

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2017 年 9 月 21 日 (21.09.2017)



(10) 国际公布号
WO 2017/157116 A1

- (51) 国际专利分类号:
H04L 12/70 (2013.01) *H04L 12/801* (2013.01)
- (21) 国际申请号: PCT/CN2017/073368
- (22) 国际申请日: 2017 年 2 月 13 日 (13.02.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610143703.9 2016 年 3 月 14 日 (14.03.2016) CN
- (71) 申请人: 深圳市中兴微电子有限公司 (SANE-CHIPS TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省深圳市南山区西丽街道留仙大道中兴工业园, Guangdong 518055 (CN)。
- (72) 发明人: 薛红娟 (XUE, Hongjuan); 中国广东省深圳市南山区西丽街道留仙大道中兴工业园, Guangdong 518055 (CN)。 曾敏 (ZENG, Min); 中国广东省深圳市南山区西丽街道留仙大道中兴工业园, Guangdong 518055 (CN)。 贺志强 (HE, Zhiqiang); 中国广东省深圳市南山区西丽街道留仙大道中兴工业园, Guangdong 518055 (CN)。 冯越 (FENG,

Yue); 中国广东省深圳市南山区西丽街道留仙大道中兴工业园, Guangdong 518055 (CN)。

- (74) 代理人: 北京派特恩知识产权代理有限公司 (CHINA PAT INTELLECTUAL PROPERTY OFFICE); 中国北京市海淀区海淀南路 21 号中关村知识产权大厦 B 座 2 层, Beijing 100080 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TS, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE,

[见续页]

(54) Title: TRAFFIC CONGESTION CONTROL METHOD AND DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种流量拥塞控制方法、装置及存储介质

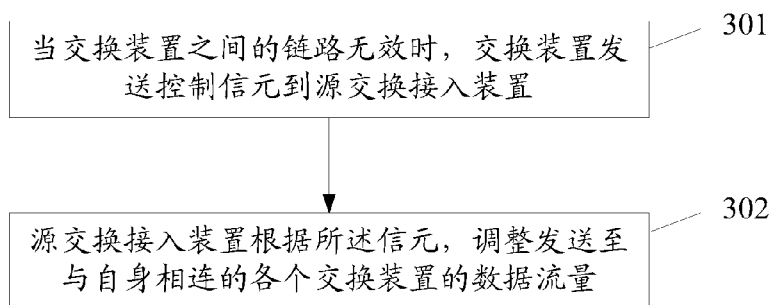


图 3

- 301 When a link between switch elements is invalid, a switch element transmits a control cell to a source switch interface
- 302 The source switch interface adjusts, according to the cell, the number of data streams transmitted to respective switch elements connected thereto

(57) Abstract: Provided is a traffic congestion control method, comprising: when a link between switch elements is invalid, a switch element transmits a control cell to a source switch interface; and the source switch interface adjusts, according to the control cell, the number of data streams transmitted to respective switch elements connected thereto. Also provided are a traffic congestion control device, and a storage medium.

(57) 摘要: 本发明提供了一种流量拥塞控制方法, 包括: 当交换装置之间的链路无效时, 交换装置发送控制信元到源交换接入装置; 源交换接入装置根据所述信元, 调整发送至与自身相连的各个交换装置的数据流量。本发明还提供了一种流量拥塞控制装置及存储介质。

WO 2017/157116 A1



IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,
RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD,
TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种流量拥塞控制方法、装置及存储介质

技术领域

本发明涉及分组交换数据网，尤其涉及一种流量拥塞控制方法、装置及存储介质。

5 背景技术

交换系统是分组交换设备的关键组成部分，现有的交换系统结构示意图如图 1 所示，交换系统由交换接入装置和交换装置组成，交换接入装置是由源交换接入装置和目的交换接入装置组成；源交换接入装置将网络报文切割成信元，通过交换装置转发给目的交换接入装置。

10 交换网络是连接路由器输入端口和输出端口的桥梁，是实现分组报文转发的核心网络。在大容量高端路由器中，为满足超大信息量的交换容量需求，通常采用多级互联的交换网络，目前，三级 CLOS（Charles Clos）交换网络是最常用的一种多级互联交换网络。

随着交换网络容量的不断增大，交换网络中交换接入装置(SI, Switch
15 Interface)与交换装置(SE, Switch Element)的数量不断增多，因此，影响交换网络流量均衡的因素也随之增多，如果不能有效的控制交换网络流量的动态平衡将会影响整个交换网络的性能。

目前交换网络的三级堆叠结构如图 2 所示，SI 为交换接入装置，SE1 为第一级交换装置，SE2 为第二级交换装置，SE3 为第三级交换装置，其中，
20 第一级交换装置和第三级交换装置在同一个芯片上，两者之间有通道相连，可以互相通信，交换接入装置与交换装置以及交换装置之间均通过串行高速链路相连。目前大多数的交换网络都支持单播和组播两种通讯模式，以

单播信元的传输方式为例，源交换接入装置将来自网络中的报文切割成信元，发给第一级交换装置，第一级交换装置将信元传输给第二级交换装置，第二级交换装置对输入的信元进行路由查找后传输给第三级交换装置，最后，第三级交换装置对接收到的信元进行路由查找后传输给目的交换接入装置，目的交换接入装置将接收的信元经过报文重组返回到网络中。如图 2 所示，在正常情况下，该交换网络的流量能够实现动态平衡，各个交换接入装置向与其相连的第一级交换装置发送的流量均等；但是，如果第一级交换装置与第二级交换装置之间的某条链路突然断开，则必然引起该第一级交换装置的输入带宽大于输出带宽，信元在第一交换装置内部发生拥塞。

10 发明内容

有鉴于此，本发明实施例期望提供一种流量拥塞控制方法、装置及存储介质，能够解决由于链路状态发生变化或系统拓扑结构变化导致的芯片级拥塞问题，提高了交换系统的性能。

为达到上述目的，本发明的技术方案是这样实现的：

15 本发明实施例提供了一种流量拥塞控制方法，其特征在于，所述方法包括：

当交换装置之间的链路无效时，交换装置发送控制信元到源交换接入装置；源交换接入装置根据所述信元，调整发送至与自身相连的各个交换装置的数据流量。

20 上述方案中，所述交换装置之间的链路无效包括：第一级交换装置监测自身与第二级交换装置之间的链路的状态，当所述链路在预设时间内接收不到任何数据时，则确定所述链路无效，并通知第三级交换装置。

上述方案中，所述当交换装置之间的链路无效时，交换装置发送控制信元到源交换接入装置包括：

25 当第一级交换装置与第二级交换装置之间的链路无效时，第三级交换

装置发送控制信元到源交换接入装置;

其中, 所述控制信元中携带有所述发生链路无效的第一级交换装置标识以及所述第一级交换装置当前无效链路与原有效链路的比例。

上述方案中, 所述源交换接入装置根据所述信元, 调整发送至与自身
5 相连的各个交换装置的数据流量包括:

当源交换接入装置接收到的控制信元中指示某一第一级交换装置与第二级交换装置之间的链路无效时, 根据所述第一级交换装置当前无效链路与原有效链路的比例, 减小发送至所述发生链路无效的第一级交换装置的数据流量, 增加发送至其他未发生链路无效的第一级交换装置的数据流量。

10 上述方案中, 当某一第一级交换装置与第二级交换装置之间的链路全部无效时, 所述源交换接入装置不向所述第一级交换装置发送数据流量。

上述方案中, 所述方法还包括:

当所述第一级交换装置与第二级交换装置之间无效的链路恢复为有效时, 恢复对所述第一级交换装置的数据流量的发送。

15 本发明实施例还提供了一种流量拥塞控制装置, 所述装置包括: 监测模块、信元发送模块、流量控制模块, 其中,

所述监测模块, 配置为监测交换装置之间是否发生链路无效;

所述信元发送模块, 配置为当交换装置之间的链路无效时, 发送控制信元到流量控制模块;

20 所述流量控制模块, 配置为根据所述信元, 调整发送至与源交换接入装置相连的各个交换装置的数据流量。

上述方案中, 所述监测模块具体配置为: 监测第一级交换装置与第二级交换装置之间的链路的状态, 当所述链路在预设时间内接收不到任何数据时, 则确定所述链路无效, 并通知信元发送模块。

25 上述方案中, 所述信元发送模块具体配置为:

当第一级交换装置与第二级交换装置之间的链路无效时，发送控制信元到流量控制模块；

其中，所述控制信元中携带有所述发生链路无效的第一级交换装置标识以及所述第一级交换装置当前无效链路与原有效链路的比例。

5 上述方案中，所述流量控制模块具体配置为：

当接收到的控制信元中指示某一第一级交换装置与第二级交换装置之间的链路无效时，根据所述第一级交换装置当前无效链路与原有效链路的比例，减小发送至所述发生链路无效的第一级交换装置的数据流量，增加发送至其他未发生链路无效的第一级交换装置的数据流量；

10 当某一第一级交换装置与第二级交换装置之间的链路全部无效时，不向所述第一级交换装置发送数据流量。

当所述第一级交换装置与第二级交换装置之间无效的链路恢复为有效时，恢复对所述第一级交换装置的数据流量的发送。

15 本发明实施例还提供一种计算机存储介质，所述计算机存储介质中存储有计算机可执行指令，该计算机可执行指令用于执行上述的流量拥塞控制方法。

20 本发明实施例所提供的流量拥塞控制方法、装置及存储介质，当交换装置之间的链路无效时，交换装置发送控制信元到源交换接入装置；源交换接入装置根据所述信元，调整发送至与自身相连的各个交换装置的数据流量。如此，使源交换接入装置能够获知交换网络中出现无效链路，从而通过调整源交换接入装置向各个交换装置发送数据流的比例，解决了交换网中流量拥塞问题。

附图说明

图 1 为本发明交换系统结构示意图；

25 图 2 为本发明交换网络的三级堆叠结构示意图；

图 3 为本发明实施例流量拥塞控制方法流程示意图;

图 4 为本发明实施例数据流和控制信息走向示意图;

图 5 为本发明实施例流量拥塞控制装置结构示意图。

具体实施方式

5 本发明实施例中, 当交换装置之间的链路无效时, 交换装置发送控制信元到源交换接入装置; 源交换接入装置根据所述信元, 调整发送至与自身相连的各个交换装置的数据流量。

下面结合附图及具体实施例, 对本发明技术方案的实施作进一步的详细描述。图 3 为本发明实施例流量拥塞控制方法流程示意图, 如图 3 所示,

10 本发明实施例流量拥塞控制包括以下步骤:

步骤 301: 当交换装置之间的链路无效时, 交换装置发送控制信元到源交换接入装置;

本发明实施例中, 所述交换装置之间的链路无效包括: 第一级交换装置监测自身与第二级交换装置之间的链路的状态, 当所述链路在预设时间内接收不到任何数据时, 则确定所述链路无效, 并通知第三级交换装置。

所述当交换装置之间的链路无效时, 交换装置发送控制信元到源交换接入装置包括: 当第一级交换装置与第二级交换装置之间的链路无效时, 第三级交换装置发送控制信元到源交换接入装置; 其中, 所述控制信元中携带有所述发生链路无效的第一级交换装置标识以及所述第一级交换装置当前无效链路与原有效链路的比例。

具体的, 可以通过修改所述第三级交换接入装置发送的控制信元中的某些字段, 在这些字段中写入发生链路无效的第一级交换装置的标识和所述第一级交换装置无效链路与原来有效链路数量的比例。第三级交换装置周期性的向每个源交换接入装置发送控制信元, 该控制信元携带上述发生链路无效的第一级交换装置的标识和所述第一级交换装置无效链路与

原来有效链路数量的比例等信息。

步骤 302: 源交换接入装置根据所述信元, 调整发送至与自身相连的各个交换装置的数据流量。

5 本发明实施例中, 所述源交换接入装置根据所述信元, 调整发送至与自身相连的各个交换装置的数据流量包括: 源交换接入装置接收到所述携带有效信息的控制信元后, 提取所述控制信元相应字段内的信息, 当源交换接入装置接收到的控制信元中指示某一第一级交换装置与第二级交换装置之间的链路无效时, 根据所述第一级交换装置当前无效链路与原有效链路的比例, 调整向与其相连的第一级交换装置发送的数据流量的比例,
10 具体的: 减小发送至所述发生链路无效的第一级交换装置的数据流量, 增加发送至其他未发生链路无效的第一级交换装置的数据流量。

本发明实施例中, 在极端情况下, 当某一第一级交换装置与第二级交换装置之间的链路全部无效时, 与所述第一级交换装置相连的所有源交换接入装置均不向所述第一级交换装置发送任何数据流量。

15 本发明实施例中, 所述方法还包括: 当所述第一级交换装置与第二级交换装置之间的链路无效的链路恢复为有效时, 与所述第一级交换装置相连的交换接入装置恢复对所述第一级交换装置的数据流量的发送。

下面结合具体应用场景, 对本发明实施例所述流量拥塞控制方法进行进一步详细说明。

20 图 4 为本发明实施例数据流和控制信息走向示意图, 其中, 带箭头的实线 “ $\longrightarrow$ ” 为数据流的流向, 带箭头的虚线 “ $\cdots\longrightarrow$ ” 为控制信元流向; 带箭头的细线 “ $\longrightarrow$ ” 表示单条链路, 带箭头的粗线 “ $\longrightarrow$ ” 表示 m 条链路, 如图 4 所示, 本发明实施例所述三级交换系统包括 N 个交换接入装置 SI, N 个第一级交换装置 SE1, N 个第二级交换装置 SE2 和 N 个第三
25 级交换装置 SE3; 如图 4 所述场景, 如果第一级交换装置 SE1 与第二级交

换装置 SE2 之间有链路突然无效, 那么第一级交换装置 SE1 的输入带宽大于输出带宽, 数据流量或信元在第一级交换装置 SE1 内部极有可能出现拥塞, 为解决这一问题, 本发明实施例通过修改第三级交换装置 SE3 向交换接入装置 SI 发送控制信元的字段, 使得所述字段包含发生链路无效的第一级交换装置 SE1 的标识和以及所述第一级交换装置 SE1 无效链路数量与原来有效链路数量的比例, 使交换接入装置 SI 调整向每个第一级交换装置发送信元数量的比例, 从而实现整个交换网流量的动态平衡。

具体的, 如图 4 所述场景, 与第一个第一级交换装置#1SE1 相连的有 N 个交换接入装置#1SI~#NSI, 在交换接入装置 SI 与交换装置 SE、以及各级交换装置 SE 之间链路有效的情况下图 4 所述三级交换网流量均衡, 每个源交换接入装置 SI 给每个第一级交换装置 SE1 发送的流量各占其发送总流量的 $1/N$;

本发明实施例中, 第一个第一级交换装置#1SE1 通过 m 条链路与 N 个第二级交换装置#1SE2~#NSE2 相连, 如果第一个第一级交换装置#1SE1 与第一个第二级交换装置#1SE2 之间的某 n ($0 < n < m$) 条链路变为无效, 如图 4 中叉号 “x” 所示, 那么第一个第一级交换装置#1SE1 的无效链路的数量与原来有效链路的数量的比例为 n/m , 由于第一级交换装置 SE1 与第三级交换装置 SE3 在同一块芯片上且有通道相连, 因此, 当第一个第一级交换装置#1SE1 与第一个第二级交换装置#1SE2 之间的 n 条链路无效时, 第三级交换装置 SE3 向源交换接入装置#1SI~#NSI 发送控制信元, 所述控制信元中包括发生链路无效的第一级交换装置的标识以及所述第一级交换装置无效链路与原有效链路的比例;

源交换接入装置#1SI~#NSI 收到所述控制信元后, 确定无效链路的数量大于 0, 且无效链路的数量与原来有效链路的数量的比例为 n/m , 所述源交换接入装置#1SI~#NSI 将发往第一个第一级交换装置#1SE1 的数据流量或

信元数量减少为原来的 n/m , 即发往第一个第一级交换装置#1SE1 的数据流量或信元数量变为 $(1/N) * (1-n/m)$, 同时, 将发往其它第一级交换装置#2SE1~#NSE1 的数据流量或信元数量增大为原来的 $(n/m) * 1/(N-1)$;

以第一个源交换接入装置#1SI 为例, 第一个源交换接入装置#1SI 发往其他第一级交换装置#2SE1~#NSE1 的数据流量或信元数量变为 $(1/N) * (1+(n/m) * (1/(N-1)))$ 。总的流量或信元数量百分比为: $(1/N) * (1-n/m) + (1/N) * (1+(n/m) * (1/(N-1))) * (N-1) = 100\%$; 可见, 进行数据流量调整后, 并没有影响总数据流量或信元数量。其他源交换接入装置#2SI~#NSI 以此类推。

需要说明的是, 上述 N 个交换接入装置和交换装置, m 条有效链路和 n 条无效链路是为了清楚的描述本发明的方案而列举的, 实际中并不限于上述几个装置和链路。

本发明实施例所述流量拥塞控制方法, 用于处理交换网络三级堆叠结构中第一级交换装置与第二级交换装置之间的链路突然断开或系统拓扑结构发生变化导致的流量拥塞的问题。实时监测链路状态, 并通过第三级交换装置向交换接入装置发送控制信元, 所述控制信元携带有发生链路无效的第一级交换装置的标识和所述第一级交换装置无效链路与原来有效链路数量的比例, 使得源交换接入装置接收到所述信元后, 能够获知第一级交换装置和第二级交换装置之间的可用链路减少, 通过调整源交换接入装置向各个第一级交换装置发送数据流量或信元数量的比例, 从而解决了交换网中发生链路无效时导致的流量拥塞问题, 提高整个交换网络的性能。

本发明实施例不仅应用于以单播方式传输的场景中, 也应用于以组播方式传输的场景中。

本发明所述的方法并不限于具体实施方式中所述的实施例, 本领域技术人员根据本发明的技术方案得出其它的实施方式, 同样属于本发明的技

术创新范围。

本发明实施例还提供了一种流量拥塞控制装置，图 5 为本发明实施例流量拥塞控制装置结构示意图，如图 5 所示，所述装置包括：监测模块 51、信元发送模块 52、流量控制模块 53，其中，所述监测模块 51 位于第一级交换装置，所述信元发送模块 52 位于第三级交换装置，所述流量控制模块 53 位于源交换接入装置，具体的，

所述监测模块 51，配置为监测交换装置之间是否发生链路无效；

本发明实施例中，所述监测模块 51 具体配置为：监测第一级交换装置与第二级交换装置之间的链路的状态，当所述链路在预设时间内接收不到任何数据时，则确定所述链路无效，并通知信元发送模块 52；

所述信元发送模块 52，配置为当交换装置之间的链路无效时，发送控制信元到流量控制模块；

本发明实施例中，所述信元发送模块 52 具体配置为：当第一级交换装置与第二级交换装置之间的链路无效时，发送控制信元到流量控制模块 53；其中，所述控制信元中携带有所述发生链路无效的第一级交换装置标识以及所述第一级交换装置当前无效链路与原有效链路的比例。

具体的，所述信元发送模块 52 可以通过修改所述第三级交换接入装置发送的控制信元中的某些字段，在这些字段中写入发生链路无效的第一级交换装置的标识和所述第一级交换装置无效链路与原来有效链路数量的比例。所述信元发送模块 52 周期性的向每个源交换接入装置发送控制信元，该控制信元携带上述发生链路无效的第一级交换装置的标识和所述第一级交换装置无效链路与原来有效链路数量的比例等信息。

所述流量控制模块 53，配置为根据所述信元，调整发送至与源交换接入装置相连的各个交换装置的数据流量。

本发明实施例中，所述流量控制模块 53 具体配置为：收到所述携带有

有效信息的控制信元后，提取所述控制信元相应字段内的信息，当接收到的控制信元中指示某一第一级交换装置与第二级交换装置之间的链路无效时，根据所述第一级交换装置当前无效链路与原有效链路的比例，调整向与其相连的第一级交换装置发送的数据流量的比例，具体的：减小发送至
5 所述发生链路无效的第一级交换装置的数据流量，增加发送至其他未发生链路无效的第一级交换装置的数据流量。

本发明实施例中，在极端情况下，所述流量控制模块 53 还配置为：当某一第一级交换装置与第二级交换装置之间的链路全部无效时，与所述第一级交换装置相连的所有源交换接入装置中的流量控制模块 53 均不向所述
10 第一级交换装置发送数据流量。

本发明实施例中，所述流量控制模块 53 还配置为：当所述第一级交换装置与第二级交换装置之间无效的链路恢复为有效时，所述第一级交换装置相连的交换接入装置中的流量控制模块 53 恢复对所述第一级交换装置的数据流量的发送。

图 5 中所示的流量拥塞控制装置中的各处理模块的实现功能，可参照前述流量拥塞控制方法的相关描述而理解。本领域技术人员应当理解，图 5 所示的流量拥塞控制装置中各处理模块的功能可通过运行于处理器上的程序而实现，也可通过具体的逻辑电路而实现，比如：可由中央处理器（CPU）、微处理器（MPU）、数字信号处理器（DSP）、或现场可编程门阵列（FPGA）
15 实现。
20

本发明实施例中，如果以软件功能模块的形式实现上述流量拥塞控制方法，并作为独立的产品销售或使用，也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，本发明实施例的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件
25 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)执行本发明各个实施例所述方法的全部或部分。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read Only Memory, ROM)、磁碟或者光盘等各种可以存储程序代码的介质。这样,本发明实施例不限制于任何特定的硬件和软件结合。

5 相应地,本发明实施例还提供一种计算机存储介质,该计算机存储介质中存储有计算机程序,该计算机程序用于执行本发明实施例的上述流量拥塞控制方法。

在本发明所提供的几个实施例中,应该理解到,所揭露的方法及装置,可以通过其他方式实现。以上所描述的装置实施例仅仅是示意性的,例如,10 所述模块的划分,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,如:多个模块或组件可以结合,或可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的各组成部分相互之间的通信连接可以是通过一些接口,设备或模块的间接耦合或通信连接,可以是电性的、机械的或其他形式的。

15 上述作为分离部件说明的模块可以是、或也可以不是物理上分开的,作为模块显示的部件可以是、或也可以不是物理模块,即可以位于一个地方,也可以分布到多个网络模块上;可以根据实际的需要选择其中的部分或全部模块来实现本实施例方案的目的。

另外,在本发明各实施例中的各功能模块可以全部集成在一个处理模20 块中,也可以是各模块分别单独作为一个模块,也可以两个或两个以上模块集成在一个模块中;上述集成的模块既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

本领域普通技术人员可以理解:实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成,前述的程序可以存储于计算机可25 读取存储介质中,该程序在执行时,执行包括上述方法实施例的步骤;而

前述的存储介质包括：移动存储设备、只读存储器（ROM，Read-Only Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

或者，本发明实施例上述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用，也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，本发明实施例的技术方案本质上或者说对现有技术做出贡献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）执行本发明各个实施例所述方法的全部或部分。而前述的存储介质包括：移动存储设备、ROM、磁碟或者光盘等各种可以存储程序代码的介质。

本发明实施例中记载的流量拥塞控制方法、装置只以上述实施例为例，但不仅限于此，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

以上所述仅为本发明的较佳实施例而已，并非用于限定本发明的保护范围。

权利要求书

1、一种流量拥塞控制方法，所述方法包括：

当交换装置之间的链路无效时，交换装置发送控制信元到源交换接入装置；源交换接入装置根据所述信元，调整发送至与自身相连的各个交换装置的数据流量。

2、根据权利要求1所述方法，其中，所述交换装置之间的链路无效包括：第一级交换装置监测自身与第二级交换装置之间的链路的状态，当所述链路在预设时间内接收不到任何数据时，则确定所述链路无效，并通知第三级交换装置。

3、根据权利要求1所述方法，其中，所述当交换装置之间的链路无效时，交换装置发送控制信元到源交换接入装置包括：

当第一级交换装置与第二级交换装置之间的链路无效时，第三级交换装置发送控制信元到源交换接入装置；

其中，所述控制信元中携带有所述发生链路无效的第一级交换装置标识以及所述第一级交换装置当前无效链路与原有效链路的比例。

4、根据权利要求1所述方法，其中，所述源交换接入装置根据所述信元，调整发送至与自身相连的各个交换装置的数据流量包括：

当源交换接入装置接收到的控制信元中指示某一第一级交换装置与第二级交换装置之间的链路无效时，根据所述第一级交换装置当前无效链路与原有效链路的比例，减小发送至所述发生链路无效的第一级交换装置的数据流量，增加发送至其他未发生链路无效的第一级交换装置的数据流量。

5、根据权利要求4所述方法，其中，当某一第一级交换装置与第二级交换装置之间的链路全部无效时，所述源交换接入装置不向所述第一级交换装置发送数据流量。

6、根据权利要求 1 至 5 任一项所述方法，其中，所述方法还包括：

当所述第一级交换装置与第二级交换装置之间无效的链路恢复为有效时，恢复对所述第一级交换装置的数据流量的发送。

7、一种流量拥塞控制装置，所述装置包括：监测模块、信元发送模块、
5 流量控制模块，其中，

所述监测模块，配置为监测交换装置之间是否发生链路无效；

所述信元发送模块，配置为当交换装置之间的链路无效时，发送控制信元到流量控制模块；

所述流量控制模块，配置为根据所述信元，调整发送至与源交换接入
10 装置相连的各个交换装置的数据流量。

8、根据权利要求 7 所述装置，其中，所述监测模块配置为：监测第一级交换装置与第二级交换装置之间的链路的状态，当所述链路在预设时间内接收不到任何数据时，则确定所述链路无效，并通知信元发送模块。

9、根据权利要求 7 所述装置，其中，所述信元发送模块配置为：

15 当第一级交换装置与第二级交换装置之间的链路无效时，发送控制信元到流量控制模块；

其中，所述控制信元中携带有所述发生链路无效的第一级交换装置标识以及所述第一级交换装置当前无效链路与原有效链路的比例。

10、根据权利要求 7 所述装置，其中，所述流量控制模块配置为：

20 当接收到的控制信元中指示某一第一级交换装置与第二级交换装置之间的链路无效时，根据所述第一级交换装置当前无效链路与原有效链路的比例，减小发送至所述发生链路无效的第一级交换装置的数据流量，增加发送至其他未发生链路无效的第一级交换装置的数据流量；

当某一第一级交换装置与第二级交换装置之间的链路全部无效时，不
25 向所述第一级交换装置发送数据流量。

当所述第一级交换装置与第二级交换装置之间无效的链路恢复为有效时，恢复对所述第一级交换装置的数据流量的发送。

11、一种计算机存储介质，所述计算机存储介质中存储有计算机可执行指令，该计算机可执行指令用于执行权利要求 1 至 6 任一项所述的流量

5 拥塞控制方法。

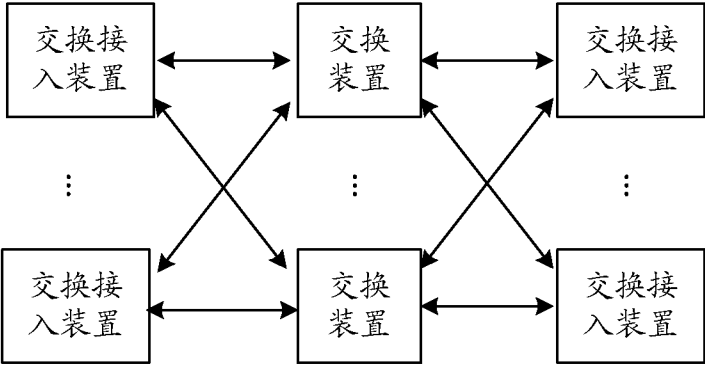


图 1

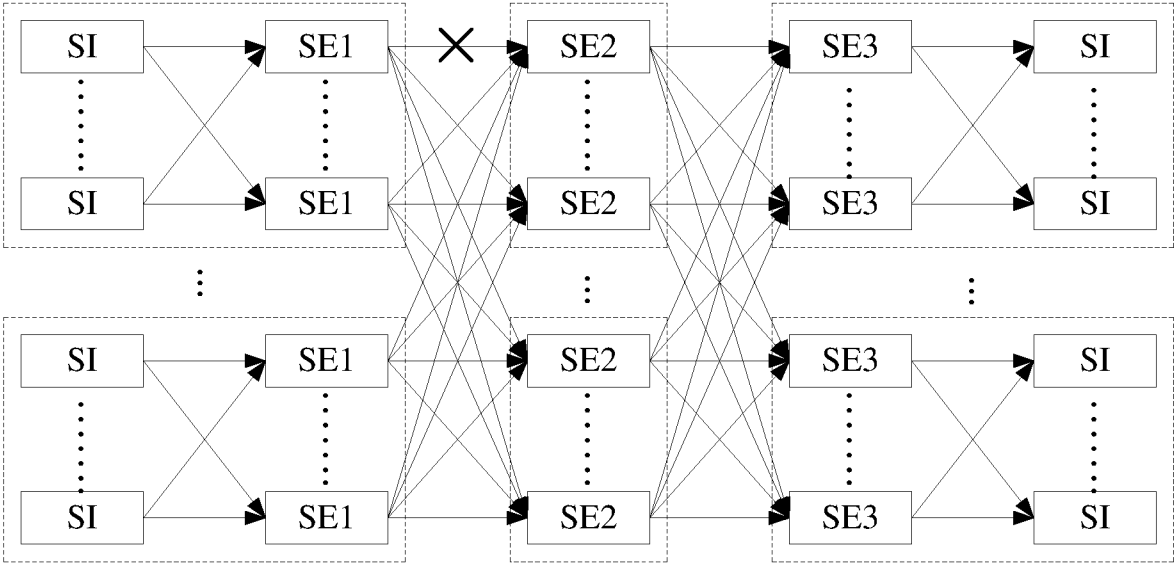


图 2

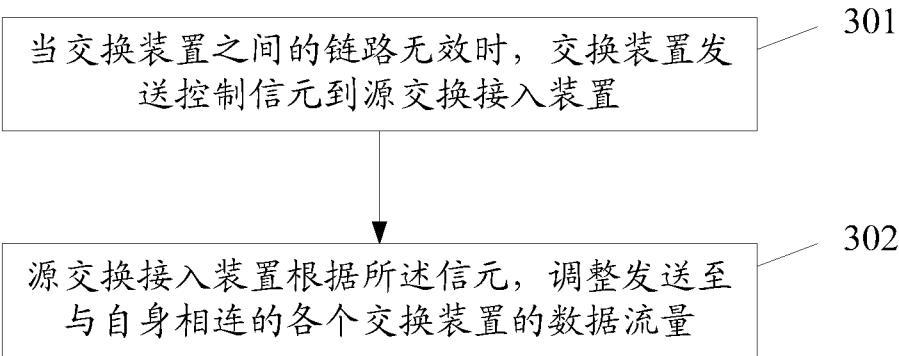


图 3

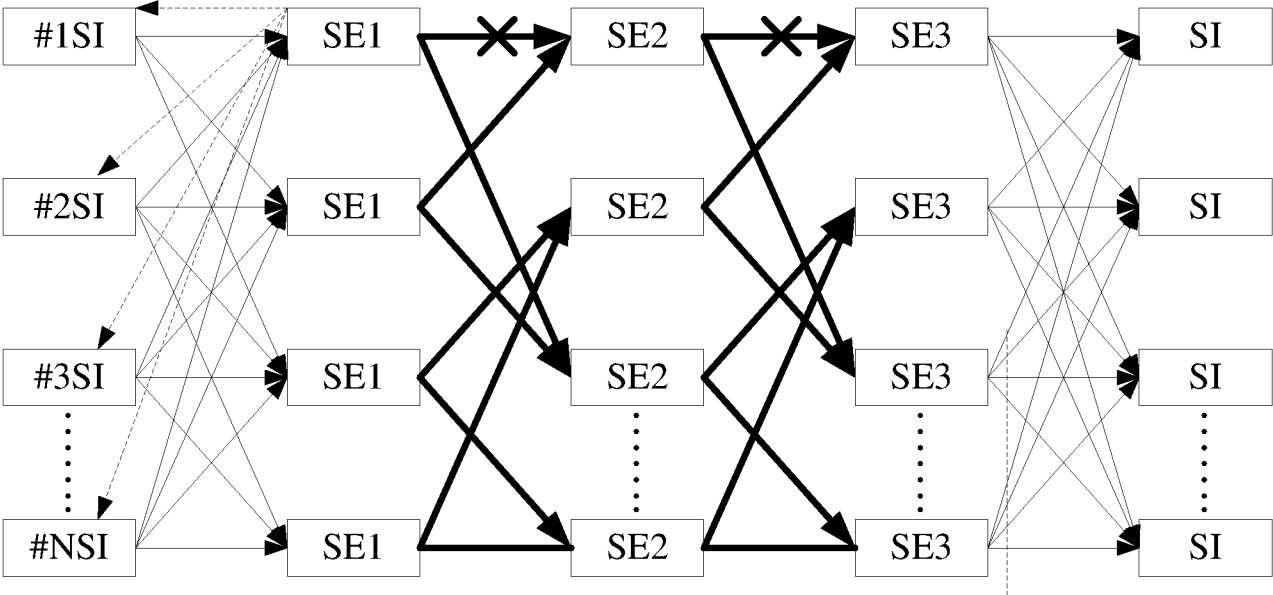


图 4

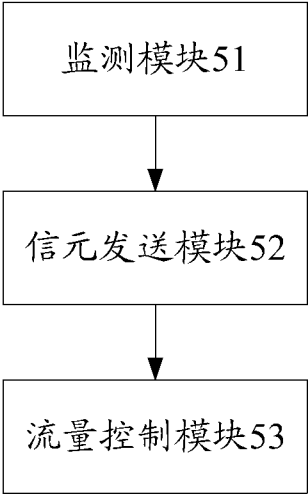


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/073368

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/70 (2013.01) i; H04L 12/801 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS; CNTXT; CNKI; VEN: exchange, flow, control, ZTE MICROELECTRONICS, switch device, invalid, fault, flow control, congestion, switch, link

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104618254 A (H3C TECHNOLOGIES CO., LIMITED), 13 May 2015 (13.05.2015), description, paragraphs [0026]-[0046], and claims 1-3	1-2, 6-8, 11
X	CN 105337883 A (UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA et al.), 17 February 2016 (17.02.2016), description, paragraphs [0046]-[0061], and claim 1	1-2, 6-8, 11
X	US 2003021230 A1 (PETASWITCH SOLUTIONS INC.), 30 January 2003 (30.01.2003), description, paragraph [0011]	1-2, 6-8, 11
X	US 2008298248 A1 (ROECK, G. et al.), 04 December 2008 (04.12.2008), description, paragraphs [0029]-[0036], and claim 1	1-2, 6-8, 11
A	CN 104184676 A (HUAWEI TECHNOLOGIES CO., LTD.), 03 December 2014 (03.12.2014), the whole text of the description	1-11
A	US 7719982 B2 (INTEL CORPORATION), 18 May 2010 (18.05.2010), the whole text of the description	1-11

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 13 April 2017 (13.04.2017)	Date of mailing of the international search report 04 May 2017 (04.05.2017)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer ZHU, Qi Telephone No.: (86-10) 62411269

International application No.
PCT/CN2017/073368

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/073368

A. 主题的分类 H04L 12/70(2013.01)i; H04L 12/801(2013.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																							
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) H04L 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CPRSABS;CNTXT;CNKI;VEN:链路, 拥塞, 交换, 流量, 控制, 中兴微电子, 交换设备, 交换装置, 无效, 故障, flow control, congestion, switch, link																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 104618254 A (杭州华三通信技术有限公司) 2015年 5月 13日 (2015 - 05 - 13) 说明书第[0026]-[0046]段, 权利要求1-3</td> <td>1-2, 6-8, 11</td> </tr> <tr> <td>X</td> <td>CN 105337883 A (电子科技大学 等) 2016年 2月 17日 (2016 - 02 - 17) 说明书第[0046]-[0061]段, 权利要求1</td> <td>1-2, 6-8, 11</td> </tr> <tr> <td>X</td> <td>US 2003021230 A1 (PETASWITCH SOLUTIONS INC) 2003年 1月 30日 (2003 - 01 - 30) 说明书第[0011]段</td> <td>1-2, 6-8, 11</td> </tr> <tr> <td>X</td> <td>US 2008298248 A1 (ROECK, GUENTER等) 2008年 12月 4日 (2008 - 12 - 04) 说明书第[0029]-[0036], 权利要求1</td> <td>1-2, 6-8, 11</td> </tr> <tr> <td>A</td> <td>CN 104184676 A (华为技术有限公司) 2014年 12月 3日 (2014 - 12 - 03) 说明书全文</td> <td>1-11</td> </tr> <tr> <td>A</td> <td>US 7719982 B2 (英特尔公司) 2010年 5月 18日 (2010 - 05 - 18) 说明书全文</td> <td>1-11</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 104618254 A (杭州华三通信技术有限公司) 2015年 5月 13日 (2015 - 05 - 13) 说明书第[0026]-[0046]段, 权利要求1-3	1-2, 6-8, 11	X	CN 105337883 A (电子科技大学 等) 2016年 2月 17日 (2016 - 02 - 17) 说明书第[0046]-[0061]段, 权利要求1	1-2, 6-8, 11	X	US 2003021230 A1 (PETASWITCH SOLUTIONS INC) 2003年 1月 30日 (2003 - 01 - 30) 说明书第[0011]段	1-2, 6-8, 11	X	US 2008298248 A1 (ROECK, GUENTER等) 2008年 12月 4日 (2008 - 12 - 04) 说明书第[0029]-[0036], 权利要求1	1-2, 6-8, 11	A	CN 104184676 A (华为技术有限公司) 2014年 12月 3日 (2014 - 12 - 03) 说明书全文	1-11	A	US 7719982 B2 (英特尔公司) 2010年 5月 18日 (2010 - 05 - 18) 说明书全文	1-11
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
X	CN 104618254 A (杭州华三通信技术有限公司) 2015年 5月 13日 (2015 - 05 - 13) 说明书第[0026]-[0046]段, 权利要求1-3	1-2, 6-8, 11																					
X	CN 105337883 A (电子科技大学 等) 2016年 2月 17日 (2016 - 02 - 17) 说明书第[0046]-[0061]段, 权利要求1	1-2, 6-8, 11																					
X	US 2003021230 A1 (PETASWITCH SOLUTIONS INC) 2003年 1月 30日 (2003 - 01 - 30) 说明书第[0011]段	1-2, 6-8, 11																					
X	US 2008298248 A1 (ROECK, GUENTER等) 2008年 12月 4日 (2008 - 12 - 04) 说明书第[0029]-[0036], 权利要求1	1-2, 6-8, 11																					
A	CN 104184676 A (华为技术有限公司) 2014年 12月 3日 (2014 - 12 - 03) 说明书全文	1-11																					
A	US 7719982 B2 (英特尔公司) 2010年 5月 18日 (2010 - 05 - 18) 说明书全文	1-11																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																							
国际检索实际完成的日期 2017年 4月 13日		国际检索报告邮寄日期 2017年 5月 4日																					
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 朱琦 电话号码 (86-10)62411269																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/073368

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	104618254	A	2015年 5月 13日	无			
CN	105337883	A	2016年 2月 17日	无			
US	2003021230	A1	2003年 1月 30日	无			
US	2008298248	A1	2008年 12月 4日	WO	2008148122	A3	2009年 1月 29日
				WO	2008148122	A2	2008年 12月 4日
CN	104184676	A	2014年 12月 3日	无			
US	7719982	B2	2010年 5月 18日	US	2007047535	A1	2007年 3月 1日