



(12)发明专利

(10)授权公告号 CN 103858105 B

(45)授权公告日 2017.08.29

(21)申请号 201280048841.9

(22)申请日 2012.09.28

(65)同一申请的已公布的文献号

申请公布号 CN 103858105 A

(43)申请公布日 2014.06.11

(30)优先权数据

1117230.1 2011.10.05 GB

(85)PCT国际申请进入国家阶段日

2014.04.03

(86)PCT国际申请的申请数据

PCT/GB2012/052413 2012.09.28

(87)PCT国际申请的公布数据

W02013/050745 EN 2013.04.11

(73)专利权人 美光科技公司

地址 美国爱达荷州

(72)发明人 马来克·派卡斯基

(74)专利代理机构 北京律盟知识产权代理有限公司 11287

代理人 孙宝成

(51)Int.Cl.

G06F 11/07(2006.01)

(56)对比文件

US 2008/0288633 A1, 2008.11.20,

US 2011/0047413 A1, 2011.02.24,

US 2002/0194319 A1, 2002.12.19,

US 2006/0136789 A1, 2006.06.22,

审查员 叶璇

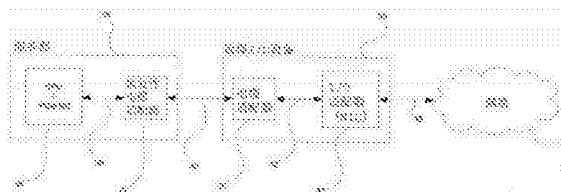
权利要求书2页 说明书7页 附图7页

(54)发明名称

连接方法

(57)摘要

一种用于管理第一设备和第二设备之间的通信故障的方法。该方法包括在所述第一设备和第二设备中间的第三设备处检测第一设备和第二设备之间的通信故障并将第一数据从第三设备传送至第一设备,该第一数据指示第二设备不能被第一设备使用。



1. 一种用于管理第一设备和第二设备之间的通信故障的方法,该方法包括:

在所述第一设备和所述第二设备中间的第三设备处拦截所述第一设备向所述第二设备传送的第二数据;

在第三设备处检测所述第一设备和所述第二设备之间的通信故障,该通信故障导致对所述第二数据的响应不被所述第一设备接收;以及

从所述第三设备向所述第一设备传送第一数据,该第一数据指示所述第二设备不能被所述第一设备使用;

其中意欲指示所述第一数据的源的所述第一数据的一部分指示所述第二设备;其中所述第一数据包括指示所述第一数据为响应于所述第二数据的数据。

2. 根据权利要求1所述的方法,其中所述第一设备和第二设备中的至少一个为服务器。

3. 根据权利要求1所述的方法,其中所述第一设备和第二设备中的至少一个为I/O设备。

4. 根据权利要求1所述的方法,其中所述第一设备经由PCI Express连接而被连接至所述第二设备。

5. 根据权利要求1所述的方法,其中所述第一数据指示所述第二设备不可达。

6. 根据权利要求1所述的方法,其中所述第一数据指示所述第二设备发生故障。

7. 根据权利要求1所述的方法,其中所述第一数据指示所述第一数据受损。

8. 根据权利要求1所述的方法,其中检测所述第一设备和所述第二设备之间的通信故障包括检测连接所述第一设备和所述第二设备的电缆的断开。

9. 根据权利要求1所述的方法,其中检测所述第一设备和所述第二设备之间的通信故障包括从所述第三设备向所述第二设备传送第三数据;以及

如果对所第三数据的响应在预定时间段中未从所述第二设备接收到,则确定所述第一设备和所述第二设备之间的通信发生故障。

10. 根据权利要求1所述的方法,其中所述第二数据包括指示该第二数据属于的交易的数字段;以及

所述方法还包括在所述第三设备处产生第一数据;所述产生包括在所述第三设备处至少部分地依赖于由所述第二数据的数据字段指示的交易而从多个已存储的错误信息中选择一错误信息,其中所选择的错误信息被包括在所述第一数据中。

11. 一种用于管理第一设备和第二设备之间的通信故障的计算机装置,该计算机装置包括:

存储器,用于存储处理器可读指令;以及

处理器,被布置成读取并执行所述存储器中存储的指令;

其中所述处理器可读指令包括被布置成控制所述计算机执行根据权利要求1-10中任一项所述的方法。

12. 一种用于管理第一设备和第二设备之间的通信故障的装置,该装置包括:

用于拦截所述第一设备向所述第二设备传送的第二数据的装置;其中该第二数据包括指示该第二数据属于的交易的数字段;

用于检测所述第一设备和所述第二设备之间的通信故障以使得对所述第二数据的响应不被所述第一设备接收的装置;以及

用于在第三设备处产生第一数据,并且将该第一数据从所述第三设备传送到所述第一设备的装置,以防止由于所述第一设备没有接收到来自所述第二设备的对所述第二数据的响应而造成的所述第一设备的服务中断,该第一数据指示所述第二设备不能被所述第一设备使用;

其中所述第一数据包括指示所述第一数据的源的部分,所述第一数据的部分指示所述第二设备从而使得所述第一数据看起来是由所述第二设备产生的;所述第一数据包括指示所述第一数据为响应于所述第二数据的数据;以及

其中所述用于在所述第三设备处产生第一数据,并且将该第一数据从所述第三设备传送到所述第一设备的装置,包括用于在所述第三设备处至少部分地依赖于由所述第二数据的数据字段指示的交易而从多个已存储的错误信息中选择一错误信息的装置,其中所选择的错误信息被包括在所述第一数据中。

13. 一种装置,包括:

恢复性电缆适配器,其被配置为:

拦截从所连接的第一设备向所连接的第二设备传送的第二数据;其中该第二数据包括指示该第二数据属于的交易的数字段;

检测所述连接的第一设备和第二设备之间的通信故障,该故障导致对所述第二数据的响应不被所述第一设备接收;以及

在第三设备处产生第一数据,并且将该第一数据从所述第三设备传送到所述第一设备,以防止由于所述第一设备没有接收到来自所述第二设备的对所述第二数据的响应而造成的所述第一设备的服务中断;该第一数据指示所述第二设备不能被所述第一设备使用;

其中所述第一数据包括指示所述第一数据的源的部分,所述第一数据的部分指示所述第二设备从而使得所述第一数据看起来是由所述第二设备产生的;所述第一数据包括指示所述第一数据为响应于所述第二数据的数据;以及

其中所述的在第三设备处产生第一数据,包括在所述第三设备处至少部分地依赖于由所述第二数据的数据字段指示的交易而从多个已存储的错误信息中选择一错误信息,其中所选择的错误信息被包括在所述第一数据中。

连接方法

技术领域

[0001] 本发明涉及管理第一设备和第二设备之间的通信故障的方法。

背景技术

[0002] 在计算机系统中经常需要在设备之间发送数据,例如,经常需要将处理设备连接至多个输入和输出设备。适当的数据通信是通过连接设备以允许设备经由链路相互发送数据包的方式来获得的,其中所述链路可以是有线链路或无线链路。

[0003] 对于连接设备的系统来说,能够管理未预料的故障或那些连接设备中的一个的不可用越来越重要。例如,在线缆意外断开(例如从而连接设备从服务器断开)时或在远程设备故障的情况下,系统需要是恢复性的(继续提供他们指定的服务)。然而,在一般情况下,操作系统为运行(“热的”)系统中的输入/输出(I/O)适配器的管理的(例如计划的)移除和替代提供一些支持(即,热插拔/热交换),但是在热系统中很少或没有提供对设备意外移除的支持。

[0004] 之前,例如PCI Express仅仅被用于服务器内I/O适配器的永久连接(即,“框内(in-box)”)。对于这种配置,其中I/O设备被永久地安装在服务器内,总体上I/O设备的故障代表服务器的故障,并且因此,该服务器的连续操作不被认为是必须的。对于远程连接PCI Express I/O设备(即连接至服务器外部的I/O设备)的越来越多的关注意味着需要服务器平稳支持连接设备的未预料热移除成为系统恢复力(resilience)的关键。

[0005] 数据包的发送和接收通常基于数据交易来描述。数据交易包含正在设备间发送的一个或多个数据包。PCI Express实施分离数据交易模式,其中源设备向目标设备传送请求数据包,并且等待来自目标设备作为响应的完成数据包。通常,操作系统不适于平稳地处理失败的PCI Express数据交易。例如,如果服务器向连接的设备发送请求数据包,并且意外地,未接收到响应于该请求的完成数据包,则服务器的操作系统可能崩溃。如此,基于PCI Express的现有的连接系统在PCI Express资源出现意外不可用的情况时可能崩溃。

[0006] PCI Express的标准实施不能提供处理PCI Express子系统或连接的PCI Express设备故障的适当手段。这导致难以建立具有一个可接受的水平的故障恢复力的远程或共享I/O系统。

发明内容

[0007] 本发明实施方式的目的解决或减轻上述一个或多个问题。

[0008] 根据本发明的第一方面,提供管理第一设备和第二设备之间的通信故障的方法,该方法包括:在第一设备和第二设备中间的第三设备处检测第一设备和第二设备之间的通信故障;从第三设备向第一设备传送第一数据,所述第一数据指示第二设备不能被第一设备使用。

[0009] 以这种方式,第一设备不经受与第二设备的通信的意外中断,该意外中断可以使得第一设备故障。通过检测第一设备和第二设备之间通信的中断,第三设备可以向第一设

备发送合适数据以防止第一设备故障。例如,第一数据可以通过通知第一设备该第二设备已经改变状态来向在第一设备上运行的软件(OS、驱动器、应用)隐藏通信故障。例如,第三设备可以模仿第二设备继续存在,但是处于第一设备理解为“有效但不能使用”的状态。第一设备可以适用于平稳处理“有效但不能使用”的状态,而无需服务中断(例如,通过与第二设备的备选通信)。

[0010] 第三设备可以在由第一设备限定的故障单元内,而第二设备可以在由第一设备限定的故障单元外。

[0011] 在指示第二设备不能使用时,第一数据可以指示第二设备是本地可用的,但由于外部问题而可能临时不能执行其指定的服务。

[0012] 该方法可以进一步包括在第三设备处拦截从第一设备至第二设备的第二数据。

[0013] 第三设备可以在第一设备和第二设备之间,以使得第一设备和第二设备之间的通信是经由第三设备的。例如,第三设备可以采取电缆适配器的形式。可替代地,从第一设备发送至第二设备的数据可以通过第一设备和第二设备之间的另一设备重新导向至第三设备。

[0014] 检测第一设备和第二设备之间的通信故障可以包括检测对第二数据的响应未被第一设备接收到。检测第一设备和第二设备之间的通信故障可以包括检测连接第一设备和第二设备的电缆已经断开。例如,检测电缆断开可以包括检测之前出现的“电缆检测”信号没有出现。可替代地,在第一设备和第二设备之间的连接为无线连接的情况下,检测通信故障可以包括检测无线连接中的中断或干扰。

[0015] 意欲指示第一数据源的第一数据的一部分可以指示第二设备。第一数据可以包括指示第一数据是响应于第二数据的数据。例如,第二数据可以包括指示第二数据属于的数据交易的数据字段,并且第一数据可以包括指示其属于相同数据交易的数据。

[0016] 第一设备和第二设备中的一个可以是服务器,和/或第一设备和第二设备中的一个可以是远程连接的第一设备或第二设备的资源。例如,第一设备和第二设备中的一个可以为I/O设备或其它远程资源。

[0017] 第一设备可以经由PCI Express连接而被连接至第二设备。

[0018] 第一数据可以指示第二设备是不可达的,第一数据可以指示第二设备已经发生故障,和/或第一数据可以指示第一数据是受损的。例如,第一数据可以包括具有“1”值的状态位,该状态位指示第二设备已经发生故障和/或第一数据是受损的。

[0019] 检测第一设备和第二设备之间的通信故障可以包括从第三设备向第二设备传送第三数据,并且如果对第三数据的响应在预定时间段中未被从第二设备接收到,则确定第一设备和第二设备之间通信故障。

[0020] 根据本发明的第二方面,提供一种用于管理第一设备和第二设备之间的通信故障的装置,该装置包括:用于在第一设备和第二设备中间的第三设备处检测第一设备和第二设备之间的通信故障的装置;用于将第一数据从第三设备传送至第一设备的装置,所述第一数据指示第二设备不能被第一设备使用。

[0021] 根据本发明的第三方面,提供一种用于第一设备和第二设备之间的连接的有恢复力的电路适配器,包括:被配置成检测连接的第一设备和第二设备之间的通信故障的检测器;被配置成将第一数据传送至连接的第一设备的传送器,第一数据指示连接的第二设备

不能被第一设备使用。

[0022] 经理解的是,关于本发明的一个方面描述的许多特征适用于关于本发明的其它方面。

[0023] 将理解的是,本发明的方面可以以任何便利的方式来实施,这些方式包括通过合适的硬件和/或软件方式。例如,被配置成实施本发明的切换设备可以使用合适的硬件部件来创建。可替代地,可编程设备可以进行编程以实施本发明的实施方式。因此本发明还提供用于实施本发明方面的合适的计算机程序。这类计算机程序可以在合适的载体媒介上被执行,其中载体媒介包括有形载体媒介(例如,硬盘、CD ROM等)和通信信号之类的无形载体媒介。

附图说明

[0024] 现在将参考附图通过示例的方式来描述本发明的实施方式,其中:

[0025] 图1和1a是I/O设备被提供为服务器部件的先前技术I/O配置的示意图;

[0026] 图2是I/O设备被远程连接至服务器的先前技术I/O配置的示意图;

[0027] 图3是多个I/O设备被远程连接至服务器的先前技术I/O配置的示意图;

[0028] 图4是根据本发明的实施方式的I/O设备被连接至服务器的I/O配置的示意图;

[0029] 图5是数据包报头的示意图;以及

[0030] 图6是根据本发明的实施方式的服务器被连接至I/O设备的I/O配置的示意图。

具体实施方式

[0031] 参考图1,服务器1包括CPU/PCI Express根组件(CPU/RC)2和网络接口控制器(NIC)3。CPU/RC2经由PCI Express芯片到芯片连接4而被连接至NIC3。NIC3经由以太网连接6(其可以是例如电缆连接或无线连接)连接至网络5。通过使用NIC3和合适的软件,服务器1被配置成向服务器1的用户提供对网络5的访问。将理解的是,为清楚起见,服务器1的与本发明不相关的其它细节已经被省略,那些被省略的细节对于本领域技术人员是显而易见的。

[0032] 一般而言,图1的服务器1被认为是单个故障单元。服务器1的内部部件(如NIC3或连接4)的故障被认为是整个服务器1的故障,因为服务器1无法再提供它的预期功能(即,促进对网络5的访问)。因此,当NIC3故障或NIC3和CPU/RC2之间的通信故障时,无法要求服务器1继续工作而服务器1将故障。

[0033] 与内部部件的故障相反,如以太网连接6或下游开关(未示出)的外部部件的故障不被认为是服务器1的故障,并且因此,服务器1需要继续操作。一般而言,考虑到替代设备可以容易地被连接(或在例如未意料的断开事件中被重新连接),服务器1的外部连接部件的故障不应该导致服务器1的故障。因此,服务器1适用于平稳处理当其被适当地通知外部资源当前不可用时的状况。例如,只要CPU/RC2和NIC3之间的通信被维持,以太网连接6的故障使得NIC3通知CPU/RC2以太网连接6不可用(通过合适的错误信息的传输)。当接收该信息后,CPU/RC2可以停止向NIC3发送通过以太网连接6传输的数据包,而不会造成整个服务器1的故障。

[0034] 当接收到来自NIC的错误信息,CPU/RC能够通过使用可用的备份资源来继续提供

其指定服务。参考图1a,服务器1a包括CPU/RC2a,该CPU/RC2a经由PCI Express芯片到芯片连接4a被连接至NIC3a并且经由PCI Express芯片到芯片连接4b被连接至NIC3b。NIC3a、3b中的每一个经由相应的以太网连接6a、6b连接至网络5。如此,NIC3a、3b中的每一个可以促进对网络5的访问。例如,当以太网连接6b故障时,相应的NIC3b向CPU/RC2a传送错误信息以指示以太网连接6b不可用。由于CPU/RC2a适用于平稳处理这种错误信息,CPU/RC2a可以使用NIC3a和以太网连接6a来继续提供其指定功能。

[0035] 在图2的布置中,NIC7被封装在服务器8的外部(即,由服务器8限定的单个故障单元外面)。参考图2,服务器8包括经由PCI Express连接11连接至CPU/RC10的PCI Express电缆适配器9。PCI Express电缆适配器9经由电缆14将CPU/RC10连接至封装在远程I/O设备13内的PCI Express电缆适配器12。远程I/O设备13还封装NIC7,NIC7通过PCI Express连接15被连接至PCI Express电缆适配器12。PCI Express电缆适配器9、12仅提供电缆驱动和通过电缆14输送PCI Express信号所需的信号调节功能。也就是说,电缆适配器9、12不提供任何逻辑功能,如此,对于服务器8(上运行的软件)来说,图2的系统在逻辑上与图1的系统相同。

[0036] 电缆14的故障或NIC7本身的故障断开CPU/RC10和NIC7之间的通信。如此,CPU/RC10将接收不到与NIC7的未完成的数据交易的响应。如上所述,缺少对未完成的数据交易的响应将导致服务器整体故障。虽然在图1的布置中,通常可以接受的是,NIC3(或NIC3和CPU/RC2之间的连接4)的故障构成服务器1整体的故障,但在图2的布置中,电缆14和NIC7的远程性质意味着这些部件可以简单地被替代或重新连接。如此,在图2的布置中,不期望的是NIC7或电缆14的故障会引起服务器8的故障。

[0037] 进一步地,如图3的布置所示,远离服务器的I/O资源的放置更容易允许多个I/O资源被提供,以使得辅助资源可以在其他连接的I/O资源故障的情况下作为备份执行。参考图3,服务器20的CPU/RC19经由电缆适配器21、经由电缆23被连接至远程I/O设备22以及经由电缆25被连接至远程I/O设备24。远程I/O设备22包括电缆适配器26和NIC27,电缆适配器26用于至电缆23的连接,NIC27经由连接28被连接至电缆适配器26。远程I/O设备24包括电缆适配器29和NIC30,电缆适配器29用于至电缆25的连接,NIC30经由连接31被连接至电缆适配器29。每个NIC27、30经由相应的连接32、33连接至网络5。将理解的是,在I/O设备22、24中的一个故障的情况下,即不能使用NIC27、30中的一个进行通信,或者在连接23、25中的一个故障时,服务器20可以使用I/O设备中的另外一个I/O设备来继续提供它的指定服务(即,至网络5的连接)。然而,当I/O设备22、24(或者连接23、25)中的一个故障时,CPU/RC19将无法接收对传送至NIC27、30中断开的NIC的数据包的响应,该断开的NIC导致服务器20的不期望的故障。

[0038] 图4示出根据本发明的实施方式修改的图2的布置。尤其,在图4的布置中,由服务器限定的故障单元内的电缆适配器使用恢复性电缆适配器来代替,恢复性电缆适配器的操作在下文中更加详细的描述。但是,一般而言,恢复性电缆适配器将严重错误(如PCI电缆的故障)转换成非严重错误。

[0039] 在图4中,服务器35包括经由PCI Express连接38连接至恢复性电缆适配器37的CPU/RC36。恢复性电缆适配器37经由电缆40将CPU/RC36连接至封装在远程I/O设备39内的PCI Express电缆适配器38。远程I/O设备39封装NIC41,NIC41通过PCI Express连接42被连接至PCI Express电缆适配器38,并且经由以太网连接43连接至网络5。

[0040] 更加详细的,恢复性电缆适配器37提供监控在恢复性电缆适配器37的下游的部件(例如电缆40、NIC41等)故障所需的逻辑和硬件。例如,恢复性电缆适配器37被配置成检测电缆何时被拔出(例如,通过提供“存在检测”信号的方式,或任何其它合适方法)并被配置成监听来自远程I/O设备39指示NIC41有问题的信息。恢复性电缆适配器37进一步适用于检查数据包以便确定和记录关于由服务器35发起的数据交易的信息,并适用于等待合适的响应。使用板上计时器,恢复性电缆适配器37可以等待预定时间段,该预定时间段到期代表故障的发生。例如,时间段的阈值可以被设置以使得其比服务器35由于未接收到响应数据包而将崩溃的时间早到期。恢复性电缆适配器37还可以被配置成不依赖于CPU/RC36生成用于传输至NIC41的数据包并被配置成等待合适的响应(由此,预定时间段内没有响应代表NIC41的故障)。将理解的是,恢复性电缆适配器37的下游部件的故障可以具有一个或多个不同的原因,并且应该理解的是,恢复性电缆适配器37可以实施任何合适方法以用于检测这类事件,这对于本领域技术人员是显而易见的。

[0041] 当检测到下游部件故障时,恢复性电缆适配器37模仿NIC41来向CPU/RC36提供合适的非严重错误信息。非严重错误信息表示NIC41为可用的但是处于不能被使用的状态(例如,模仿以太网43已经被拔出的状态)。

[0042] 为了模仿NIC41,恢复性电缆适配器37可以生成数据包,该数据包看来是由NIC41生成的。更详细地,NIC41可以具有多个独立的设备功能,最多八个由PCI Express协议支持的功能。也就是说,NIC41可以多达八个单独设备来呈现给CPU/RC36。NIC41的每个设备功能具有独特地识别该功能的相应的标识符。从NIC41的特定设备功能发送的数据包具有数据交易标识符,该数据交易标识符包括对应于发送数据包的设备功能的标识符的请求者标识符。

[0043] 参考图5描述PCI Express协议使用的数据包报头的格式。请求者标识符45识别发起数据包的设备功能并包括从0至15索引的十六个比特。可以看出,请求者标识符45包括占据前八个比特的总线号字段46、占据中间五个比特的设备号字段47和占据后三个比特的功能号字段48。当使用PCI Express协议时,总线号46、设备号47和功能号48的组合独特地识别由特定设备提供的功能。

[0044] 图5中示出的数据包报头进一步包括标记字段49,该标记字段49包含八个比特。如上所述,数据交易可以由请求数据包和一个或多个相应的完成数据包组成。每个请求数据包与存储在标记字段49中的值相关联。每个相应的完成数据包具有存储在标记字段49中的相同的值,因此将完成数据包与相关请求数据包相关联。独特的标记值被分配至未完成请求,该未完成请求要求从目标设备的一个或多个完成数据包。由于标记字段49具有八个比特,因此两百五十六(2^8)个可能标记值可以被表示。

[0045] 功能号字段48被提供有发送请求的设备的相关功能的功能号。如果设备具有的功能数量少于8,则在功能号字段48中可能存在未使用的比特。因此已知仅使用功能号字段48的足够比特来表示设备的功能并使用功能号字段48的任何未使用字段作为虚拟功能号,该虚拟功能号逻辑上与标记字段49合并。在仅提供单个功能的情况下,功能号字段48的所有比特可以逻辑上与标记字段49合并,以对多达两千零四十八(2^{11})个未完成请求提供支持。

[0046] 恢复性电缆适配器37可以检查所接收的数据包,并且尤其是标记字段和请求者标识符,以便确定该数据包的数据交易标识符。该数据可以被存储在由恢复性电缆适配器提

供的板上存储器中。由恢复性电缆适配器37创建的数据包然后可以被提供有正确的数据交易标识符(如果被提供的错误类型的信息请求数据交易标识符),以使得CPU/RC36认为信息关于CPU/RC36和NIC41之间的未完成的数据交易。类似地,数据包可以通过具有NIC41的请求者标识符的恢复性电缆适配器37来生成,以使得那些数据包看来是由NIC41生成的。

[0047] 合适的错误信息依赖于由恢复性电缆适配器37记录的关于数据交易的信息来选择。例如,在CPU/RC36向NIC41传送寄存器读取请求数据包的情况下,当下游部件故障并断开CPU/RC36和NIC41之间的通信时,恢复性电缆适配器37可以提供数据包,在该数据包中错误状态位被设置为“1”值。在服务器1上运行的软件将阐释数据包的接收,在该数据包中状态位被设置为指示错误的发生,但是该错误将不引起服务器35的故障。作为另一个示例,如果数据交易涉及通过网络5传送的数据,则恢复性电缆适配器37可以向CPU/RC36传送指示以太网连接43已经断开的信息。每个错误信息可以被硬编码至恢复性电缆适配器37的板上存储器中。恢复性电缆适配器37上运行的逻辑可以然后被用于根据正等待完成数据包的数据交易来确定存储的信息中哪一信息是合适的。

[0048] 提供至CPU/RC36的错误信息的选择还可以取决于恢复性电缆适配器37正模仿的远程连接的设备的细节。例如,在远程连接的设备为存储器资源的情况下,指示以太网连接已经断开的错误信息将是不合适的。如此,错误信息可以针对远程连接的设备来定制,用于指示该设备是本地可用的但是不能用于执行其指定服务(即,使用非严重错误来替代严重错误)。

[0049] 以上描述的本发明的实施方式是基于两个假设的:第一,任何远程连接至服务器35的设备具有服务器35认为该设备不能使用但是不会引起服务器35崩溃的状态;第二,不能使用的状态与I/O设备的完全运行状态相比是简单的。

[0050] 此外或可替代地,在连接的服务器(或至服务器的连接)故障时避免远程连接的I/O设备故障可能是可取的。这在远程I/O设备连接至多个服务器并被多个服务器共享(例如,在系统中实施多根I/O虚拟化)的系统中是特别但不排他地合适的。将理解的是,一个服务器故障不应该引起远程I/O设备故障并因此不能用于向剩余的服务器(一个或多个)提供它的服务。

[0051] 图6示出根据本发明的实施方式修改的图2的布置。尤其,在图6的布置中,远程I/O设备内的电缆适配器使用恢复性电缆适配器来代替。

[0052] 在图6中,服务器50包括经由PCI Express连接53连接至电缆适配器52的CPU/RC51。电缆适配器52经由电缆56将CPU/RC51连接至封装在远程I/O设备55内的恢复性电缆适配器54。远程I/O设备55还封装NIC57,该NIC57通过PCI Express连接58被连接至恢复性电缆适配器54,并且经由以太网连接59连接至网络5。

[0053] 恢复性电缆适配器54适用于检测在恢复性电缆适配器54的上游的部件故障。当检测到上游部件故障时,恢复性电缆适配器54被用于模仿服务器50的存在。例如,恢复性电缆适配器54可以被配置成响应由NIC57发出的未完成存储读/写请求及时地发出具有非严重错误状态的完成数据包。

[0054] 如上参考恢复性电缆适配器37的描述,提供至NIC57的错误信息将依赖于要求完成数据包的数据交易。

[0055] 将理解的是,服务器和远程资源的配置可以包括恢复性电缆适配器37和恢复性电

缆适配器54二者。也就是说,在一些配置中,恢复性电缆适配器可以被提供在由一个或多个服务器限定的一个或多个相应的故障单元中和由一个或多个相应的远程连接资源限定的一个或多个故障单元中。

[0056] 已经参考包括PCI Express以太网接口卡(NIC)的系统描述了先前所述的示例实施方式,该系统被远程连接至服务器。然而,将理解的,本发明具有更广泛的应用性。确实,本发明可以用于使用PCI Express之外的连接器并具有NIC之外的远程设备(例如,光纤信道主机总线适配器、存储控制器等)的系统中。

[0057] 进一步地,尽管上文描述有关于包括服务器并具有远程连接至该服务器的资源的系统,但将理解的是,本发明可以被用于第一服务器和第二服务器之间的通信。

[0058] 进一步地,本发明适用于连接设备的其它配置,例如多服务器I/O虚拟系统。确实,将理解的是,在许多实施方式中,使用本发明的系统可以包括多个服务器,其中每个服务器具有至多个独立远程设备的连接。这种布置将允许每个服务器在关联的一个远程设备故障的情况下通过使用另一个远程连接的设备来继续有效运行。在多个服务器或多个远程资源被提供的情况下,恢复性电缆适配器37、54保持哪个服务器或哪个远程资源已经故障的记录以便选择合适的非严重错误信息(在必要的情况下考虑故障服务器或远程资源的活动的数据交易)。

[0059] 将进一步理解的是,尽管恢复性电缆适配器在上文被描述为包括本发明功能和标准电缆适配器功能的单一设备,但是本发明可以在专用设备中实施。例如,参考图2,实施本发明的设备可以放置在CPU/RC10和电缆适配器9之间、或放置在电缆适配器9和电缆14之间。

[0060] 恢复性电缆适配器37、54可以以现场可编程门阵列(FPGA)或特定用途集成电路(ASIC)来实施。然而,将理解的是,恢复性电缆适配器37、54可以使用任何合适方式来实施。

[0061] 先前的描述已经描述了本发明的实施方式,其中数据包在服务器和I/O设备之间传送。将理解的是,术语服务器意欲广泛地覆盖任何计算设备。

[0062] 进一步地,对于本领域技术人员来说,根据于本发明的教导,在不背离所附权利要求范围的情况下,本发明的修改和应用将是显而易见的。

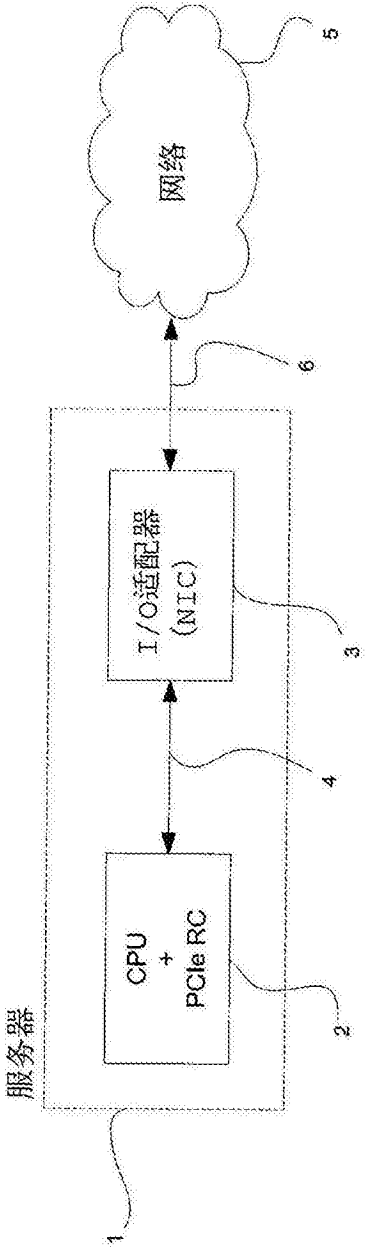


图1

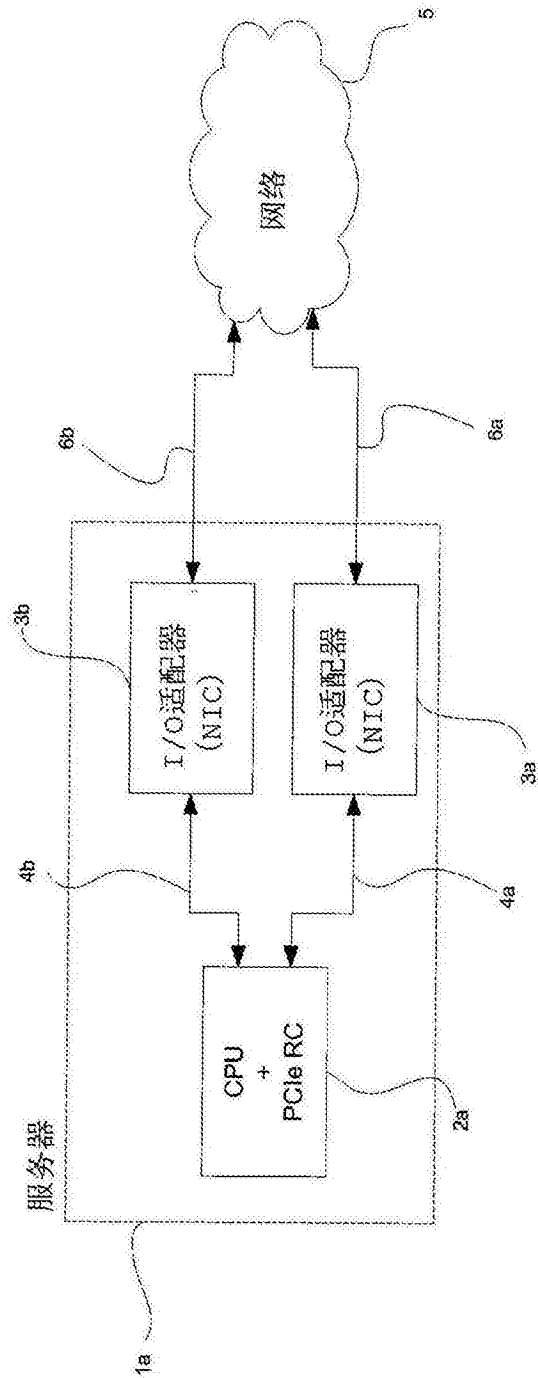


图1a

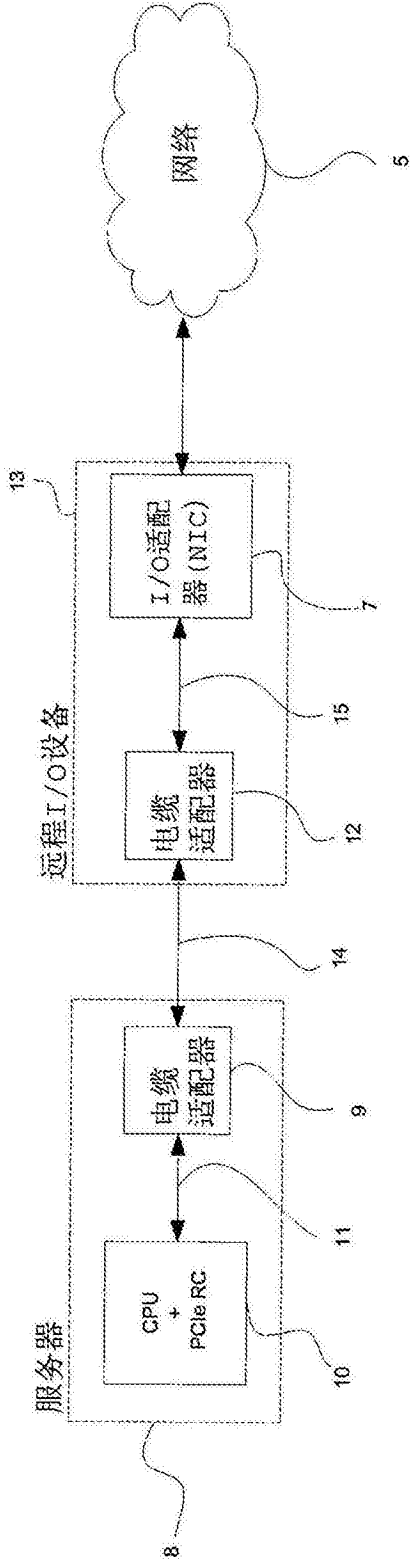


图2

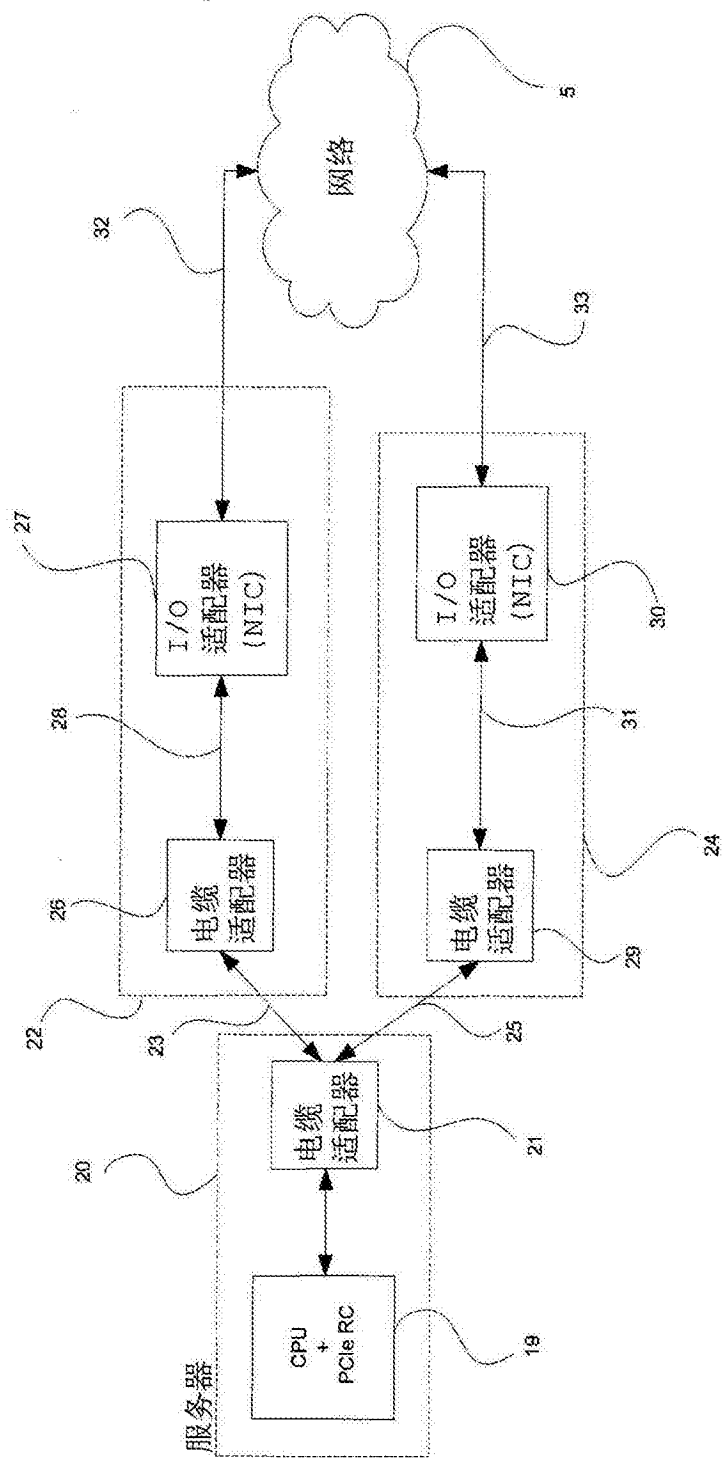


图3

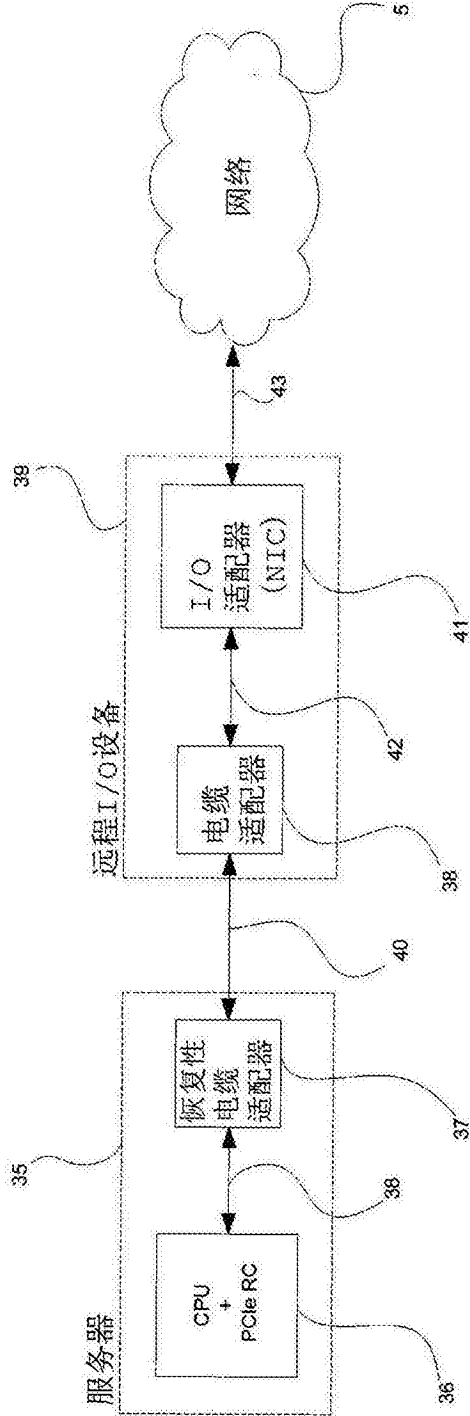


图4

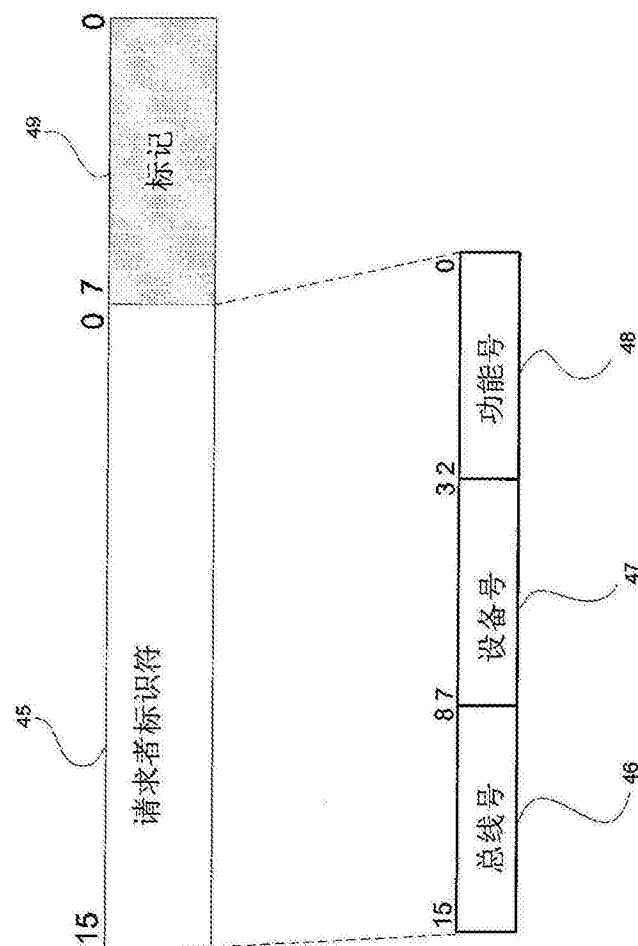


图5

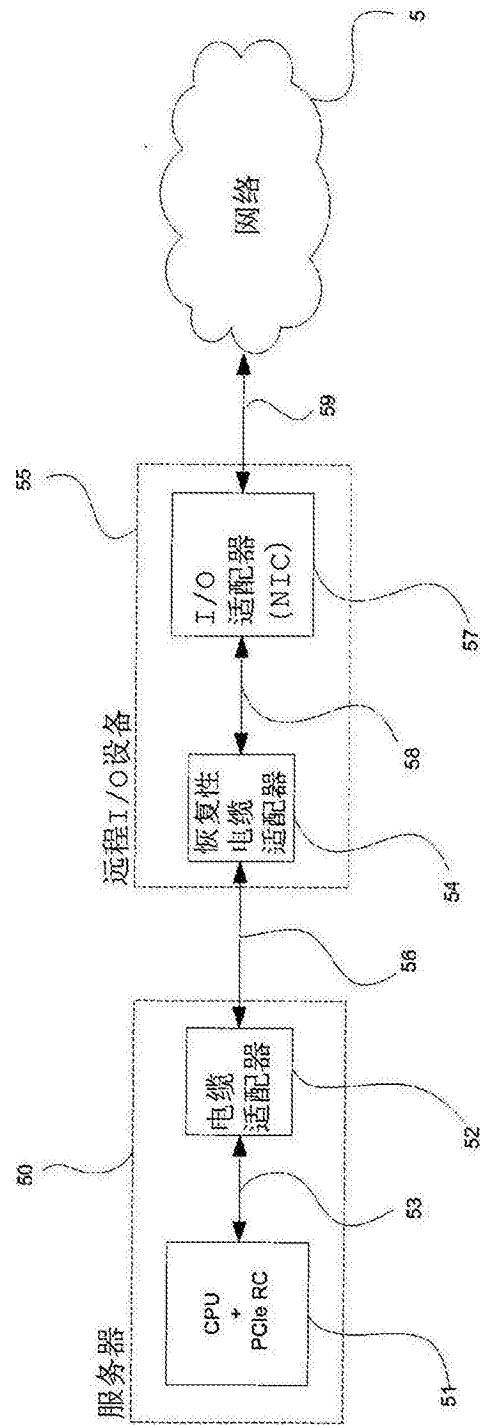


图6