象的尺寸。被请求对象的尺寸可从超高速缓存器或者在对象不在超高速缓存器中的情况通过与服务器 186 接触得到确定。

3. 若被请求对象的尺寸小于尺寸阈值，第二超高速缓存可把被请求对象发送给第一超高速缓存 172。接着第一超高速缓存 172 可把被请求对象发送给客户 110。

25 4. 若被请求对象的尺寸大于尺寸阈值，可在第一超高速缓存 172 和第二超高速缓存 178 之间进行转交，例如 TCP/HTTP 转交。然后第二超高速缓存 178 可从超高速缓存器或从服务器 186 检索被请求对象并把被请求对象发送到客户 110。

基于内容的路由选择

30 在这种情况下，图 1(a) 的网络调度器 160 可充当图 8 的基于内容的路由器 165。尤其，某个网络调度器，例如在特定方式下运行的图

2-7 的网络调度器 160 可充当基于内容的路由器 165。例如利用散列函数可在超高速缓存器组 170 的超高速缓存节点 172、174、176、178 之间分配被请求对象的地址（例如 URL）。换言之，可利用散列函数形成超高速缓存节点和被请求节点之间的对应。被请求对象的地址可被散列。散列函数的输出可以是和一个或多个服务器相关的超高速缓存节点，而该一个或多个服务器可生成或取得被请求对象或者存储着被请求对象。换言之，散列函数的输出可以是作为被请求对象的主物主的超高速缓存部件。这并不必须意味着在超高速节点和服务器之间存在特殊关系。虽然可选择存在一个这样的关系。

当某客户，例如客户 110、120、130，请求检索某对象时，基于内容的路由器 165 可把该请求以及例如 TCP 连接信息转交给由其选择的某超高速缓存节点。由基于内容的路由器 165 选择的该超高速缓存是被请求对象的主物主。然后，该选定的超高速缓存节点可从超高速缓存器或从与它关联的服务器检索被请求对象，并且把被请求对象经网络 150 发送到请求客户。

由于以特定方式运行的网络调度器可充当基于内容的路由器，有可能使网络调度器利用基于内容的和不基于内容的路由选择的组合路由请求。当采用基于内容的路由选择时，运行成基于内容的路由器 165 的网络调度器 160 可研究由客户发送的请求，以确定哪个超高速缓存节点是被请求对象的主物主。由于基于内容的路由选择可避免不命中超高速缓存部件，基于内容的路由选择可减少超高速缓存器阵消耗的处理机周期。然后，基于内容的路由选择会增加充当基于内容的路由器 165 的网络调度器 160 消耗的处理机周期。以基于内容进行路由选择的网络调度器 160 由于建立和客户的连接，会消耗更多的处理机周期。

只要其不变成系统瓶颈，使用基于内容的路由器 165 是有好处的。例如，由 IBM 2216 实现的基于内容的路由器只要每秒请求数少于 1 万次就不会成为瓶颈。备选地，可采用带有集成超高速缓存的基于内容的路由器。在相对低和中等系统负载下，基于内容的路由器可利用集成的超高速缓存为请求服务。在相对高的系统负载下，这种路由器可恢复网络调度器功能。

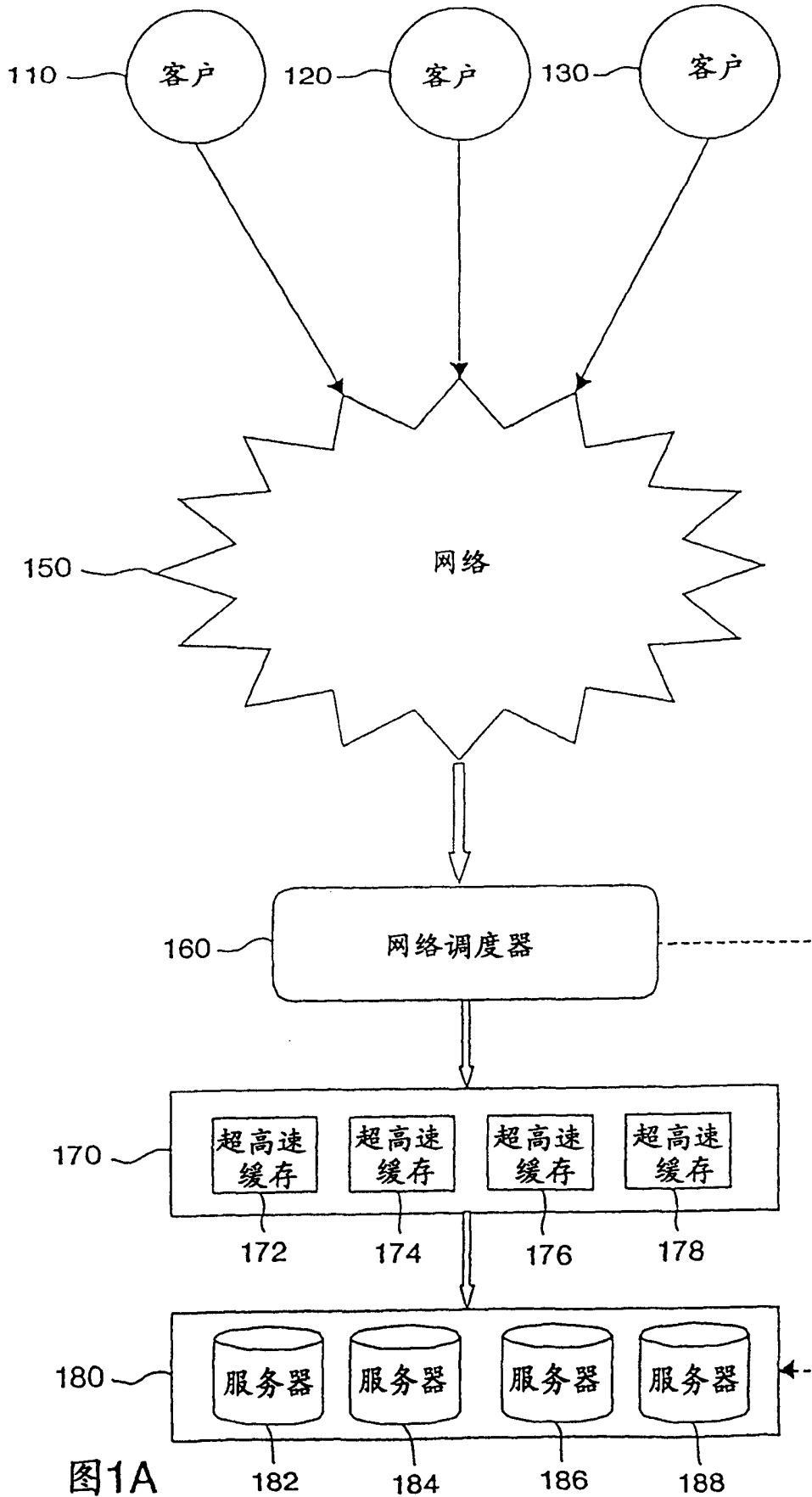
例如, IBM 2216 作为网络调度器的能力是每秒约可对 15,000 个请求进行路由选择。充当超高速缓存节点的 IBM 2216 的能力是在约为 100 % 超高速缓存速率下每秒可约为 5,000 个对象服务。充当基于内容的路由器并带有转交能力的 IBM 2216 每秒约能为 $5,000/0.5 = 10,000$ 个请求服务。

作为例如上述图 2-7 中示出的本发明的各示范实施例的一种变型, 可把网络调度器 160 路由到超高速缓存器组 170 的请求中一定百分比执行成为基于内容的路由选择方式。基于内容的路由选择会导致超高速缓存器组 170 上的较少开销, 但会导致网络调度器 160 上的较多开销。采用基于内容的路由选择的请求百分数可选择成均衡网络调度器和超高速缓存器阵资源的使用。另一种变型可以是当超高速缓存器阵变为瓶颈时增加采用基于内容的路由选择方式的进行路由的请求的百分比。然而, 网络调度器 160 变为瓶颈下, 可降低采用基于路由选择的进行路由的请求的百分比。

即, 可选择地, 可把网络调度器, 例如图 1(a) 的网络调度器的功能集成到超高速缓存节点中。

一般的业内人士可想到上述各示范实施例的各种修改。例如, 客户请求不必由第一节点引导到被请求对象所属的主物主上。若主物主过载, 请求可发送到第二节点。第二节点可向第一节点回送被请求对象, 或者, 第二节点利用转交直接向客户回送被请求对象。可选地, 可把请求转交给服务器节点。

尽管参照某些示范实施例对本发明进行了描述和说明, 但决非意味着本发明受限于所示的细节。相反, 在权利要求书的等价的范围内和在不背离本发明的精神下, 在细节上可作出各种细节。



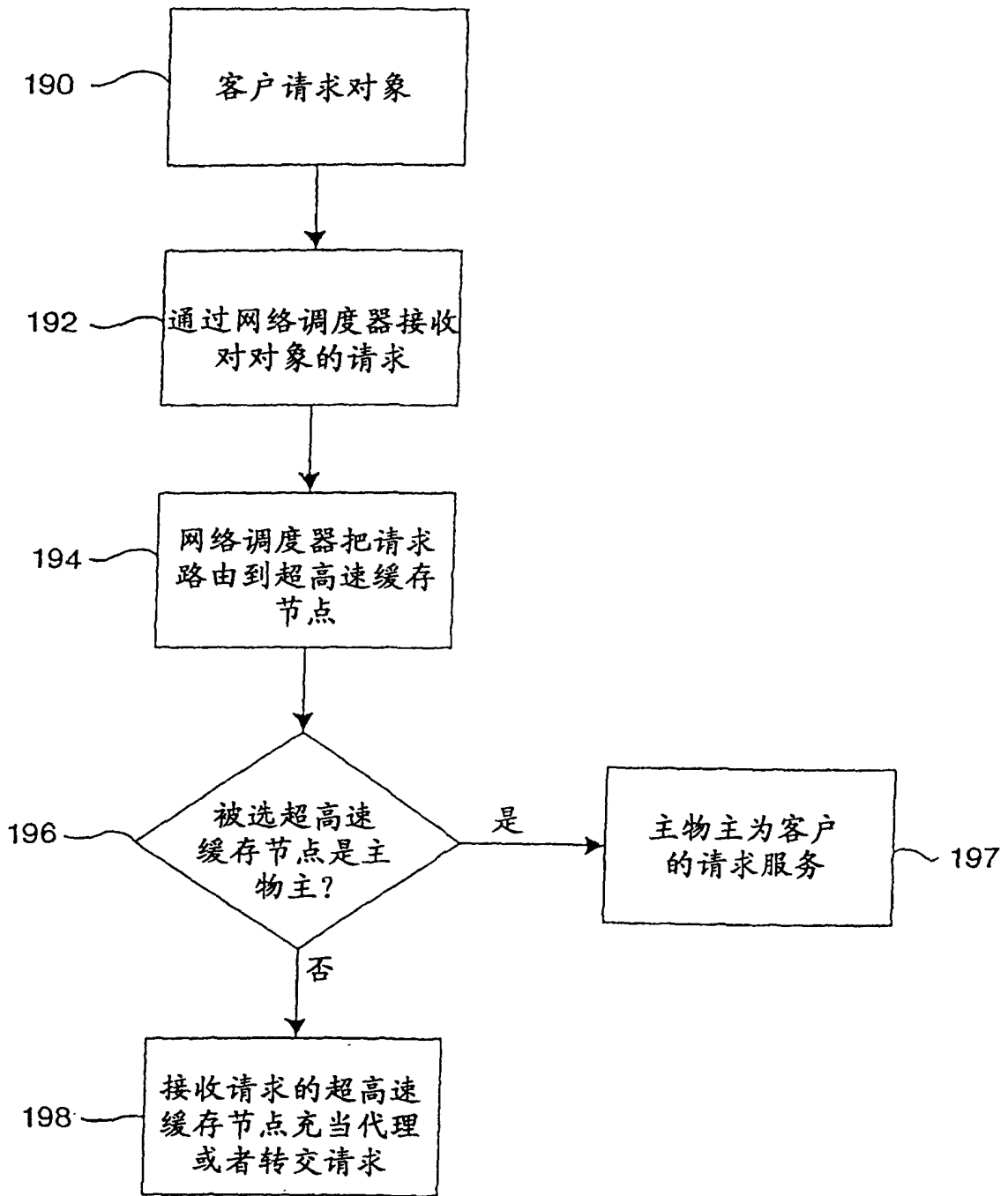
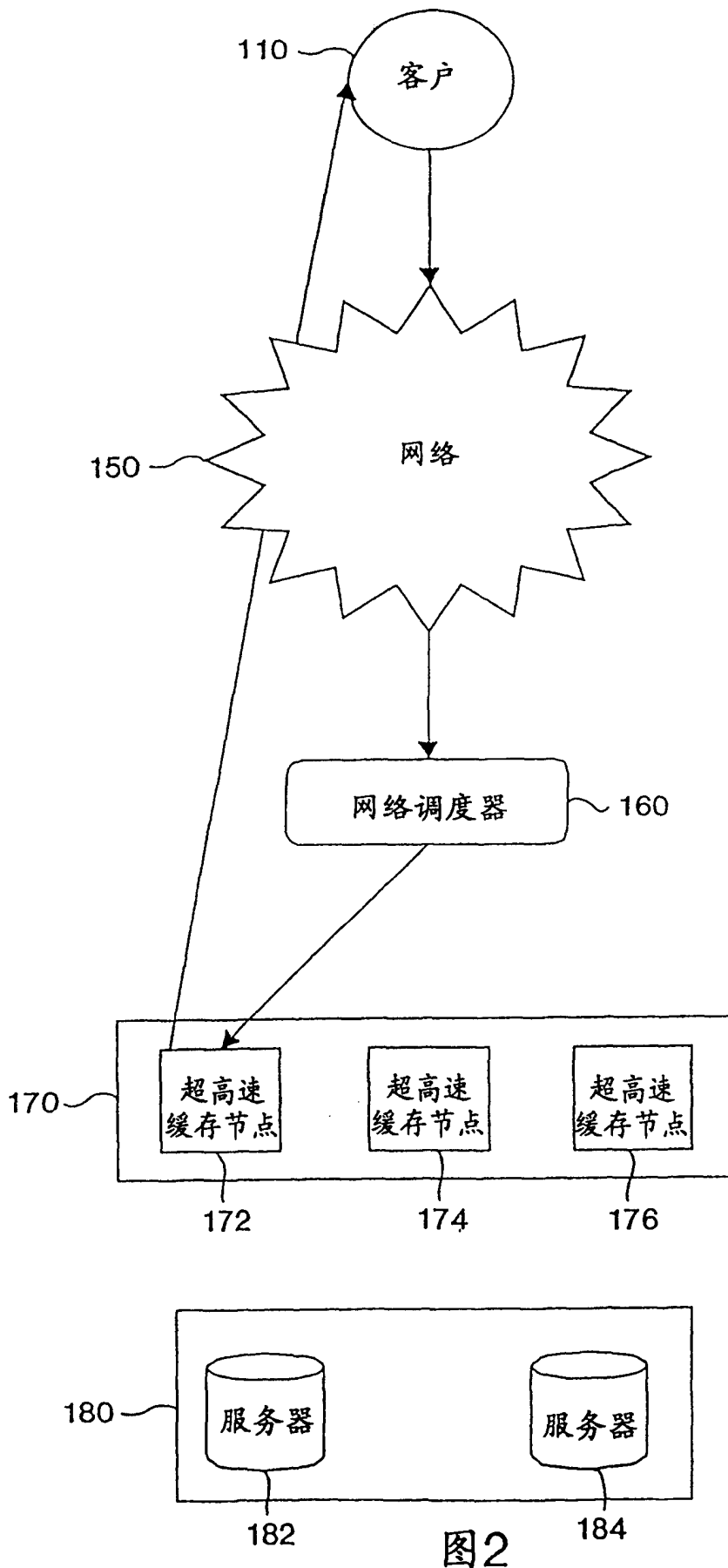


图1B



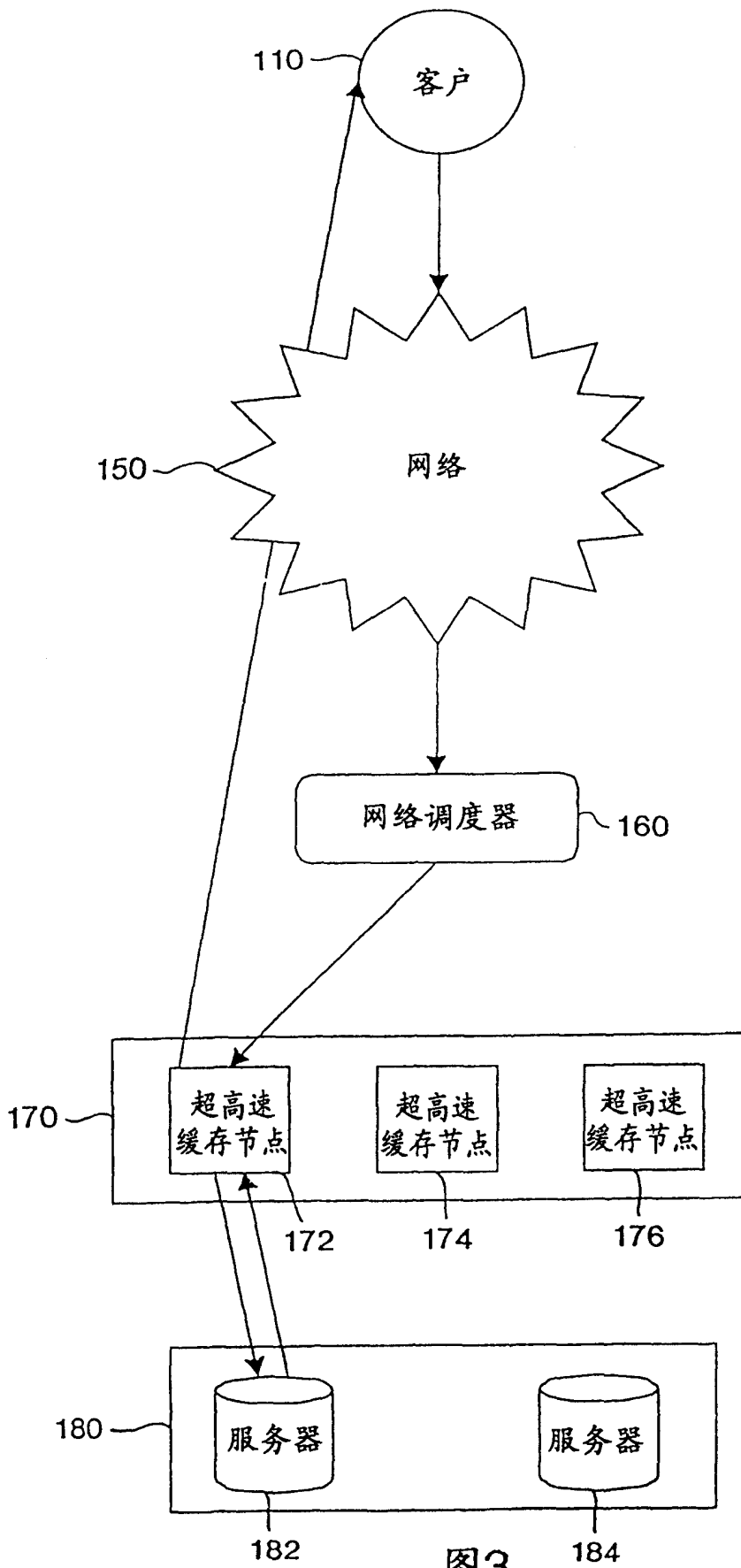


图3

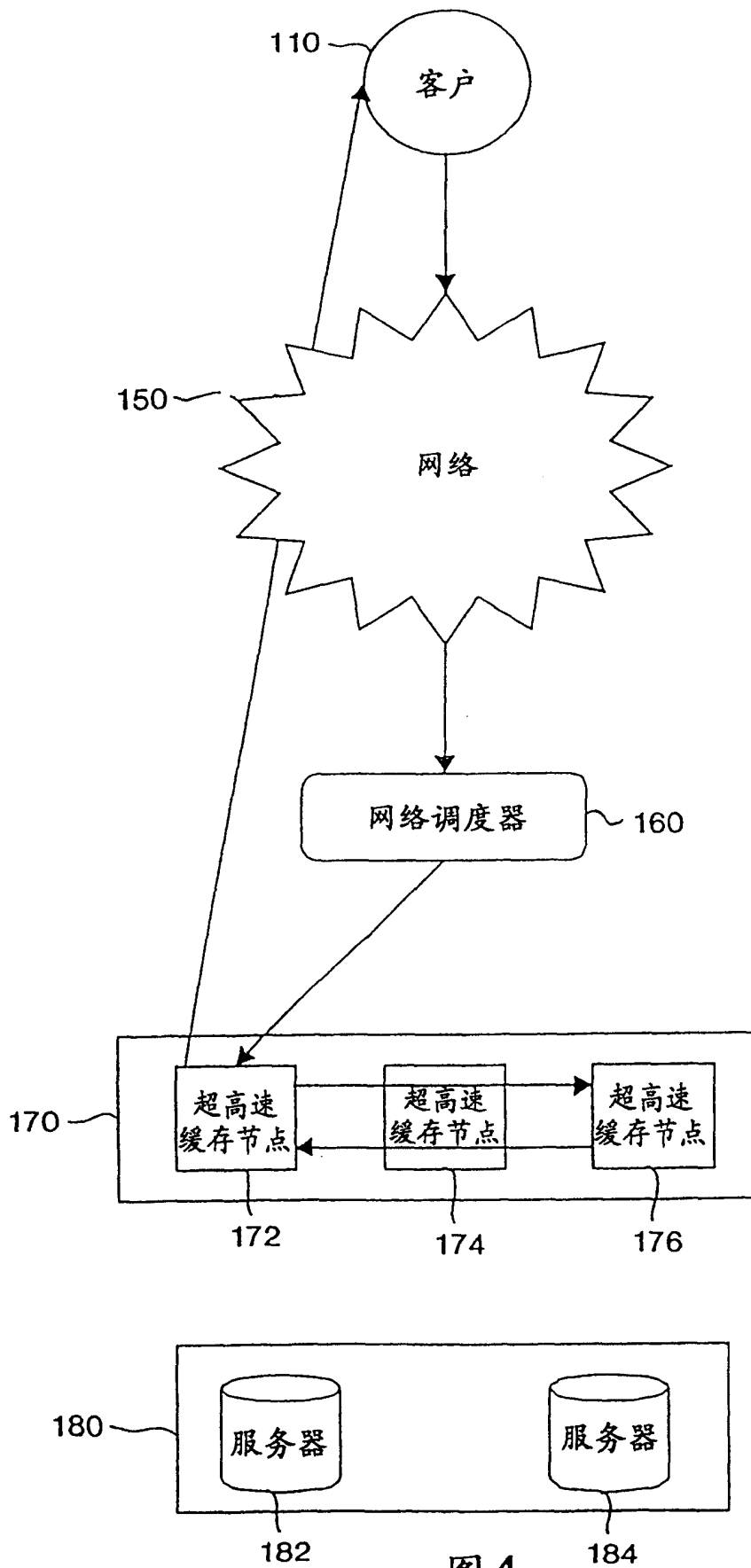


图4

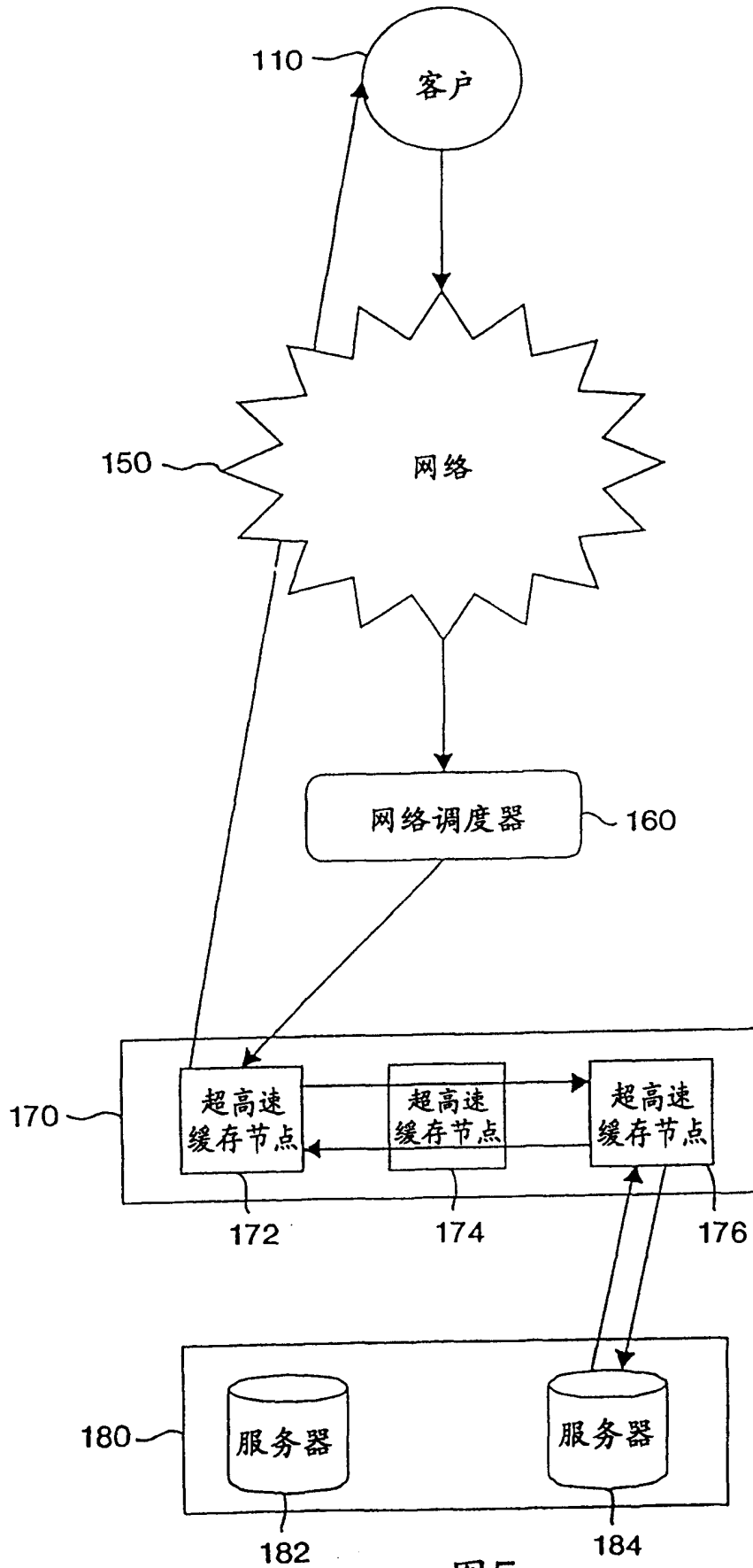


图5

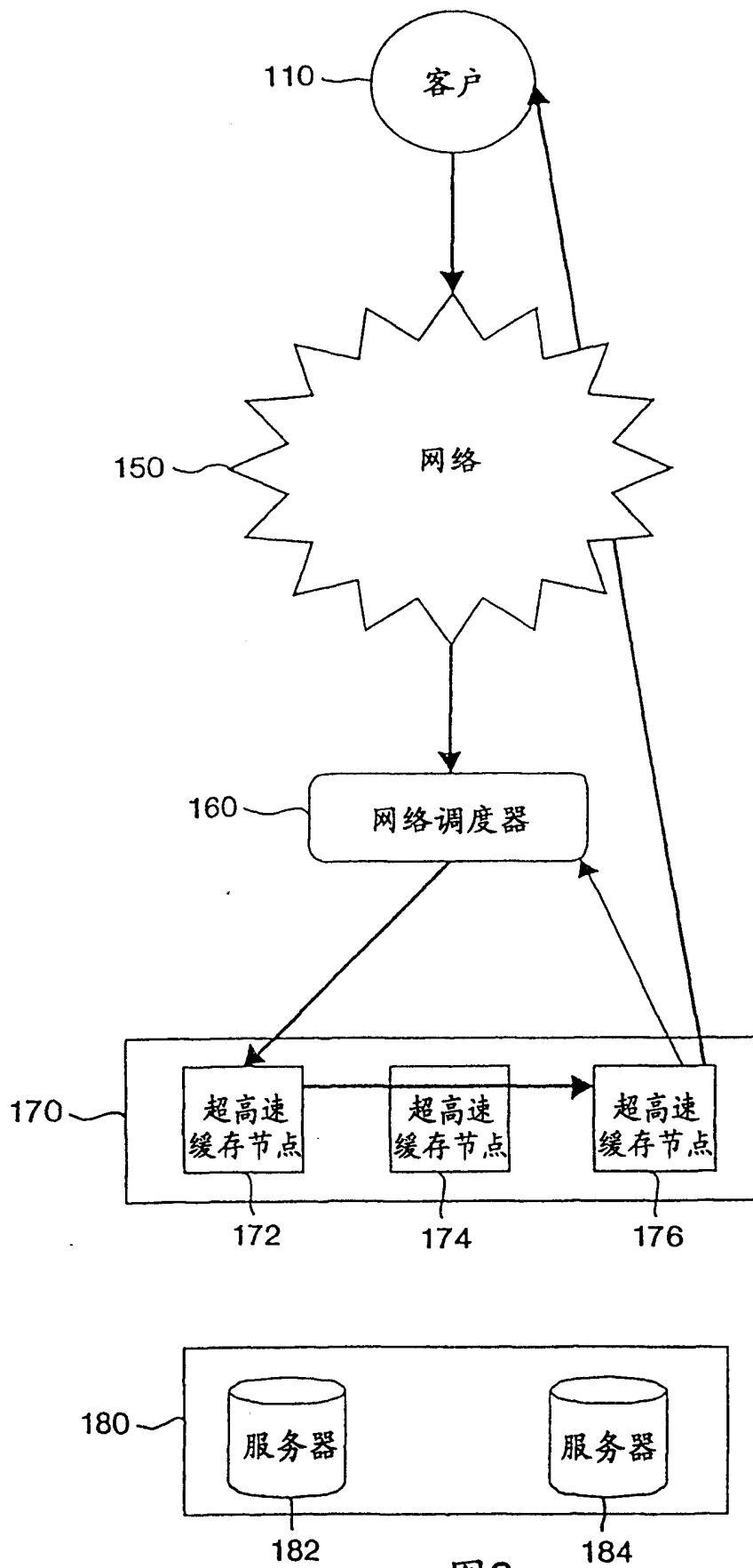


图6

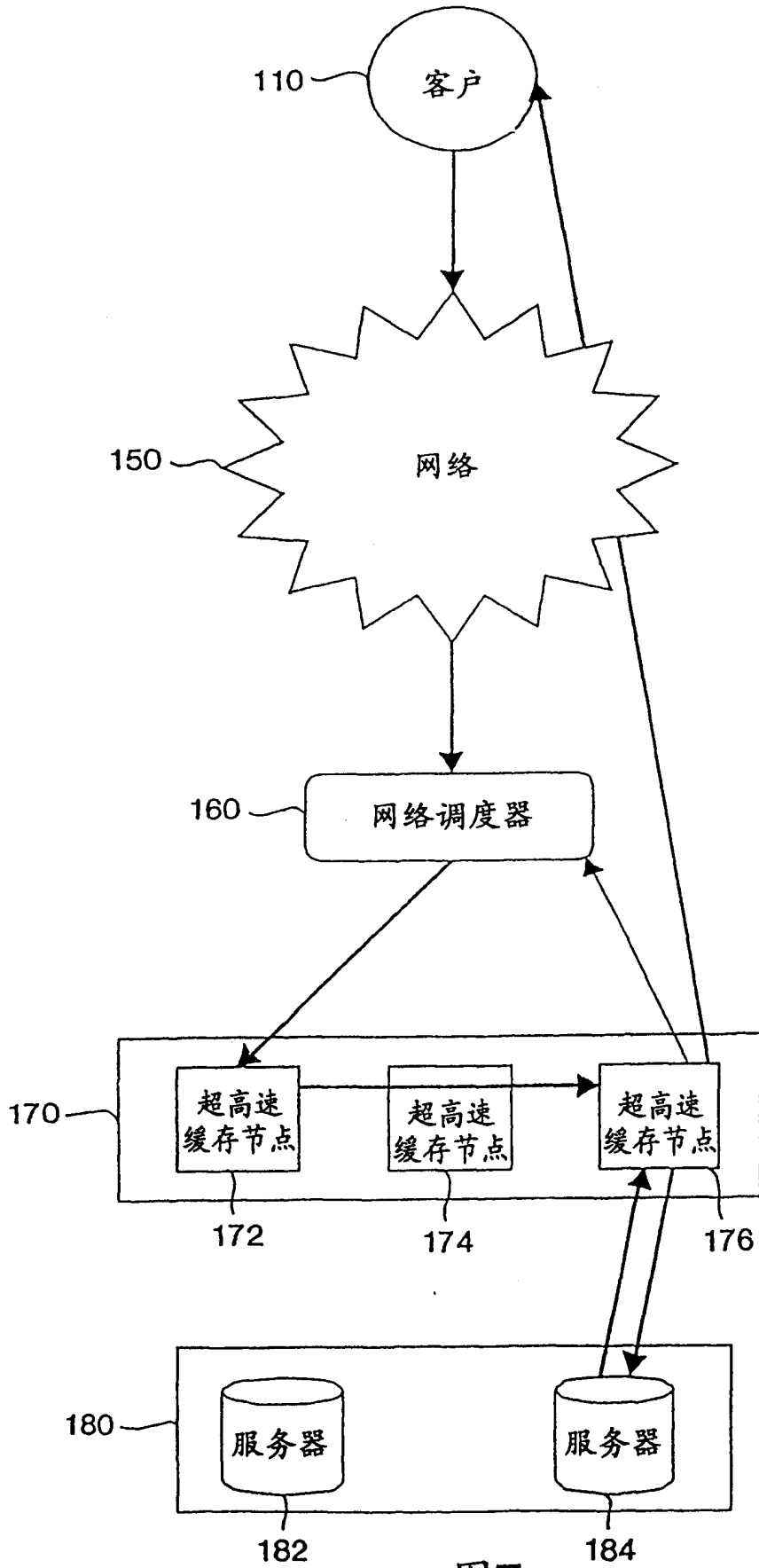


图7

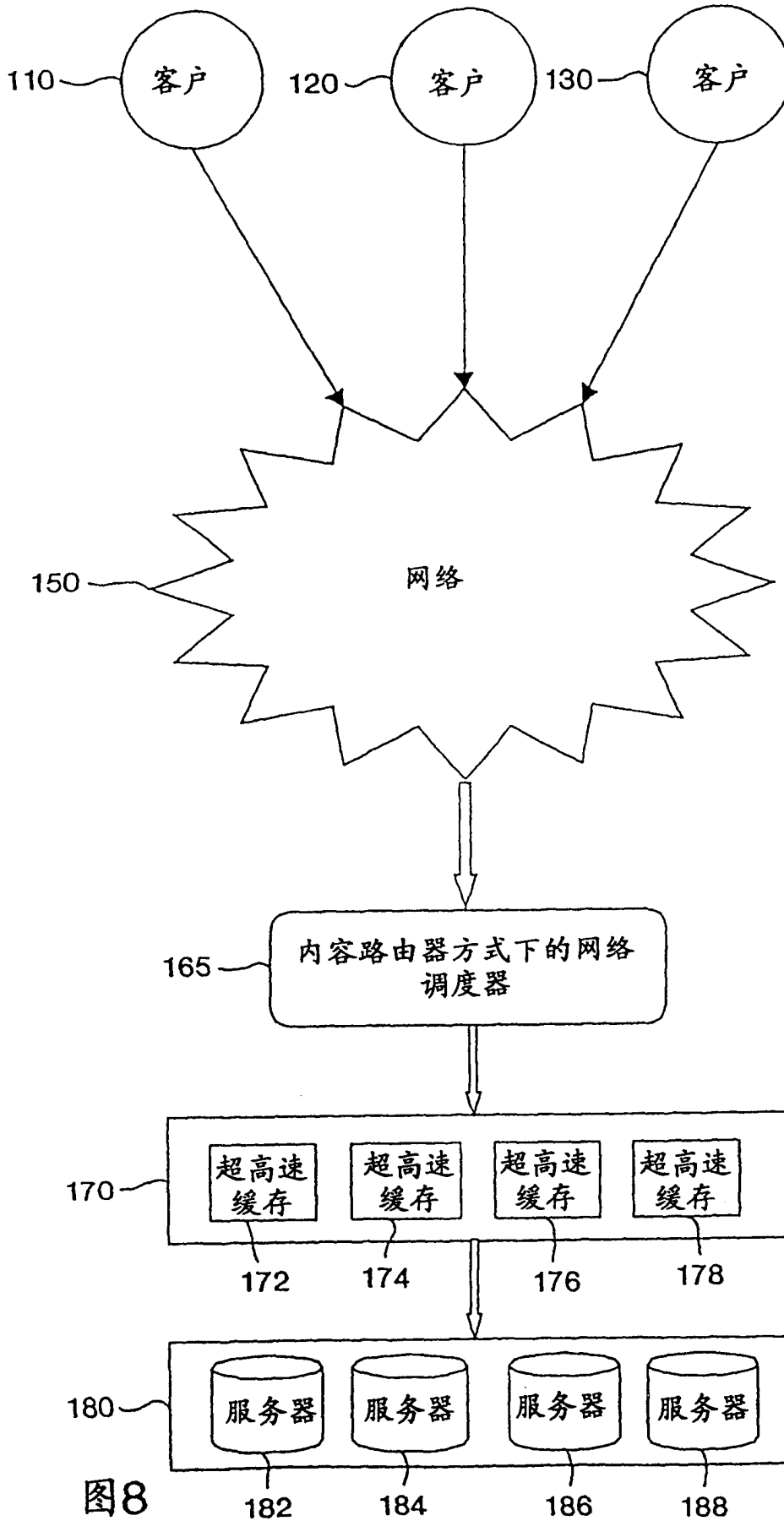


图8