

一种链路故障处理方法、装置、电子设备及介质

技术领域

本申请涉及通信领域，特别是涉及一种链路故障处理方法、装置、电子设备及介质。

背景技术

5 大规模的神经网络模型在训练时，需要使用多台主机的多个图形处理器（Graphics Processing Unit，GPU），在训练过程中各 GPU 之间需要频繁交互。例如，图 1 的左图为训练传统的数据中心（Data Center，DC）流量模型时各 GPU 之间的通信流量图，右图为训练大规模的人工智能（Artificial Intelligence，AI）流量模型时各 GPU 之间的通信流量图，图 1 中的两个流量图的横轴均为时间，纵轴均为带宽。可以看出，训练传统的 DC 流量模型涉及的流数多，且训练总耗时短，即 GPU 之间的通信链路多，且每条通信链路的通信时间短，因此通信链路的通信质量对于 DC 流量模型的收敛影响较小。而训练 AI 流量模型时涉及的流数少，但是具有网络利用率高和带宽高的特点，且训练总耗时长，例如训练耗时通常为 10-90 天。在训练过程中需要网络持续保持高性能的状态，训练过程中若存在一条通信链路故障，将会延长一轮训练过程所需的时间，甚至可能导致训练流程出错，导致需要重启训练流程，因此在训练 AI 流量模型的过程中，网络质量对于训练效率有重要影响。

10 在训练过程中，各主机之间基于交换机或路由器等网络设备进行通信。目前在网络设备之间出现链路故障的情况下，网络设备感知到链路故障后，还需通过路由更新的方式向其他网络设备进行故障通知。网络设备接收到路由更新消息后，网络设备的控制面再重新计算路由，并对转发芯片中的转发表项进行修改，以避免继续通过故障的链路通信。但上述过程中，故障通知以及重新计算路由所需的时间较长，将会导致故障收敛不及时，从而影响 AI 流量模型的训练效率。

20 发明内容

本申请实施例的目的在于提供一种链路故障处理方法、装置、电子设备及介质，以实现在网络设备之间的链路发生故障时，更及时地进行路由切换。具体技术方案如下：

第一方面，本申请实施例提供了一种链路故障处理方法，应用于第一网络设备的转发芯片，所述第一网络设备包括第一接口，并通过所述第一接口与第二网络设备相连，所述第二网络设备与第三网络设备相连，所述第三网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述方法包括：

25 接收所述第二网络设备发送的第一报文，所述第一报文包括所述第三网络设备的设备标识以及状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

若所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

30 将所述第一转发表项的状态设置为无效状态。

可选地，在所述接收所述第二网络设备发送的第一报文之前，所述方法还包括：

通过所述第一接口接收所述第二网络设备转发的路由更新报文，所述路由更新报文包括所述第三网络设备连接的宿主机的第一 IP 地址以及所述第三网络设备的设备标识；

35 基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

可选地，所述第二网络设备包括第二接口，并通过所述第二接口与所述第一网络设备相连；在所述接收所述第二网络设备发送的第一报文之前，所述方法还包括：

向所述第二网络设备发送第二报文，所述第二报文包括所述第一网络设备的设备标识，以使得所述

第二网络设备记录所述第二接口与所述第一网络设备的设备标识之间的第一对应关系；

接收所述第二网络设备发送的第三报文，所述第三报文包括所述第二网络设备的设备标识，记录所述第一接口与所述第二网络设备的设备标识之间的第二对应关系。

可选地，所述将所述第一转发表项的状态设置为无效状态，包括：

5 将所述第一转发表项的状态标志位修改为第一值，所述第一值用于表示无效状态。

可选地，在所述接收所述第二网络设备发送的第一报文之后，所述方法还包括：

若所述状态信息表示所述第二网络设备与所述第三网络设备之间可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

将所述第一转发表项的状态设置为有效状态。

10 可选地，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

15 可选地，所述第二报文为第二链路层广播报文，所述第二链路层广播报文的载荷部分包括所述第一网络设备的设备标识；或者，所述第二报文为第二 IP 组播报文，所述第二 IP 组播报文的载荷部分包括所述第一网络设备的设备标识；

所述第三报文为第三链路层广播报文，所述第三链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第三报文为第三 IP 组播报文，所述第三 IP 组播报文的载荷部分包括所述第二网络设备的设备标识。

20 可选地，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

第二方面，本申请实施例提供了一种链路故障处理方法，应用于第二网络设备的转发芯片，所述第二网络设备与第三网络设备相连，所述第二网络设备连接第一网络设备的所述第一接口，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述方法包括：

25 获取状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

向所述第一网络设备发送第一报文，所述第一报文包括所述第三网络设备的设备标识以及所述状态信息，以使得所述第一网络设备在所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达时，查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项，并将所述

30 第一转发表项的状态设置为无效状态。

可选地，在所述获取状态信息之前，所述方法还包括：

接收所述第三网络设备发送的路由更新报文，所述路由更新报文包括所述第三网络设备连接的

35 的第一 IP 地址以及所述第三网络设备的设备标识；

向所述第一网络设备转发所述路由更新报文，以使得所述第一网络设备基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

可选地，所述第二网络设备包括第三接口，并通过所述第三接口与所述第三网络设备相连，所述第三网络设备包括第四接口，并通过所述第四接口与所述第二网络设备相连；在所述获取状态信息之前，所述方法还包括：

向所述第三网络设备发送第四报文，所述第四报文包括所述第二网络设备的设备标识，以使得所述第三网络设备记录所述第四接口与所述第二网络设备的设备标识之间的第三对应关系；

接收所述第三网络设备发送的第五报文，所述第五报文包括所述第三网络设备的设备标识，记录所述第三接口与所述第三网络设备的设备标识之间的第四对应关系。

5 可选地，在向所述第一网络设备发送第一报文之前，所述方法还包括：

从所述第四对应关系中获取所述第三接口对应的所述第三网络设备的设备标识。

可选地，所述第二网络设备与所述第三网络设备之间存在多条链路；所述获取状态信息，包括：

10 若所述多条链路均故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达；或者，若所述多条链路中的任意一条链路未故障，则确定所述状态信息表示所述第二网络设备与
所述第三网络设备之间可达。

可选地，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

15 可选地，所述第四报文为第四链路层广播报文，所述第四链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第四报文为第四 IP 组播报文，所述第四 IP 组播报文的载荷部分包括
所述第二网络设备的设备标识；

所述第五报文为第五链路层广播报文，所述第五链路层广播报文的载荷部分包括所述第三网络设备的设备标识；或者，所述第五报文为第五 IP 组播报文，所述第五 IP 组播报文的载荷部分包括所述第三
20 网络设备的设备标识。

可选地，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

第三方面，本申请实施例提供了一种链路故障处理装置，应用于第一网络设备的转发芯片，所述第一网络设备包括第一接口，并通过所述第一接口与第二网络设备相连，所述第二网络设备与第三网络设备
25 相连，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述装置包括：

接收模块，用于接收所述第二网络设备发送的第一报文，所述第一报文包括所述第三网络设备的设备标识以及状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

30 查找模块，用于若所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

设置模块，用于将所述第一转发表项的状态设置为无效状态。

可选地，所述装置还包括：

所述接收模块，还用于在所述接收所述第二网络设备发送的第一报文之前，通过所述第一接口接收
35 所述第二网络设备转发的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

生成模块，用于基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

可选地，所述第二网络设备包括第二接口，并通过所述第二接口与所述第一网络设备相连；该装置

还包括：

发送模块，用于在所述接收所述第二网络设备发送的第一报文之前，向所述第二网络设备发送第二报文，所述第二报文包括所述第一网络设备的设备标识，以使得所述第二网络设备记录所述第二接口与所述第一网络设备的设备标识之间的第一对应关系；

- 5 所述接收模块，还用于接收所述第二网络设备发送的第三报文，所述第三报文包括所述第二网络设备的设备标识，记录所述第一接口与所述第二网络设备的设备标识之间的第二对应关系。

可选地，所述设置模块，具体用于：

将所述第一转发表项的状态标志位修改为第一值，所述第一值用于表示无效状态。

- 10 可选地，所述查找模块，还用于在所述接收所述第二网络设备发送的第一报文之后，若所述状态信息表示所述第二网络设备与所述第三网络设备之间可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

所述设置模块，还用于将所述第一转发表项的状态设置为有效状态。

可选地，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

- 15 所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

可选地，所述第二报文为第二链路层广播报文，所述第二链路层广播报文的载荷部分包括所述第一网络设备的设备标识；或者，所述第二报文为第二 IP 组播报文，所述第二 IP 组播报文的载荷部分包括所述第一网络设备的设备标识；

- 20 所述第三报文为第三链路层广播报文，所述第三链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第三报文为第三 IP 组播报文，所述第三 IP 组播报文的载荷部分包括所述第二网络设备的设备标识。

可选地，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

- 25 第四方面，本申请实施例提供了一种链路故障处理装置，应用于第二网络设备的转发芯片，所述第二网络设备与第三网络设备相连，所述第二网络设备连接第一网络设备的第一接口，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述装置包括：

- 30 获取模块，用于获取状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

发送模块，用于向所述第一网络设备发送第一报文，所述第一报文包括所述第三网络设备的设备标识以及所述状态信息，以使得所述第一网络设备在所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达时，查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项，并将所述第一转发表项的状态设置为无效状态。

- 35 可选地，所述装置还包括：

接收模块，用于在所述获取状态信息之前，接收所述第三网络设备发送的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

所述发送模块，还用于向所述第一网络设备转发所述路由更新报文，以使得所述第一网络设备基于

所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

可选地，所述第二网络设备包括第三接口，并通过所述第三接口与所述第三网络设备相连，所述第三网络设备包括第四接口，并通过所述第四接口与所述第二网络设备相连；所述装置还包括：

5 所述发送模块，还用于在所述获取状态信息之前，向所述第三网络设备发送第四报文，所述第四报文包括所述第二网络设备的设备标识，以使得所述第三网络设备记录所述第四接口与所述第二网络设备的设备标识之间的第三对应关系；

接收模块，用于接收所述第三网络设备发送的第五报文，所述第五报文包括所述第三网络设备的设备标识，记录所述第三接口与所述第三网络设备的设备标识之间的第四对应关系。

10 可选地，所述获取模块，还用于在向所述第一网络设备发送第一报文之前，从所述第四对应关系中获取所述第三接口对应的所述第三网络设备的设备标识。

可选地，所述第二网络设备与所述第三网络设备之间存在多条链路；所述获取模块，具体用于：

若所述多条链路均故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达；或者，若所述多条链路中的任意一条链路未故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间可达。

15 可选地，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

20 可选地，所述第四报文为第四链路层广播报文，所述第四链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第四报文为第四 IP 组播报文，所述第四 IP 组播报文的载荷部分包括所述第二网络设备的设备标识；

所述第五报文为第五链路层广播报文，所述第五链路层广播报文的载荷部分包括所述第三网络设备的设备标识；或者，所述第五报文为第五 IP 组播报文，所述第五 IP 组播报文的载荷部分包括所述第三网络设备的设备标识。

25 可选地，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

第五方面，本申请实施例提供了一种第一网络设备，所述第一网络设备包括第一接口，并通过所述第一接口与第二网络设备相连，所述第二网络设备与第三网络设备相连，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述第一网络设备包括：处理器；收发器；机器可读存储介质，所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令；所述机器可执行指令促使所述处理器执行以下步骤：

30 接收所述第二网络设备发送的第一报文，所述第一报文包括所述第三网络设备的设备标识以及状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

35 若所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

将所述第一转发表项的状态设置为无效状态。

可选地，所述机器可执行指令还促使所述处理器执行以下步骤：

在所述接收所述第二网络设备发送的第一报文之前，通过所述第一接口接收所述第二网络设备转发的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

5 可选地，所述第二网络设备包括第二接口，并通过所述第二接口与所述第一网络设备相连；所述机器可执行指令还促使所述处理器执行以下步骤：

在所述接收所述第二网络设备发送的第一报文之前，向所述第二网络设备发送第二报文，所述第二报文包括所述第一网络设备的设备标识，以使得所述第二网络设备记录所述第二接口与所述第一网络设备的设备标识之间的第一对应关系；

10 接收所述第二网络设备发送的第三报文，所述第三报文包括所述第二网络设备的设备标识，记录所述第一接口与所述第二网络设备的设备标识之间的第二对应关系。

可选地，所述机器可执行指令还促使所述处理器执行以下步骤：

将所述第一转发表项的状态标志位修改为第一值，所述第一值用于表示无效状态。

可选地，所述机器可执行指令还促使所述处理器执行以下步骤：

15 在所述接收所述第二网络设备发送的第一报文之后，若所述状态信息表示所述第二网络设备与所述第三网络设备之间可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

将所述第一转发表项的状态设置为有效状态。

20 可选地，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

25 可选地，所述第二报文为第二链路层广播报文，所述第二链路层广播报文的载荷部分包括所述第一网络设备的设备标识；或者，所述第二报文为第二 IP 组播报文，所述第二 IP 组播报文的载荷部分包括所述第一网络设备的设备标识；

所述第三报文为第三链路层广播报文，所述第三链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第三报文为第三 IP 组播报文，所述第三 IP 组播报文的载荷部分包括所述第二网络设备的设备标识。

30 可选地，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

第六方面，本申请实施例提供了一种第二网络设备，所述第二网络设备与第三网络设备相连，所述第二网络设备连接第一网络设备的第一接口，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述第二网络设备包括：处理器；收发器；机器可读存储介质，所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令；所述机器可执行指令促使所述处理器执行以下步骤：

35 获取状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

向所述第一网络设备发送第一报文，所述第一报文包括所述第三网络设备的设备标识以及所述状态信息，以使得所述第一网络设备在所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达

时，查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项，并将所述第一转发表项的状态设置为无效状态。

可选地，所述机器可执行指令还促使所述处理器执行以下步骤：

5 在所述获取状态信息之前，接收所述第三网络设备发送的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

向所述第一网络设备转发所述路由更新报文，以使得所述第一网络设备基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

10 可选地，所述第二网络设备包括第三接口，并通过所述第三接口与所述第三网络设备相连，所述第三网络设备包括第四接口，并通过所述第四接口与所述第二网络设备相连；所述机器可执行指令还促使所述处理器执行以下步骤：

在所述获取状态信息之前，向所述第三网络设备发送第四报文，所述第四报文包括所述第二网络设备的设备标识，以使得所述第三网络设备记录所述第四接口与所述第二网络设备的设备标识之间的第三对应关系；

15 接收所述第三网络设备发送的第五报文，所述第五报文包括所述第三网络设备的设备标识，记录所述第三接口与所述第三网络设备的设备标识之间的第四对应关系。

可选地，所述机器可执行指令还促使所述处理器执行以下步骤：

在向所述第一网络设备发送第一报文之前，从所述第四对应关系中获取所述第三接口对应的所述第三网络设备的设备标识。

20 可选地，所述第二网络设备与所述第三网络设备之间存在多条链路；所述机器可执行指令还促使所述处理器执行以下步骤：

若所述多条链路均故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达；或者，若所述多条链路中的任意一条链路未故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间可达。

25 可选地，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

30 可选地，所述第四报文为第四链路层广播报文，所述第四链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第四报文为第四 IP 组播报文，所述第四 IP 组播报文的载荷部分包括所述第三网络设备的设备标识；

所述第五报文为第五链路层广播报文，所述第五链路层广播报文的载荷部分包括所述第三网络设备的设备标识；或者，所述第五报文为第五 IP 组播报文，所述第五 IP 组播报文的载荷部分包括所述第三网络设备的设备标识。

35 可选地，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

第七方面，本申请实施例提供一种机器可读存储介质，存储有机器可执行指令，在被处理器调用和执行时，所述机器可执行指令促使所述处理器：实现第一方面或第二方面所述的方法。

第八方面，本申请实施例提供一种计算机程序产品，所述计算机程序产品促使所述处理器：实现第

一方面或第二方面所述的方法。

本申请实施例的有益效果：本申请实施例提供的链路故障处理方法、装置、电子设备及介质，第一网络设备的转发芯片可以通过第二网络设备发送的第一报文，获得第二网络设备与第三网络设备之间是否可达，从而及时感知第二网络设备与第三网络设备之间的连接状态。而且第一网络设备的转发芯片还能够

5 能够在第二网络设备与第三网络设备之间不可达时，及时将下一跳地址为第一报文的源地址、设备标识为第三网络设备的设备标识的第一转发表项设置为无效状态，即禁用包括第二网络设备和第三网络设备之间的链路的转发路径，以便后续可以选择其他转发表项向第三网络设备连接的主机转发报文。即，本申请实施例能够在两个网络设备之间存在链路故障时，及时通知其他网络设备，而且网络设备还可以直接通过转发芯片将故障链路涉及的转发表项的状态设置为无效状态，不涉及控制面对接收到的报文的处理，因此速度更快，能够更及时地进行路由切换，从而提高 AI 流量模型的训练效率。

10

附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。

- 图 1 为本申请实施例提供的一种网络流量对比图；
- 15 图 2 为本申请实施例提供的一种通信系统结构示意图；
- 图 3 为本申请实施例提供的一种链路故障处理方法的流程图；
- 图 4 为本申请实施例提供的一种扩展团体属性的格式示意图；
- 图 5 为本申请实施例提供的一种第一链路层广播报文的报文格式的示例性示意图；
- 图 6 为本申请实施例提供的一种第一 IP 组播报文的报文格式的示例性示意图；
- 20 图 7 为本申请实施例提供的一种第一报文包括的载荷部分的格式示意图；
- 图 8 为本申请实施例提供的一种拓扑发现报文包括的载荷部分的格式示意图；
- 图 9 为本申请实施例提供的另一种链路故障处理方法的流程图；
- 图 10 为本申请实施例提供的另一种通信系统的示例性示意图；
- 图 11 为本申请实施例提供的一种链路故障处理装置的结构示意图；
- 25 图 12 为本申请实施例提供的另一种链路故障处理装置的结构示意图；
- 图 13 为本申请实施例提供的一种第一网络设备的结构示意图；
- 图 14 为本申请实施例提供的一种第二网络设备的结构示意图。

具体实施方式

为使本发明的目的、技术方案、及优点更加清楚明白，以下参照附图并举实施例，对本发明进一步

30 详细说明。显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员所获得的所有其他实施例，都属于本发明保护的范围。

神经网络模型的训练规模按照训练时使用的 GPU 数量划分，可以分为单规模（Single-Scale）、小规模（Small-Scale）、中规模（Medium-Scale）、大规模（Large-Scale）和特大规模（Extra-Large-Scale）等。其中，Single-Scale 表示使用单个 GPU 训练模型，Small-Scale 表示至少使用 8 个 GPU 训练模型，

35 Medium-Scale 表示至少使用 100 个 GPU 训练模型，Large-Scale 表示至少使用 1000 个 GPU 训练模型，Extra-Large-Scale 表示至少使用 10000 个 GPU 训练模型。

随着神经网络模型规模的不断扩大，单 GPU 已经无法满足训练神经网络模型时对计算资源和存储资源的需求，例如，AI 流量模型的训练规模一般为 Large-Scale 或者 Extra-Large-Scale。因此目前通常

采用多 GPU 联合训练神经网络模型方式，以加快模型的训练速度。例如，可以将各训练数据拆分为多份，并将每一份训练数据分配给不同的 GPU，以便各 GPU 利用 AI 流量模型对不同的训练数据进行处理，之后联合各 GPU 的处理结果计算损失值，并利用损失值调整 AI 流量模型的网络参数。又例如，利用不同的 GPU 分别使用 AI 流量模型处理每个训练数据中不同维度的数据，之后联合各 GPU 的处理结果计算损失值，并利用损失值调整 AI 流量模型的网络参数。这种联合训练的方式通常涉及大量的 GPU，因此各 GPU 之间需要频繁交互。

GPU 所属的主机之间可基于网络设备进行通信，网络设备之间存在链路故障时，需要进行路由切换。以下结合图 2 对目前的路由切换过程进行简要说明，图 2 所示的通信系统包括：主机 H1、主机 H2、网络设备 R11、网络设备 R12、网络设备 R21 和网络设备 R22。当网络设备 R11 和网络设备 R22 之间的链路故障时，网络设备 R11 通过双向转发检测（Bidirectional Forwarding Detection, BFD）或者连接故障管理（Connectivity Fault Management, CFM）的方式，感知与网络设备 R22 之间的链路故障。之后网络设备 R11 通过等价多路径路由（Equal-Cost Multi-Path Routing, ECMP）或者快速重路由（Fast Reroute, FRR）的方式，对自身存储的路由进行切换。然后网络设备 R11 通过内部网关协议（Internet Gateway Protocol, IGP）链路态传播（LinkState propagation）或者边界网关协议（Border Gateway Protocol, BGP）路由更新（route updates）的方式，向网络设备 R21 发送通知消息。网络设备 R21 接收到通知消息后，通过控制面重新计算路由，并对转发芯片中故障链路涉及的转发表项进行删除，并通过 BGP 可编程中断控制器（Programmable Interrupt Controller, PIC）或者 IGP 路线计算收敛（route calculation convergence）的方式，确定已被删除的转发表项的等价路由，使得后续网络设备 R21 使用从网络设备 R21 到网络设备 R12 再到网络设备 R22 之间的链路，向主机 H2 发送报文。

其中，BFD 以及 CFM 的收敛时间为几毫秒至几秒。ECMP 以及 FRR 的收敛时间为几毫秒，且收敛时间受故障检测时间的影响。IGP LinkState propagation 以及 BGP route updates 的收敛时间为几毫秒至几秒，且收敛时间受到通信系统中的网络设备规模以及通信路线数目的影响。BGP PIC 以及 IGP route calculation convergence 的收敛时间为几百毫秒至几秒，且收敛时间受到故障通知时间和通信系统中的网络设备规模的影响。由于故障通知以及重新计算路由所需的时间较长，将会导致故障收敛不及时，从而影响 AI 流量模型的训练效率。

为了在网络设备之间的连接发生故障时，更及时地进行路由切换，本申请实施例提供了一种链路故障处理方法，应用于第一网络设备，第一网络设备包括第一接口，并通过第一接口与第二网络设备相连，第二网络设备与第三网络设备相连。第一网络设备包括多条转发表项，转发表项包括目的地址、出接口和设备标识。第一网络设备、第二网络设备和第三网络设备均可以为交换机或者路由器等具备报文转发能力的设备。本申请实施例可以应用于叶脊（spine-leaf）网络架构中。

如图 3 所示，本申请实施例提供的链路故障处理方法包括如下步骤：

S301、接收第二网络设备发送的第一报文。

其中，第一报文包括第三网络设备的设备标识（identity document, ID）以及状态信息。其中，状态信息用于表示第二网络设备与第三网络设备之间是否可达，即表示第二网络设备与第三网络设备之间的链路是否发生故障。

可选地，第一报文还可以包括第二网络设备的设备标识。

S302、若状态信息表示第二网络设备与第三网络设备之间不可达，则查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项。

其中，第一网络设备包括多条转发表项，且每条转发表项包括目的地址、出接口和设备标识，其中设备标识为目的地址所属的主机接入的网络设备的标识。可选的，主机可以为笔记本电脑、台式电脑或者服务器等具备报文收发能力的设备。

其中，转发表项还包括下一跳地址，例如第一转发表项的下一跳地址为第二网络设备的网际互连协议（Internet Protocol，IP）地址。

5

S303、将第一转发表项的状态设置为无效状态。

由于第一转发表项表示的转发路径需要经过第二网络设备和第三网络设备之间的链路，而第二网络设备和第三网络设备之间不可达，即第二网络设备和第三网络设备之间的链路故障，因此该转发路径目前无法使用，所以可以将第一转发表项的状态设置为无效状态。

10

可以理解的是，后续第一网络设备转发业务报文时，若确定业务报文的地址与第一转发表项的地址相同，也不会使用该第一转发表项进行报文转发，而是直接查找其他与业务报文的地址匹配的转发表项，并利用查找的转发表项转发该业务报文。相当于在将第一转发表项的状态设置为无效状态后，即完成了第一网络设备的路由切换。

15

本申请实施例提供的链路故障处理方法，第一网络设备的转发芯片可以通过第二网络设备发送的第一报文，获得第二网络设备与第三网络设备之间是否可达，从而及时感知第二网络设备与第三网络设备之间的连接状态。而且第一网络设备的转发芯片还能够在第二网络设备与第三网络设备之间不可达时，及时将下一跳地址为第一报文的源地址、设备标识为第三网络设备的设备标识的第一转发表项设置为无效状态，即禁用包括第二网络设备和第三网络设备之间的链路的转发路径，以便后续可以选择其他转发表项向第三网络设备连接的主机转发报文。即，本申请实施例能够在两个网络设备之间存在链路故障时，及时通知其他网络设备，而且网络设备还可以直接通过转发芯片将故障链路涉及的转发表项的状态设置为无效状态，不涉及控制面对接收到的报文的处理，因此速度更快，能够更及时地进行路由切换，从而提高 AI 流量模型的训练效率。

20

在本申请实施例中，在上述 S301 第一网络设备接收第二网络设备发送的第一报文之前，第一网络设备还可以生成第一转发表项，生成方式包括如下步骤：

25

步骤 1、第一网络设备通过第一接口接收第二网络设备转发的路由更新（route update）报文。

其中，路由更新报文包括第三网络设备连接的主机的第一网际互连协议（Internet Protocol，IP）地址以及第三网络设备的设备标识。

30

作为示例，该路由更新报文具体可以为 BGP route update 报文，在第三网络设备学习到自身连接的主机的 IP 地址后，可以发送携带该 IP 地址的 BGP route update 报文。本申请实施例中，在该 BGP route update 报文中新增了第三网络设备的设备标识，相当于 BGP route update 报文中携带了发出该 BGP route update 报文的原始网络设备的设备标识。以使得其他网络设备接收到该 BGP route update 报文后，获知该 BGP route update 报文的原始来源设备。

步骤 2、第一网络设备基于第一接口、第一 IP 地址和第三网络设备的设备标识生成第一转发表项。

35

可以将第一 IP 地址作为目的地址，将第一接口作为出接口，并结合第三网络设备的设备标识，生成第一转发表项。可以理解的是，该第一转发表项中还包括下一跳地址，下一跳地址为第二网络设备的 IP 地址。

在本申请实施例中，路由更新报文包括扩展团体属性，扩展团体属性包括第三网络设备的设备标识。

扩展团体属性的格式如图 4 所示，扩展团体属性包括：类型（Type）、子类型（Sub-Type）、预留

位 (Reserved, Res) 和设备标识 (Device ID)。

通过上述方法, 本申请实施例可以在第二网络设备向第一网络设备发送的路由更新报文中携带第三网络设备的设备标识, 使得第一网络设备可以得知路由更新报文包括的第一 IP 地址所属的主机接入的网络设备的设备标识, 并记录于转发表项中, 即网络设备可以动态学习远端网络设备的设备标识, 以便
5 后续确定第二网络设备与第三网络设备之间的链路故障时, 更快速地查找到故障链路对应的第一转发表项, 从而更快速地进行路由切换, 提升了故障收敛效率。

而且, 第三网络设备的设备标识可以携带于路由更新报文的扩展团体属性中, 无需额外新增用于承载设备标识的报文, 实现较为便捷。

在本申请实施例中, 第二网络设备包括第二接口, 并通过第二接口与第一网络设备相连。在上述
10 S301 第一网络设备接收第二网络设备发送的第一报文之前, 第一网络设备还可以向第二网络设备发送第二报文。其中, 第二报文包括第一网络设备的设备标识, 以使得第二网络设备记录第二接口与第一网络设备的设备标识之间的第一对应关系。可选的, 第一网络设备可以在与第二网络设备之间的连接建立时, 向第二网络设备发送第二报文。

相应的, 第一网络设备还可以接收第二网络设备发送的第三报文, 其中第三报文包括第二网络设备的设备标识。然后第一网络设备可以记录第一接口与第二网络设备的设备标识之间的第二对应关系。可选的, 第二网络设备可以在与第一网络设备之间的连接建立时, 向第一网络设备发送第三报文, 即第一
15 网络设备可以在与第二网络设备之间的连接建立时, 接收第二网络设备发送的第三报文。

第一网络设备与第二网络设备之间的连接建立时, 第一网络设备设置第一接口的接口状态为运行 (up) 状态, 并向第二网络设备发送第二报文。且第一网络设备与第二网络设备之间的连接建立时, 第
20 二网络设备设置第二接口的接口状态为 up 状态, 并向第一网络设备发送第三报文。

通过上述方法, 第一网络设备可以得知自身的第一接口连接的网络设备的设备标识, 且第二网络设备可以得知自身的第二接口连接的网络设备的设备标识, 即网络设备可以动态学习对端网络设备的设备标识, 维护网络设备间逻辑拓扑关系。使得后续第二网络设备通过报文向其他网络设备通告与第一网络设备之间是否可达时, 可以携带第一网络设备的设备标识, 实现故障快速通知。

25 其他的每两个网络设备之间在连接建立时, 也可以通过上述方式发送携带自身设备标识的报文, 并记录对应关系。

或者, 可以通过静态配置的方式, 预先在每个网络设备中配置该网络设备的每个接口连接的对端网络设备的设备标识。

可选的, 第三报文还可以包括序列号, 序列号用于区分报文的新旧, 即第二网络设备每通过第二接口向第一网络设备发送一次第三报文, 第三报文携带的序列号加 1。在此基础上, 上述第一网络设备记录第一接口与第二网络设备的设备标识之间的第二对应关系的方式, 可以实现为: 记录第一接口、第二网络设备的设备标识与序列号之间的第二对应关系。
30

在本申请实施例中, 上述 S301 中第一网络设备接收的第一报文可以为第一链路层广播报文, 其中第一链路层广播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息。

35 参见图 5, 第一链路层广播报文包括: 第一以太网报头 (Eth)、载荷部分 (payload) 和循环冗余校验 (Cyclic Redundancy Check, CRC) 码。其中, 第一 Eth 包括: 目的媒体访问控制地址 (Destination Source Media Access Control Address, DMAC)、源媒体访问控制地址 (Source Media Access Control Address, SMAC) 和 Type, SMAC 为发送第一链路层广播报文的网络设备自身的 MAC, DMAC 是预

设的组播 MAC。其中，第一链路层广播报文中，DMAC 占用 6 字节 (Bytes)，SMAC 占用 6 Bytes，Type 占用 2 Bytes，payload 占用 46~1500 Bytes，CRC 码占用 4Bytes。其中，本申请实施例中，对第一 Eth 中的 Type 字段的取值进行了改进，从而通过 Type 字段的取值定义该第一链路层广播报文为用于故障通知的报文。Type 字段的取值可根据实际情况设置，本申请实施例对此不作限定。

5 或者，第一报文为第一 IP 组播报文，第一 IP 组播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息。

参见图 6，第一 IP 组播报文包括：第二 Eth、载荷部分和帧检验序列 (Frame Check Sequence, FCS)。其中，第二 Eth 包括：目的地址 (Destination Address)、源地址 (Source Address)、802.1Q Tag 以及长度 (Length)/Type。802.1Q Tag 包括：Type、优先级 (Priority, PRI)、标准格式指示位 (Canonical Format Indicator, CFI) 和虚拟局域网标识 (VLAN ID, VID)。FCS 可以通过 CRC-32 算法得到。其中，
10 第一 IP 组播报文中，Destination Address 占用 6 Bytes；Source Address 占用 6 Bytes；802.1Q Tag 共占用 4Bytes，其中的 Type 占用 2Bytes，PRI 占用 3 比特 (bits)，CFI 占用 1bit，VID 占用 12 bits；Length/Type 占用 2 Bytes；载荷部分占用 45~1500 Bytes；FCS 占用 4 Bytes。其中，本申请实施例中，对第二 Eth 中的 Type 字段的取值进行了改进，从而通过 Type 字段的取值定义该第一 IP 组播报文为用于故障通知的
15 报文。Type 字段的取值可根据实际情况设置，本申请实施例对此不作限定。

可选的，第一报文的载荷部分还可以包括序列号，序列号用于区分报文的新旧。

第一报文的载荷部分的报文格式如图 7 所示，载荷部分包括：版本 (version)、Type、报文长度 (packet-Length)、源设备标识 (SRC Device ID)、目的设备标识 (Dest Device ID)、序列号 (Serial Number)、目的端口标识 (DPortID) 和状态 (Status)。其中，SRC Device ID 为第二网络设备的设备
20 标识，Dest Device ID 为第三网络设备的设备标识，DportID 为第二网络设备中用于连接第三网络设备的第三接口标识，或者虚拟接口标识，虚拟接口表示第二网络设备中与第三网络设备连接的各接口。Status 表示第二网络设备与第三网络设备之间是否可达，具体用于表示 DportID 所表示的接口的接口状态。

在第二网络设备与第三网络设备之间存在一条链路的情况下，第二网络设备通过第三接口与第三网
25 络设备相连，此时 DportID 为第三接口标识。当 Status=0 时，表示第三接口的接口状态为故障状态，即第二网络设备与第三网络设备之间的链路故障；当 Status=1 时，表示第三接口为非故障状态，即第二网络设备与第三网络设备之间的链路未发生故障。

在第二网络设备与第三网络设备之间存在多条链路的情况下，第二网络设备通过多个接口分别与第三网络设备相连，第二网络设备包括的每个接口对应一条链路，此时 DportID 为第二网络设备的虚拟接
30 口标识。当 Status=0 时，表示虚拟接口的接口状态为故障状态，即第二网络设备与第三网络设备之间的链路均存在故障；当 Status=1 时，表示虚拟接口为非故障状态，即第二网络设备与第三网络设备之间存在非故障链路。

其中，第一报文可以是通过 BGP 或者 IGP 生成的报文。

在本申请实施例中，第二报文为第二链路层广播报文，其中第二链路层广播报文的载荷部分包括第
35 一网络设备的设备标识；或者，第二报文为第二 IP 组播报文，其中第二 IP 组播报文的载荷部分包括第一网络设备的设备标识。

同样的，第三报文为第三链路层广播报文，其中第三链路层广播报文的载荷部分包括第二网络设备的设备标识；或者，第三报文为第三 IP 组播报文，其中第三 IP 组播报文的载荷部分包括第二网络设备的设备标识。

的设备标识。

第二报文和第三报文的报头均与第一报文的报头相同，可参考上述描述。

第二报文和第三报文包括的载荷部分的格式相同，如图 8 所示，该载荷部分包括： version, Type, packet-Length、SRC Device ID 和源接口标识（Src PortID）。可选的，参见图 8，第二报文和第三报文的载荷部分还可以包括序列号（Serial Number）。其中，SRC Device ID 为发送报文的源网络设备自身的设备标识，Src PortID 表示源网络设备发送该报文时使用的接口的标识。

在本公开实施例中，第一报文可以称为快速通知报文，第二报文和第三报文可以称为拓扑发现报文，快速通知报文和拓扑发现报文均为本申请实施例新增的报文，两种报文的报头可以相同。网络设备可以通过报文的载荷部分包括的 Type，对这两种报文进行区分。例如，参见图 7，快速通知报文的载荷部分包括的 Type=2，参见图 8，拓扑发现报文的载荷部分包括的 Type=1。

在本申请实施例中，第一网络设备包括至少一个等价多路径路由（Equal-cost multi-path routing, ECMP）组，ECMP 组包括至少两条转发表项，且 ECMP 组包括的各转发表项的目的地址均相同。

本申请实施例中，上述 S303 第一网络设备将第一转发表项的状态设置为无效状态的方式，可以实现在：将第一转发表项的状态标志位修改为第一值。其中，第一值用于表示无效状态，例如第一值为 0。

本申请实施例中，第一网络设备可以直接通过转发芯片管理转发表项，因此提高了路由切换的速度。

在本申请实施例中，在上述 S301 第一网络设备接收第二网络设备发送的第一报文之后，若状态信息表示第二网络设备与第三网络设备之间可达，则第一网络设备可以查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项，并将第一转发表项的状态设置为有效状态。

第一网络设备查找第一转发表项的方式可以参考上述 S302 中的具体实现方式，此处不再赘述。

其中，第一网络设备可以将第一转发表项的状态标志位修改为第二值，其中第二值用于表示有效状态，例如第二值为 1。

通过上述方法，网络设备可以根据接收到的第一报文，对自身存储的转发表项的状态进行及时更新，以便及时进行路由切换。

基于相同的构思，本申请实施例还提供了一种链路故障处理方法，该方法应用于第二网络设备，第二网络设备与第三网络设备相连，第二网络设备连接第一网络设备的第一接口，第一网络设备包括多条转发表项，转发表项包括目的地址、出接口和设备标识，设备标识为目的地址所属的主机接入的网络设备的标识。如图 9 所示，该方法包括如下步骤：

S901、获取状态信息。其中状态信息用于表示第二网络设备与第三网络设备之间是否可达。

S902、向第一网络设备发送第一报文，其中第一报文包括第三网络设备的设备标识以及状态信息，以使得第一网络设备在状态信息表示第二网络设备与第三网络设备之间不可达时，查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项，并将第一转发表项的状态设置为无效状态。

需要说明的是，第一网络设备为：第二网络设备连接的网络设备中，与第二网络设备之间的链路未发生故障的网络设备。第一网络设备可以为一个或多个，本申请实施例以第一网络设备为一个为例进行说明。

第二网络设备以组播或广播的方式向第一网络设备发送第一报文。可选的，第二网络设备可以周期性地发送第一报文。或者第二网络设备可以在与自身连接的网络设备之间的链路的连接状态发生变化时，每间隔预设时长发送一次第一报文，并连续发送预设次数。例如，预设时长为 10 毫秒，预设次数为 10 次。

第一网络设备查找第一转发表项，并将第一转发表项的状态设置为无效状态的具体方式，可以参考上述描述，此处不再赘述。

本申请实施例提供的链路故障处理方法，第一网络设备的转发芯片可以通过第二网络设备发送的第一报文，获得第二网络设备与第三网络设备之间是否可达，从而及时感知第二网络设备与第三网络设备之间的连接状态。而且第一网络设备的转发芯片还能够在第二网络设备与第三网络设备之间不可达时，及时将下一跳地址为第一报文的源地址、设备标识为第三网络设备的设备标识的第一转发表项设置为无效状态，即禁用包括第二网络设备和第三网络设备之间的链路的转发路径，以便后续可以选择其他转发表项向第三网络设备连接的主机转发报文。即，本申请实施例能够在两个网络设备之间存在链路故障时，及时通知其他网络设备，而且网络设备还可以直接通过转发芯片将故障链路涉及的转发表项的状态设置为无效状态，不涉及控制面对接收到的报文的处理，因此速度更快，能够更及时地进行路由切换，从而提高 AI 流量模型的训练效率。

在本申请实施例中，在 S901 第二网络设备获取状态信息之前，第二网络设备还可以：接收第三网络设备发送的路由更新报文，并向第一网络设备转发路由更新报文，以使得第一网络设备基于第一接口、第一 IP 地址和第三网络设备的设备标识生成第一转发表项。其中，路由更新报文包括第三网络设备连接的主机的第一 IP 地址以及第三网络设备的设备标识。

其中，路由更新报文包括扩展团体属性，扩展团体属性包括第三网络设备的设备标识。路由更新报文的格式可参考上述描述。

第一网络设备生成第一转发表项的具体方式，可以参考上述描述，此处不再赘述。

在本申请实施例中，第二网络设备包括第三接口，并通过第三接口与第三网络设备相连，第三网络设备包括第四接口，并通过第四接口与第二网络设备相连。在上述 S901 第二网络设备获取状态信息之前，第二网络设备可以向第三网络设备发送第四报文，以使得第三网络设备记录第四接口与第二网络设备的设备标识之间的第三对应关系。其中第四报文包括第二网络设备的设备标识。可选的，第二网络设备可以在与第三网络设备之间的连接建立时，向第三网络设备发送第四报文。

在上述 S901 第二网络设备获取状态信息之前，第二网络设备还可以接收第三网络设备发送的第五报文，记录第三接口与第三网络设备的设备标识之间的第四对应关系。其中第五报文包括第三网络设备的设备标识。可选的，第三网络设备可以在与第二网络设备之间的连接建立时，向第二网络设备发送第五报文；即，第二网络设备可以在与第三网络设备之间的连接建立时，接收第三网络设备发送的第五报文。

第四报文和第五报文可以称为拓扑发现报文，拓扑发现报文的格式可以参考上述描述。

在本申请实施例中，第二网络设备与第三网络设备之间存在多条链路时，上述 S901 第二网络设备获取状态信息的方式，包括以下两种情况：

情况 1、若第二网络设备与第三网络设备之间的多条链路均故障，则确定状态信息表示第二网络设备与第三网络设备之间不可达。

情况 2、若第二网络设备与第三网络设备之间的多条链路中的任意一条链路未故障，则确定状态信息表示第二网络设备与第三网络设备之间可达。

第二网络设备可以通过 BFD 或者 CFM 的方式，检测每条链路是否故障，或者还可以通过其他方式检测，本申请实施例对此不作具体限定。

在上述 S902 第二网络设备向第一网络设备发送第一报文之前，第二网络设备还可以从第四对应关

系中获取第三接口对应的第三网络设备的设备标识。

在本申请实施例中，第一报文为第一链路层广播报文，第一链路层广播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息；或者，

5 第一报文为第一 IP 组播报文，第一 IP 组播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息。

在本申请实施例中，第四报文为第四链路层广播报文，第四链路层广播报文的载荷部分包括第二网络设备的设备标识；或者，第四报文为第四 IP 组播报文，第四 IP 组播报文的载荷部分包括第二网络设备的设备标识。

10 第五报文为第五链路层广播报文，第五链路层广播报文的载荷部分包括第三网络设备的设备标识；或者，第五报文为第五 IP 组播报文，第五 IP 组播报文的载荷部分包括第三网络设备的设备标识。

其中，第四报文和第五报文均可以称为拓扑发现报文，拓扑发现报文的格式可以参考上述描述。

参见图 10，以下对本申请实施例提供的链路故障处理的完整流程进行说明：

图 10 所示的通信系统包括：主机 (Host) -1、Host-2、Host-3、Host-4、叶 (Leaf) -1、Leaf-2、Leaf-3、Leaf-4、脊 (Spine) -1 和 Spine-2。Leaf-1 的设备 ID 为 11，Leaf-2 的设备 ID 为 12，Leaf-3 的设备 ID 为 13，Leaf-4 的设备 ID 为 14，Spine-1 的设备 ID 为 21，Spine-2 的设备 ID 为 22。Leaf-1 连接 Host-1，Leaf-2 连接 Host-2、Leaf-3 连接 Host-3、Leaf-4 连接 Host-4。Spine-1 和 Spine-2 分别与 Leaf-1~Leaf-4 连接，且 Spine 和 Leaf 之间基于 BGP 协议通信。每个 Leaf 通过自身不同的接口连接 Spine-1 和 Spine-2，例如 Leaf-1 中连接 Spine-1 的接口为 leaf-1-interface-1，Leaf-1 中连接 Spine-2 的接口为 leaf-1-interface-2。每个 Leaf 中存储有除自身连接的 Host 以外的其他 Host 的转发表项。

20 例如，Leaf-1 中存储的转发表项如下：

目的地址为 Host-2 地址的 ECMP 1 包括的转发表项：

下一跳 1 指向 Spine-1，即下一跳地址 1 为 Spine-1 的地址，出接口为 leaf-1-interface-1，设备 ID 为 12；

下一跳 2 指向 Spine-2，出接口为 leaf-1-interface-2，设备 ID 为 12。

25 目的地址为 Host-3 地址的 ECMP 2 包括的转发表项：

下一跳 1 指向 Spine-1，出接口为 leaf-1-interface-1，设备 ID 为 13；

下一跳 2 指向 Spine-2，出接口为 leaf-1-interface-2，设备 ID 为 13。

目的地址为 Host-4 地址的 ECMP 3 包括的转发表项：

下一跳 1 指向 Spine-1，出接口为 leaf-1-interface-1，设备 ID 为 14；

30 下一跳 2 指向 Spine-2，出接口为 leaf-1-interface-2，设备 ID 为 14。

假设上述第一网络设备为 Leaf-1，第二网络设备为 Spine-1，第三网络设备为 Leaf-4，即，假设 Spine-1 和 Leaf-4 之间的链路故障，Spine-1 可以向 Leaf-1 发送第一报文，第一报文包括：设备 ID 21、设备 ID 14 和 Status=0。

35 Leaf-1 接收到第一报文后，查找设备 ID 为 14，且出接口为 leaf-1-interface-1 的转发表项，即 ECMP 3 包括的第一条转发表项，将该转发表项的状态设置为无效状态。后续 Leaf-1 需要向 Host-4 转发报文时，可以基于 ECMP 3 包括的其他转发表项，即基于“下一跳 2 指向 Spine-2，出接口为 leaf-1-interface-2，设备 ID 为 14”进行报文转发。

通过上述方式，本申请实施例可以在两个网络设备之间的链路故障，即两个网络设备之间不可达时，

快速通知其他网络设备,以便其他网络设备能够根据出接口标识和设备标识快速查找到故障链路对应的转发表项,从而实现快速的路由切换,提高了故障收敛效率,从而提高了 AI 流量模型的训练效率。

基于相同的构思,本申请实施例还提供了一种链路故障处理装置,应用于第一网络设备的转发芯片,第一网络设备包括第一接口,并通过第一接口与第二网络设备相连,第二网络设备与第三网络设备相连,第一网络设备包括多条转发表项,转发表项包括目的地址、出接口和设备标识,设备标识为目的地址所属的主机接入的网络设备的标识;如图 11 所示,该装置包括:接收模块 1101、查找模块 1102 和设置模块 1103;

接收模块 1101,用于接收第二网络设备发送的第一报文,第一报文包括第三网络设备的设备标识以及状态信息,状态信息用于表示第二网络设备与第三网络设备之间是否可达;

查找模块 1102,用于若状态信息表示第二网络设备与第三网络设备之间不可达,则查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项;

设置模块 1103,用于将第一转发表项的状态设置为无效状态。

可选地,装置还包括:

接收模块 1101,还用于在接收第二网络设备发送的第一报文之前,通过第一接口接收第二网络设备转发的路由更新报文,路由更新报文包括第三网络设备连接的主机的第一 IP 地址以及第三网络设备的设备标识;

生成模块,用于基于第一接口、第一 IP 地址和第三网络设备的设备标识生成第一转发表项。

可选地,第二网络设备包括第二接口,并通过第二接口与第一网络设备相连;该装置还包括:

发送模块,还用于在接收第二网络设备发送的第一报文之前,向第二网络设备发送第二报文,第二报文包括第一网络设备的设备标识,以使得第二网络设备记录第二接口与第一网络设备的设备标识之间的第一对应关系;

接收模块 1101,还用于接收第二网络设备发送的第三报文,第三报文包括第二网络设备的设备标识,记录第一接口与第二网络设备的设备标识之间的第二对应关系。

可选地,设置模块 1103,具体用于:

将第一转发表项的状态标志位修改为第一值,第一值用于表示无效状态。
可选地,查找模块 1102,还用于在接收第二网络设备发送的第一报文之后,若状态信息表示第二网络设备与第三网络设备之间可达,则查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项;

设置模块 1103,还用于将第一转发表项的状态设置为有效状态。

可选地,第一报文为第一链路层广播报文,第一链路层广播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息;或者,第一报文为第一 IP 组播报文,第一 IP 组播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息。

可选地,第二报文为第二链路层广播报文,第二链路层广播报文的载荷部分包括第一网络设备的设备标识;或者,第二报文为第二 IP 组播报文,第二 IP 组播报文的载荷部分包括第一网络设备的设备标识;

第三报文为第三链路层广播报文,第三链路层广播报文的载荷部分包括第二网络设备的设备标识;或者,第三报文为第三 IP 组播报文,第三 IP 组播报文的载荷部分包括第二网络设备的设备标识。

可选地,路由更新报文包括扩展团体属性,扩展团体属性包括第三网络设备的设备标识。

基于相同的构思，本申请实施例还提供一种链路故障处理装置，应用于第二网络设备的转发芯片，第二网络设备与第三网络设备相连，第二网络设备连接第一网络设备的第一接口，第一网络设备包括多条转发表项，转发表项包括目的地址、出接口和设备标识，设备标识为目的地址所属的主机接入的网络设备的标识；如图 12 所示，该装置包括：获取模块 1201 和发送模块 1202；

5 获取模块 1201，用于获取状态信息，状态信息用于表示第二网络设备与第三网络设备之间是否可达；

发送模块 1202，用于向第一网络设备发送第一报文，第一报文包括第三网络设备的设备标识以及状态信息，以使得第一网络设备在状态信息表示第二网络设备与第三网络设备之间不可达时，查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项，并将第一转发表项的状态设置为
10 无效状态。

可选地，该装置还包括：

接收模块，用于在获取状态信息之前，接收第三网络设备发送的路由更新报文，路由更新报文包括第三网络设备连接的主机的第一 IP 地址以及第三网络设备的设备标识；

发送模块 1202，还用于向第一网络设备转发路由更新报文，以使得第一网络设备基于第一接口、
15 第一 IP 地址和第三网络设备的设备标识生成第一转发表项。

可选地，第二网络设备包括第三接口，并通过第三接口与第三网络设备相连，第三网络设备包括第四接口，并通过第四接口与第二网络设备相连；装置还包括：

发送模块 1202，还用于在获取状态信息之前，向第三网络设备发送第四报文，第四报文包括第二网络设备的设备标识，以使得第三网络设备记录第四接口与第二网络设备的设备标识之间的第三对应关
20 系；

接收模块，用于接收第三网络设备发送的第五报文，第五报文包括第三网络设备的设备标识，记录第三接口与第三网络设备的设备标识之间的第四对应关系。

可选地，获取模块 1201，还用于在向第一网络设备发送第一报文之前，从第四对应关系中获取第三接口对应的第三网络设备的设备标识。

25 可选地，第二网络设备与第三网络设备之间存在多条链路；获取模块 1201，具体用于：

若多条链路均故障，则确定状态信息表示第二网络设备与第三网络设备之间不可达；或者，

若多条链路中的任意一条链路未故障，则确定状态信息表示第二网络设备与第三网络设备之间可达。

可选地，第一报文为第一链路层广播报文，第一链路层广播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息；或者，

30 第一报文为第一 IP 组播报文，第一 IP 组播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息。

可选地，第四报文为第四链路层广播报文，第四链路层广播报文的载荷部分包括第二网络设备的设备标识；或者，第四报文为第四 IP 组播报文，第四 IP 组播报文的载荷部分包括第二网络设备的设备标识；

35 第五报文为第五链路层广播报文，第五链路层广播报文的载荷部分包括第三网络设备的设备标识；或者，第五报文为第五 IP 组播报文，第五 IP 组播报文的载荷部分包括第三网络设备的设备标识。

可选地，路由更新报文包括扩展团体属性，扩展团体属性包括第三网络设备的设备标识。

基于相同的构思，本申请实施例提供了一种第一网络设备，第一网络设备包括第一接口，并通过第

一接口与第二网络设备相连，第二网络设备与第三网络设备相连，第一网络设备包括多条转发表项，转发表项包括目的地址、出接口和设备标识，设备标识为目的地址所属的主机接入的网络设备的标识；如图 13 所示，第一网络设备包括：处理器 1301；收发器 1304；机器可读存储介质 1302，机器可读存储介质 1302 存储有能够被处理器 1301 执行的机器可执行指令；机器可执行指令促使处理器 1301 执行以下步骤：

5 接收第二网络设备发送的第一报文，第一报文包括第三网络设备的设备标识以及状态信息，状态信息用于表示第二网络设备与第三网络设备之间是否可达；

 若状态信息表示第二网络设备与第三网络设备之间不可达，则查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项；

10 将第一转发表项的状态设置为无效状态。

 可选地，机器可执行指令还促使处理器 1301 执行以下步骤：

 在接收第二网络设备发送的第一报文之前，通过第一接口接收第二网络设备转发的路由更新报文，路由更新报文包括第三网络设备连接的主机的第一 IP 地址以及第三网络设备的设备标识；

 基于第一接口、第一 IP 地址和第三网络设备的设备标识生成第一转发表项。

15 可选地，第二网络设备包括第二接口，并通过第二接口与第一网络设备相连；机器可执行指令还促使处理器 1301 执行以下步骤：

 在接收第二网络设备发送的第一报文之前，向第二网络设备发送第二报文，第二报文包括第一网络设备的设备标识，以使得第二网络设备记录第二接口与第一网络设备的设备标识之间的第一对应关系；

20 接收第二网络设备发送的第三报文，第三报文包括第二网络设备的设备标识，记录第一接口与第二网络设备的设备标识之间的第二对应关系。

 可选地，机器可执行指令还促使处理器 1301 执行以下步骤：

 将第一转发表项的状态标志位修改为第一值，第一值用于表示无效状态。

 可选地，机器可执行指令还促使处理器 1301 执行以下步骤：

25 在接收第二网络设备发送的第一报文之后，若状态信息表示第二网络设备与第三网络设备之间可达，则查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项；

 将第一转发表项的状态设置为有效状态。

 可选地，第一报文为第一链路层广播报文，第一链路层广播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息；或者，

30 第一报文为第一 IP 组播报文，第一 IP 组播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息。

 可选地，第二报文为第二链路层广播报文，第二链路层广播报文的载荷部分包括第一网络设备的设备标识；或者，第二报文为第二 IP 组播报文，第二 IP 组播报文的载荷部分包括第一网络设备的设备标识；

35 第三报文为第三链路层广播报文，第三链路层广播报文的载荷部分包括第二网络设备的设备标识；或者，第三报文为第三 IP 组播报文，第三 IP 组播报文的载荷部分包括第二网络设备的设备标识。

 可选地，路由更新报文包括扩展团体属性，扩展团体属性包括第三网络设备的设备标识。

 在图 13 中，还可以包括通信总线 1303。处理器 1301、机器可读存储介质 1302 及收发器 1304 之间通过通信总线 1303 完成相互间的通信，通信总线 1303 可以是外设部件互连标准（Peripheral Component

Interconnect, PCI) 总线或扩展工业标准结构 (Extended Industry Standard Architecture, EISA) 总线等。该通信总线可以分为地址总线、数据总线、控制总线等。收发器 1304 可以为无线通信模块, 收发器 1304 在处理器 1301 的控制下, 与其他设备进行数据交互。

5 机器可读存储介质 1302 可以包括随机存取存储器 (Random Access Memory, RAM), 也可以包括非易失性存储器 (Non-Volatile Memory, NVM), 例如至少一个磁盘存储器。另外, 机器可读存储介质还可以是至少一个位于远离前述处理器的存储装置。

10 处理器 1301 可以是通用处理器, 包括中央处理器 (Central Processing Unit, CPU)、网络处理器 (Network Processor, NP) 等; 还可以是数字信号处理器 (Digital Signal Processing, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 或其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

基于相同的构思, 本申请实施例提供一种第二网络设备, 第二网络设备与第三网络设备相连, 第二网络设备连接第一网络设备的第一接口, 第一网络设备包括多条转发表项, 转发表项包括目的地址、出接口和设备标识, 设备标识为目的地址所属的主机接入的网络设备的标识; 如图 14 所示, 第二网络设备包括: 处理器 1401; 收发器 1404; 机器可读存储介质 1402, 机器可读存储介质 1402 存储有能够被
15 处理器 1401 执行的机器可执行指令; 机器可执行指令促使处理器 1401 执行以下步骤:

获取状态信息, 状态信息用于表示第二网络设备与第三网络设备之间是否可达;

向第一网络设备发送第一报文, 第一报文包括第三网络设备的设备标识以及状态信息, 以使得第一网络设备在状态信息表示第二网络设备与第三网络设备之间不可达时, 查找出接口为第一接口、设备标识为第三网络设备的设备标识的第一转发表项, 并将第一转发表项的状态设置为无效状态。

20 可选地, 机器可执行指令还促使处理器 1401 执行以下步骤:

在获取状态信息之前, 接收第三网络设备发送的路由更新报文, 路由更新报文包括第三网络设备连接的主机的第一 IP 地址以及第三网络设备的设备标识;

向第一网络设备转发路由更新报文, 以使得第一网络设备基于第一接口、第一 IP 地址和第三网络设备的设备标识生成第一转发表项。

25 可选地, 第二网络设备包括第三接口, 并通过第三接口与第三网络设备相连, 第三网络设备包括第四接口, 并通过第四接口与第二网络设备相连; 机器可执行指令还促使处理器 1401 执行以下步骤:

在获取状态信息之前, 向第三网络设备发送第四报文, 第四报文包括第二网络设备的设备标识, 以使得第三网络设备记录第四接口与第二网络设备的设备标识之间的第三对应关系;

30 接收第三网络设备发送的第五报文, 第五报文包括第三网络设备的设备标识, 记录第三接口与第三网络设备的设备标识之间的第四对应关系。

可选地, 机器可执行指令还促使处理器 1401 执行以下步骤: 在向第一网络设备发送第一报文之前, 从第四对应关系中获取第三接口对应的第三网络设备的设备标识。

可选地, 第二网络设备与第三网络设备之间存在多条链路; 机器可执行指令还促使处理器 1401 执行以下步骤:

35 若多条链路均故障, 则确定状态信息表示第二网络设备与第三网络设备之间不可达; 或者,

若多条链路中的任意一条链路未故障, 则确定状态信息表示第二网络设备与第三网络设备之间可达。

可选地, 第一报文为第一链路层广播报文, 第一链路层广播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息; 或者,

第一报文为第一 IP 组播报文，第一 IP 组播报文的载荷部分包括第二网络设备的设备标识、第三网络设备的设备标识和状态信息。

可选地，第四报文为第四链路层广播报文，第四链路层广播报文的载荷部分包括第二网络设备的设备标识；或者，第四报文为第四 IP 组播报文，第四 IP 组播报文的载荷部分包括第二网络设备的设备标识；

第五报文为第五链路层广播报文，第五链路层广播报文的载荷部分包括第三网络设备的设备标识；或者，第五报文为第五 IP 组播报文，第五 IP 组播报文的载荷部分包括第三网络设备的设备标识。

可选地，路由更新报文包括扩展团体属性，扩展团体属性包括第三网络设备的设备标识。

在图 14 中，还可以包括通信总线 1403。处理器 1401、机器可读存储介质 1402 及收发器 1404 之间通过通信总线 1403 完成相互间的通信，通信总线 1403 可以是外设部件互连标准（Peripheral Component Interconnect, PCI）总线或扩展工业标准结构（Extended Industry Standard Architecture, EISA）总线等。该通信总线可以分为地址总线、数据总线、控制总线等。收发器 1404 可以为无线通信模块，收发器 1404 在处理器 1401 的控制下，与其他设备进行数据交互。

机器可读存储介质 1402 可以包括随机存取存储器（Random Access Memory, RAM），也可以包括非易失性存储器（Non-Volatile Memory, NVM），例如至少一个磁盘存储器。另外，机器可读存储介质还可以是至少一个位于远离前述处理器的存储装置。

处理器 1401 可以是通用处理器，包括中央处理器（Central Processing Unit, CPU）、网络处理器（Network Processor, NP）等；还可以是数字信号处理器（Digital Signal Processing, DSP）、专用集成电路（Application Specific Integrated Circuit, ASIC）、现场可编程门阵列（Field-Programmable Gate Array, FPGA）或其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

基于同一种构思，根据上述本申请实施例提供的链路故障处理方法，本申请实施例还提供了一种机器可读存储介质，机器可读存储介质存储有能够被处理器执行的机器可执行指令。处理器被机器可执行指令促使实现上述任一链路故障处理方法的步骤。

在本申请提供的又一实施例中，还提供了一种包含指令的计算机程序产品，当其在计算机上运行时，使得计算机执行上述实施例中任一链路故障处理方法的步骤。

需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

本说明书中的各个实施例均采用相关的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

以上所述仅为本发明的较佳实施例，并不用以限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明保护的范围之内。

权 利 要 求

- 1、一种链路故障处理方法，其特征在于，应用于第一网络设备的转发芯片，所述第一网络设备包括第一接口，并通过所述第一接口与第二网络设备相连，所述第二网络设备与第三网络设备相连，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述方法包括：
- 5 接收所述第二网络设备发送的第一报文，所述第一报文包括所述第三网络设备的设备标识以及状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；
- 若所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；
- 10 将所述第一转发表项的状态设置为无效状态。
- 2、根据权利要求1所述的方法，其特征在于，在所述接收所述第二网络设备发送的第一报文之前，所述方法还包括：
- 通过所述第一接口接收所述第二网络设备转发的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一IP地址以及所述第三网络设备的设备标识；
- 15 基于所述第一接口、所述第一IP地址和所述第三网络设备的设备标识生成所述第一转发表项。
- 3、根据权利要求1所述的方法，其特征在于，所述第二网络设备包括第二接口，并通过所述第二接口与所述第一网络设备相连；在所述接收所述第二网络设备发送的第一报文之前，所述方法还包括：
- 向所述第二网络设备发送第二报文，所述第二报文包括所述第一网络设备的设备标识，以使得所述第二网络设备记录所述第二接口与所述第一网络设备的设备标识之间的第一对应关系；
- 20 接收所述第二网络设备发送的第三报文，所述第三报文包括所述第二网络设备的设备标识，记录所述第一接口与所述第二网络设备的设备标识之间的第二对应关系。
- 4、根据权利要求1-3任一项所述的方法，其特征在于，所述将所述第一转发表项的状态设置为无效状态，包括：
- 将所述第一转发表项的状态标志位修改为第一值，所述第一值用于表示无效状态。
- 25 5、根据权利要求1所述的方法，其特征在于，在所述接收所述第二网络设备发送的第一报文之后，所述方法还包括：
- 若所述状态信息表示所述第二网络设备与所述第三网络设备之间可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；
- 将所述第一转发表项的状态设置为有效状态。
- 30 6、根据权利要求1所述的方法，其特征在于，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，
- 所述第一报文为第一IP组播报文，所述第一IP组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。
- 35 7、根据权利要求3所述的方法，其特征在于，所述第二报文为第二链路层广播报文，所述第二链路层广播报文的载荷部分包括所述第一网络设备的设备标识；或者，所述第二报文为第二IP组播报文，所述第二IP组播报文的载荷部分包括所述第一网络设备的设备标识；
- 所述第三报文为第三链路层广播报文，所述第三链路层广播报文的载荷部分包括所述第二网络设备

的设备标识；或者，所述第三报文为第三 IP 组播报文，所述第三 IP 组播报文的载荷部分包括所述第二网络设备的设备标识。

8、根据权利要求 2 所述的方法，其特征在于，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

5 9、一种链路故障处理方法，其特征在于，应用于第二网络设备的转发芯片，所述第二网络设备与第三网络设备相连，所述第二网络设备连接第一网络设备的第一接口，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述方法包括：

获取状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

10 向所述第一网络设备发送第一报文，所述第一报文包括所述第三网络设备的设备标识以及所述状态信息，以使得所述第一网络设备在所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达时，查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项，并将所述第一转发表项的状态设置为无效状态。

10、根据权利要求 9 所述的方法，其特征在于，在所述获取状态信息之前，所述方法还包括：

15 接收所述第三网络设备发送的路由更新报文，所述路由更新报文包括所述第三网络设备连接的宿主机的第一 IP 地址以及所述第三网络设备的设备标识；

向所述第一网络设备转发所述路由更新报文，以使得所述第一网络设备基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

20 11、根据权利要求 9 所述的方法，其特征在于，所述第二网络设备包括第三接口，并通过所述第三接口与所述第三网络设备相连，所述第三网络设备包括第四接口，并通过所述第四接口与所述第二网络设备相连；在所述获取状态信息之前，所述方法还包括：

向所述第三网络设备发送第四报文，所述第四报文包括所述第二网络设备的设备标识，以使得所述第三网络设备记录所述第四接口与所述第二网络设备的设备标识之间的第三对应关系；

25 接收所述第三网络设备发送的第五报文，所述第五报文包括所述第三网络设备的设备标识，记录所述第三接口与所述第三网络设备的设备标识之间的第四对应关系。

12、根据权利要求 11 所述的方法，其特征在于，在向所述第一网络设备发送第一报文之前，所述方法还包括：

从所述第四对应关系中获取所述第三接口对应的所述第三网络设备的设备标识。

30 13、根据权利要求 9 所述的方法，其特征在于，所述第二网络设备与所述第三网络设备之间存在多条链路；所述获取状态信息，包括：

若所述多条链路均故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达；或者，

若所述多条链路中的任意一条链路未故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间可达。

35 14、根据权利要求 9 所述的方法，其特征在于，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备

标识、所述第三网络设备的设备标识和所述状态信息。

15、根据权利要求 11 所述的方法，其特征在于，所述第四报文为第四链路层广播报文，所述第四链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第四报文为第四 IP 组播报文，所述第四 IP 组播报文的载荷部分包括所述第二网络设备的设备标识；

5 所述第五报文为第五链路层广播报文，所述第五链路层广播报文的载荷部分包括所述第三网络设备的设备标识；或者，所述第五报文为第五 IP 组播报文，所述第五 IP 组播报文的载荷部分包括所述第三网络设备的设备标识。

16、根据权利要求 10 所述的方法，其特征在于，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

10 17、一种链路故障处理装置，其特征在于，应用于第一网络设备的转发芯片，所述第一网络设备包括第一接口，并通过所述第一接口与第二网络设备相连，所述第二网络设备与第三网络设备相连，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述装置包括：

接收模块，用于接收所述第二网络设备发送的第一报文，所述第一报文包括所述第三网络设备的设备标识以及状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

15 查找模块，用于若所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

设置模块，用于将所述第一转发表项的状态设置为无效状态。

18、根据权利要求 17 所述的装置，其特征在于，所述装置还包括：

20 所述接收模块，还用于在所述接收所述第二网络设备发送的第一报文之前，通过所述第一接口接收所述第二网络设备转发的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

生成模块，用于基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

19、根据权利要求 17 所述的方法，其特征在于，所述第二网络设备包括第二接口，并通过所述第二接口与所述第一网络设备相连；所述装置还包括：

发送模块，用于在所述接收所述第二网络设备发送的第一报文之前，向所述第二网络设备发送第二报文，所述第二报文包括所述第一网络设备的设备标识，以使得所述第二网络设备记录所述第二接口与所述第一网络设备的设备标识之间的第一对应关系；

30 所述接收模块，还用于接收所述第二网络设备发送的第三报文，所述第三报文包括所述第二网络设备的设备标识，记录所述第一接口与所述第二网络设备的设备标识之间的第二对应关系。

20、根据权利要求 17-19 任一项所述的装置，其特征在于，所述设置模块，具体用于：

将所述第一转发表项的状态标志位修改为第一值，所述第一值用于表示无效状态。

21、根据权利要求 17 所述的装置，其特征在于，

35 所述查找模块，还用于在所述接收所述第二网络设备发送的第一报文之后，若所述状态信息表示所述第二网络设备与所述第三网络设备之间可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

所述设置模块，还用于将所述第一转发表项的状态设置为有效状态。

22、根据权利要求 17 所述的装置，其特征在于，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

5 所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

23、根据权利要求 19 所述的装置，其特征在于，所述第二报文为第二链路层广播报文，所述第二链路层广播报文的载荷部分包括所述第一网络设备的设备标识；或者，所述第二报文为第二 IP 组播报文，所述第二 IP 组播报文的载荷部分包括所述第一网络设备的设备标识；

10 所述第三报文为第三链路层广播报文，所述第三链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第三报文为第三 IP 组播报文，所述第三 IP 组播报文的载荷部分包括所述第二网络设备的设备标识。

24、根据权利要求 18 所述的装置，其特征在于，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

25、一种链路故障处理装置，其特征在于，应用于第二网络设备的转发芯片，所述第二网络设备与第三网络设备相连，所述第二网络设备连接第一网络设备的第一接口，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述装置包括：

获取模块，用于获取状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

20 发送模块，用于向所述第一网络设备发送第一报文，所述第一报文包括所述第三网络设备的设备标识以及所述状态信息，以使得所述第一网络设备在所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达时，查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项，并将所述第一转发表项的状态设置为无效状态。

26、根据权利要求 25 所述的装置，其特征在于，所述装置还包括：

25 接收模块，用于在所述获取状态信息之前，接收所述第三网络设备发送的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

所述发送模块，还用于向所述第一网络设备转发所述路由更新报文，以使得所述第一网络设备基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

30 27、根据权利要求 25 所述的装置，其特征在于，所述第二网络设备包括第三接口，并通过所述第三接口与所述第三网络设备相连，所述第三网络设备包括第四接口，并通过所述第四接口与所述第二网络设备相连；所述装置还包括：

所述发送模块，还用于在所述获取状态信息之前，向所述第三网络设备发送第四报文，所述第四报文包括所述第二网络设备的设备标识，以使得所述第三网络设备记录所述第四接口与所述第二网络设备的设备标识之间的第三对应关系；

35 接收模块，用于接收所述第三网络设备发送的第五报文，所述第五报文包括所述第三网络设备的设备标识，记录所述第三接口与所述第三网络设备的设备标识之间的第四对应关系。

28、根据权利要求 27 所述的装置，其特征在于，

所述获取模块，还用于在向所述第一网络设备发送第一报文之前，从所述第四对应关系中获取所述

第三接口对应的所述第三网络设备的设备标识。

29、根据权利要求 25 所述的装置，其特征在于，所述第二网络设备与所述第三网络设备之间存在多条链路；所述获取模块，具体用于：

5 若所述多条链路均故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达；或者，

若所述多条链路中的任意一条链路未故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间可达。

30、根据权利要求 25 所述的装置，其特征在于，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

10 所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

31、根据权利要求 27 所述的装置，其特征在于，所述第四报文为第四链路层广播报文，所述第四链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第四报文为第四 IP 组播报文，所述第四 IP 组播报文的载荷部分包括所述第二网络设备的设备标识；

所述第五报文为第五链路层广播报文，所述第五链路层广播报文的载荷部分包括所述第三网络设备的设备标识；或者，所述第五报文为第五 IP 组播报文，所述第五 IP 组播报文的载荷部分包括所述第三网络设备的设备标识。

32、根据权利要求 26 所述的装置，其特征在于，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

33、一种第一网络设备，其特征在于，所述第一网络设备包括第一接口，并通过所述第一接口与第二网络设备相连，所述第二网络设备与第三网络设备相连，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述第一网络设备包括：

25 处理器；

收发器；

机器可读存储介质，所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令；所述机器可执行指令促使所述处理器执行以下步骤：

30 接收所述第二网络设备发送的第一报文，所述第一报文包括所述第三网络设备的设备标识以及状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

若所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

将所述第一转发表项的状态设置为无效状态。

34、根据权利要求 33 所述的第一网络设备，其特征在于，所述机器可执行指令还促使所述处理器执行以下步骤：

35 在所述接收所述第二网络设备发送的第一报文之前，通过所述第一接口接收所述第二网络设备转发的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

35、根据权利要求 33 所述的第一网络设备，其特征在于，所述第二网络设备包括第二接口，并通过所述第二接口与所述第一网络设备相连；所述机器可执行指令还促使所述处理器执行以下步骤：

5 在所述接收所述第二网络设备发送的第一报文之前，向所述第二网络设备发送第二报文，所述第二报文包括所述第一网络设备的设备标识，以使得所述第二网络设备记录所述第二接口与所述第一网络设备的设备标识之间的第一对应关系；

接收所述第二网络设备发送的第三报文，所述第三报文包括所述第二网络设备的设备标识，记录所述第一接口与所述第二网络设备的设备标识之间的第二对应关系。

10 36、根据权利要求 33-35 任一项所述的第一网络设备，其特征在于，所述机器可执行指令还促使所述处理器执行以下步骤：

将所述第一转发表项的状态标志位修改为第一值，所述第一值用于表示无效状态。

37、根据权利要求 33 所述的第一网络设备，其特征在于，所述机器可执行指令还促使所述处理器执行以下步骤：

15 在所述接收所述第二网络设备发送的第一报文之后，若所述状态信息表示所述第二网络设备与所述第三网络设备之间可达，则查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项；

将所述第一转发表项的状态设置为有效状态。

20 38、根据权利要求 33 所述的第一网络设备，其特征在于，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

25 39、根据权利要求 35 所述的第一网络设备，其特征在于，所述第二报文为第二链路层广播报文，所述第二链路层广播报文的载荷部分包括所述第一网络设备的设备标识；或者，所述第二报文为第二 IP 组播报文，所述第二 IP 组播报文的载荷部分包括所述第一网络设备的设备标识；

所述第三报文为第三链路层广播报文，所述第三链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第三报文为第三 IP 组播报文，所述第三 IP 组播报文的载荷部分包括所述第二网络设备的设备标识。

30 40、根据权利要求 34 所述的第一网络设备，其特征在于，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

41、一种第二网络设备，其特征在于，所述第二网络设备与第三网络设备相连，所述第二网络设备连接第一网络设备的第二接口，所述第一网络设备包括多条转发表项，所述转发表项包括目的地址、出接口和设备标识，所述设备标识为所述目的地址所属的主机接入的网络设备的标识；所述第二网络设备包括：

35 处理器；

收发器；

机器可读存储介质，所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令；所述机器可执行指令促使所述处理器执行以下步骤：

获取状态信息，所述状态信息用于表示所述第二网络设备与所述第三网络设备之间是否可达；

向所述第一网络设备发送第一报文，所述第一报文包括所述第三网络设备的设备标识以及所述状态信息，以使得所述第一网络设备在所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达时，查找出接口为所述第一接口、设备标识为所述第三网络设备的设备标识的第一转发表项，并将所述
5 第一转发表项的状态设置为无效状态。

42、根据权利要求 41 所述的第二网络设备，其特征在于，所述机器可执行指令还促使所述处理器执行以下步骤：

在所述获取状态信息之前，接收所述第三网络设备发送的路由更新报文，所述路由更新报文包括所述第三网络设备连接的主机的第一 IP 地址以及所述第三网络设备的设备标识；

10 向所述第一网络设备转发所述路由更新报文，以使得所述第一网络设备基于所述第一接口、所述第一 IP 地址和所述第三网络设备的设备标识生成所述第一转发表项。

43、根据权利要求 41 所述的第二网络设备，其特征在于，所述第二网络设备包括第三接口，并通过所述第三接口与所述第三网络设备相连，所述第三网络设备包括第四接口，并通过所述第四接口与所述第二网络设备相连，所述机器可执行指令还促使所述处理器执行以下步骤：

15 在所述获取状态信息之前，向所述第三网络设备发送第四报文，所述第四报文包括所述第二网络设备的设备标识，以使得所述第三网络设备记录所述第四接口与所述第二网络设备的设备标识之间的第三对应关系；

接收所述第三网络设备发送的第五报文，所述第五报文包括所述第三网络设备的设备标识，记录所述第三接口与所述第三网络设备的设备标识之间的第四对应关系。

20 44、根据权利要求 43 所述的第二网络设备，其特征在于，所述机器可执行指令还促使所述处理器执行以下步骤：

在向所述第一网络设备发送第一报文之前，从所述第四对应关系中获取所述第三接口对应的所述第三网络设备的设备标识。

25 45、根据权利要求 41 所述的第二网络设备，其特征在于，所述第二网络设备与所述第三网络设备之间存在多条链路；所述机器可执行指令还促使所述处理器执行以下步骤：

若所述多条链路均故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间不可达；或者，

若所述多条链路中的任意一条链路未故障，则确定所述状态信息表示所述第二网络设备与所述第三网络设备之间可达。

30 46、根据权利要求 41 所述的第二网络设备，其特征在于，所述第一报文为第一链路层广播报文，所述第一链路层广播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息；或者，

所述第一报文为第一 IP 组播报文，所述第一 IP 组播报文的载荷部分包括所述第二网络设备的设备标识、所述第三网络设备的设备标识和所述状态信息。

35 47、根据权利要求 43 所述的第二网络设备，其特征在于，所述第四报文为第四链路层广播报文，所述第四链路层广播报文的载荷部分包括所述第二网络设备的设备标识；或者，所述第四报文为第四 IP 组播报文，所述第四 IP 组播报文的载荷部分包括所述第二网络设备的设备标识；

所述第五报文为第五链路层广播报文，所述第五链路层广播报文的载荷部分包括所述第三网络设备

的设备标识；或者，所述第五报文为第五 IP 组播报文，所述第五 IP 组播报文的载荷部分包括所述第三网络设备的设备标识。

48、根据权利要求 42 所述的第二网络设备，其特征在于，所述路由更新报文包括扩展团体属性，所述扩展团体属性包括所述第三网络设备的设备标识。

5 49、一种机器可读存储介质，其特征在于，存储有机器可执行指令，在被处理器调用和执行时，所述机器可执行指令促使所述处理器：实现权利要求 1-8 或 9-16 任一项所述的方法。

50、一种计算机程序产品，其特征在于，所述计算机程序产品促使所述处理器：实现权利要求 1-8 或 9-16 任一项所述的方法。