[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
H04L 12/28
H04L 12/26 H04L 12/24
H04Q 1/20

[12] 发明专利申请公开说明书

[21] 申请号 200310103121.0

[43] 公开日 2004 年 5 月 26 日

[11] 公开号 CN 1499780A

[22] 申请日 2003.10.31
[21] 申请号 200310103121.0
[30] 优先权
[32] 2002.10.31 [33] US [31] 10/284，856
[71] 申请人 加林克半导体 V.N. 有限公司
地址 美国加利福尼亚州
[72] 发明人 王林萧 常荣峰 埃里克·林
詹姆斯·诚寿·易

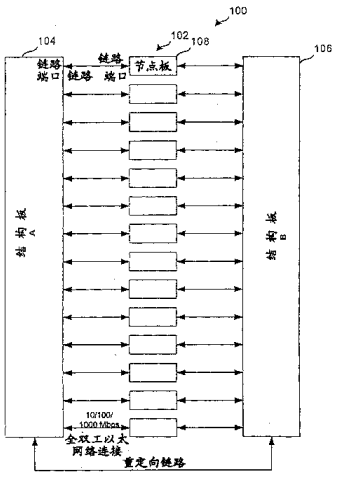
[74] 专利代理机构 北京邦信阳专利商标代理有限公司
代理人 黄泽雄 张耀丽

权利要求书 5 页 说明书 16 页 附图 6 页

[54] 发明名称 具有高可行性的以太网背板结构

[57] 摘要

本发明公开了一种具有高可行性的背板结构。该背板系统包括多个与冗余交换结构板进行通信的冗余节点板。节点板上的上行链路端口在与交换结构板通信的一端被逻辑分组为多个中继端口，节点板和交换板在正常模式运行时例行地执行链路完整性检查，在发现链路故障时每一节点板和交换板能够独立地启动故障切换从而切换到可用端口。链路在预定时间间隔内没有信息传送或者接收到了预定连续数量的无效数据包后，通过发送一个链路脉动信号来检测链路故障。一旦链路故障排除，运行重新回到正常模式。



ISSN1008-4274

知识产权出版社出版

1. 一种用于确定通信链路故障的方法，其步骤包括:

设定一个定时器来预定一个时间间隔;

在预定的时间间隔内等待接收一个链路脉动数据包;

只要接收到一个有效的数据包，就重置定时器至该预定时间间隔; 及

当定时器到期时，将通信链路的状态转换为故障模式。

2. 如权利要求 1 的方法，其步骤进一步包括，在接收到预定数量的连续的无效数据包后，将通信链路的状态转换为故障模式。

3. 一种用于节点板处理网络故障的方法，其步骤包括:

创建多个冗余端口的一个中继端口;

在每一冗余端口上执行一个链路完整性检查;

当检测到一个端口故障时，更新一个媒体访问控制表; 及

将送往故障端口的业务映射到一个工作端口。

4. 如权利要求 3 的方法，进一步包括以下步骤:

当故障端口接收到一个有效数据包时，更新该媒体访问控制表。

5. 如权利要求 3 的方法，执行步骤包括:

设定一个定时器来预定一个时间间隔;

在预定的时间间隔内等待接收一个链路脉动数据包;

只要接收到一个有效的数据包，就重置定时器至该预定时间间隔; 及

当定时器过期时，将通信链路的状态转换为故障模式。

6. 如权利要求 5 的方法，执行步骤进一步包括，在接收到预定数量的连续的无效数据包后，将通信链路的状态转换为故障模式。

7. 一种用于具有一个重定向表的交换结构板处理链路故障的方法，其步骤包括:

对连接到该交换结构板的每一链路执行一个链路故障完整性检

查；

一旦故障完整性检查检测到一个链路故障，激活一个链路故障切换模式，该链路故障切换模式的步骤包括：

将链路故障数据发送给至少一个其它的交换结构板，

用该链路故障数据更新一个媒体访问控制表，及

将故障端口映射到一个重定向端口。

8. 如权利要求 7 的方法，其执行步骤包括：

设定一个定时器来预定一个时间间隔；

在预定的时间间隔内等待接收一个链路脉动数据包；

只要接收到一个有效的数据包，就重置定时器至该预定时间间隔；及

当定时器过期时，将通信链路的状态转换为故障模式。

9. 如权利要求 8 的方法，其执行步骤进一步包括，在接收到预定数量的连续的无效数据包后，将通信链路的状态转换为故障模式。

10. 一种发送一个链路脉动信息的方法，其步骤包括：

设定一个定时器；和

当定时器到期时，发送一个链路脉动数据包。

11. 如权利要求 10 的方法，其步骤进一步包括：

接收一个数据包，用于发送；

发送该数据包；和

只要数据包被发出，就重置该定时器。

12. 一个媒体访问控制数据包，包括：

一个目的媒体访问控制地址；

一个源媒体访问控制地址；

一个表示媒体访问控制格式的以太类型域；及

一个操作码

其中，十六进制下的以太类型值为 88-08。

13. 如权利要求 12 的媒体访问控制数据包，其中，用于一个发送器和一

个接收器的操作码是一个相等的预定的两个字节的值。

14. 如权利要求 13 的媒体访问控制数据包，进一步包括足以使数据包长度达到 64 位的填充字节。

15. 一个背板系统，其包括:

多个节点板;

多个交换结构板，每个交换结构板具有一个重定向表;

其中多个节点板的每个节点具有与多个交换结构板中的每一个之间的单一链路;

其中当一个交换结构板检测到一个链路故障时，该交换结构板将该故障链路信息传达到其它交换结构板，促使其它交换结构板更新其重定向表。

16. 如权利要求 15 的系统，其中，如在接收到上一数据包后开始计时的一个预定时间间隔之后没有接收到一个媒体访问控制层链路脉动信息数据包，则该链路故障被检测到。

17. 如权利要求 15 的系统，其中，在接收到预定数量的连续的无效数据包后，该链路故障被检测到。

18. 一种指令的计算机可读媒体，包括:

用于设定一个定时器来预定一个时间间隔的装置;

用于在预定的时间间隔内等待接收一个链路脉动数据包的装置;

用于只要接收到一个有效的数据包、就重置定时器至该预定时间间隔的装置; 及

用于当定时器过期时将通信链路的状态转换为故障模式的装置。

19. 如权利要求 18 的指令的计算机可读媒体，进一步包括用于在接收到预定数量的连续的无效数据包后将通信链路的状态转换为故障模式的装置。

20. 指令的计算机可读媒体，包括:

用于创建多个冗余端口的一个中继端口的装置;

用于在每一冗余端口上执行一个链路完整性检查的装置;

用于当检测到一个端口故障时更新一个媒体访问控制表的装置；及

用于将送往故障端口的业务映射到一个工作端口的装置。

21. 如权利要求 20 的指令的计算机可读媒体，进一步包括

用于当故障端口接收到一个有效数据包时更新该媒体访问控制表的装置。

22. 如权利要求 20 的指令的计算机可读媒体，用于执行的装置进一步包括:

用于设定一个定时器来预定一个时间间隔的装置;

用于在预定的时间间隔内等待接收一个链路脉动数据包的装置;

用于只要接收到一个有效的数据包、就重置定时器至该预定时间间隔的装置；及

用于当定时器过期时将通信链路的状态转换为故障模式的装置。

23. 如权利要求 22 的指令的计算机可读媒体，其用于执行的装置进一步包括用于在接收到预定数量的连续的无效数据包后将通信链路的状态转换为故障模式的装置。

24. 一种指令的计算机可读媒体，包括:

用于对连接到该交换结构板的每一链路执行一个链路故障完整性检查的装置;

用于一旦故障完整性检查检测到一个链路故障时便激活一个链路故障切换模式的装置，该装置包括:

用于将链路故障数据发送给至少一个其它的交换结构板的装置，

用于使用该数据更新一个媒体访问控制表的装置，及

用于将故障端口映射到一个重定向端口的装置。

25. 如权利要求 24 的指令的计算机可读媒体，其用于执行的装置进一步包括:

用于设定一个定时器来预定一个时间间隔的装置;

用于在预定的时间间隔内等待接收一个链路脉动数据包的装置；

只要接收到一个有效的数据包、就重置定时器至该预定时间间隔的装置；及

用于当定时器过期时将通信链路的状态转换为故障模式的装置。

26. 如权利要求 25 的指令的计算机可读媒体，其用于执行的装置进一步包括用于在接收到预定数量的连续的无效数据包后将通信链路的状态转换为故障模式的装置。

具有高可行性的以太网背板结构

发明背景

本发明涉及网络装置领域，尤其涉及一种以太网网络装置，该装置具有用于检测链路故障并响应该故障而切换到一个运行良好的端口的背板结构。

背景简介

任何基于全天 24 小时、每周 7 日（24/7）运行的业务都不能承受超过只是两三分钟的或者也许不超过半小时的停机情况的发生。意外的停机可能严重地阻碍数据的运行并且就因此造成的收入损失及雇佣人力纠正此类情况的支出而言，代价及其昂贵。近来的两项 1995 年的研究表明，由于意外停机造成的平均业务损失在每小时 8 万美元与每小时 35 万美元之间，考虑到这些损失，很明显，建立起一个冗余信息技术结构的成本要比甚至一个短暂的停机所造成的损失要小。尤其在考虑到计算机运行成本要相对小于停机时间的成本时，就更为如此。而且，管理者们确切地知道为此添加额外设备、软件和培训操作员的花费有多少，但意外停机造成的损失却难以事先估计。

因为其成本低、配置与安装简易，以太网网络已被大量应用于局域网（LAN）中。以太网网络技术经过多年改进之后，以太网网络的应用如今已经从局域网延伸到了广域网/城域网（WAN/MAN）。更近些时期以来，以太网网络技术还被结合到基于底座的背板系统中，这是因为它的成本低、来源广而且具有检测嵌入式错误的能力。

在背板式系统中，要求背板为链路卡与模块之间提供牢固可靠的连接。然而，以太网网络最初是在局域网的环境下被开发的，局域网应用中的“可用性”要求大大不同于在背板应用中的要求。例如，在传统的局域

网环境下，当网络检测到一个链路或端口通信故障时，生成树协议（spanning tree protocol）在网络检测到一个链路或端口故障时通过重构动态拓扑结构来提供“故障切换（failover）”功能。但是，会聚（convergence）时间却相对要长。从发现一个故障到完成拓扑结构的改变并恢复到正常运行状态可能需要20到50秒的时间。即使使用传统的“改进”协议，快速生成树在检测到开关或链路的故障后恢复正常运行可能需要50毫秒。

根据电气电子工程师协会802.3标准，链路聚合已发展到通过将多个链路聚集形成一个链路组来增加带宽和可用性，媒体访问控制层（MAC）可将该多重链路当作一个单一逻辑链路。当此集合中的一个链路出现故障时，业务可通过送回工作链路被重新分配（或重新路由）。但是，链路聚合只在由相同的终端节点所共享的平行连接之间提供故障切换。

在背板应用中，以太网网络通常具有非常简单的结构，比如一个星形拓扑结构，意味着从每一个插卡槽处将一个第一总线连接到一个第一交换结构（switch fabric）并将一个第二总线连接到一个第二交换结构。如果第一总线不能工作，那么装置自动转换至使用第二总线。但是，生成树的恢复需使用20至50秒的会聚时间，这在背板环境下是不能接受的。此外，如前所述的链路聚合只在相同的终端节点所共享的平行连接之间提供故障切换。这就是说，故障链路的相同终端不能共享一个备份的链路。因此，链路聚合在以太网网络背板环境下不能得到应用。

因此，需要一个简单、快速并强有力的解决方案来实现以太网背板环境下链路错误检测和故障切换转换的高可用性。

发明简介

在此公开并要求权利的本发明，其一方面包括了一种高可用性的背板结构。该背板系统包括运行地与冗余交换结构板进行通信的冗余节点板。节点板上的上行链路端口在与交换结构板通信的一端被逻辑归组为中继端口。节点板和交换结构板在正常模式运行时对链路的完整性进行例行检

查，这样在发现链路故障时每一节点板和交换结构板能够独立地启动故障切换交换至可用端口。一旦链路故障排除后，运行重新回到正常模式。

附图说明

为了更全面地了解本发明及其优点，结合附图进行以下的说明，附图中：

图 1 描述了一个双重结构信息包交换式背板的拓扑结构的总体框图；

图 2 根据一个公开的实施方案描述了背板系统的一个更为详细的框图；

图 3 根据一个公开的实施方案描述了一个节点板的故障切换作业的流程图；

图 4 根据一个公开的实施方案描述了一个结构板的故障切换作业的流程图；

图 5 描述了来自发送端的脉动信号的状态图；

图 6 描述了来自接收端的脉动信号的状态图。

发明详述

本发明公开的背板结构通过自动检测链路故障并执行“故障切换”至一个备份链路从而提供了一种高可用性以太网网络背板。故障切换被定义为“关掉”一个出现故障的冗余部分并“打开”可用的备份部分的程序或机制。本发明的一个方面是使得快速、简单的故障切换变得容易实现。此外，使节点间控制信息交换最小化从而减轻 CPU（中央处理器）的数据处理负担。

在此公开了两种链路故障检测方案。第一种方案包括在 MAC（媒体访问控制）模块发送“脉动（heart beat）”信息；第二种方案则包括一种帧错误率的使用，两种方案之一或该两种方案可实施用来检测链路故障。一旦检测到链路故障，则马上利用 CPU 来执行故障切换程序。

当一个节点板的逻辑电路已经检测到一个故障端口时，对该节点板的

通信业务进行故障切换（或重定向）至可用端口（假设所有节点具有至少两个端口并且每个端口都连接至一个交换结构节点）。在每个交换结构节点上，都有一个转换链路来连接两个交换结构节点。当交换结构节点检测到一个故障端口时，指定发往该故障端口的通信业务将被转换到转换（或备份）端口。然后，其他的交换结构节点将故障切换业务转发到其目标装置。

现在参照图 1，示出了一个双重结构信息包交换式背板（PSB）100 的拓扑结构的总体框图。该背板 100 用来在基于背板的系统中将众多的插件和模块连接在一起。背板 100 通常的拓扑结构为星型的拓扑结构。由于可靠性和可用性对于背板系统来说是重要的设计要求，因此高可用性系统中通常使用双重链路。例如，CompactPCI®信息包交换式背板规范（通常也记作 PICMG®2.16 规范），在此声明其内容结合进本申请以供参考，定义了基于以太网技术的多达 24 个节点板的信息包交换式背板标准，并且采用了星型的拓扑结构。

在本具体实施方案中，PSB100 由 14 个节点板 102、一个第一交换结构板（SFB）104 和一个第二 SFB106 构成。但是，本发明却可扩展到使用任意数量的节点板或结构板。与所有节点板 102 的运行相似，一个节点板 108 运行地连接到第一 SFB104 从而由此进行数据包传送的通信。为了提高可用性，将第二 SFB106 以可行的连接方式增加到节点板 108 用于数据包由此进行通信。每一节点板 102 具有两个链路端口，分别连接到第一 SFB104 及第二 SFB106。该双重结构的 PSB 系统 100 被称为双重星型拓扑结构。该链路端口是一个全双工以太网网络连接（full duplex Ethernet connection），其速度通常约为 10/100/1000Mbps（兆比特每秒），并且只要帧为以太网帧，链路端口的速度可以是任意的。

下表 1 定义了图 1 中高可用性背板系统的主要部分：

表 1. 高可用性背板主要部分

节点板	由一个子系统构成，可以产生和接收数据包
链路端口	一个物理端口，为一个连接到一个节点板和一个交换结构板的链路终点
链路	一个节点板与一个交换结构板之间的一个物理连接
结构板	由多个链路端口构成的一个节点，它在节点板之间提供数据交换功能
重定向链路	连结两个结构板的一个链路，用于对故障切换业务重新路由

现在参照图 2，示出了根据所公开的实施方案的以太网络背板系统 100 的一个更为详尽的框图。一个第一节点板 200 包括两个（或冗余的）上行链路端口（例如，以太网）；一个第一 PHY 上行链路端口 204 和一个第二 PHY 上行链路端口 206，两个端口均分别提供与第一 SFB 104 和与第二 SFB 106 间的通信连接。这样该第一 SFB 104 和第二 SFB 106 就被平行地分别连接到第一节点板 200 上。该节点板 200 包括一个功能执行子系统 210，一个业务分配控制器 208 和两个上行链路端口接口 204 及 206。业务分配控制器 208 执行缓冲和调度，然后基于端口中继算法将业务从功能执行子系统 210 中分派至上行链路端口接口 204 和 206 之中的一个。

每一个节点板的 PHY 上行链路端口（204 和 206）都被归组为一个称为中继端口 207 的逻辑端口。当中继端口 207 接收到数据包时，背板系统 100 并不区分哪一物理的上行链路端口（204 或 206）应接收此数据包。然而当数据包从第一节点板 200 的中继端口 207 发送出来时，业务分配控制器 208 便决定该数据包将发送至哪一物理的上行链路端口（204 和 206），并将该数据包传送到那一上行链路端口。选择中继端口 207 中的输出上行链路端口（204 或 206）所使用的数据可以基于源和/或目标 MAC 地址，或者任何其他的数据包信息的组合。例如，它可以基于来自源和目标 MAC 地址的一个散列键。

CPU 在业务分配控制器 208 中使用并维持一个中继表，以此来决定使用物理的上行链路端口（204 或 206）中的哪一个用于输出数据包业务。该中继表为输出数据包存储了当前的从中继端口到物理的上行链路端口的映射信息。通过访问该中继表的映射信息并决定应使用哪一中继端口和物理的上行链路端口用于数据包业务，背板系统 100 控制第一 SFB104 和第二 SFB106 之间的数据包业务分配。中继表中的这一关联信息根据正常模式运行和故障切换运行情况而动态地变化。

该背板系统 100 还包括了一个具有两个上行链路端口（例如以太网）的第二节点板 202：一个第一上行链路端口 214 和一个第二上行链路端口 216，它们提供与第一 SFB 104 和与第二 SFB 106 之间的一个通信连接，这样该第一 SFB 104 和第二 SFB 106 也被平行地分别连接到第二节点板 202 上。该第二节点板 202 同样包含一个业务分配控制器 218（例如，在本实施方案中为一个以太网络交换装置），它将选择第一和第二上行链路端口（214 和 216）中哪一个将向下链接到第二节点板 202 中的功能执行子系统 220。该第一和第二上行链路端口（214 和 216）为冗余系统。

该第一和第二 SFB（104 和 106）提供了节点板 102 和 202 之间的通信方式。在本具体实施方案中，第一 SFB 104 包括一个结构交换装置 224 和多个 PHY 端口装置 226（例如以太网类型）。交换装置 224 则包括一个在故障切换时通过访问而提供用于重定向数据包的重定向信息的重定向位图（也被称作执行不到位图）229 和一个用于存储一些中继端口中任何一个的状态信息的中继表 231。交换装置 224 则通过 AB PHY 端口 230 与第二 SFB 106 相连。在本具体实施方案中，该第二 SFB 106 包括一个结构交换装置 232 和多个 PHY 端口装置 234（例如以太网类型）。该结构交换装置 232 同样包括一个在故障切换时通过被访问而提供用于重定向数据包的重定向信息的重定向位图 236 和一个存储一些中继端口中任何一个的状态信息的中继表 233。交换装置 232 则通过 AB PHY 端口 238 并经由 AB PHY 端口 240 与第一 SFB 104 相连。

在本实施方案中，背板系统 100 通过一个第一链路 209 将第一节点板

200 上的第一 PHY 上行链路端口 204 连接到第一 SFB104 上的一个 PHY 端口 242。通过一个第二链路 211，第二 PHY 上行链路端口 206 连接到第二 SFB 106 上的一个 PHY 端口 244。第二节点板 202 上的第一 PHY 上行链路端口 214 通过一个第三链路 213 连接到第一 SFB104 上的一个 PHY 端口 246，与此同时第二 PHY 上行链路端口 216 则通过一个第四链路 215 连接到第二 SFB 106 上的一个 PHY 端口 248。

在一个实施方案中，节点板 200 和 202 之间的节点板信号通过第一 SFB104 经由各自的第一上行链路端口（204 和 214）在第一节点板 200 的功能执行子系统 210 与第二节点板 202 的功能执行子系统之间进行通信。相似地，作为对第一链路 209 的检测出的故障的响应，进行故障切换，节点板信号通过第二 SFB 106 经由各自的第二上行链路端口（206 和 216）在第一节点板 200 的功能执行子系统 210 与第二节点板 202 的功能执行子系统之间进行通信。一旦第一链路 209 的故障被排除后，通过第一上行链路端口（204 和 214）恢复到正常模式的运行状态。

链路故障检测可以在不同层下进行，例如，IEEE 802.3 规范为以太网 PHY 制定了一个 PHY-层的检测机制。在没有数据业务时，一个发送的 PHY 装置会定期（如每 16±8 毫秒）发出一个简单的脉动（HB）脉冲，称为正常链路脉动（NLP），如果接收的 PHY 装置在预定时间（如 50-150 毫秒）内既没有检测到数据包的到来也没有检测到 NLP 的到来，则接收的 PHY 装置就认为链路出现了故障。

在系统层，在节点板上的或者是附加在一个交换结构板上的本地 CPU 可用来通过定期向系统的另一个 CPU 发送脉动数据包的方法来检查链路的完整性。但是，这一方法却要占用 CPU 更多的处理功率和时间来处理信息以询问链路从而检测到已发生的链路故障。即使在链路繁忙的情况下，该方法还要求额外的带宽。由于决策的路径长，用此方法来恢复链路显得较慢。

在本发明中，链路故障检测是在 MAC 层中实施的。优选在背板系统 100 中在 MAC 层进行检测是由于以下原因。在背板环境下，不是所有执行

的PHY装置都能够像以太网络PHY装置一样能够嵌入一个链路故障检测机制（例如，LVDS装置就不能应用这一检测技术）。因此，一个MAC模块需提供链路故障检测。而且，在 PHY 实施中的快速链路故障检测需要约50-150 毫秒的处理时间，而 MAC 模块却能够快得多地检测链路故障，其检测速度依端口速度而定。对于一个千兆位的端口，检测时间不到 1 毫秒，而对于一个 100Mbps（兆比特秒）的端口，检测时间为几个毫秒。此外，PHY 层的链路故障检测不能检测到由 MAC 模块失灵造成的链路故障。然而，应注意到，当执行一个 PHY 方案时，所公开的 MAC 层的检测方案能够兼容该 PHY 层的链路故障检测方案。

在讨论故障切换操作时，在此具体实施方案中，假设背板控制逻辑控制数据包的传送路线为：数据包从第一节点板 200 中通过其第一 PHY 上行链路端口 204、经过第一链路 209，进入结构端口 242，通过结构装置 228 转换为结构端口 246 的输出，经过第三链路 213，进入第二节点板 202 的第一上行链路端口 214，并由交换装置 218 转换到第二节点板 202 的第一子系统 220 中。由此，当第一节点板 200 检测到第一链路 209 出现故障时，背板控制逻辑启动数据包业务的故障切换，从第一 PHY 上行链路端口 204 经过到第二上行链路端口 SFB106 到达第二 PHY 上行链路端口 206。故障切换是通过改变中继表并迫使中继端口 207 的所有数据包业务（原先使用现有的出现故障的第一上行链路端口 204）仅使用第二上行链路端口 206 来完成的。

首先，假设与第一节点板 200（图 2）的第一上行链路端口 204 互连的第一链路 209 已出现故障。当第一 SFB 104 检测到该第一链路 209 的故障时，所有第一上行链路端口 204 输出的数据包业务就被重定向，从而发送到重定向链路 240。然后，第二 SFB 106 接收从重定向端口 238 发来的数据包（或数据帧），并且通过第二链路 213 将它们发送给第一 SFB 104。

在运行过程中，节点板使用端口中继来执行故障切换。如前所述，节点板上的各上行链路端口将归组为一个逻辑中继端口。当从功能子系统 210 发出的数据包到达后，业务分配控制器 208 将首先在本地 MAC 地址表

中搜索该数据包的目标MAC地址。该MAC表显示出MAC地址的相关信息并且该目标端口可能是上行链路端口或逻辑中继端口之一。如果MAC与一个上行链路端口204或206相关，则该业务将总是被传送到那一特定端口而且故障切换将不适用于该特定业务。如果目标是上行链路中继，那么业务分配控制器208将执行中继分配算法将数据包派送至上行链路端口204或206其中的一个。

可基于对源MAC和/或目标MAC地址进行散列而产生的散列键来选择物理端口。

第一节点板的MAC表

MAC 地址	控制信息	状态	端口/中继端口
MAC_b	…		中继端口 1
…			

第一节点板上中继端口1的中继端口表

散列键	物理端口
0	端口 1a
1	端口 1b
2	端口 1a
3	端口 1b

第一节点板200的CPU通过在中继表中适当地为物理端口赋值“ON”来控制中继端口207中上行链路端口（204和206）之间的数据包业务的

负载分配。

当节点板 200 的 CPU 被通知例如在中继端口 207 的链路 209 出现故障时，CPU 就会为中继表中两个冗余节点板（200 和 202）改变所有第一链路端口（204 和 214，在中继表中也被标记为端口 1a）和第二链路端口（206 和 216）的状态。这样，将迫使使用中继端口 207 的所有数据包业务使用第二上行链路端口（206 和 216，同样在中继表中标记为端口 1b）。故障切换由此实现。

当结构节点 104 检测到它的任何一个端口的链路故障时，该节点板的 CPU 会收到通知并启动故障切换程序。该结构板将故障信息传达给其他结构板节点。例如，第一 SFB 104 的 CPU 将与该错误链路相连的节点板（如，现在为一个不可到达节点）告知第二 SFB106。第一 SFB104 中有一个重定向位图，指示哪一个端口将不能通过其他节点到达。当链路故障通知被收到后，CPU 会更新重定向位图 229 的信息并传回一个 ACK。重定向位图 229 用作一个从重定位链路接收到的业务的转发区域，这样，两个上行链路都在工作的节点板不会收到两份相同的广播数据包。

如重定向节点位图所示，重定位链路所接收到的数据包只能被转发给与节点连接的端口。通过提供该重定向节点位图，可防止节点接收到重复的广播数据包。如果没有重定向位图，一个广播数据包将被转发给包括重定向端口在内的所有端口。第二 SFB 106 也将会广播此数据包。结果，除了源节点板 204 上的一个上行链节点端口以外，所有节点都会收到分别由两个结构板各自发出的相同的数据包。通过使用重定向位图，第二 SFB106 就只将数据包转发给第一 SFB104 上的不可到达节点，而不向从第一 SFB104 接收数据包的其他节点发送。

由于来自有故障链路节点的业务会被重定向到可用的链路，发生故障链路的交换结构板将不再访问与故障端口相关的 MAC 地址。于是该节点板上的 MAC 词条将最终失去时效。结果，目标为 A 的数据包将被滥用。因此，接收到链路故障通知的交换结构板的 CPU 应将与故障链路端口有关的所有 MAC 词条设为“静止”，这些具有“静止”状态的表格词条将不会失去

时效。

当第一 SFB104 从第二 SFB106 收到故障链路信息的确认（ACK）之后，第一 SFB104 的 CPU 通过对故障端口到重定向端口间进行重新映射开始对预定发送到故障端口的数据包故障切换至重定向端口。

在发送端，如果没有数据包被实时发送，一个 MAC 传送模块就会定时发出一个 MAC 脉动信息。脉动信息的持续时间是可被设置的。在目前的实施中，时间间隔的单位为一个“时间槽”，512 字节的传送时间，也就是以 10 兆比特每秒的速度传输 51.2 微秒和以 100 兆比特每秒的速度传输 5.12 微秒。如果链路在繁忙地发送常规数据包业务，链路脉动信息数据包就不会被发送，从而使得在链路繁忙时的带宽最优化。这是较之以 CPU 类生成树方式进行链路故障检测而言的一个优点。

应注意到上行链路端口和交换结构板所用的 PHY 装置并不局限于以太网装置，其他传统的背板 PHY 装置如 LVDS（低压差分信号）也是可用的。（LVDS 是一种低功率、低噪音的高速传送差分技术。）

参照图 3，根据一个公开的实施方案描述了一个节点板的故障切换程序的流程图。当某一装置检测到一个链路故障，它将马上进入故障切换模式，并将故障链路上的业务重定向至一个可用的链路。流程从功能程序块 300 开始，此程序块中创建一个冗余 PHY 上行链路端口的中继端口。在功能程序块 302 中，对所有端口启动链路完整性检查。程序进入一个决策程序块 304，在这里如果没有检测到链路故障，程序从路径“N”流出并回到程序块 302 执行下一链路完整性检查。反之，如果检测到一个链路故障，程序从决策程序块 304 路径“Y”流出，到达功能程序块 308，对数据包业务进行故障切换到一个可用端口。然后程序进入决策程序块 310 以判断故障切换条件是否已解决。如果没有，则从路径“N”流出并回到功能程序块 308，继续对数据包业务进行故障切换。如果已解决，则从决策程序块 310 的路径“Y”出来，进入功能程序块 312，以重新开始正常模式。程序接下来回到功能程序块 302 开始下一完整性检查。

现在参看图 4，根据一个公开的实施方案描述了一个结构板的故障切

换程序的流程图。程序始于功能程序块 400，此程序块中，背板系统 100 执行正常模式下的各项功能。链路完整性检测在功能程序块 402 中进行。在决策程序块 404 中，如果没有检测到链路故障，程序从路径“N”流出，回到程序块 402 执行下一检测功能。如果检测到一个链路故障，程序从决策程序块 404 路径“Y”流出，到达功能程序块 406 启动故障切换模式。在故障切换模式下，如功能程序块 408 中所示出的那样，将链路故障数据传送到其他的结构板。然后如功能程序块 410 中所示出的那样，用故障端口状态信息更新 MAC 表。然后程序进入决策程序块 412，以判断其是否接收到了其他结构板的确认（ACK）信号。如果没有，程序从路径“N”流出，继续检查是否接收到 ACK 信号。如果已收到 ACK 信号；程序从决策程序块 412 路径“Y”流出，到达功能程序块 414，以便根据重定向位图中包含的重定向信息将故障端口映射到一个重定向端口。数据包业务从而相应地被重新定向，直至故障切换得到解决。决策程序块 416 中，检查确定故障切换是否已解决。如果没有，程序将由路径“N”流出，进入功能程序块 418 继续在故障切换模式下运行。然后程序将循环回到决策程序块 416 的输入端，以执行下一个故障切换是否已恢复的检查。如果链路已恢复，程序从决策程序块 416 的路径“Y”流出，进入功能程序块 420，在这里链路恢复数据将被转发给其他结构板。然后，如功能程序块 422 中所示出的那样，MAC 表中的数据将相应被更新，以反映链路的恢复。之后，如功能程序块 424 中所示出的那样，更新重定向位图，将重定向信息转移到冗余端口。然后，如功能程序块 426 中所示出的那样，背板系统 100 恢复正常运行模式。程序接下来循环回到功能程序块 400，开始执行正常的运行转换功能。

参照图 5，描述了来自发送端的脉动信号的状态图。程序始于功能程序块 500，MAC 链路脉动信号（LHB）在此被激活，状态指示为“Ready（准备）”。如功能程序块 502 中所示出的那样，如果接收到一个 LHB 信号，其状态就被标示为“OK（正常）”。在决策程序块 504 中，将判断链路状态是否为“Pending（挂起）”状态。如果是，程序由路径“Y”流出，进入功能程序块 506，以便仅转发 CPU 数据包业务。在决策程序块 508 中，检查链路故障，如果没有链路故障被确定，程序由路径“Y”流出，进入功能程序块 500，以继续激活 MAC LHB 信号。反之，如果已检测到链路故障，

程序会从决策程序块 508 的路径“N”流出而终止。

如果一个链路状态检查结果为非挂起状态，程序就会从决策程序块 504 路径“N”流出，进入决策程序块 510，以判断数据包是否已准备传送。如果是，程序就从路径“Y”流出，进入功能程序块 512，以转发该数据包，并将 LHB 定时器重置。然后，程序从功能程序块 512 循环回到决策程序块 504 再次判断系统是否为挂起状态以便更新链路状态。如果数据包还未准备好传送，程序从决策程序块 510 的路径“N”流出，进入决策程序块 514，以判断是否 LHB 的定时器已过期。如果没有，程序从路径“N”流出回到功能程序块 504，以检查链路状态。如果 LHB 的定时器已到期，程序从决策程序块 514 的路径“Y”流出，进入功能程序块 516，以发送一个或多个 LHB 数据包。然后程序回到功能程序块 504，再次检查链路状态。

以下是一个 64 字节链路脉动（HB）信息的格式（十六进制）。

Dest_MAC_address(6)	SRC_MAC_address(6)	Ethetype(2)	Opcode(2)	Pad(44)	CRC(4)
01-80-C2-00-00-01	Port MAC Address	88-08	00-02	Pad 44 “00”	CRC-32

如上表所示，目标 MAC 地址域是一个 6 个字节的值，为 01-80-C2-00-00-01。IEEE 802.3x 标准全双工暂停（IEEE Std. 802.3x Full Duplex PAUSE）操作下的流程控制信息的地址是共享的。端口 MAC 地址的域被用作源 MAC 地址，为 6 个字节。以太类型（Ethertype）域为两个字节，为 88-08，表示了 MAC 控制格式。该 2 字节操作码（Opcode）的值是一个可编程的值（例如：值 00-02），但是不能为“00-01”，因为“00-01”已经被定义为 IEEE 802.3X 标准的流程控制帧。传送和接收的末端都必须使用相同的操作码值。长度为 44 字节的填充（Pad）域添加 44 个字节的零“00”，以达到以太网数据帧的最小长度：64 字节。

下面是HB控制信息的格式。

0	1	2	3	4	5	6	7
01	80	C2	00	00	01	00	00
00	00	00	00	88	08	00	02
00	00	00	00	00	00	00	00
....				CRC32			

参照图6，它描述了来自接收端的脉动信号的状态图。为启动MAC模块，接通电源，这时所有端口被重置且状态变为“链路准备就绪（LINK READY）”。CPU此时激活脉动特征，同时MAC开始发送MAC LHB信号。MAC等待接收器从远程装置中发出显示相同能力的第一LHB信号，然后将状态变为“链路完好（LINK OK）”，然后，转发被转换的业务。在接收端，MAC链路检测模块监控到达的数据包，如果一个好的数据包在从接收到上一数据帧之后开始计时的预先定义的窗口时间（其时间窗被标识为LINK_FAIL））内到达，则该链路处于工作模式。LINK_FAIL时间窗的值是可设的，通常将其设成约为LHB信息传送周期的两倍。一个好的数据包表明，一个含有MAC脉动信息的好的数据帧或者控制帧已被传送。应注意到，MAC LHB信号在MAC模块被吸收，且不能被转发到CPU或其他的端口。一旦MAC在LINK_FAIL窗口内未检测到一个数据帧，它就会进入“链路挂起（LINK PENDING）”状态，并发给CPU一个脉动丢失（HARTBEAT LOST）信息。当该MAC收到N个连续的坏的数据帧时，它也会进入“链路挂起”状态。在“链路挂起”状态下，MAC模块停止传送被切换的业务。然而，CPU数据包和LHB数据包在该状态下仍被连续传送。CPU可通过尽量与远程装置通信而验证最终的链路状态。如果连接不能恢复，该端口进入链路故障状态。

程序从决策程序块 600 开始，判断 LHB 是否已被激活。如果没有，程序由路径“N”流出，到达一个结束终点。否则，程序由路径“Y”流出，进入功能程序块 602，以接收该第一 LHB 数据包。然后链路状态被设为“OK”。程序流入功能程序块 604，将定时器和错误计数器重置。在决策程序块 606 中，系统判断一个帧结束部分（end-of-frame）是否已被接收到。如果是，程序由路径“Y”流出，进入决策程序块 608，以判断所接收到的帧是否是好的。如果是，程序由路径“Y”流出到功能程序块 604，将定时器和计数器重置，以准备接收下一个数据包（或帧）。如果所接收到的帧不是好的，程序就由决策程序块 608 的路径“N”流出到达功能程序块 610，以增加错误计数。然后程序进入决策程序块 612，判断错误累计次数。如果错误数小于或等于预定值 N，则程序由路径“N”流出，到达决策程序块 606 的输入端，判断下一个帧结束部分是否已被接收到。如果错误计数器的错误数大于或等于预定值“N”，程序由路径“Y”流出，进入功能程序块 614，以将状态改为“Pending（挂起）”。接着，程序进入决策程序块 616，判断 CPU 验证是否失败。如果是，程序由路径“Y”流出到达结束端点。如果 CPU 验证没有失败，程序由路径“N”流出到功能程序块 604，将定时器和错误计数器重置。

如果没有收到帧结束部分，程序由决策程序块 606 路径“N”流出，进入决策程序块 618，以判断 LHB 定时器是否过期。如果没有，程序由路径“N”流出回到决策程序块 606 的数据输入端。如果 LHB 定时器已过期，程序由路径“Y”流出，进入决策程序块 620，以判断是否正在进行数据包接收。若没有，程序流入功能程序块 614，将状态改变为“Pending（挂起）”，然后进入决策程序块 616，以判断 CPU 验证是否失败。如果数据包接收正在进行中，程序由决策程序块 620 的路径“Y”流出，进入决策程序块 622，以判断是否收到一个好的帧（或数据包）。如是，程序由路径“Y”流出到功能程序块 604，将 LHB 定时器和错误计数器重置。如果没有收到一个好的数据包，程序流入功能程序块 614，将状态改变为“Pending（挂起）”，接着进入决策程序块 616，以判断 CPU 验证是否失败。

尽管以上详细地描述了优选实施方案，但是应当理解，在不脱离所附权利要求书所限定的本发明精神和范围的情况下，可以对其作出各种不同的变化、替换和可供选择的方案。

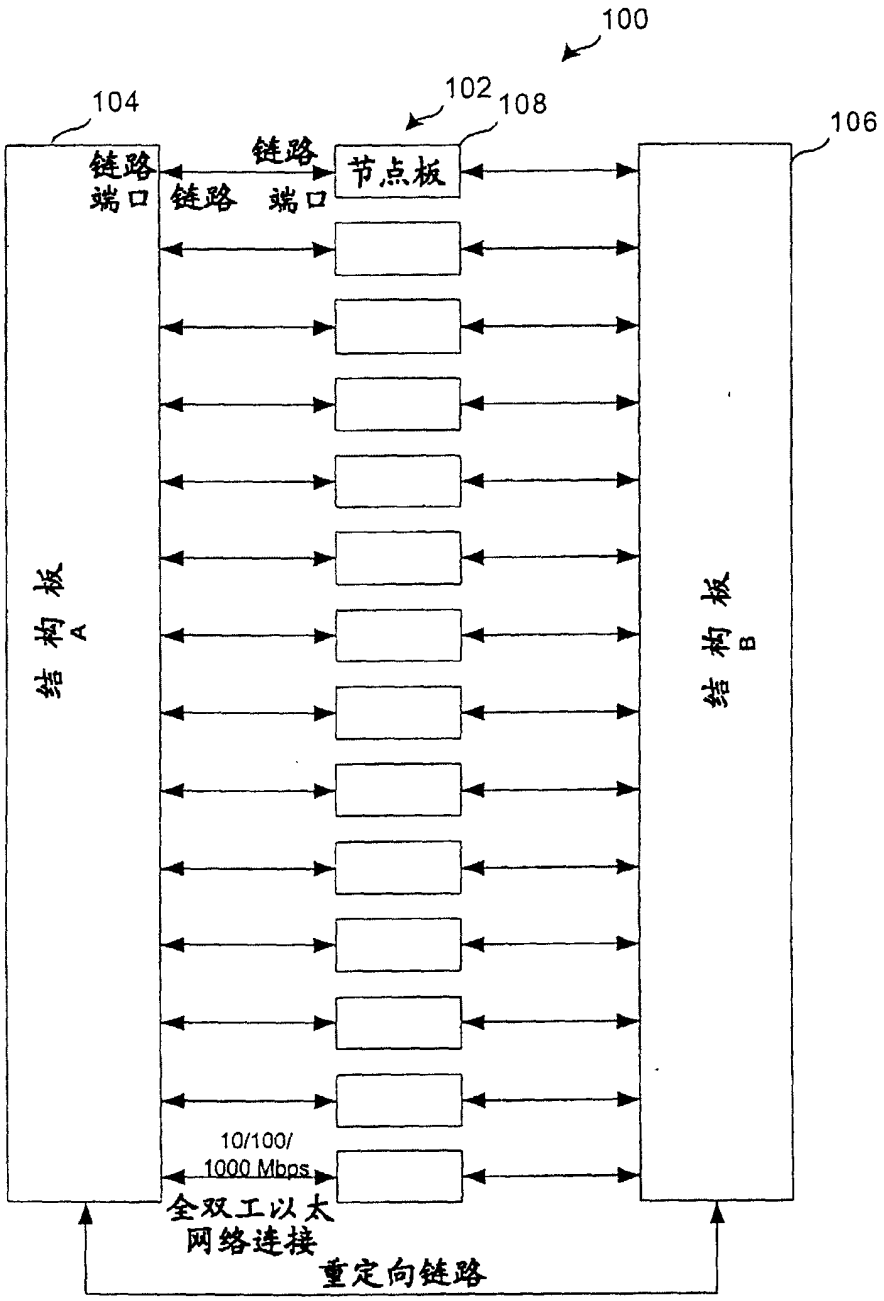

100
102
104
106
108
链路端口
链路
链路端口
节点板
结构板A
结构板B
10/100/1000 Mbps
全双工以太网络连接
重定向链路

图 1

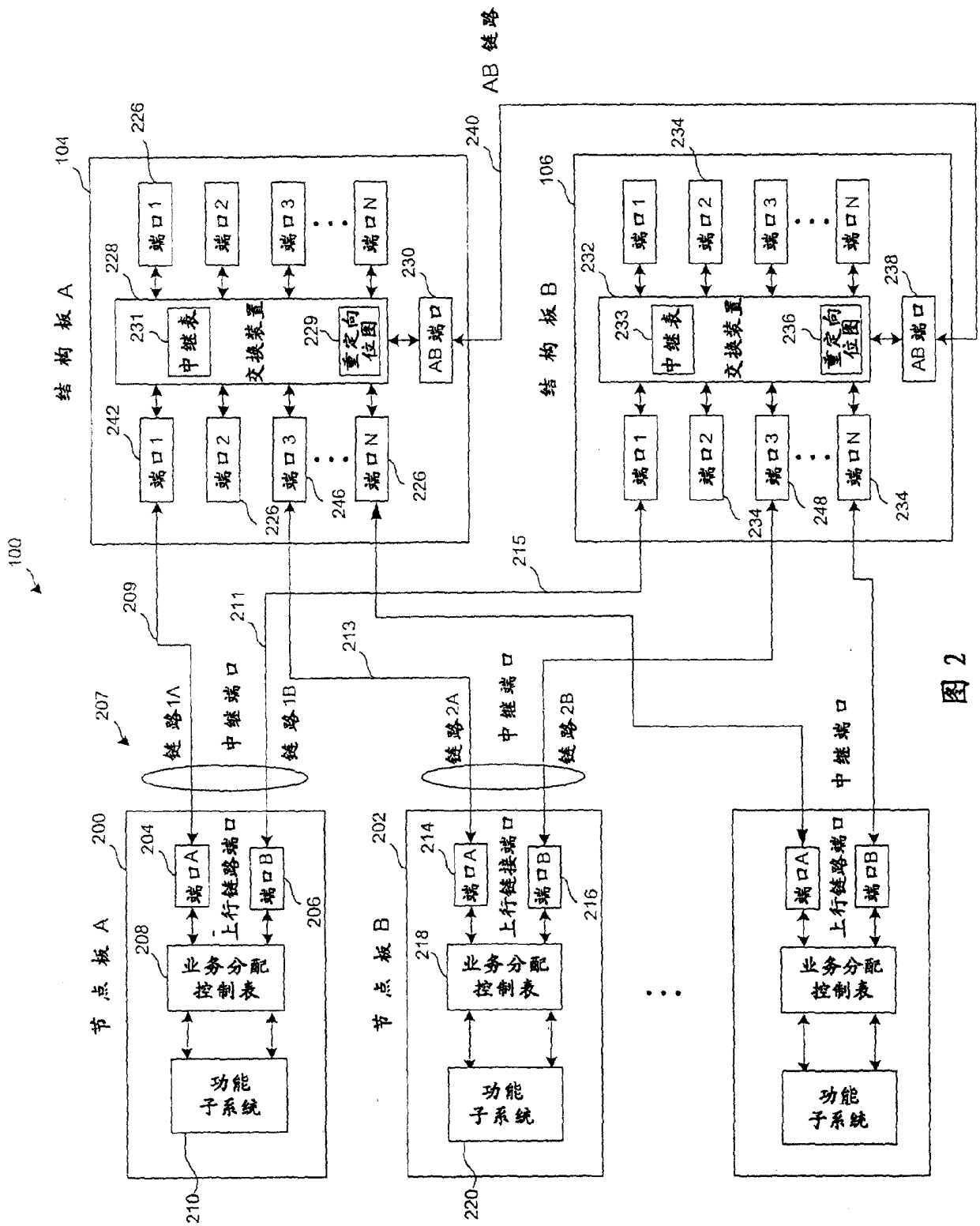

AB 链路
结构板 A
结构板 B
交换装置
中继表
重定向位图
端口 1
端口 2
端口 3
端口 N
AB 端口
中继端口
链路1A
链路1B
链路2A
链路2B
节点板 A
节点板 B
上行链路端口
端口A
端口B
业务分配控制表
功能子系统
100
104
106
200
202
204
206
207
208
209
210
211
213
214
215
216
218
220
226
228
229
230
231
232
233
234
236
238
240
242
246
248

图 2

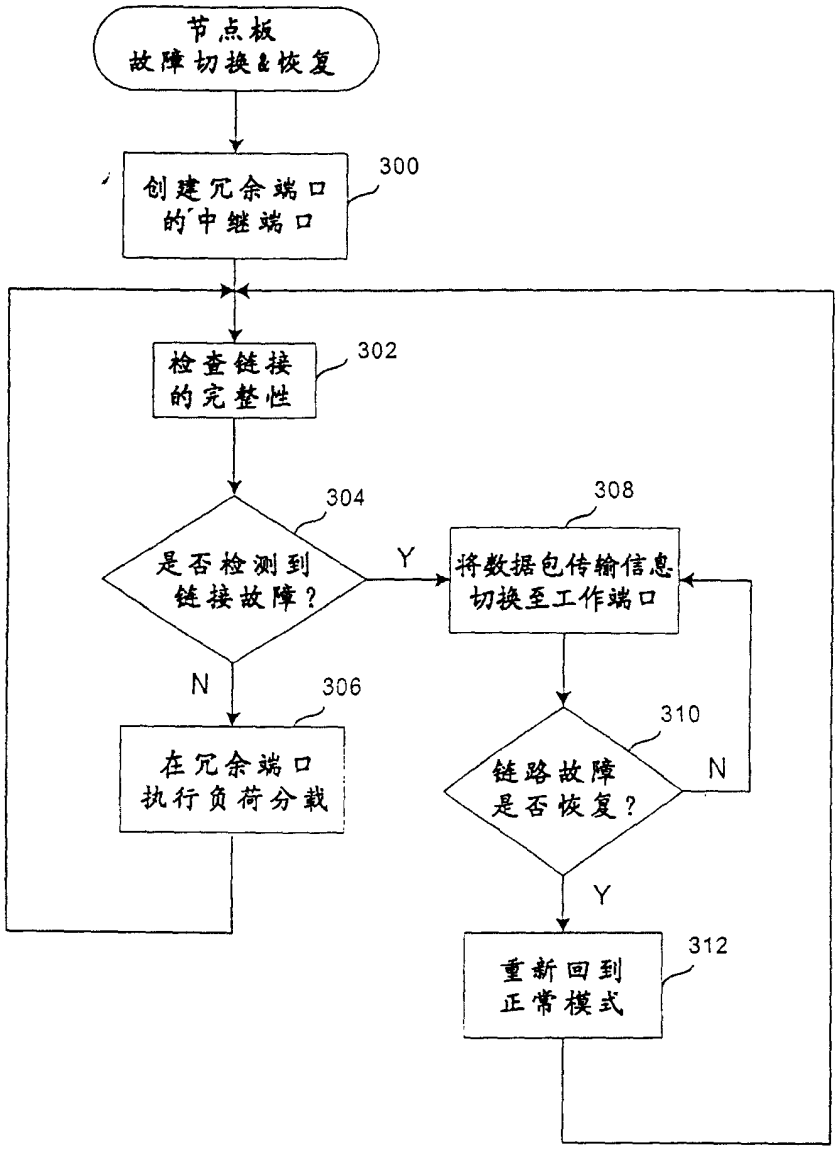

节点板
故障切换&恢复
创建冗余端口
的中继端口
300
检查链接
的完整性
302
是否检测到
链接故障？
304
Y
N
将数据包传输信息
切换至工作端口
308
在冗余端口
执行负荷分载
306
链路故障
是否恢复？
310
N
Y
重新回到
正常模式
312

图 3

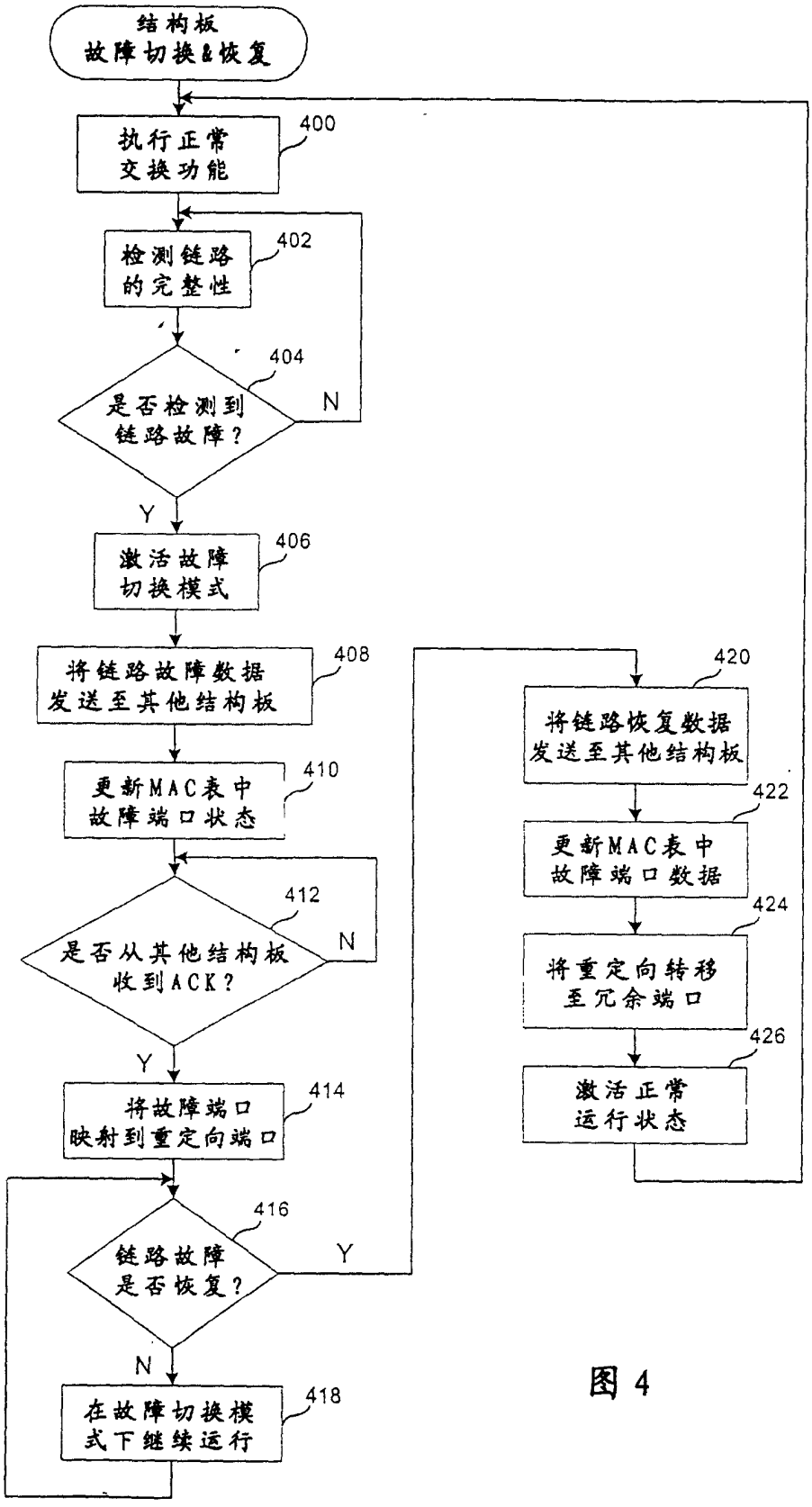

结构板
故障切换&恢复
执行正常
交换功能
400
检测链路
的完整性
402
是否检测到
链路故障？
404
N
Y
激活故障
切换模式
406
将链路故障数据
发送至其他结构板
408
更新MAC表中
故障端口状态
410
是否从其他结构板
收到ACK？
412
N
Y
将故障端口
映射到重定向端口
414
链路故障
是否恢复？
416
Y
N
在故障切换模
式下继续运行
418
将链路恢复数据
发送至其他结构板
420
更新MAC表中
故障端口数据
422
将重定向转移
至冗余端口
424
激活正常
运行状态
426

图 4

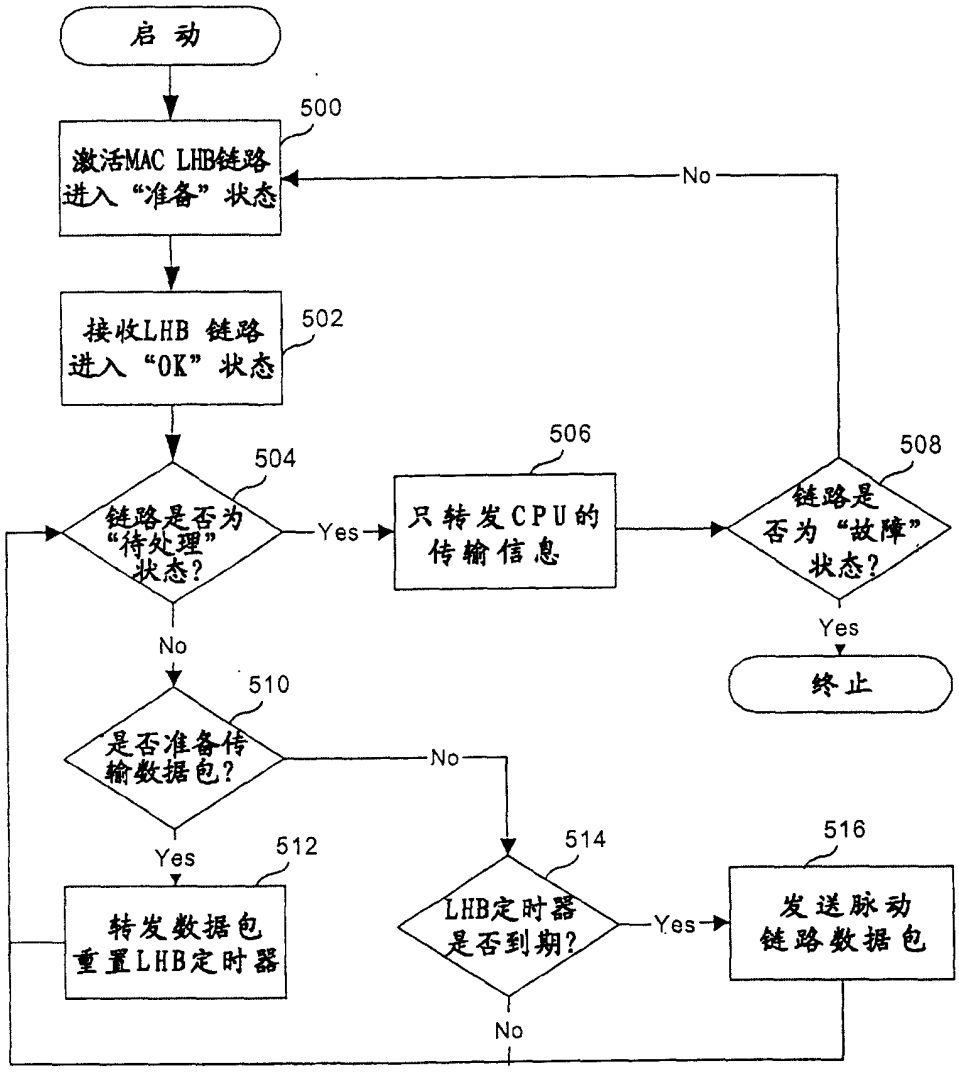

启 动
500
激活MAC LHB链路
进入“准备”状态
No
接收LHB 链路
进入“OK”状态
502
504
链路是否为
“待处理”
状态?
Yes
506
只转发CPU的
传输信息
508
链路是
否为“故障”
状态?
Yes
终止
No
510
是否准备传
输数据包?
No
Yes
512
转发数据包
重置LHB定时器
514
LHB定时器
是否到期?
Yes
516
发送脉动
链路数据包
No

图 5

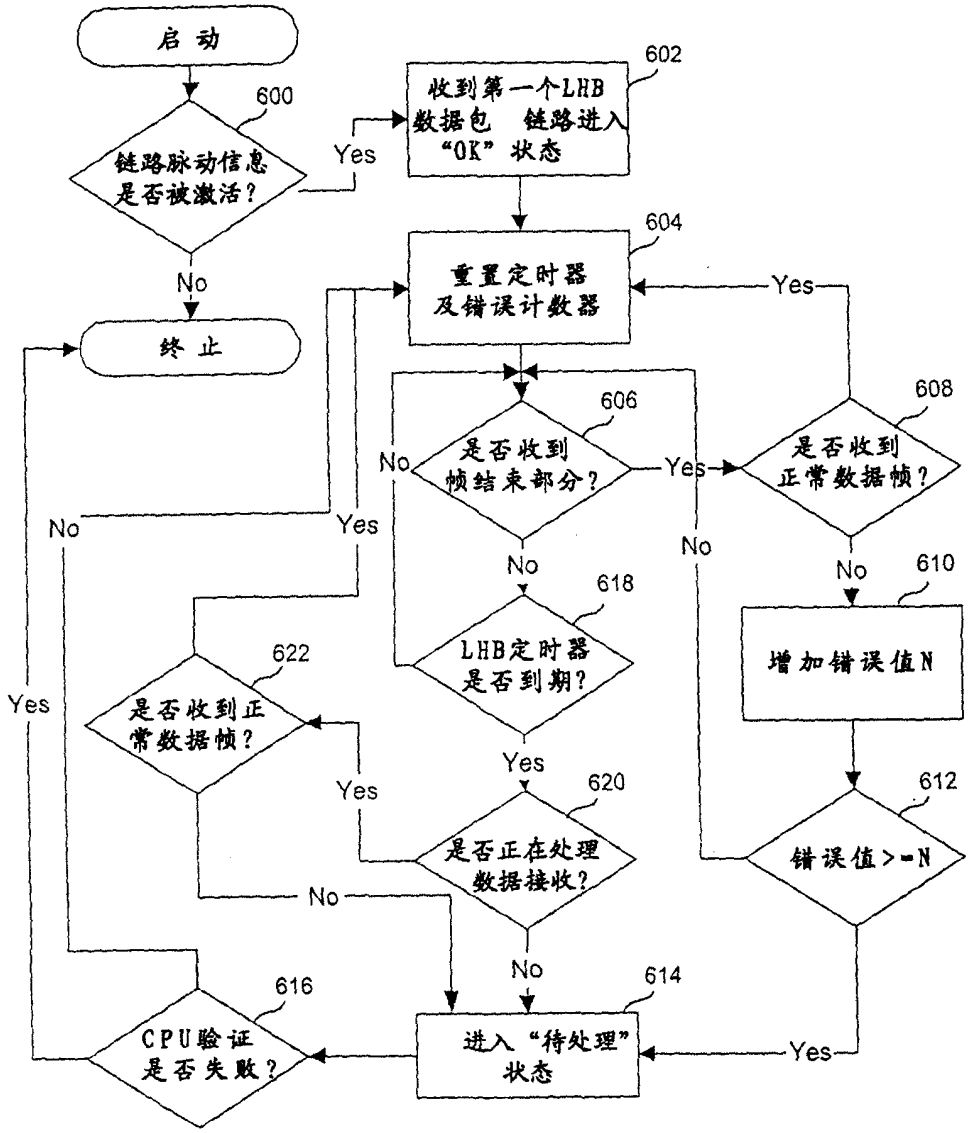

启 动
600
链路脉动信息是否被激活?
Yes
No
终 止
602
收到第一个LHB数据包 链路进入"OK"状态
604
重置定时器及错误计数器
606
是否收到帧结束部分?
No
Yes
608
是否收到正常数据帧?
No
Yes
610
增加错误值N
612
错误值>=N
Yes
No
618
LHB定时器是否到期?
No
Yes
620
是否正在处理数据接收?
Yes
No
622
是否收到正常数据帧?
Yes
No
614
进入"待处理"状态
616
CPU验证是否失败?
Yes
No

图 6