(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103533018 A
(43)申请公布日 2014.01.22

(21)申请号 201310262856.1

(22)申请日 2013.04.27

(30)优先权数据
13/541,792 2012.07.05 US

(71)申请人 A10网络股份有限公司
地址 美国加利福尼亚州

(72)发明人 L·韩

(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 袁逸

(51)Int.Cl.
H04L 29/08(2006.01)

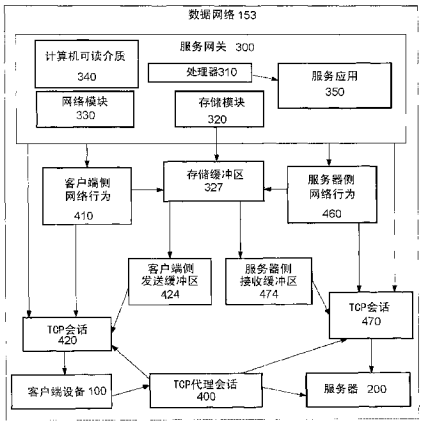
权利要求书6页 说明书12页 附图11页

(54)发明名称

基于动态网络条件为TCP代理会话分配缓冲区的方法

(57)摘要

由服务网关为客户端和服务器之间的TCP代理会话分配缓冲区，包括：监控TCP代理会话的服务器和客户端侧会话的动态网络行为；以及基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区的容量的加权平均来在所述存储缓冲区中分配服务器侧缓冲区的容量和客户端侧缓冲区的容量。在一个分配方法中，所述网关确定服务器或客户端侧缓冲区的可用容量是否足够用于存储数据分组。如果不足，则基于动态网络行为的测量和所述加权平均来增加所述服务器或客户端侧缓冲区的所分配的容量，并相应地调整所述服务器或客户端侧缓冲区的可用容量。



CN 103533018 A

1. 一种由包括处理器的服务网关为客户端和服务器之间的 TCP 代理会话分配缓冲区的方法，包括：

(a) 由所述处理器监控所述 TCP 代理会话的服务器侧会话的动态服务器侧网络行为和所述 TCP 代理会话的客户端侧会话的动态客户端侧网络行为；以及

(b) 由所述处理器基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区的容量的加权平均来在所述存储缓冲区中分配服务器侧缓冲区的容量和客户端侧缓冲区的容量。

2. 根据权利要求 1 所述的方法，其中所述分配 (b) 包括：

(b1) 基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量；以及

(b2) 基于所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量来分配所述服务器侧缓冲区的容量和所述客户端侧缓冲区的容量。

3. 根据权利要求 1 所述的方法，其中所述分配 (b) 包括：

(b1) 从所述服务器接收服务器侧会话数据，或者

从所述客户端接收客户端侧会话数据分组；

(b2) 确定所述服务器侧缓冲区的可用容量是否足够用于存储所述服务器侧会话数据分组，或者

确定所述客户端侧缓冲区的可用容量是否足够用于存储所述客户端侧会话数据分组；

(b3) 响应于确定所述服务器侧缓冲区的可用容量不足以用于存储所述服务器侧会话数据分组，增加所述服务器侧缓冲区的所分配容量，或者

响应于确定所述客户端侧缓冲区的可用容量不足以用于存储所述客户端侧会话数据分组，增加所述客户端侧缓冲区的所分配容量；以及

(b4) 根据所述服务器侧缓冲区的所分配容量的增加来调整所述服务器侧缓冲区的可用容量，或者

根据所述客户端侧缓冲区的所分配容量的增加来调整所述客户端侧缓冲区的可用容量。

4. 根据权利要求 3 所述的方法，其中所述增加 (b3) 包括：

(b3i) 响应于确定所述服务器侧缓冲区的可用容量不足以用于存储所述服务器侧会话数据分组，确定所述服务器侧缓冲区的所分配容量是否小于所述服务器侧缓冲区的最大容量，或者

响应于确定所述客户端侧缓冲区的可用容量不足以用于存储所述客户端侧会话数据分组，确定所述客户端侧缓冲区的所分配容量是否小于所述客户端侧缓冲区的最大容量；以及

(b3ii) 响应于确定所述服务器侧缓冲区的所分配容量小于所述服务器侧缓冲区的最大容量，确定在所述服务器侧缓冲区的最大容量以内是否存在足够的存储空间以用于存储所述服务器侧会话数据分组，或者

响应于确定所述客户端侧缓冲区的所分配容量小于所述客户端侧缓冲区的最大容量，

确定所述客户端侧缓冲区的最大容量以内是否存在足够的存储空间以用于存储所述客户端侧会话数据分组；

(b3iii) 响应于确定在所述服务器侧缓冲区的最大容量以内存在足够的存储空间以用于存储所述服务器侧会话数据分组，将所述存储空间分配给所述服务器侧缓冲区并根据所分配的存储空间来调整所述服务器侧缓冲区的可用容量，或者

响应于确定在所述客户端侧缓冲区的最大容量以内存在足够的存储空间以用于存储所述客户端侧会话数据分组，将所述存储空间分配给所述客户端侧缓冲区并根据所分配的存储空间来调整所述客户端侧缓冲区的可用容量。

5. 根据权利要求 2 所述的方法，其中所述监控 (a) 和所述计算 (b1) 包括：

(a1) 测量服务器侧数据分组往返时间 (RTT) 和客户端侧数据分组 RTT；以及

(b1i) 使用所述服务器侧数据分组 RTT、客户端侧 RTT、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

6. 根据权利要求 5 所述的方法，其中，在计算所述服务器侧缓冲区的最大容量时，使对所述存储缓冲区的容量的加权平均偏向于所述服务器，或者

其中，在计算所述客户端侧缓冲区的最大容量时，使对所述存储缓冲区的容量的加权平均偏向于所述客户端。

7. 根据权利要求 5 所述的方法，其中所述监控 (a) 进一步包括：

(a3) 获取所述服务器侧数据分组 RTT 和所述客户端侧数据分组 RTT 的新的测量；以及

(a4) 使用新的服务器侧数据分组 RTT、新的客户端侧数据分组 RTT、以及对所述存储缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

8. 根据权利要求 5 所述的方法，其中所述测量 (a1) 和所述计算 (a2) 包括：

(a1i) 测量所述服务器侧数据分组 RTT、所述客户端侧数据分组 RTT、服务器侧传输速率、以及客户端侧传输速率；以及

(a2i) 使用所述服务器侧数据分组 RTT、所述客户端侧 RTT、所述服务器侧传输速率、所述客户端侧传输速率、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

9. 根据权利要求 8 所述的方法，其中所述监控 (a) 进一步包括：

(a3) 获取所述服务器侧传输速率和所述客户端侧传输速率的新的测量；以及

(a4) 使用所述服务器侧数据分组 RTT、所述客户端数据分组 RTTT、新的服务器侧传输速率、新的客户端侧传输速率、以及对所述存储缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

10. 一种用于为客户端和服务器之间的 TCP 代理会话分配缓冲区的计算机程序产品，所述计算机程序产品包括：

具有嵌入在其中的计算机可读程序代码的非瞬时性的计算机可读存储介质，所述计算机可读程序代码配置用于：

(a) 监控所述 TCP 代理会话的服务器侧会话的动态服务器侧网络行为和所述 TCP 代理会话的客户端侧会话的动态客户端侧网络行为；以及

(b) 基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区

的容量的加权平均来在所述存储缓冲区中分配服务器侧缓冲区的容量和客户端侧缓冲区的容量。

11. 根据权利要求 10 所述的计算机程序产品，其中所述配置用于分配 (b) 的计算机可读程序代码进一步配置用于：

(b1) 基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量；以及

(b2) 基于所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量来分配所述服务器侧缓冲区的容量和所述客户端侧缓冲区的容量。

12. 根据权利要求 10 所述的计算机程序产品，其中所述配置用于分配 (b) 的计算机可读程序代码进一步配置用于：

(b1) 从所述服务器接收服务器侧会话数据，或者

从所述客户端接收客户端侧会话数据分组；

(b2) 确定所述服务器侧缓冲区的可用容量是否足够用于存储所述服务器侧会话数据分组，或者

确定所述客户端侧缓冲区的可用容量是否足够用于存储所述客户端侧会话数据分组；

(b3) 响应于确定所述服务器侧缓冲区的可用容量不足以用于存储所述服务器侧会话数据分组，增加所述服务器侧缓冲区的所分配容量，或者

响应于确定所述客户端侧缓冲区的可用容量不足以用于存储所述客户端侧会话数据分组，增加所述客户端侧缓冲区的所分配容量；以及

(b4) 根据所述服务器侧缓冲区的所分配容量的增加来调整所述服务器侧缓冲区的可用容量，或者

根据所述客户端侧缓冲区的所分配容量的增加来调整所述客户端侧缓冲区的可用容量。

13. 根据权利要求 12 所述的计算机程序产品，其中所述配置用于增加 (b3) 的计算机可读程序代码进一步配置用于：

(b3i) 响应于确定所述服务器侧缓冲区的可用容量不足以用于存储所述服务器侧会话数据分组，确定所述服务器侧缓冲区的所分配容量是否小于所述服务器侧缓冲区的最大容量，或者

响应于确定所述客户端侧缓冲区的可用容量不足以用于存储所述客户端侧会话数据分组，确定所述客户端侧缓冲区的所分配容量是否小于所述客户端侧缓冲区的最大容量；

(b3ii) 响应于确定所述服务器侧缓冲区的所分配容量小于所述服务器侧缓冲区的最大容量，确定在所述服务器侧缓冲区的最大容量以内是否存在足够的存储空间以用于存储所述服务器侧会话数据分组，或者

响应于确定所述客户端侧缓冲区的所分配容量小于所述客户端侧缓冲区的最大容量，确定所述客户端侧缓冲区的最大容量以内是否存在足够的存储空间以用于存储所述客户端侧会话数据分组；以及

(b3iii) 响应于确定在所述服务器侧缓冲区的最大容量以内存在足够的存储空间以用

于存储所述服务器侧会话数据分组，将所述存储空间分配给所述服务器侧缓冲区并根据所分配的存储空间来调整所述服务器侧缓冲区的可用容量，或者

响应于确定在所述客户端侧缓冲区的最大容量以内存在足够的存储空间以用于存储所述客户端侧会话数据分组，将所述存储空间分配给所述客户端侧缓冲区并根据所分配的存储空间来调整所述客户端侧缓冲区的可用容量。

14. 根据权利要求 11 所述的计算机程序产品，其中所述配置用于监控 (a) 和用于计算 (b1) 的计算机可读程序代码进一步配置用于：

(a1) 测量服务器侧数据分组往返时间 (RTT) 和客户端侧数据分组 RTT；以及

(b1i) 使用所述服务器侧数据分组 RTT、客户端侧 RTT、和对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

15. 根据权利要求 14 所述的计算机程序产品，其中，在计算所述服务器侧缓冲区的最大容量时，使对所述存储缓冲区的容量的加权平均偏向于所述服务器，或者

其中，在计算所述客户端侧缓冲区的最大容量时，使对所述存储缓冲区的容量的加权平均偏向于所述客户端。

16. 根据权利要求 14 所述的计算机程序产品，其中，所述配置用于监控 (a) 的计算机可读程序代码进一步配置用于：

(a3) 获取所述服务器侧数据分组 RTT 和所述客户端侧数据分组 RTT 的新的测量；以及

(a4) 使用新的服务器侧数据分组 RTT、新的客户端侧数据分组 RTT、以及对所述存储缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

17. 根据权利要求 14 所述的计算机程序产品，其中，所述配置用于测量 (a1) 和用于计算 (a2) 的计算机可读程序代码进一步配置用于：

(a1i) 测量所述服务器侧数据分组 RTT、所述客户端侧数据分组 RTT、服务器侧传输速率、以及客户端侧传输速率；以及

(a2i) 使用所述服务器侧数据分组 RTT、所述客户端侧 RTT、所述服务器侧传输速率、所述客户端侧传输速率、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

18. 根据权利要求 17 所述的计算机程序产品，其中，所述配置用于监控 (a) 的计算机可读程序代码进一步配置用于：

(a3) 获取所述服务器侧传输速率和所述客户端侧传输速率的新的测量；以及

(a4) 使用所述服务器侧数据分组 RTT、所述客户端数据分组 RTT、所述新的服务器侧传输速率、所述新的客户端侧传输速率、以及对所述存储缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

19. 一种用于为客户端和服务器之间的 TCP 代理会话分配缓冲区的系统，包括：

处理器；以及

具有嵌入在其中的计算机可读程序代码的非瞬时性的计算机可读存储介质，所述计算机可读程序代码配置用于：

(a) 监控所述 TCP 代理会话的服务器侧会话的动态服务器侧网络行为和所述 TCP 代理会话的客户端侧会话的动态客户端侧网络行为；以及

(b) 基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区的容量的加权平均来在所述存储缓冲区中分配服务器侧缓冲区的容量和客户端侧缓冲区的容量。

20. 根据权利要求 19 所述的系统，其中，所述配置用于分配 (b) 的计算机可读程序代码进一步配置用于：

(b1) 基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量；以及

(b2) 基于所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量来分配所述服务器侧缓冲区的容量和所述客户端侧缓冲区的容量。

21. 根据权利要求 19 所述的系统，其中，所述配置用于分配 (b) 的计算机可读程序代码进一步配置用于：

(b1) 从所述服务器接收服务器侧会话数据，或者

从所述客户端接收客户端侧会话数据分组；

(b2) 确定所述服务器侧缓冲区的可用容量是否足够用于存储所述服务器侧会话数据分组，或者

确定所述客户端侧缓冲区的可用容量是否足够用于存储所述客户端侧会话数据分组；

(b3) 响应于确定所述服务器侧缓冲区的可用容量不足以用于存储所述服务器侧会话数据分组，增加所述服务器侧缓冲区的所分配容量，或者

响应于确定所述客户端侧缓冲区的可用容量不足以用于存储所述客户端侧会话数据分组，增加所述客户端侧缓冲区的所分配容量；以及

(b4) 根据所述服务器侧缓冲区的所分配容量的增加来调整所述服务器侧缓冲区的可用容量，或者

根据所述客户端侧缓冲区的所分配容量的增加来调整所述客户端侧缓冲区的可用容量。

22. 根据权利要求 21 所述的系统，其中所述配置用于增加 (b3) 的计算机可读程序代码进一步配置用于：

(b3i) 响应于确定所述服务器侧缓冲区的可用容量不足以用于存储所述服务器侧会话数据分组，确定所述服务器侧缓冲区的所分配容量是否小于所述服务器侧缓冲区的最大容量，或者

响应于确定所述客户端侧缓冲区的可用容量不足以用于存储所述客户端侧会话数据分组，确定所述客户端侧缓冲区的所分配容量是否小于所述客户端侧缓冲区的最大容量；

(b3ii) 响应于确定所述服务器侧缓冲区的所分配容量小于所述服务器侧缓冲区的最大容量，确定在所述服务器侧缓冲区的最大容量以内是否存在足够的存储空间用于存储所述服务器侧会话数据分组，或者

响应于确定所述客户端侧缓冲区的所分配容量小于所述客户端侧缓冲区的最大容量，确定在所述客户端侧缓冲区的最大容量以内是否存在足够的存储空间用于存储所述客户端侧会话数据分组；以及

(b3iii) 响应于确定在所述服务器侧缓冲区的最大容量以内存在足够的存储空间用于存储所述服务器侧会话数据分组，将所述存储空间分配给所述服务器侧缓冲区并根据所分配的存储空间来调整所述服务器侧缓冲区的可用容量，或者

响应于确定在所述客户端侧缓冲区的最大容量以内存在足够的存储空间用于存储所述客户端侧会话数据分组，将所述存储空间分配给所述客户端侧缓冲区并根据所分配的存储空间来调整所述客户端侧缓冲区的可用容量。

23. 根据权利要求 20 所述的系统，其中所述配置用于监控 (a) 和用于计算 (b1) 的计算机可读程序代码进一步配置用于：

(a1) 测量服务器侧数据分组往返时间 (RTT) 和客户端侧数据分组 RTT；以及

(b1i) 使用所述服务器侧数据分组 RTT、客户端侧 RTT、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

24. 根据权利要求 23 所述的系统，其中，在计算所述服务器侧缓冲区的最大容量时，使对所述存储缓冲区的容量的加权平均偏向于所述服务器，或者

其中，在计算所述客户端侧缓冲区的最大容量时，使对所述存储缓冲区的容量的加权平均偏向于所述客户端。

25. 根据权利要求 23 所述的系统，其中，所述配置用于监控 (a) 的计算机可读程序代码进一步配置用于：

(a3) 获取所述服务器侧数据分组 RTT 和所述客户端侧数据分组 RTT 的新的测量；以及

(a4) 使用所述新的服务器侧数据分组 RTT、所述新的客户端侧数据分组 RTT、以及对所述存储缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

26. 根据权利要求 23 所述的系统，其中，所述配置用于测量 (a1) 和计算 (a2) 的计算机可读程序代码进一步配置用于：

(a1i) 测量所述服务器侧数据分组 RTT、所述客户端侧数据分组 RTT、服务器侧传输速率、以及客户端侧传输速率；以及

(a2i) 使用所述服务器侧数据分组 RTT、所述客户端侧 RTT、所述服务器侧传输速率、所述客户端侧传输速率、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

27. 根据权利要求 26 所述的系统，其中所述配置用于监控 (a) 的计算机可读程序代码进一步配置用于：

(a3) 获取所述服务器侧传输速率和所述客户端侧传输速率的新的测量；以及

(a4) 使用所述服务器侧数据分组 RTT、所述客户端侧数据分组 RTTT、所述新的服务器侧传输速率、所述新的客户端侧传输速率、以及对所述存储缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

基于动态网络条件为 TCP 代理会话分配缓冲区的方法

技术领域

[0001] 本发明一般涉及数据通信，并且更具体地，涉及服务网关。

背景技术

[0002] 一段时间以来，诸如是防火墙和服务器负载均衡器的众多服务网关提供了传输控制协议 (TCP) 代理功能性。TCP 代理的典型使用包括网络分析、安全、和由于不对称客户端和服务器条件而引起的话务适配。TCP 代理服务器典型地分配固定量的存储缓冲区来处理在客户端设备和服务器之间的 TCP 代理会话的数据分组缓冲。存储缓冲区用于处理客户端侧会话和服务器侧会话的数据分组缓冲区。在客户端侧会话发送和接收缓冲区、以及服务器侧会话发送和接收缓冲区之间的存储空间分配并不经常考虑性能。通常的分配方法包括在会话发送和接收缓冲区之间等分存储缓冲区，或者在发送和接收缓冲区之间的固定的加权平均，诸如在服务器侧接收缓冲区和客户端侧发送缓冲区之间 2 比 1 的比例。

[0003] 虽然 TCP 代理在一段时间以来是已知的，但是其还没有以任何显著的方式使用。在缓冲区分配方法和 TCP 代理会话的性能之间的相关性没有得到很好的理解。现今，广泛使用诸如是 HTTP 代理、内容过滤应用、高级服务器负载均衡方法的众多服务网关应用。这些应用的性能部分地取决于 TCP 代理会话缓冲区分配方法的性能。

[0004] 由于客户端设备和服务器不同的和动态的网络特性，缓冲区空间的预定分配并不能在不同的网络条件上均性能良好。因此，需要基于一个或多个动态网络条件来为 TCP 代理会话分配缓冲区空间。

发明内容

[0005] 根据本发明的一个实施例，一种由包括处理器的服务网关为客户端和服务器之间的 TCP 代理会话分配缓冲区的方法，包括 :(a) 由处理器监控 TCP 代理会话的服务器侧会话的动态服务器侧网络行为和该 TCP 代理会话的客户端侧会话的动态客户端侧网络行为 ;以及 (b) 由处理器基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区的容量的加权平均来在所述存储缓冲区中分配服务器侧缓冲区的容量和客户端侧缓冲区的容量。

[0006] 在本发明的一个方面中，所述分配 (b) 包括 :(b1) 基于所述动态服务器侧网络行为、所述动态客户端侧网络行为、以及对存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量 ;以及 (b2) 基于所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量来分配所述服务器侧缓冲区的容量和所述客户端侧缓冲区的容量。

[0007] 在本发明的一个方面中，所述分配 (b) 包括 :(b1) 从所述服务器接收服务器侧会话数据，或者从所述客户端接收客户端侧会话数据分组 ;(b2) 确定所述服务器侧缓冲区的可用容量是否足够用于存储所述服务器侧会话数据分组，或者确定所述客户端侧缓冲区的可用容量是否足够用于存储所述客户端侧会话数据分组 ;(b3) 响应于确定所述服务器侧

缓冲区的可用容量不足够用于存储所述服务器侧会话数据分组，增加所述服务器侧缓冲区的所分配的容量，或者响应于确定所述客户端侧缓冲区的可用容量不足够用于存储所述客户端侧会话数据分组，增加所述客户端侧缓冲区的所分配的容量；以及 (b4) 根据所述服务器侧缓冲区的所分配的容量的增加来调整所述服务器侧缓冲区的可用容量，或者根据所述客户端侧缓冲区的所分配的容量的增加来调整所述客户端侧缓冲区的可用容量。

[0008] 在本发明的一个方面中，所述增加 (b3) 包括：(b3i) 响应于确定所述服务器侧缓冲区的可用容量不足够用于存储所述服务器侧会话数据分组，确定所述服务器侧缓冲区的所分配的容量是否小于所述服务器侧缓冲区的最大容量，或者响应于确定所述客户端侧缓冲区的可用容量不足够用于存储所述客户端侧会话数据分组，确定所述客户端侧缓冲区的所分配的容量是否小于所述客户端侧缓冲区的最大容量；以及 (b3ii) 响应于确定所述服务器侧缓冲区的所分配的容量小于所述服务器侧缓冲区的最大容量，确定在所述服务器侧缓冲区的最大容量以内是否存在足够的存储空间来存储所述服务器侧会话数据分组，或者响应于确定所述客户端侧缓冲区的所分配的容量小于所述客户端侧缓冲区的最大容量，确定所述客户端侧缓冲区的最大容量以内是否存在足够的存储空间来存储所述客户端侧会话数据分组；(b3iii) 响应于确定在所述服务器侧缓冲区的最大容量以内存在足够的存储空间来存储所述服务器侧会话数据分组，将所述存储空间分配给所述服务器侧缓冲区并根据所分配的存储空间来调整所述服务器侧缓冲区的可用容量，或者响应于确定在所述客户端侧缓冲区的最大容量以内存在足够的存储空间用于存储所述客户端侧会话数据分组，将所述存储空间分配给所述客户端侧缓冲区并根据所分配的存储空间来调整所述客户端侧缓冲区的可用容量。

[0009] 在本发明的一个方面中，所述监控 (a) 和所述计算 (b1) 包括：(a1) 测量服务器侧数据分组往返时间 (RTT) 和客户端侧数据分组 RTT；以及 (b1i) 使用所述服务器侧数据分组 RTT、客户端侧 RTT、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

[0010] 在本发明的一个方面中，在计算所述服务器侧缓冲区的最大容量时，使所述存储缓冲区的容量的加权平均偏向于所述服务器，或者其中，在计算所述客户端侧缓冲区的最大容量时，使所述存储缓冲区的容量的加权平均偏向于所述客户端。

[0011] 在本发明的一个方面中，所述监控 (a) 进一步包括：(a3) 获取所述服务器侧数据分组 RTT 和所述客户端侧数据分组 RTT 的新的测量；以及 (a4) 使用新的服务器侧数据分组 RTT、新的客户端侧数据分组 RTT、以及对所述存储缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

[0012] 在本发明的一个方面中，所述测量 (a1) 和所述计算 (a2) 包括：(a1i) 测量所述服务器侧数据分组 RTT、所述客户端侧数据分组 RTT、服务器侧传输速率、以及客户端侧传输速率；以及 (a2i) 使用所述服务器侧数据分组 RTT、所述客户端侧 RTT、所述服务器侧传输速率、所述客户端侧传输速率、以及对所述存储缓冲区的容量的加权平均来计算所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

[0013] 在本发明的一个方面中，所述监控 (a) 进一步包括：(a3) 获取所述服务器侧传输速率和所述客户端侧传输速率的新的测量；以及 (a4) 使用所述服务器侧数据分组 RTT、所述客户端数据分组 RTT、新的服务器侧传输速率、新的客户端侧传输速率、以及对所述存储

缓冲区的容量的加权平均来修改所述服务器侧缓冲区的最大容量和所述客户端侧缓冲区的最大容量。

[0014] 在此还描述和要求保护了对应于以上概述的方法的系统和计算机程序产品。

附图说明

[0015] 图 1 说明了根据本发明的用于提供在客户端设备和服务器之间的 TCP 代理会话的服务网关的实施例。

[0016] 图 2a 说明了根据本发明的用于从服务器接收数据分组的服务网关的实施例。

[0017] 图 2b 说明了根据本发明的用于发送数据分组到客户端设备的服务网关的实施例。

[0018] 图 3 说明了根据本发明的用于监控服务器侧网络行为和客户端侧网络行为的服务网关的实施例。

[0019] 图 4 说明了根据本发明的用于服务器侧接收缓冲区和客户端侧发送缓冲区的存储空间分配过程的实施例。

[0020] 图 5a 说明了根据本发明的用于将数据分组存储到服务器侧接收缓冲区中的服务网关的实施例。

[0021] 图 5b 说明了根据本发明的用于将数据分组存储到客户端侧发送缓冲区中的服务网关的实施例。

[0022] 图 6 说明了根据本发明的用于从客户端设备接收数据分组的服务网关的实施例。

[0023] 图 7 说明了根据本发明的用于客户端侧接收缓冲区和服务器侧发送缓冲区的存储空间分配过程的实施例。

[0024] 图 8a 说明了根据本发明的用于将数据分组存储到客户端侧接收缓冲区中的服务网关的实施例。

[0025] 图 8b 说明了根据本发明的用于将数据分组存储到服务器侧发送缓冲区中的服务网关的实施例。

具体实施方式

[0026] 本发明能够采用完全硬件实施例、完全软件实施例或包含硬件和软件元素两者的实施例的形式。在优选实施例中，本发明以软件来实现，其包括但不限于固件、常驻软件、微码等等。

[0027] 进一步，本发明能够采用从计算机可使用或计算机可读介质可访问的计算机程序产品的形式，其提供了由计算机或任何指令执行系统使用或者与计算机或任何指令执行系统一起使用的程序代码。为了本说明书的目的，计算机可使用或计算机可读介质能够是能够包含、存储、通信、传播、或传输该程序以由指令执行系统、装置或设备使用或与指令执行系统、装置或设备一起使用的任何装置。

[0028] 所述介质能够是电子、磁、光、电磁、红外、或者半导系统（或装置或设备）或传播介质。计算机可读介质的示例包括半导体或固态存储器、磁带、可移动计算机软磁盘、随机访问存储器 (RAM)、只读存储器 (ROM)、硬磁盘和光盘。光盘的当前示例包括压缩盘 - 只读存储器 (CD-ROM)、压缩盘 - 读 / 写 (CD-R/W) 和 DVD。

[0029] 适用于存储和 / 或执行程序代码的数据处理系统将包括直接或间接通过系统总线耦合到存储元件的至少一个处理器。所述存储元件能够包括在程序代码实际执行期间所采用的本地存储器、大容量存储器、以及提供暂时对至少某些程序代码的存储从而减少在执行期间必须从大容量存储器取回代码的次数的高速缓存。

[0030] 能够直接地或通过居间的 I/O 控制器将输入 / 输出或 I/O 设备（包括但不限于键盘、显示器、定点设备等）耦合到系统。

[0031] 还可以将网络适配器耦合到所述系统以便使得所述数据处理系统能够成为通过居间的专用或公共网络耦合到其它数据处理系统或远程打印机或存储设备。调制解调器、电缆调制解调器和以太网卡仅仅是当前可用的网络适配器类型之中的一些。

[0032] 图中的流程图和方框图说明了根据本发明的各种实施例的系统、方法和计算机程序产品的可能实现的架构、功能性和操作。在这一点上，流程图和方框图中的每个块可以代表代码的模块、片段或部分，其包括用于实现特定的局部功能的一个或多个可执行指令。还应当注意，在某些可替代实现中，在方框中标注的功能可以不按图中标注的次序发生。例如，取决于所涉及的功能性，连续示出的两个块实际上可以基本上同时地执行，或者有时可以以相反的次序来执行这些块。还将注意，方框图和 / 或流程图说明中的每个框以及方框图和 / 或流程图说明中的方框的组合能够由执行特定功能或动作的基于专用硬件的系统、或专用硬件和计算机指令的组合来实现。

[0033] 在此所使用的术语仅仅是为了描述特定实施例的目的，并且不意图成为对本发明的限制。如在此使用的，除非上下文中清楚指示的另外情形，否则单数形式“一”、“某”和“该”同样意图包括复数形式。将进一步理解，当在本说明书中使用时，术语“包括”和 / 或“具有”具体说明了阐明的特征、整体、步骤、操作、元素、和 / 或组件的存在，但是并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和 / 或其群组的存在或添加。

[0034] 图 1 说明了经由数据网络 153 提供客户端设备 100 和服务器设备 200 之间的 TCP 代理会话 400 的服务网关 300 的实施例。

[0035] 在一个实施例中，数据网络 153 包括网际协议 (IP) 网络、公司数据网络、区域公司数据网络、和因特网服务提供商网络、住宅数据网络、诸如是以太网的有线网络、诸如是 WiFi 网络的无线网络、或者蜂窝网络。在一个实施例中，数据网络 153 驻留于数据中心，或者连接到网络或应用网络云。

[0036] 客户端设备 100 典型地是具有网络接入能力的计算设备。在一个实施例中，客户端设备 100 是工作站、台式个人计算机或膝上型个人计算机、个人数据助理 (PDA)、平板计算设备、智能电话、或蜂窝电话、机顶盒、因特网媒体浏览器、因特网媒体播放器、智能传感器、智能医疗设备、网顶盒、联网电视机、联网 DVR、联网蓝光播放器、联网手持游戏设备、或媒体中心。

[0037] 在一个实施例中，客户端设备 100 是住宅宽带网关、商业因特网网关、商业 Web 代理服务器、网络顾客前端设备 (CPE)、或因特网接入网关。

[0038] 在一个实施例中，客户端设备 100 包括宽带远程接入服务器 (BRAS)、数字订户线接入复用器 (DSLAM)、电缆调制解调器终接系统 (CMTS)、或服务提供商接入网关。

[0039] 在一个实施例中，客户端设备 100 包括移动宽带接入网关，诸如，网关 GPRS 支持节点 (GGSN)、归属代理 (HA)、或 PDN 网关 (PGW)。

[0040] 在一个实施例中，客户端设备 100 包括服务器负载均衡器、应用投递控制器、话务管理器、防火墙、VPN 服务器、远程接入服务器、或企业或数据中心接入网关。

[0041] 在一个实施例中，客户端设备 100 是执行类似于服务网关 300 的功能的设备。

[0042] 客户端设备 100 经由服务网关 300 向服务器 200 发起 TCP 会话 400。

[0043] 服务器 200 是典型情况下耦合到处理器和存储了计算机可读程序代码的计算机可读介质的计算设备。服务器 200 与处理器和计算机可读程序代码一起实现了使用 TCP 会话 400 来将基于 TCP 的服务或应用服务提供给客户端设备 100 的 Web 服务器、文件服务器、视频服务器、数据库服务器、应用服务器、语音系统、会议服务器、媒体网关、媒体中心、app 服务器或网络服务器的功能性。

[0044] 在一个实施例中，服务器 200 是执行类似于服务网关 300 的功能的设备。

[0045] 在一个实施例中，TCP 会话 400 包括 HTTP 会话、FTP 文件传递会话、基于 TCP 的视频流送会话、基于 TCP 的音乐流送会话、文件下载会话、群组会议会话、数据库访问会话、远程终端接入会话、远程登陆会话、电子商务交易、远程规程呼叫、或基于 TCP 的网络通信会话。

[0046] 服务网关 300 操作地耦合到处理器 310、存储模块 320、网络接口模块 330、以及计算机可读介质 340。计算机可读介质 340 存储计算机可读程序代码，其在由处理器 310 使用存储模块 320 来执行时实现了如这里所述的本发明的各种实施例。在某些实施例中，将服务网关 300 实现为服务器负载均衡器、应用投递控制器、服务投递平台、话务管理器、安全网关、防火墙系统的组件、虚拟专用网络 (VPN) 的组件、视频服务器的负载均衡器、将负载分发给一个或多个服务器的网关、Web 或 HTTP 服务器、网络地址转译 (NAT) 网关、或 TCP 代理服务器。

[0047] 在客户端设备 100 和服务器 200 之间提供 TCP 代理会话 400 的一个实施例中，服务网关 300 建立与客户端设备 100 的客户端侧 TCP 会话 420，以及与服务器 200 的服务器侧 TCP 会话 470。

[0048] 在一个实施例中，计算机可读介质 340 包括用于服务应用 350 和执行服务应用 350 的处理器 310 的指令。

[0049] 在一个实施例中，服务应用 350 实现 VPN 防火墙、网关安全应用、HTTP 代理、基于 TCP 的音频或视频流送会话代理、Web 会话代理、内容过滤、服务器负载均衡、防火墙、或网络应用会话代理的功能性。

[0050] 在一个实施例中，服务网关 300 包括用于处理 TCP 代理会话 400 的存储缓冲区 327。在一个实施例中，存储缓冲区 327 的容量是预先确定的，例如，1MB、400KB、200KB 或者 2MB。在一实施例中，服务网关 300 保留存储模块 320 的一部分作为存储缓冲区 327 以处理 TCP 代理会话 400。

[0051] 在一个实施例中，服务网关 300 为客户端侧 TCP 会话 420 分配客户端侧发送缓冲区 424，并且为服务器侧 TCP 会话 470 分配服务器侧接收缓冲区 474，其中，客户端侧发送缓冲区 424 和服务器侧接收缓冲区 474 驻留在存储缓冲区 327 中。

[0052] 在一个实施例中，服务网关 300 监控服务器侧会话 470 的动态服务器侧网络行为 460 和客户端侧会话 420 的动态客户端侧网络行为 410。服务网关 300 基于网络行为 410 和网络行为 460 来分配客户端侧发送缓冲区 424 以及分配服务器侧接收缓冲区 474。

[0053] 在如图 2a 中所说明的一个实施例中，服务网关 300 接收来自服务器 200 的服务器侧 TCP 会话 470 的数据分组 479。服务网关 300 将数据分组 479 存储在服务器侧接收缓冲区 474 中。服务应用 350 从服务器侧接收缓冲区 474 中取回数据分组 479 并处理数据分组 479。在一个实施例中，服务应用 350 使用数据分组 479 来创建数据分组 429。服务应用 350 通过将数据分组 429 置于客户端侧发送缓冲区 424 中来发送数据分组 429 到客户端设备 100。服务网关 300 指令网络模块 330 从客户端侧发送缓冲区 424 传送数据分组 429 到客户端设备 100。

[0054] 在一个实施例中，服务网关 300 确定服务器侧接收缓冲区 474 的可用容量 484 足够用于存储数据分组 479。服务网关 300 将数据分组 479 存储到服务器侧接收缓冲区 474 中并将可用容量 484 减小达存储数据分组 479 所需的存储空间量。

[0055] 在一个实施例中，服务网关 300 确定可用容量 484 不足，服务网关 300 增加服务器侧接收缓冲区 474 的所分配的容量 494，并根据所分配的容量 494 的增加来调整可用容量 484。如果所调整的可用容量 484 足够，则服务网关 300 将数据分组 479 存储在服务器侧接收缓冲区 474 中并根据存储数据分组 479 所需的存储空间来减小可用容量 484。

[0056] 在服务网关 300 确定可用容量 484 不足的一个实施例中，即使在调整之后，服务网关 300 也会丢弃数据分组 479。

[0057] 在一个实施例中，服务应用 350 从服务器侧接收缓冲区 474 取回数据分组 479。服务网关 300 从服务器侧接收缓冲区 474 中移除数据分组 479 并根据数据分组 479 所消耗的存储空间来增加可用容量 484。

[0058] 在如图 2b 中所说明的一个实施例中，在服务网关 300 将数据分组 429 存储到客户端侧发送缓冲区 424 中之前，服务网关 300 检查客户端侧发送缓冲区 424 的可用容量是否足够用于存储数据分组 429。如果该检查是肯定的，则服务网关 300 将数据分组 429 存储到客户端侧发送缓冲区 424 中并根据存储数据分组 429 所需的存储空间来减小可用容量 434。

[0059] 在一个实施例中，服务网关 300 确定可用容量 434 不足，服务网关 300 增加客户端侧发送缓冲区 424 的所分配的容量 444，并根据所分配的容量 444 的增加来调整可用容量 434。如果所调整的可用容量 434 足够，则服务网关 300 将数据分组 429 存储到客户端侧发送缓冲区 424 中并根据存储数据分组 429 所需的存储空间来减小可用容量 434。

[0060] 在服务网关 300 确定可用容量 434 不足的一个实施例中，即使在调整之后，服务网关 300 也丢弃数据分组 429。在一个实施例中，服务网关 300 等待直到可用容量 434 在调整之后变得足够。

[0061] 在一个实施例中，网络模块 330 成功传送来自客户端侧发送缓冲区 424 的数据分组 429。服务网关 300 从客户端侧发送缓冲区 424 移除数据分组 429 并根据数据分组 429 所消耗的存储空间来增加可用容量 434。

[0062] 在图 3 所说明的实施例中，服务网关 300 为服务器侧会话 470 监控动态网络行为 460。在一个实施例中，网络行为 460 包括服务器侧数据分组往返时间 (RTT)462、服务器侧传输速率 463、或影响服务网关 300 接收数据分组 479 的动态服务器侧网络属性。在一个实施例中，服务器侧 RTT462 影响通过数据网络 153 传送数据分组 479 的时间。在一个实施例中，服务器侧传输速率 463 影响网络模块 330 接收数据分组 479 的速率。在一个实施例中，RTT462 和传输速率 463 是相关的。

[0063] 服务网关300为服务器侧会话470测量或估计服务器侧RTT462。在一个示例实施例中，服务网关300基于在服务网关300发送服务器侧会话470的数据分组到服务器200的时刻和服务网关300接收到对该已发送数据分组的确认的时刻之间的持续时间来测量服务器侧RTT462。在一个实施例中，服务网关300在服务器侧TCP会话470期间不时地测量RTT462。在一个实施例中，服务网关300基于与服务器200的一个或多个先前的服务器侧TCP会话来估计服务器侧RTT462。在一个实施例中，服务网关300估计服务器侧RTT462为10毫秒、100毫秒、3毫秒、22毫秒、或3秒。

[0064] 在一个实施例中，服务网关300计算服务器侧传输速率463为网络模块330接收服务器侧TCP会话470的数据分组所使用的网络接口的带宽。在一个实施例中，服务器侧传输速率463是虚拟网络接口的带宽。在一个实施例中，服务网关300通过计算一段时间期间服务器侧TCP会话470的一个或多个数据分组的一个或多个传输速率来计算服务器侧传输速率463。在一个实施例中，服务网关300不时地计算新的传输速率463。在一个实施例中，服务网关300计算服务器侧传输速率463为1Mbps、43Mbps、100Mbps、1Gbps、600Mbps、25Mbps或10Gbps。

[0065] 在一个实施例中，服务网关300监控客户端侧会话420的动态网络行为410。在一个实施例中，网络行为410包括客户端侧数据分组往返时间(RTT)412、客户端侧传输速率413、或影响服务网关300发送数据分组429的动态网络属性。在一个实施例中，客户端侧RTT412影响通过数据网络153传送数据分组429的时间。在一个实施例中，客户端侧传输速率413影响网络模块330发送数据分组429的速率。在一个实施例中，RTT412和传输速率413是相关的。

[0066] 服务网关300测量或估计客户端侧RTT412。在一个示例实施例中，服务网关300基于在服务网关300发送客户端侧会话420的数据分组到客户端设备100的时刻和服务网关300接收到对该已发送数据分组的确认的时刻之间的持续时间来测量客户端侧RTT412。在一个实施例中，服务网关300在客户端侧TCP会话420期间不时地测量RTT412。在一个实施例中，服务网关300基于与客户端设备100的一个或多个先前的服务器侧TCP会话来估计客户端侧RTT412。在一个实施例中，服务网关300估计服务器侧RTT412为10毫秒、100毫秒、3毫秒、22毫秒、或3秒。

[0067] 在一个实施例中，服务网关300计算客户端侧传输速率413为网络模块330传送客户端侧TCP会话420的数据分组所使用的网络接口的带宽。在一个实施例中，客户端侧传输速率413是虚拟网络接口的带宽。在一个实施例中，服务网关300通过计算一段时间期间客户端侧TCP会话410的一个或多个数据分组的一个或多个传输速率来计算客户端侧传输速率413。在一个实施例中，服务网关300不时地计算新的传输速率413。在一个实施例中，服务网关300计算客户端侧传输速率413为1Mbps、43Mbps、100Mbps、1Gbps、600Mbps、25Mbps或10Gbps。

[0068] 图4说明了服务网关300从存储缓冲区327分配存储空间给服务器侧接收缓冲区474和客户端侧发送缓冲区424的实施例。在一个实施例中，当服务网关300开始处理TCP代理会话400时，服务网关300为服务器侧接收缓冲区474保留存储缓冲区327的一部分，并为客户端侧发送缓冲区424保留存储缓冲区327的一部分。在一个实施例中，当服务网关300建立服务器侧TCP会话470和/或客户端侧TCP会话420时，服务网关300开始处

理 TCP 代理会话 400。在一个实施例中，服务网关 300 初始分配预定容量 494 的服务器侧接收缓冲区 474。例如，所分配的容量 494 初始被预定为在存储缓冲区 327 的 10％、8％、1/8 或存储缓冲区 327 的另一预定分数。在另一实施例中，将初始分配的容量 494 预设为 30KB、32KB、4KB、64KB 或 100KB。同样地，服务网关 300 预先确定初始分配容量 444 的客户端发送缓冲区 424。服务网关 300 进一步将服务器侧接收缓冲区 474 的可用容量 484 设置为与所分配的容量 494 相同，并将客户端侧发送缓冲区 424 的可用容量 434 设置为与所分配的容量 444 相同。

[0069] 在一个实施例中，服务网关 300 计算服务器侧接收缓冲区 474 的最大容量 492 和客户端侧发送缓冲区 424 的最大容量 442。在一个实施例中，服务网关 300 根据公式 1 和公式 2 来使用 RTT462 和 RTT412、并使用基于 RTT462 和 RTT412 对存储缓冲区 327 的容量 328 的加权平均来计算。所述加权平均具有朝服务器侧接收缓冲区 474 的偏向。

[0070] 公式 1

[0071]

$$\text{最大容量 492} = \frac{2*\text{RTT 462}}{\text{RTT 412} + 2*\text{RTT 462}} * \text{容量 328}$$

[0072] 公式 2

[0073]

$$\text{最大容量 442} = \frac{\text{RTT 412}}{\text{RTT 412} + 2*\text{RTT 462}} * \text{容量 328}$$

[0074] 公式 1 和公式 2 说明了其中最大容量 492 和最大容量 442 的和等于容量 328 的实施例。在该实施例中，选择 RTT462 的因子为 2 使得服务网关 300 为服务器侧 TCP 会话 470 通告服务器侧接收缓冲区 474 最大容量 492 的 1/2 或一半。使用与为服务器侧 TCP 会话 470 所通告的服务器侧接收缓冲区 474 的该部分相关的不同因子来调整因子 2 是合适的。

[0075] 在一个实施例中，服务网关 300 获取服务器侧 RTT462 的新的测量和 / 或客户端侧 RTT412 的新的测量。服务网关 300 应用公式 1 和公式 2 以修正最大容量 492 和最大容量 442。

[0076] 在一个实施例中，服务网关 300 进一步考虑服务器侧传输速率 463 和客户端侧传输速率 413。服务网关使用公式 3 和公式 4、并使用基于传输速率 463 和 RTT462 以及传输速率 413 和 RTT412 的组合对容量 328 的加权平均来计算。所述加权平均具有朝服务器侧接收缓冲区 474 的偏向。

[0077] 公式 3

[0078]

$$\text{最大容量 492} = \frac{2 * \text{传输速率 463} * \text{RTT 462}}{\text{传输速率 413} * \text{RTT 412} + 2*\text{传输速率 463} * \text{RTT 462}} * \text{容量 328}$$

[0079] 公式 4

[0080]

$$\text{最大容量 442} = \frac{\text{传输速率 413} * \text{RTT 412}}{\text{传输速率 413} * \text{RTT 412} + 2 * \text{传输速率 463} * \text{RTT 462}} * \text{容量 328}$$

[0081] 公式 3 和公式 4 说明了其中最大容量 492 和最大容量 442 的和等于容量 328 的实施例。

[0082] 在一个实施例中，服务网关 300 获取新的服务器侧传输速率 463 和 / 或新的客户端侧传输速率 413。服务网关 300 应用公式 3 和 4 以修正最大容量 492 和最大容量 442。在一个实施例中，服务网关 300 获取服务器侧 RTT462 的新的测量和 / 或客户端侧 RTT412 的新的测量。服务网关 300 应用公式 3 和公式 4 以修正最大容量 492 和最大容量 442。

[0083] 在图 5a 中，进一步说明了图 2a 中的实施例，其中，服务网关 300 将服务器数据分组 479 存储到服务器侧接收缓冲区 474 中并发现服务器侧接收缓冲区 474 的可用容量 484 不足。服务网关 300 检查服务器侧接收缓冲区 474 的所分配容量 494 是否小于服务器侧接收缓冲区 474 的最大容量 492。如果所分配的容量 494 至少是最大容量 492 或者容量 494 和最大容量 492 之间的差小于存储数据分组 479 所需的存储，则服务网关 300 丢弃数据分组 479。

[0084] 在一个实施例中，服务网关 300 确定在最大容量 492 以内存在足够的存储空间用于存储数据分组 479。服务网关 300 从存储缓冲区 327 分配存储空间 321。在一个实施例中，存储空间 321 具有容量 322。服务网关 300 将服务器侧接收缓冲区 474 的所分配容量 494 增加达容量 322 的量，并将存储空间 321 包括到服务器侧接收缓冲区 474 中。服务网关 300 分配存储空间 321 以使得经调整的所分配容量 494 不超过服务器侧接收缓冲区 474 的最大容量 492。在将数据分组 479 存储到经更新的服务器侧接收缓冲区 474 之后，服务网关 300 使用容量 322、以及数据分组 479 所消耗的空间来更新服务器侧接收缓冲区 474 的可用容量 484。

[0085] 在图 5b 中，进一步说明了图 2b 中的实施例，其中，服务网关 300 将数据分组 429 存储到客户端侧发送缓冲区 424 中并发现客户端侧发送缓冲区 424 的可用容量 434 不足。服务网关 300 检查客户端侧发送缓冲区 424 的所分配容量 444 是否小于客户端侧发送缓冲区 424 的最大容量 442。如果容量 444 至少是最大容量 442 或者所分配容量 444 和最大容量 442 之间的差小于存储数据分组 429 所需的存储，则服务网关 300 丢弃数据分组 429。

[0086] 在一个实施例中，服务网关 300 确定在最大容量 442 以内存在足够的存储空间用于存储数据分组 429。服务网关 300 从存储缓冲区 327 分配存储空间 325。在一个实施例中，存储空间 325 具有容量 326。服务网关 300 将客户端侧发送缓冲区 424 的所分配容量增加达容量 326 的量，并将存储空间 325 包括到客户端侧发送缓冲区 424 中。服务网关 300 分配存储空间 325 以使得经调整的所分配容量 444 不超过客户端侧发送缓冲区 424 的最大容量 442。在将数据分组 429 存储到经更新的客户端侧发送缓冲区 424 之后，服务网关 300 使用容量 326、以及数据分组 429 所消耗的空间来更新可用容量 434。

[0087] 在一个实施例中，客户端设备 100 经由客户端侧 TCP 会话 420 发送数据分组 421 到服务网关 300。如图 6 中所说明地，在一个实施例中，服务网关 300 包括存储缓冲区 377，其包括客户端侧接收缓冲区 426 和服务器侧发送缓冲区 476。存储缓冲区 377 与容量 378

相关联。客户端侧接收缓冲区 426 与所分配的容量 446、可用容量 436 和最大容量 448 相关联。服务器侧发送缓冲区 476 与所分配的容量 496、可用容量 486 和最大容量 498 相关联。

[0088] 服务网关 300 使用客户端侧接收缓冲区 426 和服务器侧发送缓冲区 476 来处理数据分组 421。服务网关 300 接收数据分组 421 并将数据分组 421 存储到客户端侧接收缓冲区 426 中。服务应用 350 从客户端侧接收缓冲区 426 取回数据分组 421，并基于数据分组 421 来生成数据分组 471。服务应用 350 将数据分组 471 置于服务器侧发送缓冲区 476 中。服务网关 300 将数据分组 471 从服务器侧发送缓冲区 476 发送到服务器 200。

[0089] 在一个实施例中，服务网关 300 保留预定容量 378 的存储缓冲区 377 用于处理 TCP 代理会话 400。在一个实施例中，存储缓冲区 377 的容量 378 与存储缓冲区 327 的容量 328 相同。在一个实施例中，容量 378 不同于容量 328。

[0090] 图 7 说明了服务网关 300 将来自存储缓冲区 377 的存储空间分配给服务器侧发送缓冲区 476 和客户端侧接收缓冲区 426 的实施例。在一个实施例中，当服务网关 300 开始处理 TCP 代理会话 400 时，服务网关 300 为服务器侧发送缓冲区 476 保留存储缓冲区 377 的一部分，并为客户端侧接收缓冲区 426 保留存储缓冲区 377 的一部分。在一个实施例中，当服务网关 300 建立服务器侧 TCP 会话 470 和 / 或客户端侧 TCP 会话 420 时，服务网关 300 开始处理 TCP 代理会话 400。在一个实施例中，服务网关 300 初始分配预定容量 496 的服务器侧发送缓冲区 476。例如，所分配容量 496 被初始地预定为在存储缓冲区 377 的 10％、8％、1/8、或存储缓冲区 377 的另一分数。在另一实施例中，将所分配容量 496 的初始值预设为 30KB、32KB、4KB、64KB 或 100KB。同样地，服务网关 300 预先确定客户端侧接收缓冲区 426 的初始分配容量 446。服务网关 300 进一步将服务器侧发送缓冲区 476 的可用容量 486 设置为与所分配容量 496 相同，并将客户端侧接收缓冲区 426 的可用容量 436 设置为与所分配容量 446 相同。

[0091] 在一个实施例中，网络条件 460 的 RTT462 影响数据分组 471 的发送。在一个实施例中，网络条件 460 包括服务器侧传输速率 464。服务网关 300 计算服务器侧传输速率 464 为网络模块 330 传送服务器侧 TCP 会话 470 的数据分组所使用的网络接口的带宽。在一个实施例中，服务器侧传输速率 464 是虚拟网络接口的带宽。在一个实施例中，服务网关 300 通过计算一段时间内服务器侧 TCP 会话 470 的一个或多个数据的一个或多个传输速率来计算服务器侧传输速率 464。在一个实施例中，服务网关 300 不时地计算新的传输速率 464。在一个实施例中，服务网关 300 计算服务器侧传输速率 464 为 1Mbps、43Mbps、100Mbps、1Gbps、600Mbps、25Mbps 或 10Gbps。

[0092] 在一个实施例中，网络条件 410 的 RTT412 还影响数据分组 421 的接收。在一个实施例中，网络条件 410 包括客户端侧传输速率 414。服务网关 300 计算客户端侧传输速率 414 为网络模块 330 接收客户端侧 TCP 会话 420 的数据分组所使用的网络接口的带宽。在一个实施例中，客户端侧传输速率 414 是虚拟网络接口的带宽。在一个实施例中，服务网关 300 通过计算在一段时间内客户端侧 TCP 会话 410 的一个或多个数据分组的一个或多个传输速率来计算客户端侧传输速率 414。在一个实施例中，服务网关 300 不时地计算新的传输速率 414。在一个实施例中，服务网关 300 计算客户端侧传输速率 414 为 1Mbps、43Mbps、100Mbps、1Gbps、600Mbps、25Mbps 或 10Gbps。

[0093] 在一个实施例中，服务网关 300 根据公式 5 和公式 6 使用网络条件 460 和网络条

件 410，并使用基于 RTT462 和 RTT412 对容量 378 的加权平均来计算最大容量 498 和最大容量 448。所述加权平均具有朝客户端侧接收缓冲区 426 的偏向。

[0094] 公式 5

[0095]

$$\text{最大容量 498} = \frac{\text{RTT 462}}{\text{2*RTT 412 + RTT 462}} * \text{容量 378}$$

[0096] 公式 6

[0097]

$$\text{最大容量 448} = \frac{\text{2*RTT 412}}{\text{2*RTT 412 + RTT 462}} * \text{容量 378}$$

[0098] 公式 5 和公式 6 说明了其中最大容量 498 和最大容量 448 的和等于容量 378 的实施例。在这一实施例中，选择 RTT412 的因子为 2 以使得服务网关 300 为客户端侧 TCP 会话 420 通告客户端侧接收缓冲区 426 的最大容量 448 的 1/2 或一半。使用与为客户端侧 TCP 会话 420 通告的客户端侧接收缓冲区 426 的该部分相关的不同因子来调整因子 2 是合适的。

[0099] 在一个实施例中，服务网关 300 获取服务器侧 RTT462 的新的测量和 / 或客户端侧 RTT412 的新的测量。服务网关 300 应用公式 5 和公式 6 来修正最大容量 498 和最大容量 448。

[0100] 在一个实施例中，服务网关 300 进一步考虑服务器侧传输速率 464 和客户端侧传输速率 414。服务网关使用公式 7 和公式 8、使用基于传输速率 464 和 RTT462、以及传输速率 414 和 RTT412 的组合对容量 378 的加权平均来计算。所述加权平均具有朝客户端侧接收缓冲区 426 的偏向。

[0101] 公式 7

[0102]

$$\text{最大容量 498} = \frac{\text{传输速率 464 * RTT 462}}{\text{2 * 传输速率 414 * RTT 412 + 传输速率 464 * RTT 462}} * \text{容量 378}$$

[0103] 公式 9

[0104]

$$\text{最大容量 448} = \frac{\text{2 * 传输速率 414 * RTT 412}}{\text{2 * 传输速率 414 * RTT 412 + 传输速率 464 * RTT462}} * \text{容量 378}$$

[0105] 公式 7 和公式 8 说明了其中最大容量 498 和最大容量 448 的和等于容量 378 的实施例。

[0106] 在一个实施例中，服务网关 300 获取新的服务器侧传输速率 464 和 / 或新的客户

端侧传输速率414，服务网关300应用公式7和公式8来修正最大容量498和最大容量448。在一个实施例中，服务网关300获取服务器侧RTT462的新的测量和/或客户端侧RTT412的新的测量，服务网关300应用公式7和公式8来修正最大容量492和最大容量442。

[0107] 在图8a中，服务网关300接收来自客户端设备100的数据分组421并将数据分组421存储到客户端侧接收缓冲区426中，并发现客户端侧接收缓冲区426的可用容量436不足。服务网关300检查客户端侧接收缓冲区426的所分配容量446是否小于客户端侧接收缓冲区426的最大容量448。如果所分配的容量446至少是最大容量448或者所分配容量446和最大容量448之间的差小于存储数据分组421所需的存储，则服务网关300丢弃数据分组421。

[0108] 在一个实施例中，服务网关300确定在最大容量448以内存在足够的存储空间用于存储数据分组421。服务网关300从存储器缓冲区377分配存储空间375。在一个实施例中，存储空间375具有容量376。服务网关300将所分配的容量446增加达容量376的量，并将存储空间375包括在客户端侧接收缓冲区426中。服务网关300分配存储空间375以使得经调整的所分配容量446不超过客户端侧发送缓冲区426的最大容量448。在将数据分组421存储到经更新的客户端侧接收缓冲区426之后，服务网关300使用容量376、以及数据分组421所消耗的空间来更新客户端侧接收缓冲区426的可用容量436。

[0109] 在图8b中，服务网关300将数据分组471存储到服务器侧发送缓冲区476中。在一个实施例中，服务网关300发现服务器侧发送缓冲区476的可用容量486不足。服务网关300检查服务器侧发送缓冲区476的所分配容量496是否小于服务器侧发送缓冲区476的最大容量498。如果所分配容量496至少是最大容量498或者在所分配的容量496和最大容量498之间的差小于存储数据分组471所需的存储，则服务网关丢弃数据分组471。

[0110] 在一个实施例中，服务网关300确定在最大容量498以内存在足够的存储空间用于存储数据分组471。服务网关300从存储缓冲区377分配存储空间371。在一个实施例中，存储空间371具有容量372。服务网关300将所分配的容量496增加达容量372的量，并将存储空间371包括到服务器侧发送缓冲区476中。服务网关300分配存储空间371以使得经调整的所分配的容量496不超过最大容量498。在将数据分组471存储到经更新的服务器侧发送缓冲区476之后，服务网关300使用容量372、以及数据分组471所消耗的空间来更新服务器侧发送缓冲区476的可用容量486。

[0111] 在一个实施例中，服务应用350从客户端侧接收缓冲区426取回数据分组471。服务网关300从客户端侧接收缓冲区426移除数据分组471并根据数据分组471所消耗的存储空间来增加可用容量436。

[0112] 在一个实施例中，网络模块330成功传送来自服务器侧发送缓冲区476的数据分组471。服务网关300从服务器侧发送缓冲区476移除数据分组471并根据数据分组471所消耗的存器空间来增加可用容量486。

[0113] 虽然已经按照所示实施例描述了本发明，但是本领域普通技术人员将容易意识到实施例可以有变化，并且那些变化将在本发明的精神和范围内。相应地，本领域普通技术人员可以作出众多修改，而不脱离所附权利要求书的精神和范围。

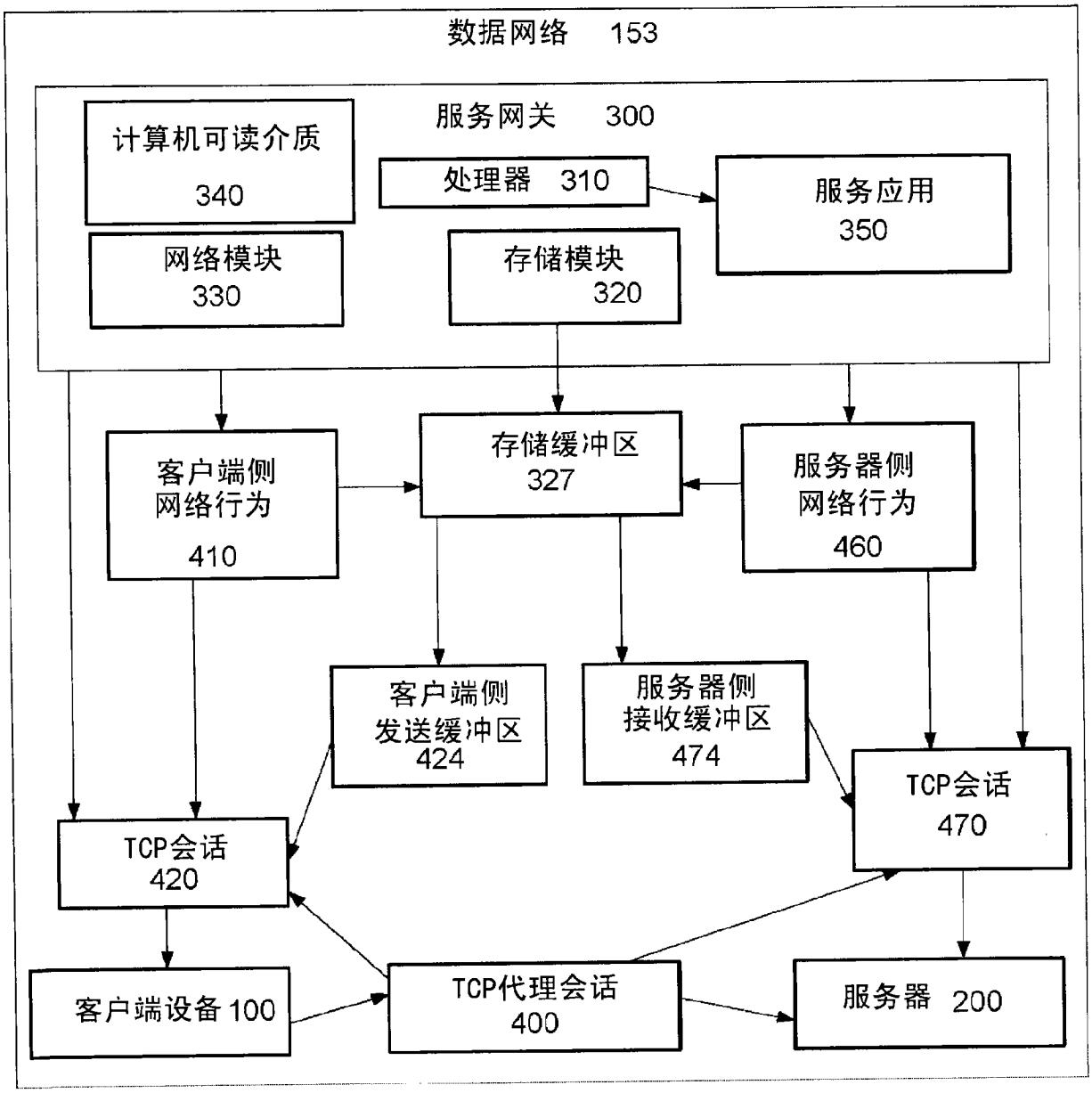

数据网络 153
服务网关 300
计算机可读介质 340
处理器 310
服务应用 350
网络模块 330
存储模块 320
客户端侧 网络行为 410
存储缓冲区 327
服务器侧 网络行为 460
客户端侧 发送缓冲区 424
服务器侧 接收缓冲区 474
TCP会话 470
TCP会话 420
客户端设备 100
TCP代理会话 400
服务器 200

图 1

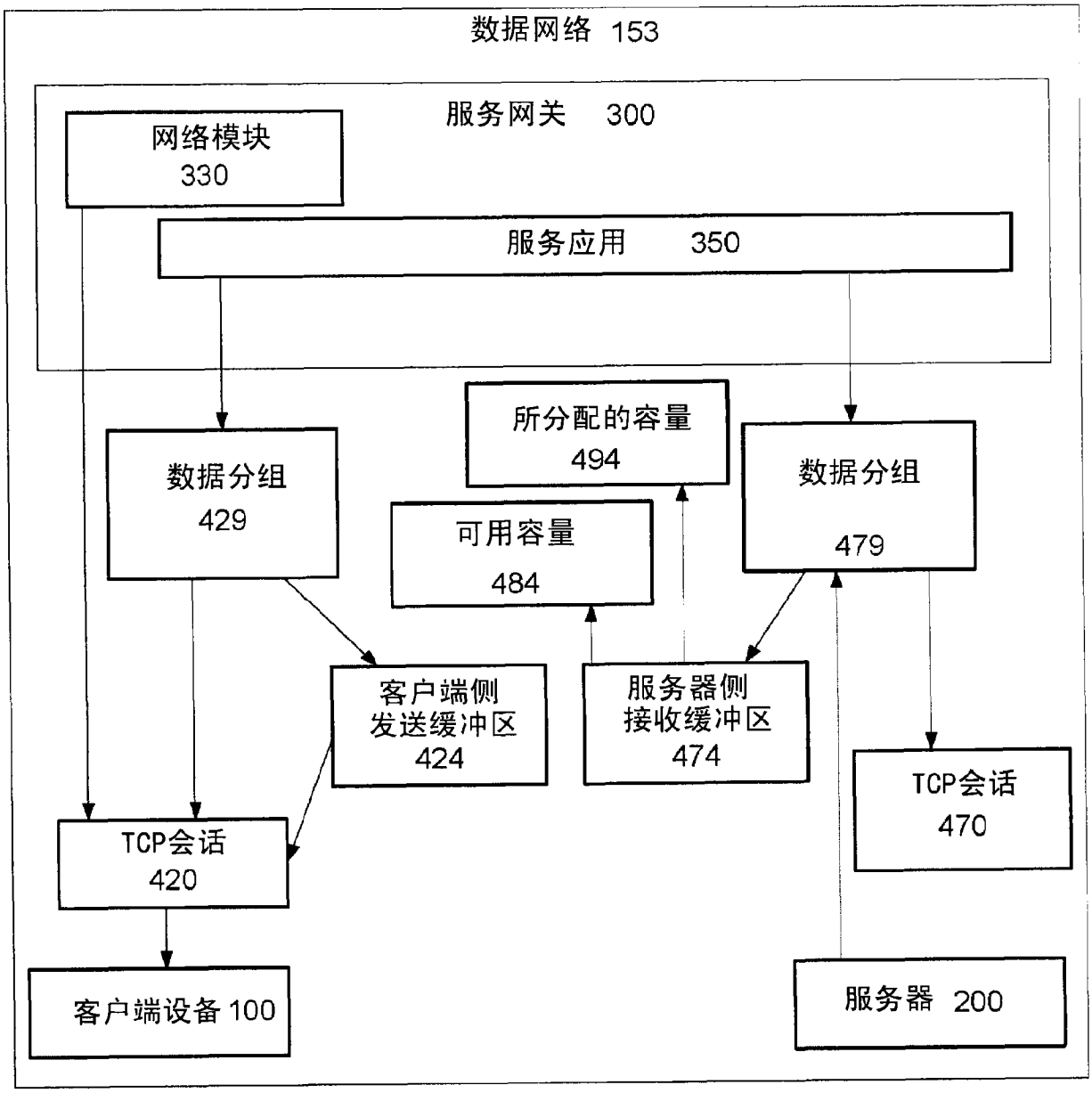

数据网络 153
服务网关 300
网络模块
330
服务应用 350
所分配的容量
494
数据分组
429
可用容量
484
数据分组
479
客户端侧
发送缓冲区
424
服务器侧
接收缓冲区
474
TCP会话
470
TCP会话
420
客户端设备 100
服务器 200

图 2a

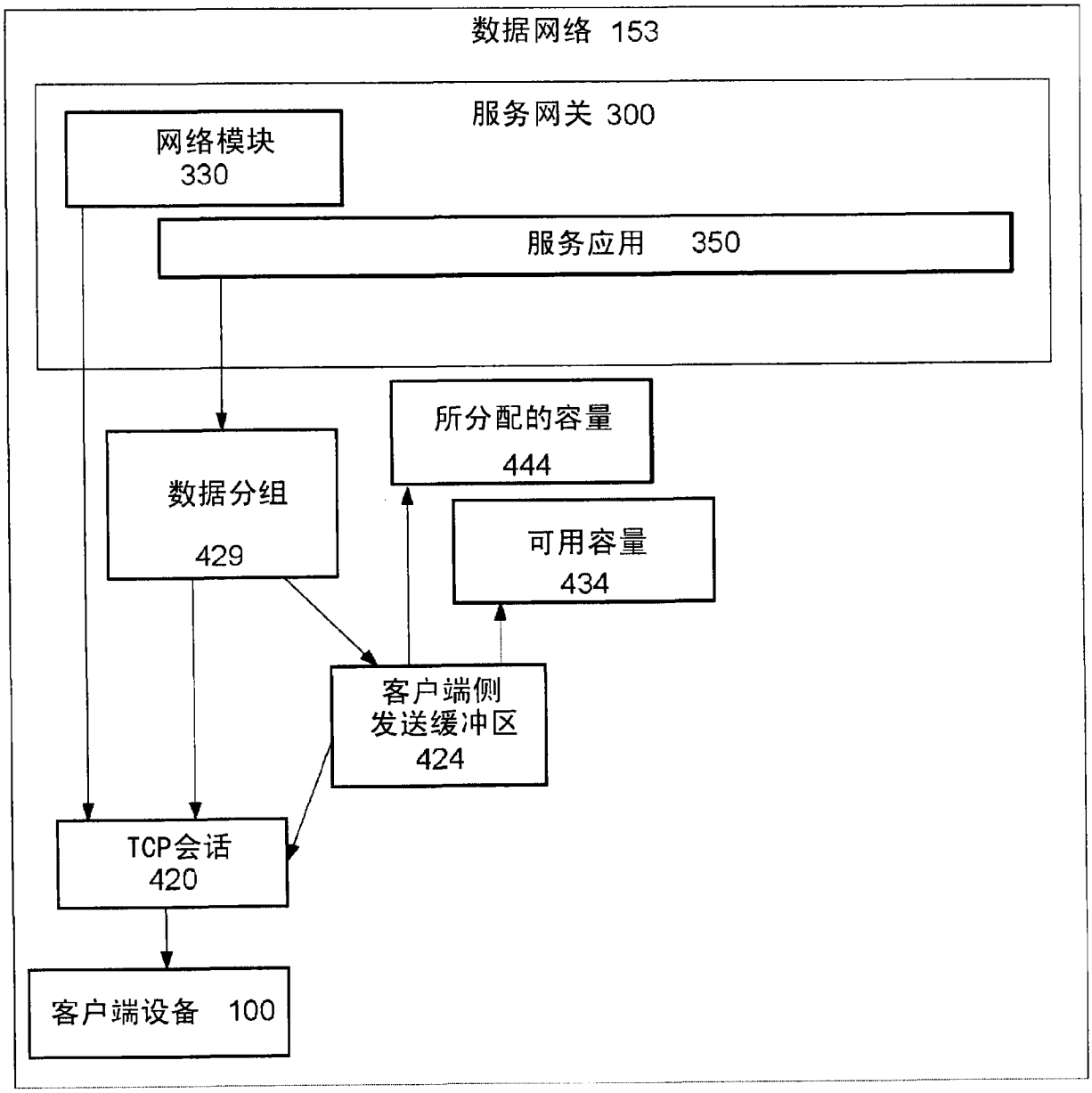

数据网络 153
服务网关 300
网络模块
330
服务应用 350
所分配的容量
444
数据分组
429
可用容量
434
客户端侧
发送缓冲区
424
TCP会话
420
客户端设备 100

图 2b

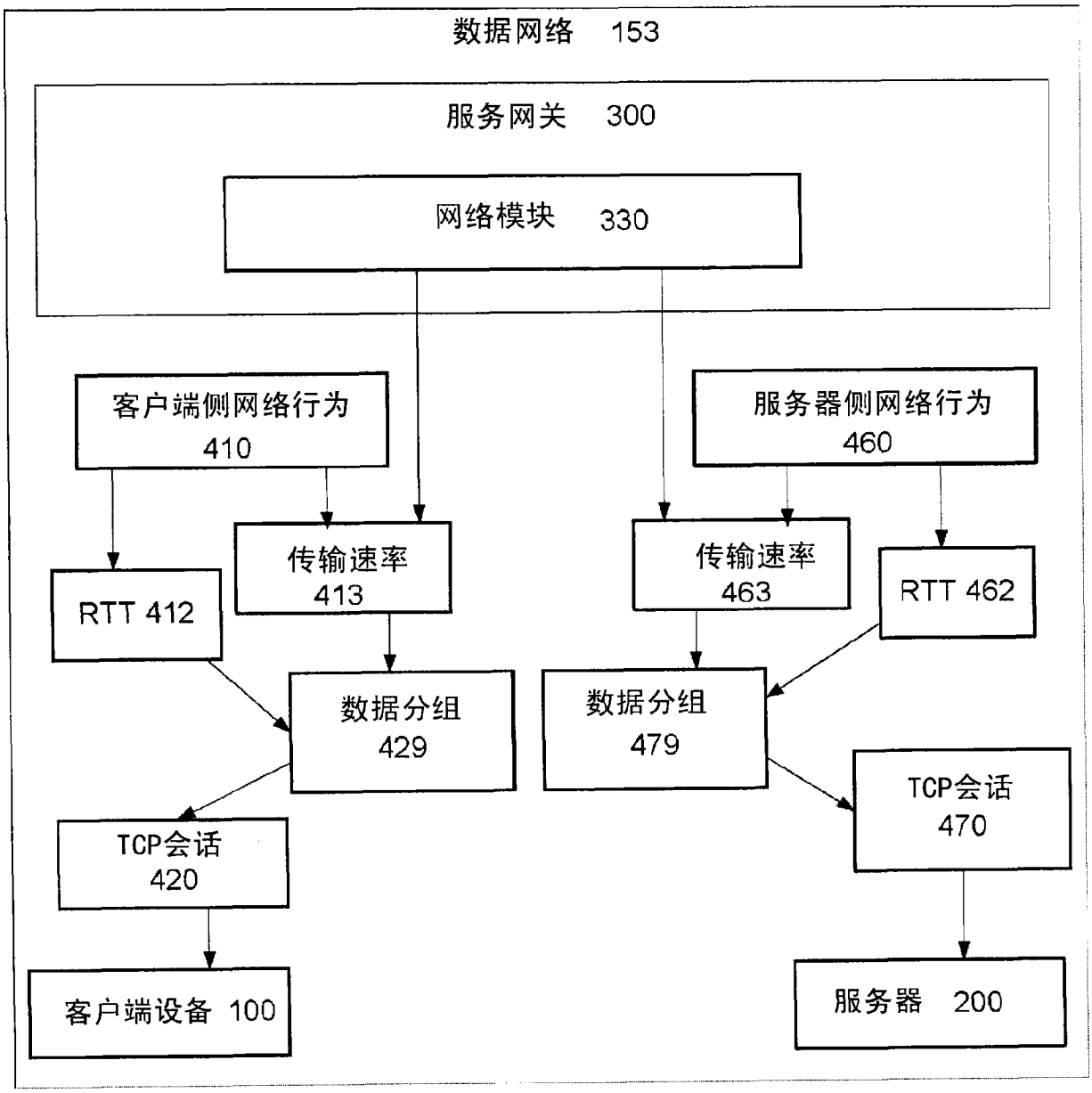

数据网络 153
服务网关 300
网络模块 330
客户端侧网络行为 410
服务器侧网络行为 460
传输速率 413
传输速率 463
RTT 412
RTT 462
数据分组 429
数据分组 479
TCP会话 420
TCP会话 470
客户端设备 100
服务器 200

图 3

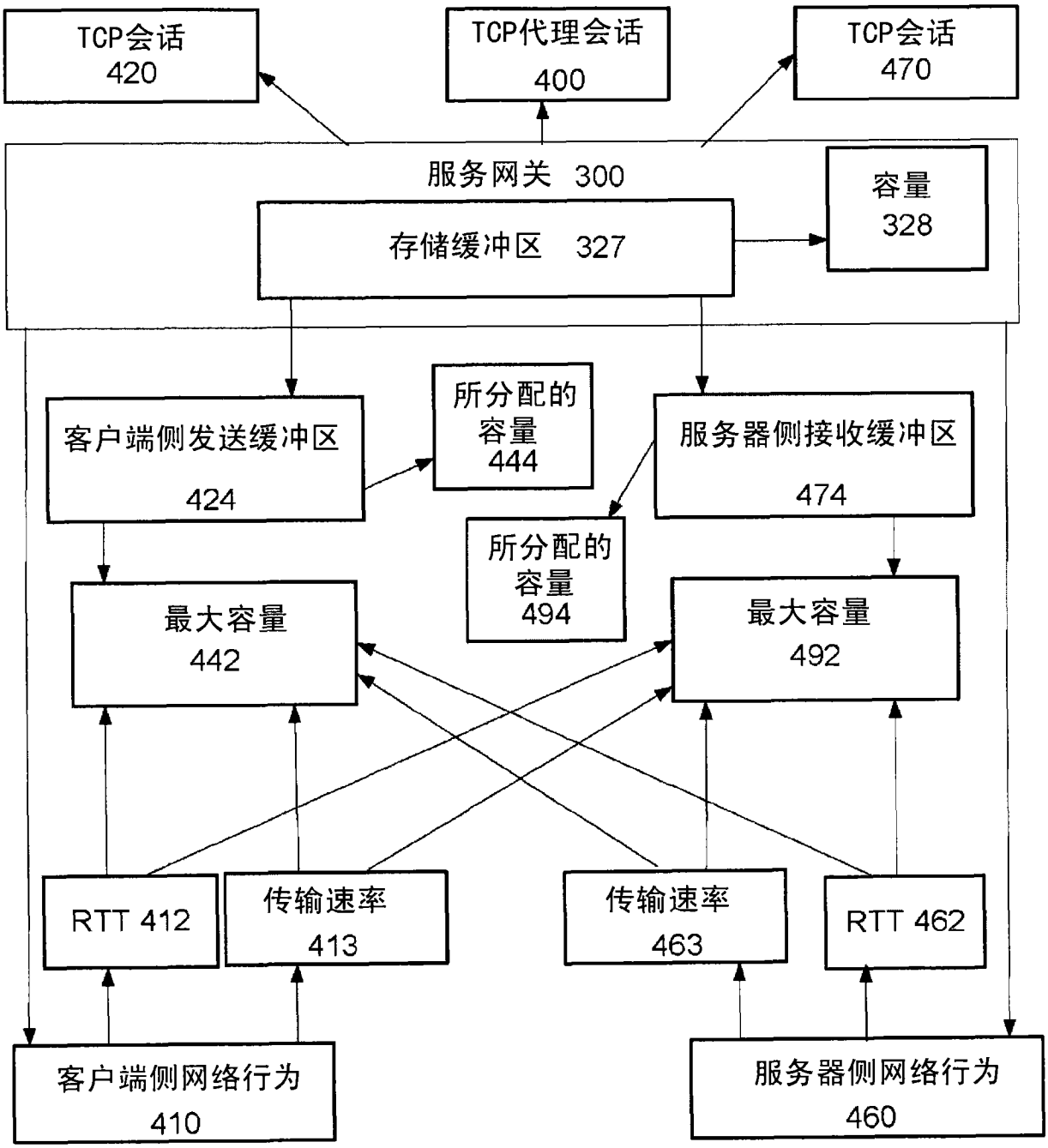

TCP会话
420
TCP代理会话
400
TCP会话
470
服务网关 300
存储缓冲区 327
容量
328
客户端侧发送缓冲区
424
所分配的
容量
444
服务器侧接收缓冲区
474
所分配的
容量
494
最大容量
442
最大容量
492
RTT 412
传输速率
413
传输速率
463
RTT 462
客户端侧网络行为
410
服务器侧网络行为
460

图 4

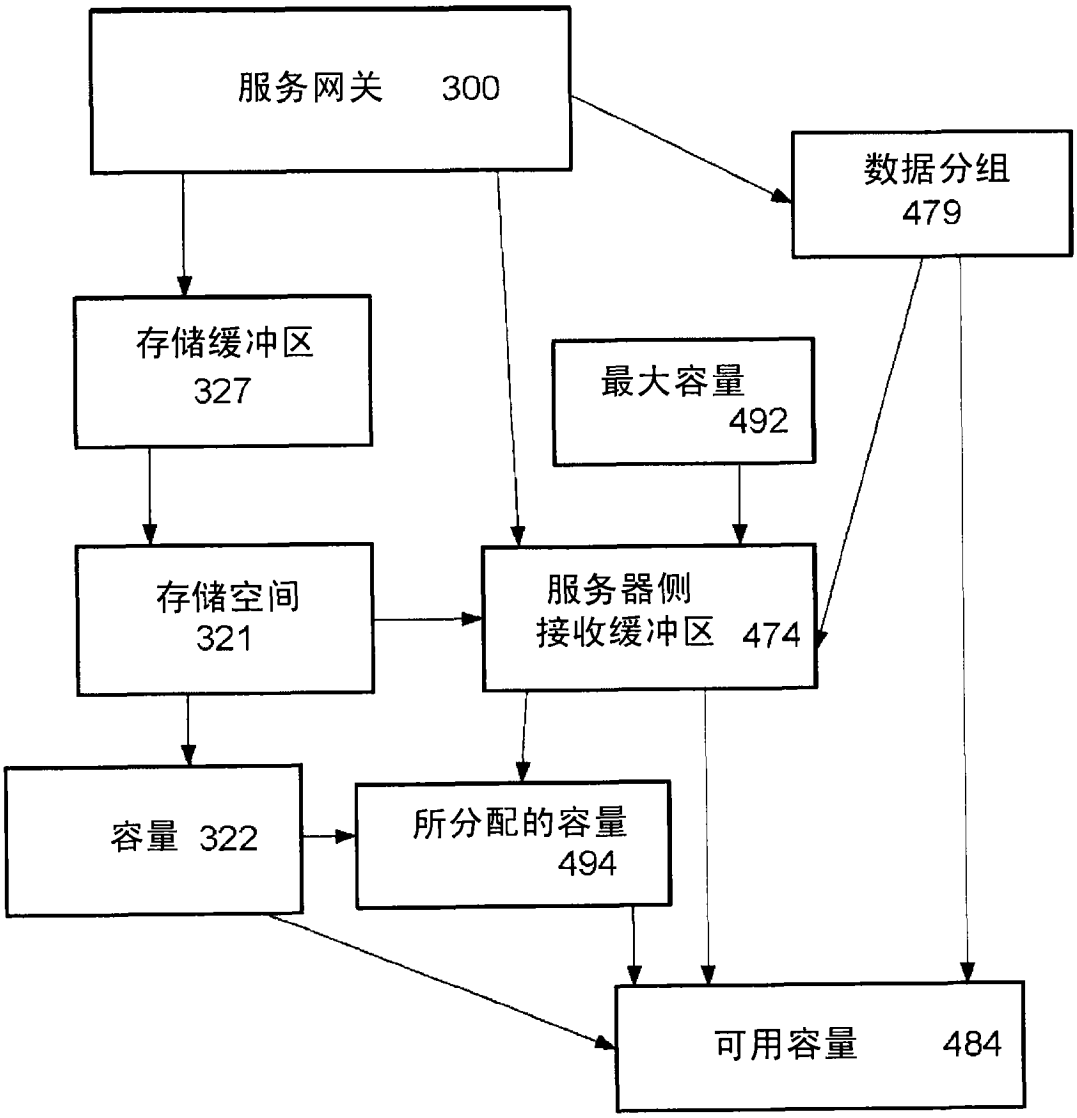

服务网关 300
数据分组 479
存储缓冲区 327
最大容量 492
存储空间 321
服务器侧 接收缓冲区 474
容量 322
所分配的容量 494
可用容量 484

图 5a

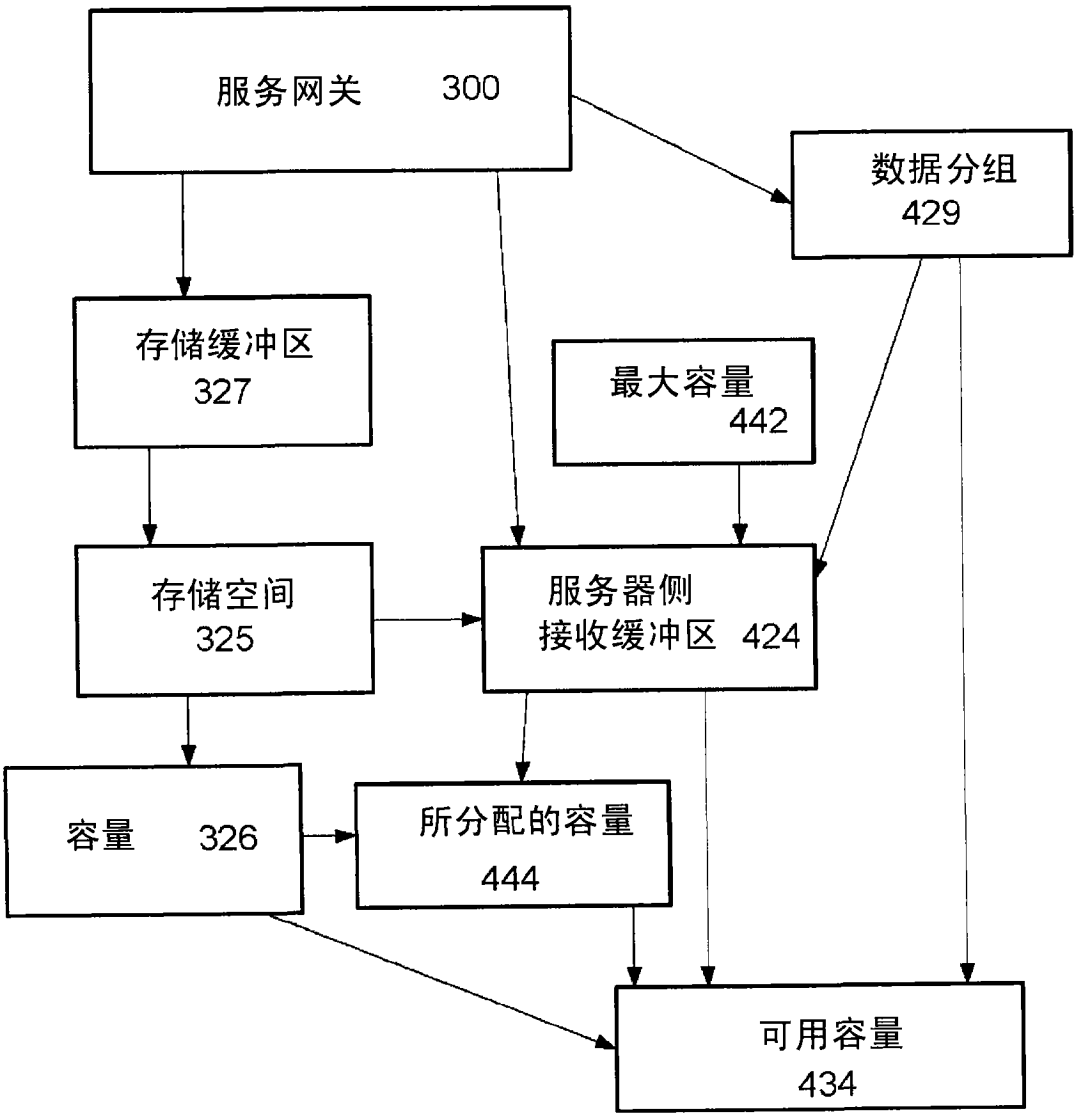

服务网关 300
数据分组 429
存储缓冲区 327
最大容量 442
存储空间 325
服务器侧 接收缓冲区 424
容量 326
所分配的容量 444
可用容量 434

图 5b

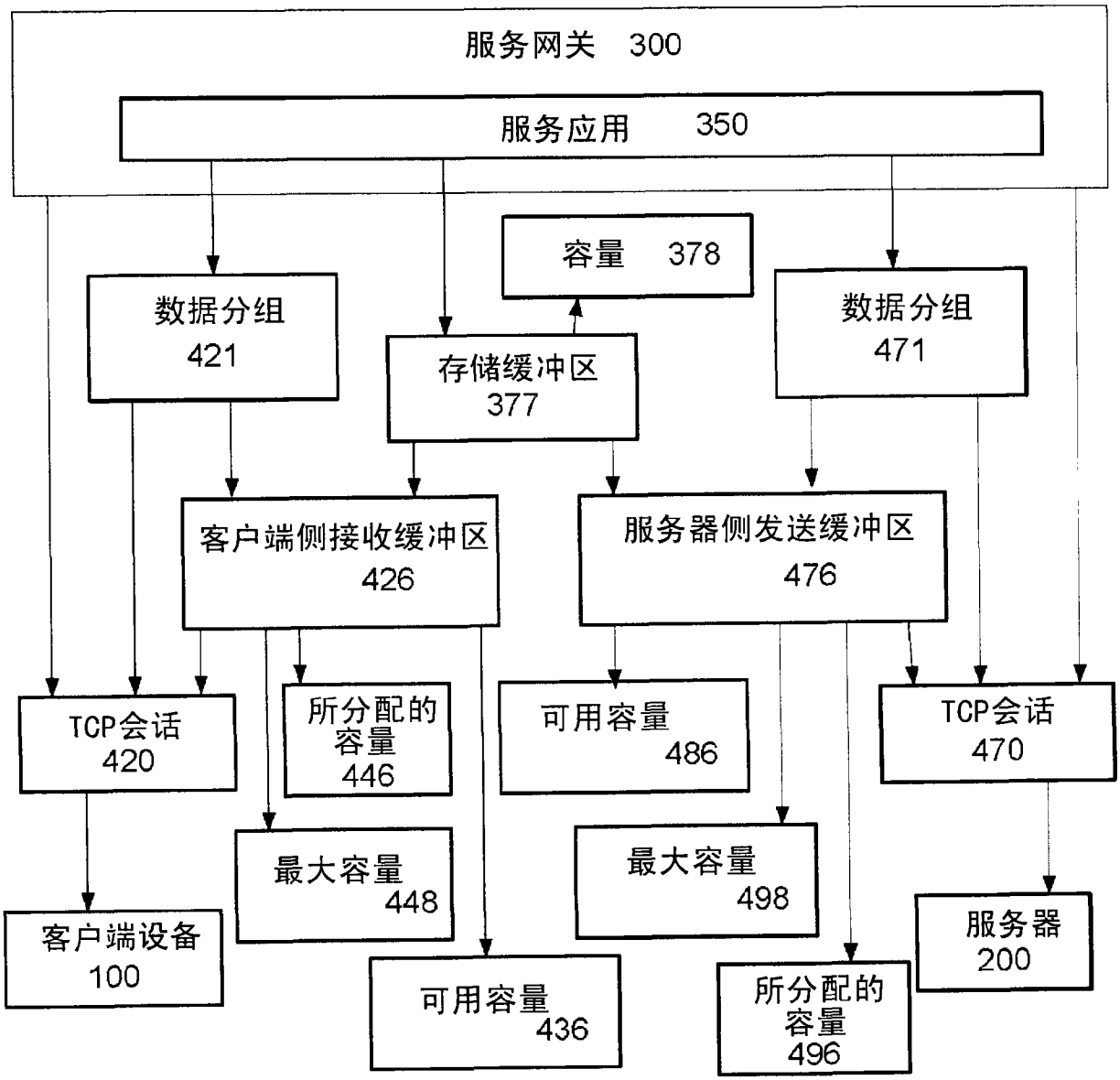

服务网关 300
服务应用 350
容量 378
数据分组 421
存储缓冲区 377
数据分组 471
客户端侧接收缓冲区 426
服务器侧发送缓冲区 476
TCP会话 420
所分配的容量 446
可用容量 486
TCP会话 470
最大容量 448
最大容量 498
客户端设备 100
可用容量 436
所分配的容量 496
服务器 200

图 6

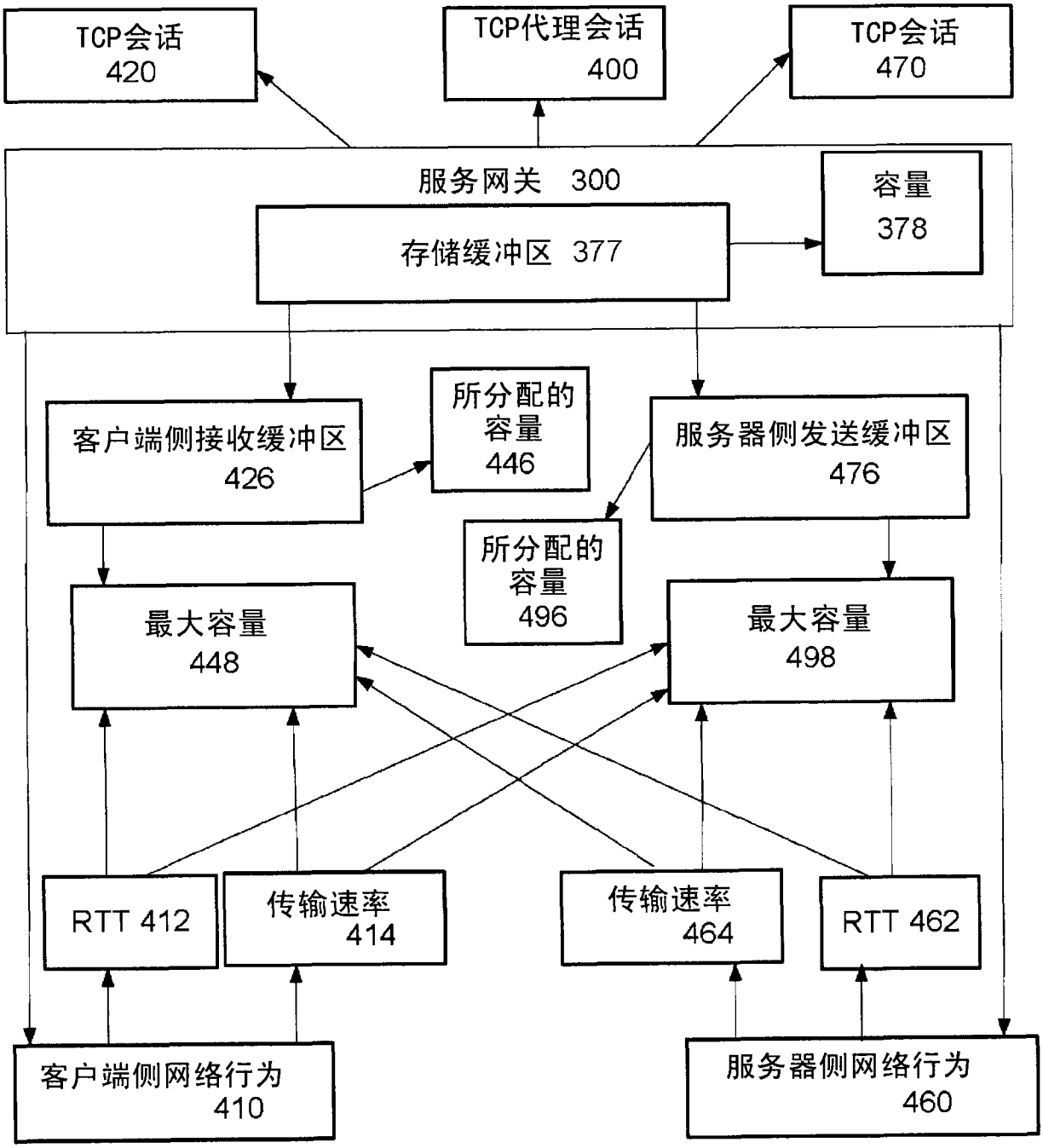

TCP会话
420
TCP代理会话
400
TCP会话
470
服务网关 300
存储缓冲区 377
容量
378
客户端侧接收缓冲区
426
所分配的
容量
446
服务器侧发送缓冲区
476
所分配的
容量
496
最大容量
448
最大容量
498
RTT 412
传输速率
414
传输速率
464
RTT 462
客户端侧网络行为
410
服务器侧网络行为
460

图 7

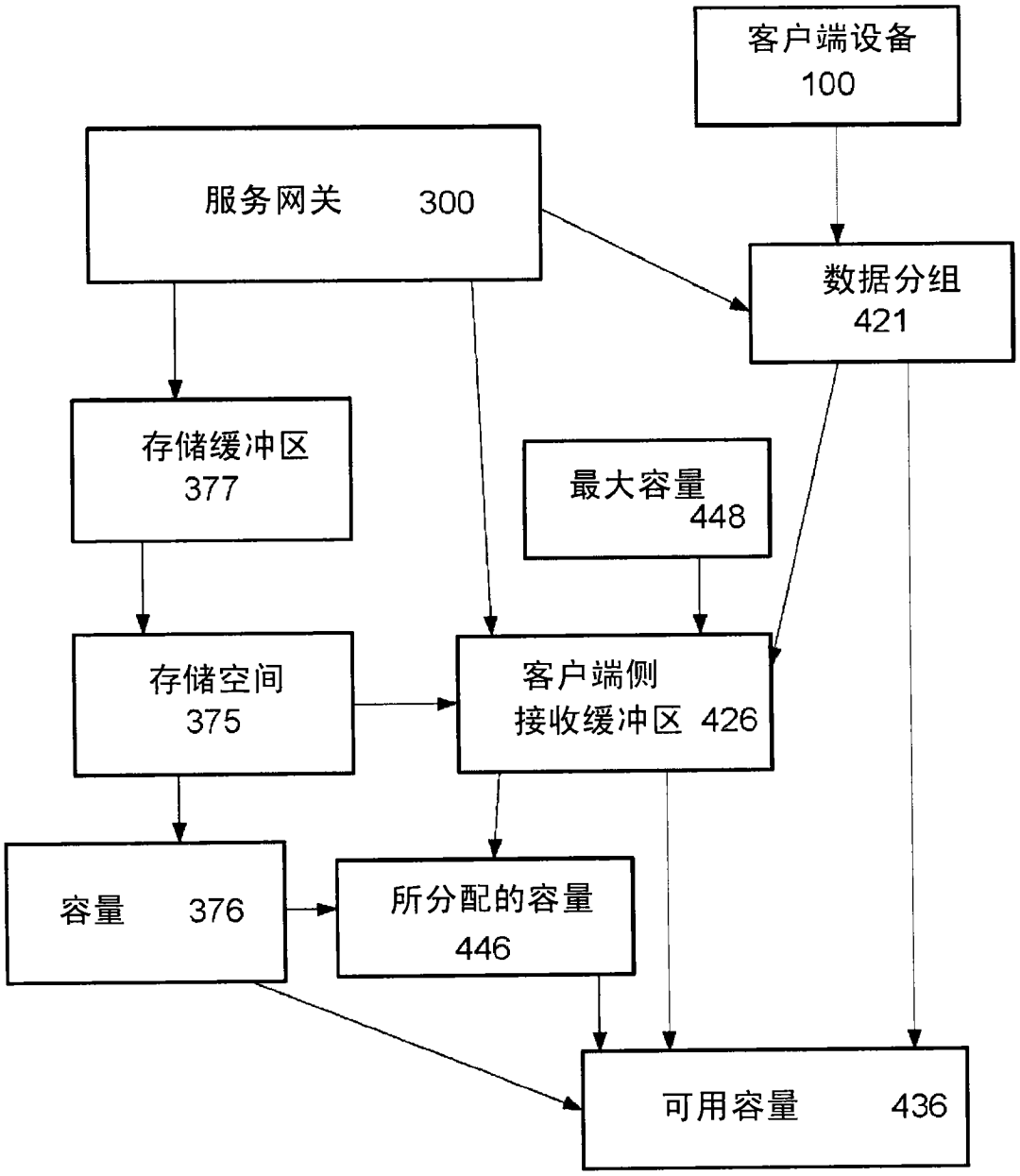

客户端设备
100
服务网关 300
数据分组
421
存储缓冲区
377
最大容量
448
存储空间
375
客户端侧
接收缓冲区 426
容量 376
所分配的容量
446
可用容量 436

图 8a

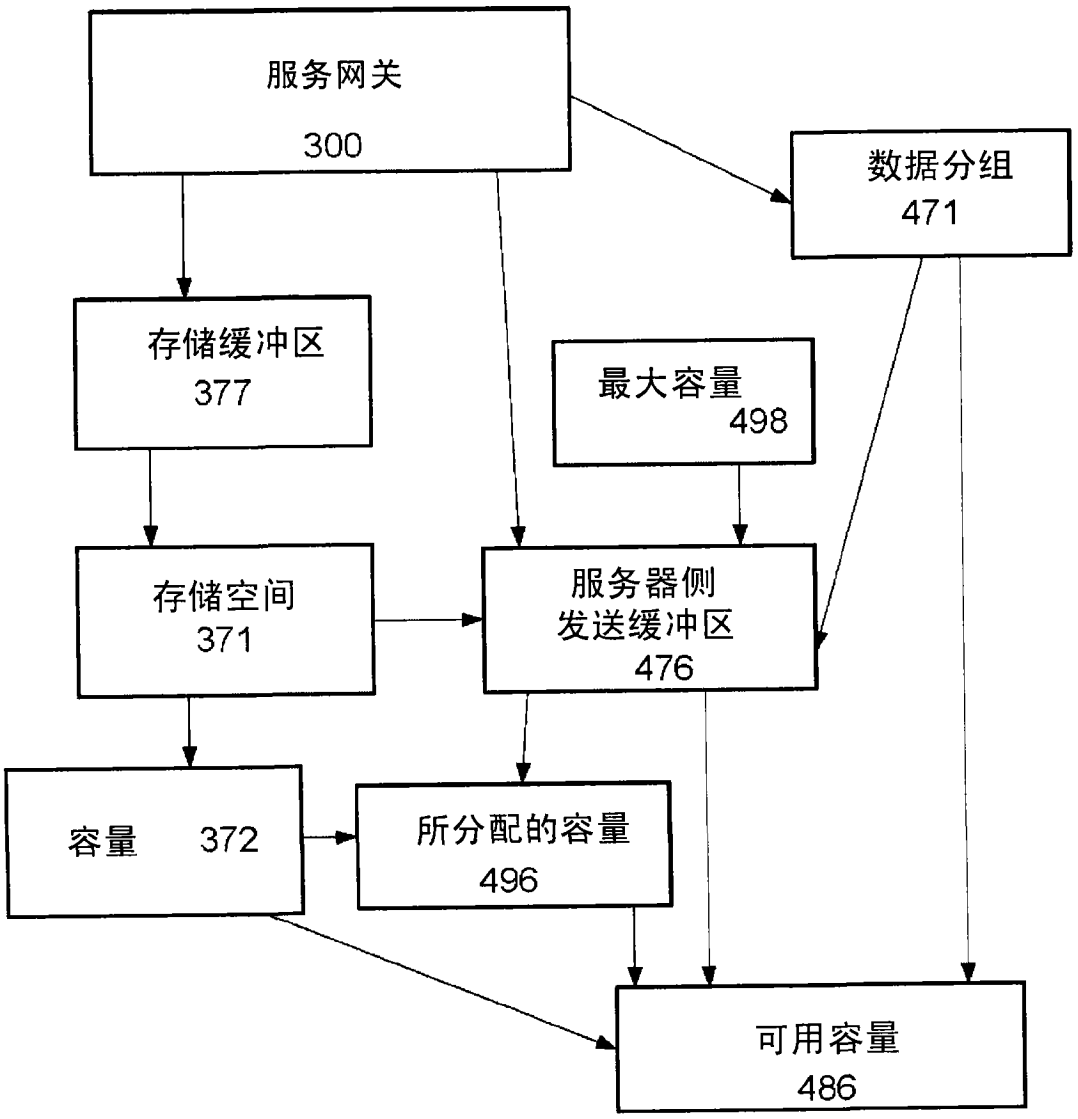

服务网关
300
数据分组
471
存储缓冲区
377
最大容量
498
存储空间
371
服务器侧
发送缓冲区
476
容量 372
所分配的容量
496
可用容量
486

图 8b