(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114095434 B
(45)授权公告日 2024.09.24

(21)申请号 202010742582.6
(22)申请日 2020.07.29
(65)同一申请的已公布的文献号
申请公布号 CN 114095434 A
(43)申请公布日 2022.02.25
(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
(72)发明人 徐聪 张海波 袁庭球
(74)专利代理机构 北京龙双利达知识产权代理有限公司 11329
专利代理师 陈洪艳 王君
(51)Int.Cl.
H04L 47/12(2022.01)
H04L 47/127(2022.01)
H04L 47/32(2022.01)
H04L 47/52(2022.01)
(56)对比文件
CN 103259696 A,2013.08.21
CN 110166366 A,2019.08.23
WO 2016128931 A1,2016.08.18
审查员 张巍

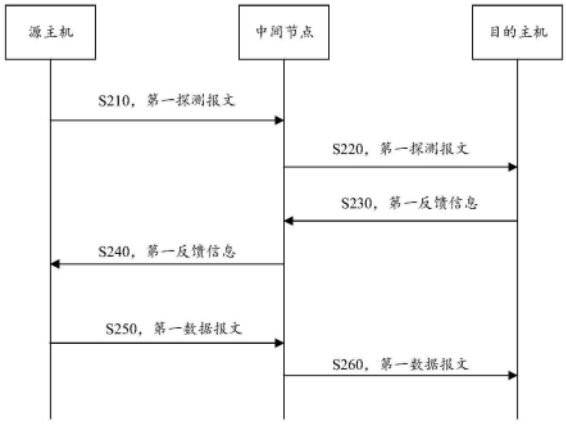
权利要求书3页 说明书19页 附图4页

(54)发明名称

控制网络拥塞的方法和相关装置

(57)摘要

本申请提供了通信领域的控制网络拥塞的方法和相关装置。本申请提出的技术方案中,在源设备向目的设备发送数据报文之前,先发送探测报文,通过中间网络设备对探测报文的传输和目的设备对探测报文的反馈,探测源设备至目的设备的传输路径的可用带宽。这样,源设备基于探测报文探测到的可用带宽来发送数据报文,有助于实现数据报文的无拥塞和无丢包传输。进一步地,本申请的技术方案中,利用探测报文不怕丢包的特性,使用相对激进的速度来传输探测报文,可以使得数据报文能够满吞吐传输。此外,本申请的技术方案中提出了重载模式和轻载模式以及各个模式下的拥塞控制方法,以及合适选择轻载模式和重载模式的切换点,可以实现传输速率的快速公平收敛。



CN 114095434 B

1.一种控制网络拥塞的方法,其特征在于,所述方法包括:

第一网络设备基于第三速率发送第一探测报文,所述第一探测报文是指源设备为所述第一网络设备且目的设备为第二网络设备的探测报文,所述探测报文用于探测源设备至目的设备之间的传输路径上的可用带宽;

所述第一网络设备接收第一反馈信息,所述第一反馈信息用于指示所述第一探测报文在所述第二网络设备上的接收情况;

所述第一网络设备基于所述第一反馈信息确定第二速率,其中,所述第二速率指示所述第二网络设备接收所述第一探测报文的速率;所述第二速率小于或等于预设的速率阈值的情况下,所述第一网络设备使用预设的目标速率发送所述第一探测报文,所述目标速率大于所述第二速率,所述第二速率大于所述速率阈值的情况下,所述第一网络设备使用所述第二速率发送所述第一探测报文;

所述第一网络设备基于所述第一反馈信息,发送第一数据报文,所述第一数据报文是指源设备为所述第一网络设备且目的设备为所述第二网络设备的数据报文。

2.如权利要求1所述的方法,其特征在于,所述第一网络设备基于所述第一反馈信息,发送第一数据报文,包括:

所述第一网络设备基于所述第一反馈信息,通过第三网络设备向所述第二网络设备发送所述第一数据报文,所述第三网络设备用于使用第一比例的带宽发送探测报文以及用于使用第二比例的带宽发送数据报文,其中,所述第一网络设备发送所述第一数据报文的第一速率与所述第二网络设备接收所述第一探测报文的第二速率的比值,等于所述第二比例与所述第一比例的比值。

3.如权利要求2所述的方法,其特征在于,所述第二网络设备每接收到一个所述第一探测报文,则发送一个所述第一反馈信息;

其中,所述第一网络设备基于所述第一反馈信息,发送第一数据报文,包括:

所述第一网络设备每接收到一个所述第一反馈信息,则发送一个所述第一数据报文。

4.如权利要求1至3中任一项所述的方法,其特征在于,所述第一网络设备基于所述第一反馈信息确定第二速率,包括:

所述第一网络设备基于所述第一反馈信息获取所述第一网络设备使用所述第三速率发送所述第一探测报文时的第一丢包率;

所述第一网络设备根据所述第一丢包率确定所述第二速率。

5.如权利要求1至3中任一项所述的方法,其特征在于,所述第一网络设备为多个网络设备中的一个,其中,所述多个网络设备中所有网络设备发送探测报文所使用的目标速率相同。

6.如权利要求1至3中任一项所述的方法,其特征在于,所述目标速率与所述速率阈值相同。

7.如权利要求2或3所述的方法,其特征在于,所述第三网络设备转发的探测报文仅包含所述第一探测报文时,所述第一网络设备使用所述目标速率在指定时长内能够将指定数量的第一探测报文从所述第一网络设备成功传输至所述第二网络设备。

8.如权利要求4所述的方法,其特征在于,所述第一网络设备根据所述第一丢包率确定所述第二速率,包括:

所述第一丢包率小于预设的丢包率阈值的情况下,所述第一网络设备基于所述第一丢包率和所述第三速率对所述第一探测报文的发送速率进行调整,得到所述第二速率。

9.如权利要求8所述的方法,其特征在于,所述丢包率阈值大于所述第一网络设备使用所述第二速率发送所述第一探测报文时所述第一探测报文的丢包率。

10.如权利要求4所述的方法,其特征在于,所述第二速率是使用结构-行为-绩效SCP策略或快速通道策略,基于所述第一丢包率和所述第三速率得到的。

11.如权利要求1至3中任一项所述的方法,其特征在于,所述第一反馈信息具体用于指示所述第二网络设备已接收到所述第一探测报文。

12.一种控制网络拥塞的装置,其特征在于,包括:

发送模块,用于基于第三速率发送第一探测报文,所述第一探测报文是指源设备为所述装置所属的第一网络设备且目的设备为第二网络设备的探测报文,所述探测报文用于探测源设备至目的设备之间的传输路径上的可用带宽;

接收模块,用于接收第一反馈信息,所述第一反馈信息用于指示所述第一探测报文在所述第二网络设备上的接收情况;

所述发送模块,用于基于所述第一反馈信息确定第二速率,其中,所述第二速率指示所述第二网络设备接收所述第一探测报文的速率;所述第二速率小于或等于预设的速率阈值的情况下,所述发送模块,还用于使用预设的目标速率发送所述第一探测报文,所述目标速率大于所述第二速率,所述第二速率大于所述速率阈值的情况下,所述发送模块,还用于使用所述第二速率发送所述第一探测报文;

所述发送模块还用于基于所述第一反馈信息,发送第一数据报文,所述第一数据报文是指源设备为所述第一网络设备且目的设备为所述第二网络设备的数据报文。

13.如权利要求12所述的装置,其特征在于,所述发送模块具体用于基于第三网络设备向所述第二网络设备发送所述第一数据报文,所述第三网络设备用于使用第一比例的带宽发送探测报文以及用于使用第二比例的带宽发送数据报文,其中,所述第一网络设备发送所述第一数据报文的第一速率与所述第二网络设备接收所述第一探测报文的所述第二速率的比值,等于所述第二比例与所述第一比例的比值。

14.如权利要求13所述的装置,其特征在于,所述第二网络设备用于:每接收到一个所述第一探测报文,则发送一个所述第一反馈信息;

其中,所述发送模块具体用于:每接收到一个所述第一反馈信息,则发送一个所述第一数据报文。

15.如权利要求12至14中任一项所述的装置,其特征在于,所述发送模块具体用于:

基于所述第一反馈信息获取所述第一网络设备使用所述第三速率发送所述第一探测报文时的第一丢包率;

根据所述第一丢包率确定所述第二速率。

16.如权利要求12至14中任一项所述的装置,其特征在于,所述第一网络设备为多个网络设备中的一个,其中,所述多个网络设备中所有网络设备发送探测报文所使用的目标速率相同。

17.如权利要求12至14中任一项所述的装置,其特征在于,所述目标速率与所述速率阈值相同。

18.如权利要求13或14所述的装置,其特征在于,所述第三网络设备转发的探测报文仅包含所述第一探测报文时,所述第一网络设备使用所述目标速率在指定时长内能够将指定数量的第一探测报文从所述第一网络设备成功传输至所述第二网络设备。

19.如权利要求15所述的装置,其特征在于,所述发送模块具体用于:

所述第一丢包率小于预设的丢包率阈值的情况下,基于所述第一丢包率和所述第三速率对所述第一探测报文的发送速率进行调整,得到所述第二速率。

20.如权利要求19所述的装置,其特征在于,所述丢包率阈值大于所述第一网络设备使用所述第二速率发送所述第一探测报文时所述第一探测报文的丢包率。

21.如权利要求15所述的装置,其特征在于,所述第二速率是使用结构-行为-绩效SCP策略或快速通道策略,基于所述第一丢包率和所述第三速率得到的。

22.如权利要求12至14中任一项所述的装置,其特征在于,所述第一反馈信息具体用于指示所述第二网络设备已接收到所述第一探测报文。

23.一种控制网络拥塞的装置,其特征在于,包括:处理器,所述处理器与存储器耦合;

所述存储器用于存储指令;

所述处理器用于执行所述存储器中存储的指令,以使得所述网络设备实现如权利要求1至11中任一项所述的方法。

24.一种计算机可读介质,其特征在于,包括指令,当所述指令在处理器上运行时,使得所述处理器实现如权利要求1至11中任一项所述的方法。

控制网络拥塞的方法和相关装置

技术领域

[0001] 本申请涉及通信领域,并且更具体地,涉及通信网络中控制网络拥塞的方法和相关装置。

背景技术

[0002] 通信网络中传输数据报文时,通常会出现网络拥塞的问题。基于网络拥塞问题,通信领域提出了拥塞控制方法。

[0003] 传统的拥塞控制方法是基于丢包等拥塞信号来检测通信网络的拥塞状况,并基于通信网络的拥塞状况从数据报文的源端来控制数据报文的发包速率,以控制通信网络的拥塞状况。具体地,当检测到数据报文未丢包时,增加数据包的发送速率;当检测到数据报文丢包时,降低数据报文的发包速率,以缓解网络拥塞。

[0004] 传统的拥塞控制方法虽然在一定程度上可以缓解通信网络拥塞状况,但是该拥塞控制方法是一种“后知后觉”的机制,当源端接收到丢包信号时,数据报文已经发生了丢包,即已经造成了性能的损失,后续的降速调整已经滞后于网络拥塞的时间点。因此,此机制下,通信网络出现拥塞成为常态,通信设备出现包的积压也成为常态,而包的积压会严重恶化数据报文传输的端到端时延,最终影响通信网络传输数据报文的性能。

发明内容

[0005] 本申请提供了控制网络拥塞的方法和相关装置。本申请的技术方案可以通过不怕丢包的探测报文来探测网络的带宽情况,并基于探测到的网络带宽来控制数据报文的传输速率。

[0006] 第一方面,本申请提供了一种控制网络拥塞的方法。所述方法包括:第一网络设备发送第一探测报文,所述第一探测报文是指源设备为所述第一网络设备且目的设备为第二网络设备的探测报文,所述探测报文用于探测源设备至目的设备之间的传输路径上的可用带宽;所述第一网络设备接收第一反馈信息,所述第一反馈信息用于指示所述第一探测报文在所述第二网络设备上的接收情况;所述第一网络设备基于所述第一反馈信息,发送第一数据报文,所述第一数据报文是指源设备为所述第一网络设备且目的设备为所述第二网络设备的数据报文。

[0007] 该方法中,因为第一网络设备在发送数据报文之前先基于探测报文探测网络了当前可以用于传输第一网络设备至第二网络设备的数据报文的带宽,并基于该探测到的可用带宽来向第二网络设备发送数据报文,所以,有助于合理利用网络带宽资源来向第二网络设备发送数据包括,从而实现网络的拥塞控制。

[0008] 在一些可能的实现方式中,所述第一网络设备基于所述第一反馈信息,发送第一数据报文,包括:

[0009] 所述第一网络设备基于所述第一反馈信息,通过第三网络设备向所述第二网络设备发送所述第一数据报文,所述第三网络设备用于使用第一比例的带宽发送探测报文以及

用于使用第二比例的带宽发送数据报文,其中,所述第一网络设备发送所述第一数据报文的第一速率与所述第二网络设备接收所述第一探测报文的第二速率的比值,等于所述第二比例与所述第一比例的比值。

[0010] 这些实现方式中,第三网络设备预设使用第一比例的带宽发送探测报文,预设使用第二比例的带宽发送数据报文。这样,第一网络设备接收到第一反馈信息之后,可以根据该第一反馈信息的接收速率获知第二网络设备接收第一探测报文的速率,并根据探测报文在第一网络设备上的接收速率以及探测报文占用的带宽比例以及数据报文能够占用的带宽比例控制此时发送数据报文的发送速率,即第二速率。

[0011] 其中,第一网络设备基于第一探测反馈信息发送第一数据报文,且第一数据报文的速率刚好为第二速率,可以使得第三网络设备刚好能够满吞吐、且无丢包地向第二网络设备转发第一网络设备发送的第一数据报文。

[0012] 在一些可能的实现方式中,所述第二网络设备用于:每接收到一个所述第一探测报文则发送一个所述第一反馈信息。其中,所述第一网络设备基于所述第一反馈信息,发送第一数据报文,包括:所述第一网络设备每接收到一个所述第一反馈信息,则发送一个所述第一数据报文。

[0013] 这些实现方式中,第二网络设备每接收一个第一探测报文则发送一个第一反馈信息,第一网络设备每接收一个第一反馈信息则发送一个第一数据报文,可以使得第一数据报文的速率刚好为第二速率,从而可以使得第一数据报文可以无丢包且满吞吐地传输至第二网络设备。

[0014] 在一些可能的实现方式中,所述第一网络设备发送第一探测报文,包括:所述第一网络设备使用第三速率发送所述第一探测报文;所述第一网络设备获取所述第一网络设备使用所述第三速率发送所述第一探测报文时的第一丢包率;所述第一网络设备根据所述第一丢包率确定所述第二速率;所述第一网络设备基于所述第二速率发送所述第一探测报文。

[0015] 也就是说,这些实现方式中,根据之前发送第一探测报文所使用的第三速率以及之前使用第三速率发送的第一探测报文时的丢包率来获知当前网络将第一网络设备发送的探测报文传输给第二网络设备的可用带宽所对应的探测报文发送速率,即第二速率,并使用第二速率重新发送第一探测报文。这样可以及时调整当前探测报文的发送速率,以实现探测报文可以满吞吐地传输至第二网络设备,从而可以实现第一数据报文可以满吞吐地传输至第二网络设备。

[0016] 在一些可能的实现方式中,所述第一网络设备基于所述第二速率发送所述第一探测报文,包括:所述第二速率小于或等于预设的速率阈值的情况下,所述第一网络设备使用预设的目标速率发送所述第一探测报文,所述目标速率大于所述第二速率;或所述第二速率大于所述速率阈值的情况下,所述第一网络设备使用所述第二速率发送所述第一探测报文。

[0017] 即第二速率较小的情况下,判断网络负载较重,可用带宽较少。此时,可以使用预设的固定目标速率来发送第一探测报文,以使得探测报文不断流,从而使得数据报文不断流。

[0018] 在第二速率较大的情况下,判断网络负载较轻,可以带宽较多。此时,可以使用调

整得到的第二速率来发送第一探测报文,以充分利用网络资源发送第一探测报文,以探测数据报文的最合理可用带宽,从而可以充分利用网络资源发送数据报文,实现第一数据报文的满吞吐和无丢包传输。

[0019] 在一些实现方式中,所述第一网络设备为多个网络设备中的一个,其中,所述多个网络设备中所有网络设备的探测报文的目标速率相同。

[0020] 每个网络设备在重载模式下使用的目标速率相同,这可以实现多个网络设备的探测报文的公平收敛。

[0021] 在一些可能的实现方式中,所述目标速率与所述速率阈值相同。这样可以使得从第三速率切换到目标速率时,探测报文在通信网络中的传输速率能够快速收敛。

[0022] 在一些可能的实现方式中,所述第一网络设备至所述第二网络设备的传输路径上的第三网络设备转发的探测报文仅包含所述第一探测报文时,所述第一网络设备使用所述目标速率在指定时长内能够将指定数量的第一探测报文从所述第一网络设备成功传输至所述第二网络设备。

[0023] 这些实现方式中,第一探测报文的目的速率相对激进,即相对来说比较大,因此可以充分利用网络资源发送第一探测报文,以探测数据报文的可用带宽,从而可以充分利用网络资源发送第一数据报文,实现第一数据报文的满吞吐和无丢包传输。

[0024] 在一些可能的实现方式中,所述第一网络设备根据所述第一丢包率确定所述第二速率,包括:所述第一丢包率小于预设的丢包率阈值的情况下,所述第一网络设备基于所述第一丢包率和所述第三速率对所述第一探测报文的发送速率进行调整,得到所述第二速率。其中,所述第一网络设备基于所述第二速率发送所述第一探测报文,包括:所述第一网络设备使用所述第二速率发送所述第一探测报文。

[0025] 即若丢包率小,则说明进入轻载模式。此时,可以根据丢包率和之前的第三速率调整得到更合适的第二速率。这样可以充分利用资源传输探测报文,从而可以充分利用资源传输数据报文。

[0026] 在一些可能的实现方式中,所述丢包率阈值大于所述第一网络设备使用所述第二速率发送所述第一探测报文时所述第一探测报文的丢包率。

[0027] 因为丢包率阈值大于第二速率时的丢包率,因此,第一网络设备从重载模式切换至轻载模式时,会进行速度调整,从而不会破坏原有多个源设备的发送速率之间的公平性。

[0028] 在一些可能的实现方式中,所述第二速率是使用结构-行为-绩效SCP策略或快速通道策略,基于所述第一丢包率和所述第三速率得到的。

[0029] 这样可以快速公平地实现各个发送设备的探测报文的发送速率的收敛,从而实现各个网络设备的数据报文的发送速率快速公平收敛。

[0030] 第二方面,本申请提供一种网络拥塞的控制方法。所述方法包括:第二网络设备接收第一探测报文,所述第一探测报文是指源设备为第一网络设备且目的设备为所述第二网络设备的探测报文,所述探测报文用于探测源设备至目的设备之间的传输路径上的可用带宽;所述第二网络设备发送第一反馈信息,所述第一反馈信息用于指示所述第二网络设备对所述第一探测报文的接收情况;所述第二网络设备接收所述第一网络设备基于所述第一反馈信息发送的第一数据报文,所述第一数据报文是指源设备为所述第一网络设备且目的设备为所述第二网络设备的数据报文。

[0031] 该方法中,第二网络设备在接收到探测报文之后,向发送该探测报文的源设备反馈该探测报文的接收情况,以便于该源设备可以根据该接收情况获知网络的可用带宽,进而使得该源设备可以基于该可用带宽发送数据报文。

[0032] 在一些可能的实现方式中,所述第二网络设备发送第一反馈信息,包括:所述第二网络设备每接收到一个所述第一探测报文,则发送一个所述第一反馈信息。

[0033] 在另一些可能的实现方式中,所述第一反馈信息具体用于指示所述第二网络设备接收到所述第一探测报文。

[0034] 第三方面,本申请提供一种网络拥塞的控制方法。所述方法包括:第三网络设备接收探测报文,所述探测报文用于探测源设备至目的设备之间的传输路径上的可用带宽;所述第三网络设备转发所述探测报文;所述第三网络设备接收反馈信息,所述反馈信息用于表示所述的目的设备接收所述探测报文的情况;所述第三设备转发所述反馈信息;所述第三设备接收所述源设备基于所述反馈信息发送的数据报文;所述第三设备转发所述数据报文。

[0035] 该方法中,第三网络设备作为发送探测报文和数据报文的源设备与接收探测报文和数据报文的目的设备之间的中间设备,辅助发送探测报文,以探测能够用于发送数据报文的可用带宽,以使得源设备可以基于该可用带宽控制数据报文的发送,从而可以控制数据报文的拥塞情况。

[0036] 在一些可能的实现方式中,所述第三网络设备转发探测报文,包括:所述第三网络设备使用第一比例的带宽发送所述探测报文;所述第三网络设备转发所述数据报文,包括:所述第三网络设备使用第二比例的带宽发送所述数据报文;其中,所述第三网络设备接收所述第一数据报文的速率与所述第三网络设备接收所述第一探测报文的速率的比值,等于所述第二比例与所述第一比例的比值。

[0037] 该实现方式中,第三网络设备使用预设的比例发送探测报文,这样,发送探测报文的源设备即可以基于该探测报文在目的设备的接收情况来获知实际用于发送探测报文的带宽,进而可以基于该带宽以及预设的第一比例和第二比例获知能够用于发送数据报文的可用带宽,最终可以基于该可用带宽实现数据报文的无丢包和满吞吐传输。

[0038] 在一些可能的实现方式中,所述第三网络设备转发所述反馈信息,包括:所述第三网络设备以最高优先级转发所述反馈信息。

[0039] 该实现方式中,第三网络设备优先传输反馈信息,即优先反馈探测报文的接收情况,从而可以提高发送探测报文的源设备获知探测报文的接收情况的成功率和准确率,进而可以提高源设备获知用于传输数据报文的可用带宽的成功率和准确率,最终可以提高网络拥塞的控制成功率和准确率。

[0040] 第四方面,本申请提供了一种网络拥塞的控制装置,该装置包括用于执行上述第一方面或其中任意一种实现方式中的方法的模块。

[0041] 第五方面,本申请提供了一种网络拥塞的控制装置,该装置包括用于执行上述第二方面或其中任意一种实现方式中的方法的模块。

[0042] 第六方面,本申请提供了一种网络拥塞的控制装置,该装置包括用于执行上述第三方面或其中任意一种实现方式中的方法的模块。

[0043] 第七方面,提供了一种网络拥塞的控制装置,该装置包括:存储器、处理器和收发

器;所述存储器用于存储程序;所述处理器用于执行所述存储器存储的程序;所述收发器用于与其他装置或设备进行通信;当所述存储器存储的程序被执行时,所述处理器和所述收发器用于执行第一方面或者其中任意一种实现方式中的方法。

[0044] 第八方面,提供了一种网络拥塞的控制装置,该装置包括:存储器、处理器和收发器;所述存储器用于存储程序;所述处理器用于执行所述存储器存储的程序;所述收发器用于与其他装置或设备进行通信;当所述存储器存储的程序被执行时,所述处理器和所述收发器用于执行第二方面或者其中任意一种实现方式中的方法。

[0045] 第九方面,提供了一种网络拥塞的控制装置,该装置包括:存储器、处理器和收发器;所述存储器用于存储程序;所述处理器用于执行所述存储器存储的程序;所述收发器用于与其他装置或设备进行通信;当所述存储器存储的程序被执行时,所述处理器和所述收发器用于执行第三方面或者其中任意一种实现方式中的方法。

[0046] 第十方面,提供一种计算机可读介质,该计算机可读介质存储用于设备执行的程序代码,该程序代码用于执行第一方面或其中任意一种实现方式中的方法。

[0047] 第十一方面,提供一种计算机可读介质,该计算机可读介质存储用于设备执行的程序代码,该程序代码用于执行第二方面或其中任意一种实现方式中的方法。

[0048] 第十二方面,提供一种计算机可读介质,该计算机可读介质存储用于设备执行的程序代码,该程序代码用于执行第三方面或其中任意一种实现方式中的方法。

[0049] 第十三方面,提供一种包含指令的计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述第一方面或其中任意一种实现方式中的方法。

[0050] 第十四方面,提供一种包含指令的计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述第二方面或其中任意一种实现方式中的方法。

[0051] 第十五方面,提供一种包含指令的计算机程序产品,当该计算机程序产品在计算机上运行时,使得计算机执行上述第三方面或其中任意一种实现方式中的方法。

[0052] 第十六方面,提供一种芯片,所述芯片包括处理器与数据接口,所述处理器通过所述数据接口读取存储器上存储的指令,执行上述第一方面或其中任意一种实现方式中的方法。

[0053] 可选地,作为一种实现方式,所述芯片还可以包括存储器,所述存储器中存储有指令,所述处理器用于执行所述存储器上存储的指令,当所述指令被执行时,所述处理器用于执行第一方面或其中任意一种实现方式中的方法。

[0054] 第十七方面,提供一种芯片,所述芯片包括处理器与数据接口,所述处理器通过所述数据接口读取存储器上存储的指令,执行上述第二方面或其中任意一种实现方式中的方法。

[0055] 可选地,作为一种实现方式,所述芯片还可以包括存储器,所述存储器中存储有指令,所述处理器用于执行所述存储器上存储的指令,当所述指令被执行时,所述处理器用于执行第二方面或其中任意一种实现方式中的方法。

[0056] 第十八方面,提供一种芯片,所述芯片包括处理器与数据接口,所述处理器通过所述数据接口读取存储器上存储的指令,执行上述第三方面或其中任意一种实现方式中的方法。

[0057] 可选地,作为一种实现方式,所述芯片还可以包括存储器,所述存储器中存储有指

令,所述处理器用于执行所述存储器上存储的指令,当所述指令被执行时,所述处理器用于执行第三方面或其中任意一种实现方式中的方法。

[0058] 第十九方面,提供了一种计算设备,该计算设备包括:存储器,用于存储程序;处理器,用于执行所述存储器存储的程序,当所述存储器存储的程序被执行时,所述处理器用于执行第一方面或者其中任意一种实现方式中的方法。

[0059] 第二十方面,提供了一种计算设备,该计算设备包括:存储器,用于存储程序;处理器,用于执行所述存储器存储的程序,当所述存储器存储的程序被执行时,所述处理器用于执行第二方面或者其中任意一种实现方式中的方法。

[0060] 第二十一方面,提供了一种计算设备,该计算设备包括:存储器,用于存储程序;处理器,用于执行所述存储器存储的程序,当所述存储器存储的程序被执行时,所述处理器用于执行第三方面或者其中任意一种实现方式中的方法。

附图说明

[0061] 图1是本申请的技术方案的一种应用场景的示例图。

[0062] 图2是本申请的控制方法的一个示例性交互流程图。

[0063] 图3是本申请的控制方法中按比例带宽传输探测报文和数据报文的一个示例图。

[0064] 图4是本申请的控制方法的另一个示例性流程图。

[0065] 图5是本申请的控制方法的又一个示例性流程图。

[0066] 图6是本申请的控制方法的再一个示例性流程图。

[0067] 图7是本申请的控制装置的一个示例性结构图。

[0068] 图8是本申请的控制装置的另一个示例性结构图。

[0069] 图9是本申请的计算机程序产品的一个示例性结构图。

具体实施方式

[0070] 下面将结合附图,对本申请中的技术方案进行描述。

[0071] 本申请实施例的技术方案可以应用于各种通信网络中,例如广域网(wide area network,WAN)、城域网(metropolitan area network,MAN)、局域网(local area network LAN)或数据中心网络等。

[0072] 本申请实施例中的源设备也可以称为源主机、源端、源节点或源侧等,是指数据的发送端。进一步地,发送数据的源设备可以理解为该数据的源互联网协议(internet protocol,IP)地址所指示的网络设备。本申请的实施例中,将发送数据的源设备简称为该探测报文或该数据报文的源设备。

[0073] 本申请实施例中的目的设备也可以称为目的主机、目的端、目的节点或目的侧等,是指数据的最终接收端,即数据的目的地。进一步地,接收数据的目的设备可以理解为该数据的目的IP地址所指示的网络设备。本申请的实施例中,将接收数据的目的设备简称为该数据的目的设备。

[0074] 本申请实施例中的数据可以包括数据报文和探测报文。其中,数据报文的含义可以参考现有技术中的数据报文的定义,此处不再赘述。

[0075] 本实施例中的探测报文用于探测该探测报文的源设备至目的设备间的传输路径

的负载情况。传输路径的负载情况可以包括传输路径的可用带宽情况,例如传输路径的可用带宽为多少。或者可以理解为,探测报文用于探测或预约下一传输周期能够用于传输数据报文的带宽。

[0076] 探测报文的长度可以是预先设定好的,例如一个探测报文可以封装为64个字节(byte,B)的以太帧。通常情况下,探测报文可以仅包含包头,且该包头中包含的内容与后续步骤中发送的数据报文的包头中的内容可以相同,例如为相同的五元组信息。

[0077] 本实施例中的中间设备也可以称为中间节点、转发设备、中转设备、中转节点、交换机、路由设备、路由器或路由节点,是指能够将源设备发送的数据转发至目的设备的通信设备。

[0078] 图1是可以应用本申请实施例的技术方案的通信网络的示意性架构图。如图1所示,可以应用本申请实施例的技术方案的通信网络中可以包括源主机110、中间节点120和目的主机130。

[0079] 可以理解的是,图1中仅示例性给出一个源主机、一个中间节点和一个目的主机。可以应用本申请实施例的技术方案的通信网络中可以包括更多的源主机、更多的中间节点和更多的目的主机。

[0080] 图1所示的通信网络中,源主机110作为数据报文的发送端,按照一定的速率发送数据报文。该数据报文到达中间节点120之后,中间节点120利用可用带宽向目的主机130发送数据报文。

[0081] 其中,若数据报文的发送速率过快,中间节点120来不及将接收的数据包转发出去时,这些来不及转发出去的数据包就会积压在中间节点120。中间节点120的缓存中的数据包积压到一定程度时,中间节点120会将接收到且来不及转发出去的数据包丢掉,从而出现数据报文的丢包现象。此时,可以认为网络出现拥塞。

[0082] 为了避免丢包,即为了避免拥塞,本申请提出了新的技术方案,以控制通信网络的拥塞。图2为本申请一个实施例的网络拥塞的控制方法的示例性流程图。该方法可以包括S210、S220、S230、S240、S250和S260。

[0083] S210,第一网络设备发送第一探测报文。其中,所述第一网络设备为所述第一探测报文的源设备,例如,所述第一探测报文中的源IP地址为所述第一网络设备的IP地址。

[0084] 该第一网络设备可以是通信网络中的任意源设备,该第一探测报文的大小可以是预设好的。

[0085] 本实施例的一些可能的实现方式中,第一网络设备可以等间隔地逐包发送第一探测报文。

[0086] 本实施例中,第一网络设备发送可以探测报文可以理解为第一网络设备向第二网络设备发送第一探测报文,其中第二网络设备为第一探测报文的目的设备,例如,第一探测报文中的目的IP地址所指示的网络设备即为第二网络设备。

[0087] 本实施例中,第一网络设备发送第一探测报文之后,第一网络设备至第二网络设备的传输路径上的第三网络设备可以接收到所述第一探测报文。

[0088] S220,第三网络设备转发所述第一探测报文。

[0089] 在一些示例中,第三网络设备可以接收一个或多个网络设备向另外一个或多个网络设备发送的探测报文,并转发这些探测报文,这些探测报文中可以包括第一网络设备向

第二网络设备发送的第一探测报文。

[0090] 第一网络设备至第二网络设备的传输路径上可以包含一个或多个第三网络设备。若第一网络设备至第二网络设备的传输路径上包含一个第三网络设备，则该第三网络设备发送探测报文之后，第二网络设备接收该探测报文中的第一探测报文；若第一网络设备至第二网络设备的传输路径上包含多个第三网络设备，则每个第三网络设备接收到上一跳设备发送的探测报文之后，向下一条设备转发探测报文，该探测报文中包含第一探测报文，直到第二网络设备接收到第一探测报文。

[0091] 第一探测报文经过一个或多个第三网络设备的转发之后，第二网络设备可以接收该第一探测报文。

[0092] S230，第二网络设备发送第一反馈信息。其中，第一反馈信息用于指示所述第一探测报文在所述第二网络设备上的接收情况。

[0093] 例如，第一反馈信息具体可以用于指示第二网络设备接收到了第一探测报文。又如，第一反馈信息具体可以用于指示第二网络设备接收到第一探测报文的时间。再如，第一反馈信息具体可以用于指示第二网络设备接收到第一探测报文以及第二网络设备接收到第一探测报文的时间。

[0094] 在一些可能的实现方式中，第二网络设备接收到第一探测报文之后，可以将第一探测报文中的目的地址修改为第一探测报文的源地址，并将修改得到的新探测报文发送出去。

[0095] 可选地，本实施例的探测报文中可以包含行程标识，该行程标识可以包括去程标识和回程标识，去程标识用于表示探测报文是从源设备发送给目的设备的，回程标识用于标识探测报文目的设备返回给源设备的。这种情况下，第二网络设备不仅修改第一探测报文的目的地址，还将第一探测报文的行程标识由去程标识修改为回程标识。

[0096] 第二网络设备发送第一反馈信息之后，第二网络设备至第一网络设备的传输路径上的第三网络设备可以接收到该第一反馈信息。

[0097] 本实施例中，第二网络设备可以每接收到一个探测报文，则发送一个该探测报文对应的反馈信息。可以理解的是，这仅是一种示例，第二网络设备也可以使用其他方式来发送第一反馈信息，例如收到多个探测报文之后，在同一个消息中携带在多个探测报文对应的反馈信息。

[0098] S240，第三网络设备发送第一反馈信息。

[0099] 可以理解的是，第一网络设备至第二网络设备的传输路径与第二网络设备至第一网络设备的传输路径可以相同，也可以不同。也就是说，该步骤中的第三网络设备与S220中的第三网络设备可以是相同的设备，也可以是不同的设备。

[0100] 例如，第一网络设备至第二网络设备的传输路径与第二网络设备至第一网络设备的传输路径不同时，该步骤中的第三网络设备与S220中的第三网络设备是不同的设备。

[0101] 在一些实现方式中，第三网络设备可以将反馈信息看作数据报文来发送。也就是说，第三网络设备转发反馈信息的方式与转发数据报文的方式或资源可以相同。

[0102] 第二网络设备至第一网络设备的传输路径上可以包含一个或多个第三网络设备。若第二网络设备至第一网络设备的传输路径上包含一个第三网络设备，则该第三网络设备发送第一反馈信息之后，第一网络设备接收第一反馈信息；若第二网络设备至第一网络设

备的传输路径上包含多个第三网络设备,则每个第三网络设备接收到上一跳设备发送的第一反馈信息之后,向下一跳设备转发第一反馈信息,直到第一网络设备接收到第一反馈信息。

[0103] S250,第一网络设备基于所述第一反馈信息,发送第一数据报文。其中,第一数据报文的源设备为所述第一网络设备,所述第一数据报文的目的设备为所述第二网络设备。

[0104] 通常来说,第一数据报文所经过的路径与第一探测报文所经过的路径是相同的。因此第一网络设备发送第一数据报文之后,将第一数据报文从第一网络设备转发至第二网络设备的第三网络设备,与将第一探测报文从第一网络设备转发至第二网络设备的第三网络设备是一致的。

[0105] S260,第三网络设备接收数据报文,并转发数据报文,该数据报文中包括第一数据报文。

[0106] 通常情况下,该步骤的第三网络设备为S220中的第三网络设备,即第一数据报文所经过的路径与第一探测报文所经过的路径相同。

[0107] 第三网络设备发送数据报文后,第二网络设备可以接收该数据报文中的第一数据报文。

[0108] 本实施例中,通过预先发送探测报文来感知通信网络的网络状况,并指导数据报文的发送,从而使得数据报文通道不会发生拥塞和丢包,即实现了“先知先觉”的无损拥塞控制。

[0109] 本申请的实施例中,可以采用多种方式将探测报文和数据报文的传输隔离开。

[0110] 在一些方式中,第三网络设备可以按照一定的比例在同一物理链路中为用于发送探测报文的探测通道和用于发送数据报文的数据通道分配带宽,以在逻辑上对两类通道进行隔离,从而在逻辑上对探测报文和数据报文进行隔离。例如,该比例可以是探测报文和数据报文各自包长在这两类报文总包长中占的比例。

[0111] 在另一些方式中,源设备可以通过构建独立信令网络的方式来传输探测报文,以实现探测报文和数据报文的物理隔离。

[0112] 采用逻辑隔离的方式传输探测报文和数据报文的一个示例中,第三网络设备可以使用第一比例的带宽发送探测报文,该探测报文中包括所述第一探测报文,并且,第三网络设备使用第二比例的带宽发送数据报文,该数据报文中包括第一数据报文。

[0113] 第一比例和第二比例可以是预先设置好的。也就是说,第三网络设备上使用多大比例的带宽来发送探测报文和使用多大比例的带宽来发送数据报文是预先设置好的。

[0114] 在一些可能的实现方式中,第三网络设备上的第一比例和第二比例可以基于探测报文和数据报文的大小来设置。

[0115] 例如,以最小以太帧84字节构建探测报文,并驱动源设备以最大以太帧1538字节来传输数据报文的场景下,第三网络设备可以为探测报文分配的带宽占该第三网络设备的总带宽的比例约为:84/(84+1538)≈5%,为数据报文分配的带宽占总带宽的比例约为95%。

[0116] 图3为本申请一个实施例按比例发送探测报和数据报文的方法的示意图。如图3所示,源主机310与目的主机320之间的中间节点330可以预留5%的带宽来发送探测报文,并使用其余95%的带宽来发送数据报文。不论中间节点330当前的可用带宽为多少,都仅能使

用可用带宽中的5%来发送探测报文,且仅能使用可用带宽中的95%来发送数据报文。

[0117] 若中间节点330的出端口的带宽为10吉比特每秒(Gbps),则中间节点330使用9.5Gbps的带宽来发送数据报文,使用0.5Gbps的带宽来发送探测报文。其中,若源主机310发送第一探测报文的速率为1Gbps,则说明第一探测报文的传输数量过快(包间隔过小),此时,中间节点330会对出端口的第一探测报文进行限速。

[0118] 具体地,中间节点330会将第一探测报文的传输速率限制在0.5Gbps,即500兆比特每秒(Mbps)。500Mbps的传输速率刚好是第一探测报文的最佳传输速率。目的主机320接收到中间节点330按照500Mbps的传输速率传输的探测报文之后,向源主机310发送第一反馈信息。源主机310接收第一反馈信息之后,可以根据第一反馈信息获知中间节点330的负载情况,即当前网络恰好可以将传输速率为500Mbps的第一探测报文传输至目的主机320。根据第一比例5%与第二比例95%的比例,可知当前网络恰好可以将使用9.5Gbps的速率发送第一数据报文最好,这样可以在不丢包无拥塞的情况下充分利用网络资源。

[0119] 中间节点按照一定比例的带宽来发送探测报文和数据报文的场景下,在一些实现方式中,第三网络设备可以使用第一比例以外的带宽来发送第一反馈信息。例如,第一反馈信息是基于第一探测报文修改得到探测报文的情况下,第三网络设备可以基于该探测报文中的回程标识确定该探测报文为第二网络设备反馈给第一网络设备的,因此使用用于发送数据报文的带宽发送该探测报文。

[0120] 本申请的实施例中,第一网络设备首轮发送第一探测报文时,可以采用相对激进的方式来发包。这样可以充分利用网络资源。

[0121] 通常来说,第一网络设备首轮发送第一探测报文时,第一探测报文的发送速率可以高于第一网络设备上的应用的带宽需求。例如,第一探测报文的发送速率可以初始化为应用带宽需求的1.5倍。

[0122] 本申请的实施例中,在第一探测报文的非首轮发送过程中,第一网络设备可以对第一探测报文的发送速率进行调节,以达到对第一探测报文的发送速率的控制的目的,从而达到对第一数据报文的发送速率的控制的目的,进而达到网络拥塞的控制的目的。

[0123] 在一些调节第一探测报文的发送速率的实现方式中,若将第一网络设备当前发送第一探测报文所使用的速率称为第三速率,则第一网络设备可以获取第一网络设备使用第三速率发送第一探测报文时的第一丢包率,并根据第一丢包率确定第二速率和基于第二速率发送第一探测报文。

[0124] 本申请的实施例中,可以使用结构-行为-绩效(structure-conduct-performance,SCP)策略,基于第三速率和第一丢包率获得第二速率。例如,可以采用SCP策略中的基于在线学习的拥塞控制方法。

[0125] SCP策略中,第一网络设备可以实时基于当前的传输速率r_C和丢包率l计算效用函数$f_{w,b}(r_C,l)$,并根据效用函数的变化方向(增大或减小)来调整第一探测报文的传输速率的变化方向(增大或减小),即增加第三速率以得到第二速率还是减少第三速率以得到第二速率。

[0126] 效用函数$f_{wb}(r_C,l)$的一种示例如下:

[0127] $$f_{w,b}(r_C,l)=r_C^s-a\cdot r_C\cdot l-w\cdot\frac{l-l_0}{r_C-r_P+b}$$

[0128] 其中,s、a,b和w为参数,具体数值可以通过在线学习获得;r_P表示上一周期的探测报文的发送速率,l_0表示目标丢包率。

[0129] 本申请的实施例中,可以基于第一丢包率和/或第二速率判断第一探测报文的传输即将是处于轻载模式还是处于重载模式。例如,若第二速率大于预设的速率阈值,则认为第一探测报文的传输即将处于轻载模式,否则认为即将处于重载模式。又如,若第一丢包率小于预设的丢包率阈值,则认为第一探测报文的传输即将处于轻载模式,否则认为第一探测报文的传输即将处于重载模式。

[0130] 若第一探测报文的传输即将处于轻载模式,则使用第二速率发送第一探测报文;若第一探测报文的传输即将处于重载模式,则使用预设的目标速率发送第一探测报文。

[0131] 在一些实现方式中,设置的目标速率应满足如下条件:第一网络设备至第二网络设备的传输路径上的第三网络设备转发的探测报文仅包含第一探测报文时,第一网络设备使用所述目标速率在指定时长内能够将指定数量的第一探测报文从第一网络设备成功传输至第二网络设备。这样可以使得第一探测报文的发送从轻载模式切换至重载模式时,例如,第三速率大于目标速率而第二速率小于或等于目标速率时,由于不降速发送第一探测报文,因此可以避免第一探测报文的断流,从而可以避免第一数据报文的断流。尤其在多个源设备高速同时发送数据报文的情况下,可以避免断流问题。

[0132] 在一些实现方式中,多个源设备中所有源设备的探测报文的目标速率可以相同。这可是实现这些源设备的探测报文的发送速率的公平收敛。

[0133] 在一些实现方式中,预设的速率阈值可以等于预设的目标速率,这使得第一网络设备从轻载模式切换到重载模式时,可以实现第一探测报文的传输速率的快速收敛。

[0134] 在轻载模式下,基于第三速率和第一丢包率获得第二速率之后,可以基于第二速率和预设的目标速率来判断是继续保持轻载模式,还是切换至重载模式。

[0135] 本申请的实施例的一些实现方式中,第一网络设备可以根据第三速率判断第一探测报文的发送已处于轻载模式还是重载模式。例如,若第三速率大于预设的速率阈值,则说明当前处于轻载模式,否则说明书当前已处于重载模式。

[0136] 若当前处于轻载模式,则使用调整得到的第二速率发送第一探测报文;若处于重载模式,则使用预设的目标速率发送第一探测报文。该实现方式中的速率阈值和目标速率的特性可以参考前述速率阈值和目标速率的相关内容。

[0137] 本申请的实施例的一些实现方式中,第一网络设备在获取第一丢包率之后,可以先基于第一丢包率和预设的丢包率阈值判断是继续使用该目标速率发送第一探测报文还是对第三速率进行调整以及使用调整得到的第二速率来发送第一探测报文。

[0138] 也就是说,根据第一丢包率判断使用第三速率传输第一探测报文之后,第一探测报文的传输当前处于轻载模式还是重载模式,若当前处于轻载模式,则使用调整得到的第二速率发送第一探测报文;若当前处于重载模式,则使用预设的目标速率发送第一探测报文。这样可以充分利用网络资源来发送探测报文,从而可以充分利用网络资源来发送数据报文。

[0139] 在该实现方式中,对第三速率进行调整,以得到第二速率时,可以使用SCP策略来基于第三速率和第一丢包率获得第二速率,获取方式可以参考前述方式。

[0140] 其中,可选地,第一丢包率可以等于效用函数中的参数l_0。这使得第一网络设备切

换至轻载模式时,可以继续第一探测报文的传输速率的迭代收敛过程,以及实现第一探测报文的快速公平收敛。

[0141] 本申请实施例的一些实现方式中,第二网络设备每收到一个第一探测报文,则发送一个第一反馈信息,第一反馈信息的目的设备为第一网络设备,第一反馈信息的源设备为第二网络设备。例如,第一反馈信息可以是对第一探测报文的目的地址和源地址进行交换修改得到的探测报文。该实现方式中,第一网络设备每收到一个第一反馈信息,则发送一个第一数据报文。

[0142] 该实现方式中,第二网络设备能够成功接收一个第一探测报文,则就相应地能够成功接收一个第一数据报文。因此,该实现方式可以避免第一数据报文引起的拥塞以及第一数据报文的丢包。

[0143] 本申请实施例的一些实现方式中,源主机的操作系统内核需要进行相应修改,写入新的传输层协议,例如可以写入与传统的传输控制协议(transmission control protocol,TCP,)、用户数据报协议(user datagram protocol,UDP)等传输层协议并列的协议。

[0144] 另外,中间节点也需要在软件层面进行相关的修改,例如,中间节点需要为探测报文预留一定的缓存来存放探测报文的突发,且中间节点的出端口需要预留一定比例的带宽专门用来发送探测报文,对超过该预留带宽的探测报文进行丢弃;正常的数据报文可以使用一定比例的带宽来发送,例如,使用为探测报文预留带宽之外的链路带宽来传输数据报文。

[0145] 由于探测报文经过了中间节点的速率调整,因此在以探测报文速率为指导的数据报文传输速率恰好合适。这样,通过源主机与中间节点的相互协作,可以实现数据报文的无损传输,即无丢包传输。

[0146] 本申请的技术方案,可以用于低时延网络的高并发数据流的拥塞控制。在一些示例中,引入双拥塞信号,分别在轻载和重载模式下用于对第一探测报文的传输速率的调整。例如,根据第一探测报文的传输速率是否低至预设的速率阈值来决定是否从轻负载模式切换到重负载模式;通过判断当前第一探测报文的丢包率是否低至预设的丢包率阈值来决定拥塞控制是否从重负载模式切换至轻负载模式。

[0147] 本申请的一个拥塞控制示例中,某个源主机有数据要传输时,首先根据初始化的速率,等间隔地在控制通道中传输探测报文,直至收到目的主机回传的探测包为止。首轮数据传输之后,源主机接下来的数据传输过程中,数据报文的发送速率会基于探测报文的发送速率和接收情况来调控,而探测报文的发送速率则以RTT为周期,以探测报文的丢包率来调控。

[0148] 例如,源主机将计时器初始化为零,并且源主机每接收一个探测报文则发送一个数据报文,同时更新计时器的时间t以及探测报文的丢包率l。之后,源主机比较当前计时时间域RTT的大小。

[0149] 如果计时器时间t没有达到一个RTT的时长,则继续循环执行“接收探测报文-发送数据报文-更新计时器-更新探测报文丢包率”的操作;如果计时器时间t达到了一个RTT的时长,则比较当前探测报文的发送速率r_C与目标速率r_0的大小,以判断是否进入重负载模式。

[0150] 如果$r_C>r_0$,则执行轻载模式的拥塞控制方法,即根据指定的算法提高或降低下一轮的探测报文发送速率,并重置报文丢包率l和计时器的时间t;反之,则进入重载拥塞控制模式。

[0151] 重载拥塞控制模式下,源主机以固定的速率r_0发送探测报文。此模式下,探测报文的发送速率不会进行调整,源主机始终按照r_0的速率等间隔地发送探测报文。同样,源主机循环执行“接收探测报文-发送数据报文-更新计时器-更新探测报文丢包率”的操作。当计时器时间t达到一个RTT时长后,源主机比较当前探测报文丢包率l与预设丢包率阈值l_0的大小,以判断是否切换回轻负载拥塞控制模式。

[0152] 如果$l<l_0$,则切换回之前的轻负载拥塞控制模式,并根据指定的拥塞控制算法调整探测报文的发送速率;否则,持续重载模式的拥塞控制流程,即继续以固定的速率r_0等间隔地发送探测报文,并重置探测报文丢包率l和计时器时间t。

[0153] 第三网络设备会预留一部分链路带宽专门用来传输源主机至目的主机的探测报文。当第三网络设备接收到数据时,先进行解析和判断。如果判断该数据为源主机至目的主机的探测报文,则使用预留链路带宽对该报文进行传输,否则使用非预留的链路带宽进行数据的传输。

[0154] 目的主机一旦收到源主机发送的报文,则进行解析和判断。如果判断该报文为探测报文,则更改报文的源目的地址,将该探测报文回传给源主机。如果该报文为数据报文,则进行正常的解析处理操作。

[0155] 为保证目的主机回传给源主机的探测报文能够顺利回传给源主机,第三网络设备可以额外提高该探测报文的传输优先级。

[0156] 图4为本申请一个实施例的网络拥塞的控制方法的示例性流程图。例如,低时延网络中可以应用如图4所示的网络拥塞的控制方法。

[0157] 下面详细介绍该控制方法的各个步骤。该控制方法可以由源主机执行,从数据传输的开始时刻执行,直至数据传输完毕。

[0158] S401,源主机需要传输数据。

[0159] S402,判断该轮数据传输是否为首轮传输,并根据判断结果进入不同分支步骤。如果该次传输为首轮数据传输,进入S403;如果该次传输不是首轮数据传输,进入S404。

[0160] S403,初始化探测报文的首轮发送速率。例如,可以将探测报文的首轮发送速率初始化为应用带宽需求的1.5倍。

[0161] S404,以初始速率发送探测报文。例如以初始速率等间隔地在控制通道中发送探测报文。其中发送探测报文的逻辑通道称为控制通道。

[0162] S405,源主机判断是否收到目的主机回传的探测报文,并根据判断结果进入不同分支步骤。如果收到了目的主机回传的探测报文,则首轮传输结束,进入S406;如果未收到目的主机回传的探测报文,则进入S404,继续首轮探测报文的传输。

[0163] 首轮传输之后,可以进行探测报文与数据报文的传输速率调节。其中,数据报文的传输速率根据回传的探测报文速率来调节;而探测报文的传输速率调节则以保证多个不同流(例如来自多个不同源主机)的探测报文的发送速率可以在共享瓶颈链路上实现公平性收敛为目的。例如,探测报文的发送速率可以以RTT为周期来调节。

[0164] S406,初始化目标速率r_0和初始化目标丢包率l_0。本实施例中的目标丢包率也可以

称为丢包率阈值。

[0165] S407,根据上轮的反馈信息计算本轮探测报文的发送速率r_C。例如,根据上轮探测报文的发送速率和丢包率l获取本轮探测报文的发送速率r_C。

[0166] 轻负载模式下,可以依据效用函数的取值变化来调整探测报文的发送速率。具体地,可以计算效用函数取值,并根据效用函数变化按照SCP方案调整探测报文的发送速率。

[0167] S408,设置计时器t=0和设置探测报文的丢包率l=0。

[0168] S409,比较计算的探测报文发送速率r_C与目标速率r_0的大小,并根据比较结果进入不同分支。如果$r_C>r_0$,则进入步骤S410,即轻载模式;如果$r_C\leqslant r_0$,则进入步骤S414,即重载模式的速率调节。

[0169] S410,使用最新计算得到的r_C发送探测报文。

[0170] S411,每接收到一个回程探测报文,发送一个数据报文。

[0171] S412,记录本次与上次收到回传探测包的时间间隔Δt,并更新计时器t=t+Δt;并且,根据回程探测报文的序号更新探测报文丢包率l。

[0172] S413,判断当前计时器时长是否达到RTT大小,并根据判断结果进入不同分支步骤。如果计时器时长达到RTT大小,则进入S414,即轻载模式的速率调节。如果计时器时长没有达到RTT大小,则继续S410。

[0173] S414,以固定的目标速率r_0发送探测报文。由于以固定的发送速率探测报文,不进行发送速率的调整,因此,天然满足公平性收敛。

[0174] S415,每收到一个回传的探测报文,发送一个数据报文。

[0175] S416,更新探测报文丢包率l,更新计时器t=t+Δt。

[0176] S417,判断当前计时器时长是否达到RTT大小,并根据判断结果进入不同分支步骤。如果计时器时长达到RTT大小,则进入S418;如果计时器时长没有达到RTT大小,则进入S414。

[0177] S418,判断$l<l_0$,即判断是否切换轻重负载模式。具体地,比较本轮探测报文的丢包率l与目标丢包率l_0的大小,并根据比较结果进入不同分支步骤。如果$l<l_0$则进入S407,即进行轻载模式的速率调控;如果$l\geqslant l_0$则进入S419。

[0178] S419,设置计时器t=0和设置探测报文的丢包率l=0,并继续执行S414,继续重载模式的速率调控。

[0179] S407至S419在首轮传输结束之后持续执行,直至整个源端到目的端的数据传输过程结束为止。

[0180] 本实施例中,经过探测报文的提前预演传输,数据报文的传输速率可以保证充分利用网络带宽同时保证绝对无丢包。其中,轻载模式下,SCP策略的调控算法可以实现快速公平收敛;重载模式下,所有源节点按照统一的速率发送探测报文,天然满足公平性收敛,进一步提高了策略的整体公平性收敛速度。另外,重负载模式切换回轻负载模式的判定条件是$l\geqslant l_0$,而轻载模式速率调控的收敛点是l=0,因此切换回轻载模式之后整个策略会继续进行速率的迭代调整,模式切换不会破坏轻载模式的公平性收敛过程。

[0181] 本申请网络拥塞的控制方法的另一个实施例的示意性流程图如图5所示。例如,低时延网络中可以应用图5所示的控制方法。

[0182] 该方法可以由目的主机执行。目的主机每收到一个探测报文或数据报文都执行方

法中的操作。

[0183] S510,目的主机接收数据。

[0184] S520,目的主机收到数据后,首先判断该数据为探测报文还是数据报文。如果收到的包是数据报文,则进入S530;如果收到的是探测报文,则进入S540。

[0185] S530,对数据包进行解析处理。数据报文的处理,可以采用传统传输层协议的工作方式,对数据包进行解封装和处理操作。

[0186] S540,回传探测报文。例如,将探测报文的源IP与目的IP对换,将探测报文回传至源主机。

[0187] 目的主机每次收到探测报文或数据报文,均执行图5所示的方法,对数据报文进行解析处理,对探测报文进行回传。

[0188] 本申请网络拥塞的控制方法的又一个实施例的示意性流程图如图6所示。例如,低时延网络中可以应用图6所示的控制方法。

[0189] 该方法可以由网络的第三网络设备执行,持续整个端到端传输过程的始终。第三网络设备为每条链路分配一定比例的带宽,专门用于传输探测报文,例如分配5%的带宽传输探测报文;使用分配带宽之外的其它链路带宽传输数据报文,例如使用其他95%的带宽传输数据报文。

[0190] S610,第三网络设备接收上一跳设备传来的数据。第三网络设备的上一跳设备可以是源主机,也可以是另一个第三网络设备。

[0191] S620,第三网络设备收到上一跳设备传来的数据后,首先判断收到的包是数据报文还是探测报文。如果收到的是数据报文,则进入步骤S630;如果收到的是探测报文,则进入S640。

[0192] S630,使用分配给探测报文以外的带宽传输数据报文。或者说,使用非预留带宽传输数据报文。

[0193] S640,判断探测报文是去程探测报文还是回程探测报文,并根据判断结果进入不同分支步骤。其中,去程探测报文是指从源主机发送给目的主机的探测报文,回程探测报文是指从目的主机发送给源主机的探测报文。

[0194] 如果该探测报文为去程探测报文,则进入步骤S650;如果该探测报文为回传探测报文,则进入步骤S660。

[0195] S650,使用预分配的带宽比例传输去程探测报文。其中,第三网络设备可以预留少量缓存,例如预留8个探测报文大小的缓存,应对探测报文的突发传输。探测报文缓存占满时,丢弃接收到的多余探测报文。

[0196] S660,使用用于传输数据报文的带宽传输回程探测报文。其中,以最高优先级回传回程探测报文。

[0197] 执行图5和图6的方法,相当于使用探测报文作为数据报文的“替身”,在当前网络环境中预演了一遍数据传输流程,通过第三网络设备探测报文的丢包和目的主机探测报文的回传,将网络准确的负载状况传递给了源主机。这样,图4中的源主机就可以根据探测报文反映的网络状况精准控制数据报文的传输速率,从而实现了整个拥塞控制过程的快速收敛和无损。整个拥塞控制过程由端到端的设备(源主机、目的主机以及网络第三网络设备)协同参与。

[0198] 图4至图5的实施例专门设计了重载模式的拥塞控制方案,使得重载模式下依然可以维持低时延、高通量、无拥塞的端到端传输。其中,利用双通道(传输探测报文的通道和传输数据报文的通道)架构的优势,在重载模式下所有源主机可以采用固定速率发送探测报文,同时发送速率相对激进,使得用于发送探测报文的控制通道持续处于丢包状态,从而可以保证数据通道在传输数据报文时可以满吞吐和无丢包。并且,重载模式下,探测报文传输不降速,源主机一直能收到回传探测报文并发送数据报文,从而可以避免断流问题。

[0199] 另外,图4所示的实施例选择了合适的模式切换点,当拥塞控制从重载模式切换到轻载模式时,轻载模式不会立刻判定达到收敛点,而是会继续向收敛点进行传输速率的迭代调控,并最终收敛至既定的稳态传输速率,保证了轻载模式和重载模式的切换不会影响轻载模式的公平性收敛过程;而拥塞控制从轻载模式切换到重载模式时,整个传输会立刻收敛至稳态速率,一定程度上加速了整个拥塞控制方案的收敛过程。

[0200] 本申请提出的拥塞控制方法还可以应用于数据中心网络中。数据中心网络的轻载模式下,可以选用2019年ACM SIGCOMM发表的学术论文“ExpressPass”中的数据中心网络拥塞控制算法。

[0201] SIGCOMM是美国计算机协会(ACM)在通信网络领域的旗舰型会议。“ExpressPass”方案中,拥塞控制算法调控每条数据流的发送速率,直至所有探测报文数据流的丢包率达到目标丢包率(target_loss)。速率调控方面,通过当前速率(cur_rate)、调控权值(w)、最大速率(max_rate)、当前探测报文丢包率(credit_loss)和目标丢包率(target_loss)几个参数来计算下一周期的探测报文发送速率。

[0202] 本申请在“ExpressPass”算法的基础上,引入重载模式下的速率调节方法,并选择轻重载模式的合适转换点,其中,cur_rate$\leqslant r_0$时由轻载模式转重载模式;credit_loss$<l_0$时由重载模式转轻载模式。

[0203] 假设网络环境依然是RTT为10微秒(us)、瓶颈链路带宽为100Gbps的低时延网络,在数据中心网络中实现拥塞控制的一个实施例如下:

[0204] 步骤701,源主机初始化目标丢包率(target_loss)、最大速率(max_rate)、权值($w=0$)和最大权值($w_{max}=0.5$)。

[0205] 步骤702,源主机初始化轻重载模式转换的性能参数:$r_0=100Mbps$,$0<l_0<$target_loss。

[0206] 步骤703,源主机根据应用需求,按照目标速率,等间隔发送探测报文,并请求带宽。其中,网络的第三网络设备预留一定比例(例如5%)的带宽,作为控制通道传输探测报文。

[0207] 步骤704,去程时,第三网络设备的逐跳出端口按照一定带宽比例(例如5%)对所有探测报文进行承诺访问速率(CAR)限速,对于超出端口能力的探测报文进行丢弃。

[0208] 步骤705,目的主机将接收到的探测报文的源地址和目的地址进行调换,并回传该探测报文。

[0209] 步骤706,第三网络设备以最高优先级将探测报文回传给源主机。回程时,第三网络设备不对探测报文进行限速。

[0210] 步骤707,源主机每收到一个探测报文立刻发送一个数据报文。

[0211] 一个RTT周期后,源主机根据探测报文丢包率和发送速率判断是否进行轻载模式

和重载模式的切换。

[0212] 如果cur_rate≤r_0,则切换为重负载模式,并以固定的速率r_0发送探测报文直至下一次状态切换;如果cur_rate>r_0,则进行轻载模式下的速率调控。

[0213] 在轻载模式下,若当前丢包率小于目标丢包率,则提速,且速率计算方式如下:cur_rate=(1-w)*cur_rate+w*max_rate*(1+target_loss);若当前丢包率大于目标丢包率,则降速,且速率计算方式如下:cur_rate=cur_rate*(1-credit_loss)*(1+target_loss)。

[0214] 重载模式下,如果credit_loss<l_0,则切换为轻负载模式,并按照目标丢包率进行速率收敛;如果credit_loss≥l_0,则继续以固定速率r_0发送探测报文。

[0215] 由上述方法可知,本申请的技术方案可以应用在时延更敏感的数据中心网络当中。其中,重载模式向轻载模式的转换点为credit_loss<l_0,由于l_0<target_loss,因此在模式切换之后,依然会进行迭代速度调整的过程,原方案的公平性收敛过程不会被破坏。同样,拥塞控制从轻载模式切换到重载模式时,整个传输会立刻收敛至稳态速率(重载模式的收敛速率为r_0),一定程度上加速了整个拥塞控制方案的收敛过程。

[0216] 该实施例进一步说明了,本申请的技术方案可以和其它任何基于传输速率和丢包率的拥塞控制方案完美兼容,唯一需要注意的是谨慎选择重载模式向轻载模式过渡的时间点。

[0217] 此外,在数据中心网络中,选用了迭代次数较少的“ExpressPass”方案作为轻负载方案,相比于基于在线学习的方案,“ExpressPass”方案的性能探测操作相对少,即无需反复计算效用函数以决定发送速率的变化,这使得轻载模式向重载模式的转换会更为迅速,更适用于时延敏感、突发性较强的数据中心网络当中。

[0218] 图7是本申请一个实施例的网络拥塞的控制装置700的示意性结构图。装置700可以包括发送模块710和接收模块720。

[0219] 在一些示例中,装置700可以用于执行图2所示方法中由源主机执行的相关步骤或操作,例如发送模块710可以用于执行S210和S250中的相关操作,接收模块720可以用于执行S240中的相关操作。

[0220] 在另一示例中,装置700可以用于执行图2所示方法中由中间节点执行的相关步骤或操作,例如发送模块710可以用于执行S220和S260中的相关操作,接收模块720可以用于执行S210、S230和S250中的相关操作。

[0221] 在又一些示例中,装置700可以用于执行图2所示方法中由目的主机执行的相关步骤或操作,例如发送模块710可以用于执行S230中的相关操作,接收模块720可以用于执行S220和S260中的相关操作。

[0222] 在其他一些示例中,装置700可以用于执行图4所示的方法,或者,装置700可以用于执行图5所示的方法,或者装置700可以用于执行图6所示的方法。

[0223] 图8为本申请一个实施例的网络拥塞的控制装置800的示意性结构图。装置800包括处理器802、通信接口803和存储器804。装置800的一种示例为芯片,装置800的另一个示例为计算设备。

[0224] 处理器802、存储器804和通信接口803之间可以通过总线通信。存储器804中存储有可执行代码,处理器802读取存储器804中的可执行代码以执行对应的方法。存储器804中

还可以包括操作系统等其他运行进程所需的软件模块。操作系统可以为LINUX™,UNIX™,WINDOWS™等。

[0225] 例如,存储器804中的可执行代码用于实现图2中由源主机、中间节点或目的主机执行的步骤或操作;处理器802读取存储器804中的该可执行代码以执行图2中由源主机、中间节点或目的主机执行的步骤或操作。

[0226] 又如,存储器804中的可执行代码用于执行图4、图5或图6所示的方法;处理器802读取存储器804中的该可执行代码以执行图4、图5或图6所示的方法。

[0227] 其中,处理器802可以为CPU。存储器804可以包括易失性存储器(volatile memory),例如随机存取存储器(random access memory,RAM)。存储器804还可以包括非易失性存储器(2non-volatile memory,2NVM),例如只读存储器(2read-only memory,2ROM),快闪存储器,硬盘驱动器(hard disk drive,HDD)或固态启动器(solid state disk,SSD)。

[0228] 在本申请的一些实施例中,所公开的方法可以实施为以机器可读格式被编码在计算机可读存储介质上的或者被编码在其它非瞬时性介质或者制品上的计算机程序指令。图9示意性地示出根据这里展示的至少一些实施例而布置的示例计算机程序产品的概念性局部视图,所述示例计算机程序产品包括用于在计算设备上执行计算机进程的计算机程序。在一个实施例中,示例计算机程序产品900是使用信号承载介质901来提供的。所述信号承载介质901可以包括一个或多个程序指令902,其当被一个或多个处理器运行时可以提供以上针对图2、图4、图5和图6任意一个图所示的方法中描述的功能或者部分功能。因此,例如,图4中所示的实施例,S401至S419的一个或多个特征可以由与信号承载介质901相关联的一个或多个指令来承担。又如,参考图6中所示的实施例,S610至S660的一个或多个特征可以由与信号承载介质901相关联的一个或多个指令来承担。

[0229] 在一些示例中,信号承载介质901可以包含计算机可读介质903,诸如但不限于,硬盘驱动器、紧密盘(CD)、数字视频光盘(DVD)、数字磁带、存储器、只读存储记忆体(read-only memory,ROM)或随机存储记忆体(random access memory,RAM)等等。在一些实施方式中,信号承载介质901可以包含计算机可记录介质904,诸如但不限于,存储器、读/写(R/W)CD、R/W DVD、等等。在一些实施方式中,信号承载介质901可以包含通信介质905,诸如但不限于,数字和/或模拟通信介质(例如,光纤电缆、波导、有线通信链路、无线通信链路、等等)。因此,例如,信号承载介质901可以由无线形式的通信介质905(例如,遵守IEEE 802.11标准或者其它传输协议的无线通信介质)来传达。一个或多个程序指令902可以是,例如,计算机可执行指令或者逻辑实施指令。在一些示例中,前述的计算设备可以被配置为,响应于通过计算机可读介质903、计算机可记录介质904、和/或通信介质905中的一个或多个传达到计算设备的程序指令902,提供各种操作、功能、或者动作。应该理解,这里描述的布置仅仅是用于示例的目的。因而,本领域技术人员将理解,其它布置和其它元素(例如,机器、接口、功能、顺序、和功能组等等)能够被取而代之地使用,并且一些元素可以根据所期望的结果而一并省略。另外,所描述的元素中的许多是可以被实现为离散的或者分布式的组件的、或者以任何适当的组合和位置来结合其它组件实施的功能词条。

[0230] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员

可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0231] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0232] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0233] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0234] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0235] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器、随机存取存储器、磁碟或者光盘等各种可以存储程序代码的介质。

[0236] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以所述权利要求的保护范围为准。

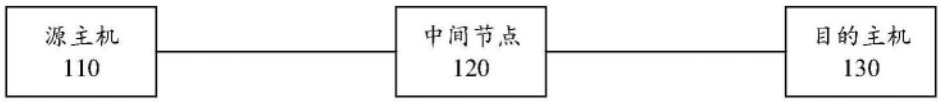

源主机
110
中间节点
120
目的主机
130

图1

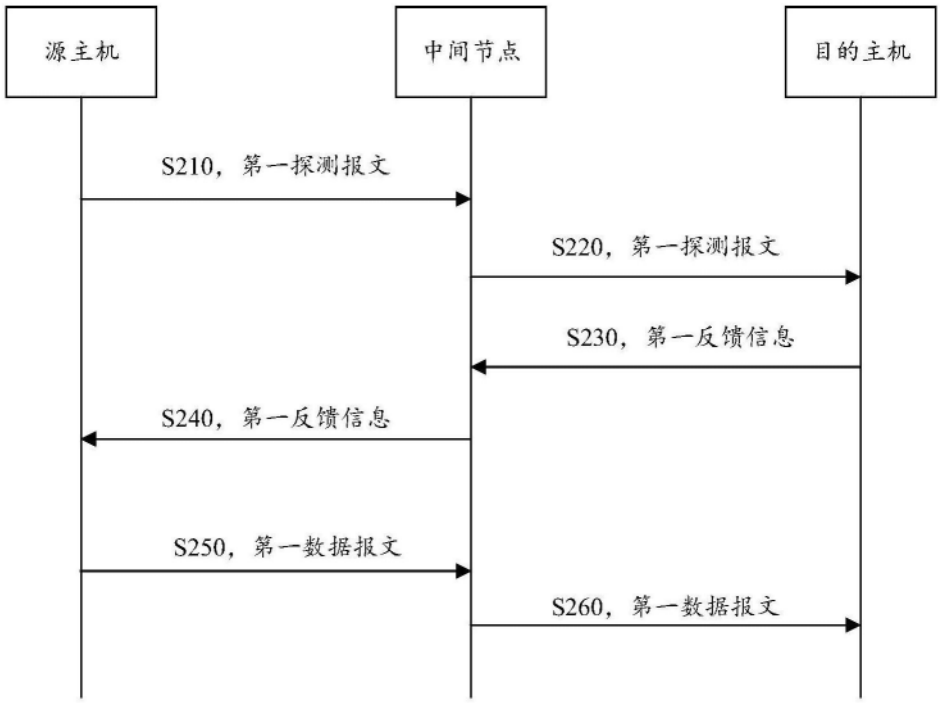

源主机
中间节点
目的主机
S210，第一探测报文
S220，第一探测报文
S230，第一反馈信息
S240，第一反馈信息
S250，第一数据报文
S260，第一数据报文

图2

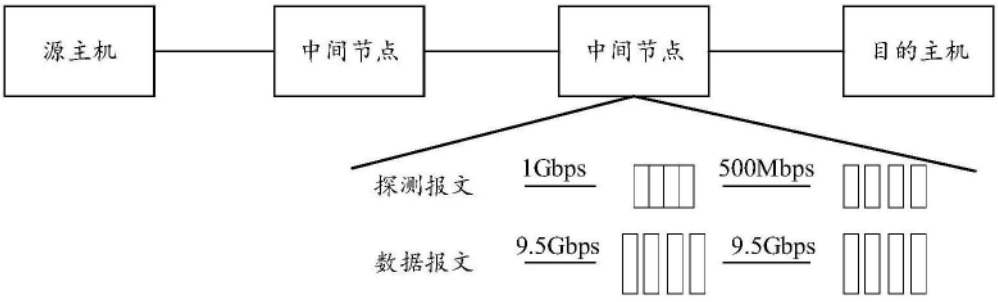

源主机
中间节点
中间节点
目的主机
探测报文
1Gbps
500Mbps
数据报文
9.5Gbps
9.5Gbps

图3

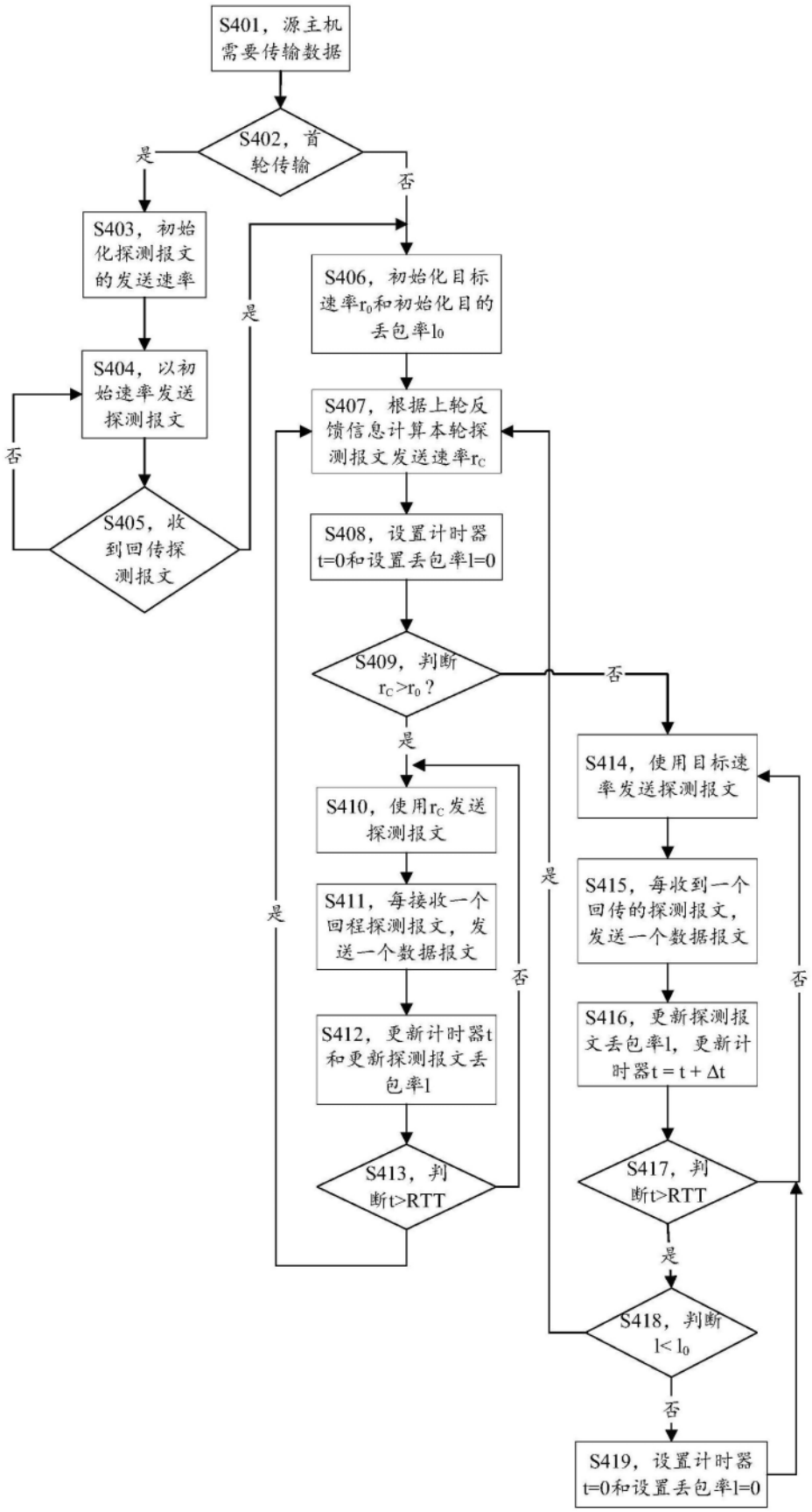

S401，源主机需要传输数据
S402，首轮传输
是
否
S403，初始化探测报文的发送速率
S404，以初始速率发送探测报文
S405，收到回传探测报文
否
是
S406，初始化目标速率r_0和初始化目的丢包率l_0
S407，根据上轮反馈信息计算本轮探测报文发送速率r_C
S408，设置计时器t=0和设置丢包率l=0
S409，判断$r_C>r_0$？
是
否
S410，使用r_C发送探测报文
S411，每接收一个回程探测报文，发送一个数据报文
S412，更新计时器t和更新探测报文丢包率l
S413，判断t>RTT
是
否
S414，使用目标速率发送探测报文
S415，每收到一个回传的探测报文，发送一个数据报文
S416，更新探测报文丢包率l，更新计时器$t=t+\Delta t$
S417，判断t>RTT
是
否
S418，判断$l<l_0$
是
否
S419，设置计时器t=0和设置丢包率l=0

图4

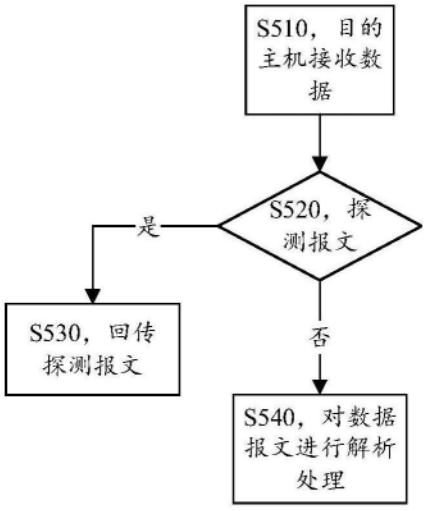

S510，目的主机接收数据
S520，探测报文
是
否
S530，回传探测报文
S540，对数据报文进行解析处理

图5

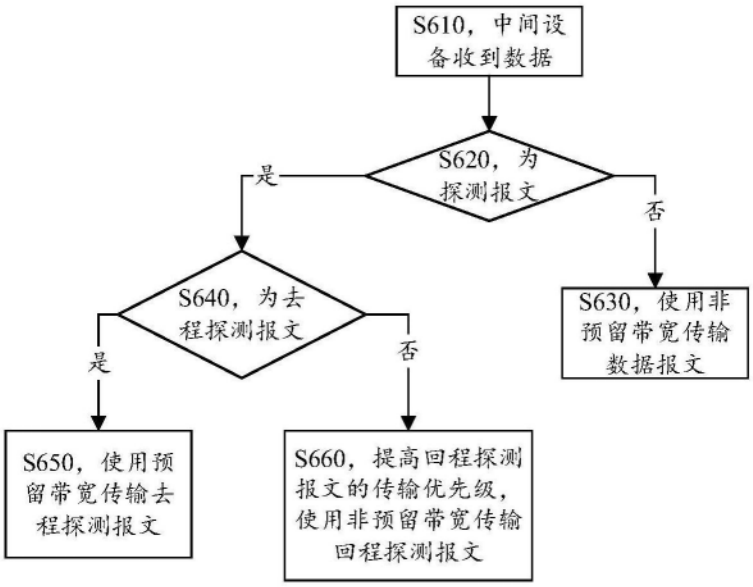

S610，中间设备收到数据
S620，为探测报文
是
否
S640，为去程探测报文
是
否
S630，使用非预留带宽传输数据报文
S650，使用预留带宽传输去程探测报文
S660，提高回程探测报文的传输优先级，使用非预留带宽传输回程探测报文

图6

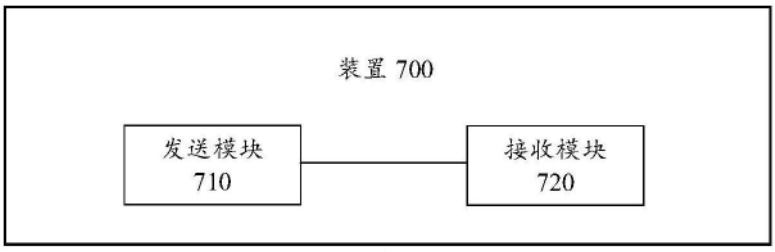

装置 700
发送模块 710
接收模块 720

图7

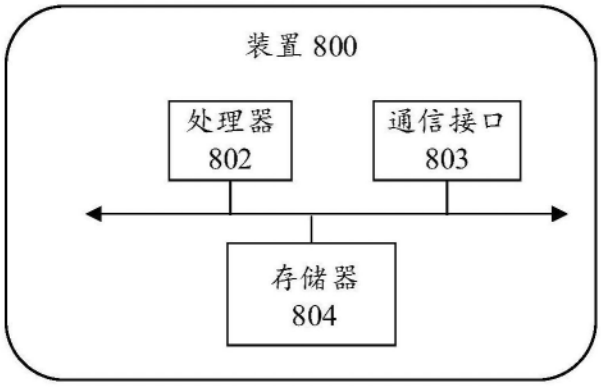

装置 800
处理器
802
通信接口
803
存储器
804

图8

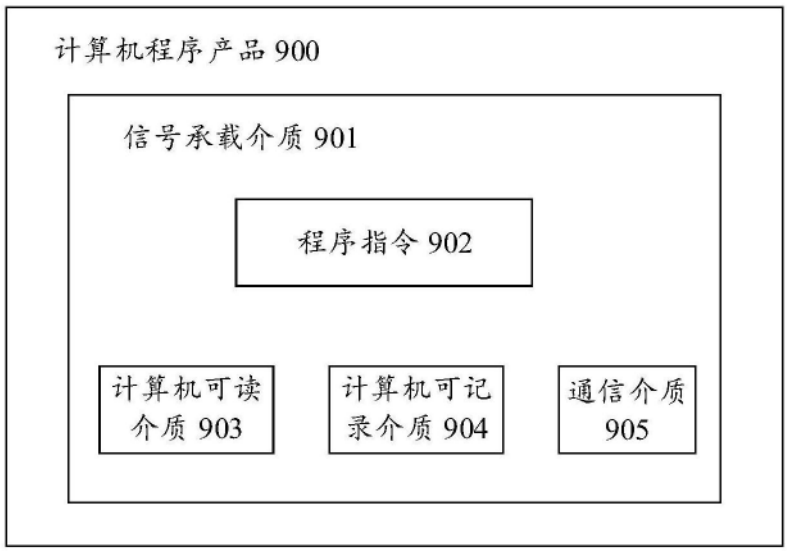

计算机程序产品 900
信号承载介质 901
程序指令 902
计算机可读
介质 903
计算机可记
录介质 904
通信介质
905

图9