(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 1722732 B
(45)授权公告日 2011.11.09

(21)申请号 200510084617.7
(22)申请日 2005.07.15
(30)优先权数据
10/893,213 2004.07.16 US
(73)专利权人 国际商业机器公司
地址 美国纽约
(72)发明人 约翰·刘易斯·哈福雷德
(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038
代理人 屠长存
(51)Int.Cl.
H04L 29/08(2006.01)
H04L 12/28(2006.01)
(56)对比文件
CN 1265244 A,2000.08.30, 全文.
CN 1231814 A,1999.10.13, 全文.
CN 1151231 A,1997.06.04, 全文.
US 6185607 B1,2001.02.06, 全文.

审查员 李彬

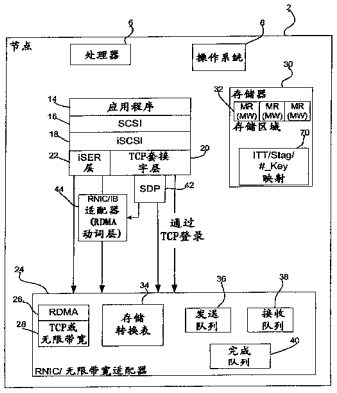
权利要求书 3 页 说明书 9 页 附图 13 页

(54)发明名称

节点之间进行通信的方法及系统

(57)摘要

本发明提供一种在本地节点处执行以与远程节点通信的方法、系统及程序。第一通信协议用来与远程节点通信，以建立第二通信协议的连接。数据结构被创建以允许与远程节点的通信，以建立第二通信协议的与远程节点的连接。扩展层对于第二通信协议被调用。数据结构传送到扩展层，以用于使用第二通信协议与远程节点通信。



CN 1722732 B

1. 一种在本地节点中实施以与远程节点通信的系统，包括：

使用第一通信协议与远程节点通信以建立第二通信协议的连接的装置；

创建数据结构以允许与远程节点的通信，以建立第二通信协议的与远程节点的连接的装置；

调用第二通信协议的扩展层的装置；以及

将数据结构传送到扩展层以用于使用第二通信协议与远程节点通信的装置。

2. 根据权利要求 1 的系统，其中，用以建立连接的与远程节点的通信通过套接字层来进行。

3. 根据权利要求 1 的系统，其中第一通信协议包括用来与远程节点通信的在 RNIC 适配器中实施的连网层，其中第二通信协议包括 RDMA 协议，并且其中扩展层包括使用 RDMA 协议处理与远程节点的通信的 iSER 层。

4. 根据权利要求 1 的系统，其中用以建立连接的与远程节点的通信使用利用第二通信协议通信的接口协议来进行，其中接口协议创建第二通信协议的数据接口，其中所述系统还包括：

通过扩展层协议发出调用以终止接口协议的装置，其中扩展层使用由接口协议创建的数据结构来与远程节点通信。

5. 根据权利要求 4 的系统，其中第一通信协议包括套接字层，其中第二通信协议包括 RDMA 协议，其中接口协议包括 SDP 层或分层于 IPoIB-RC 上的 TCP，其中扩展层包括使用 RDMA 协议来处理与远程节点的通信的 iSER 层，并且其中无限带宽适配器用来与远程节点通信。

6. 根据权利要求 1 的系统，其中本地节点包括起始节点并且远程节点包括目标节点，其中当接收到来自目标节点的用以建立连接的最后应答时扩展层被调用，其中用以终止接口协议的调用响应于调用扩展层而发出。

7. 根据权利要求 6 的系统，其中除了响应于调用扩展层之外用以终止接口协议的调用响应于使用接口协议将最后应答消息发送到目标节点而发出。

8. 根据权利要求 1 的系统，其中本地节点包括起始节点并且远程节点包括目标节点，并且其中起始节点还执行：

通过起始节点适配器接收来自目标节点的能够包括使对存储位置的直接索引无效的消息，其中直接索引与第二通信协议兼容，并且其中起始节点适配器不允许索引的直接无效；

响应于确定直接索引不匹配映射中的一个索引而调用所述起始节点适配器以使直接索引无效；以及

响应于确定直接索引不匹配映射中的一个索引或没有提供直接索引使与包含于使无效信息中的间接索引关联的、在映射中指示的至少一个直接索引无效。

9. 根据权利要求 8 的系统，其中扩展层执行调用所述起始节点适配器以使直接索引无效以及使无效消息中的间接索引关联的至少一个索引无效。

10. 根据权利要求 9 的系统，其中扩展层包括 iSER 层，其中目标节点包括 RNIC 或无限带宽适配器，并且其中被调用的所述起始节点适配器包括 RNIC 或无限带宽适配器。

11. 根据权利要求 10 的系统，其中扩展层响应于接收到来自起始节点的最后登录请求

而被调用。

12. 根据权利要求 11 的系统，其中目标节点还被配置为：

使用第一通信协议发送最后登录应答给起始节点，其中数据结构在发送最后登录应答之后传送到扩展层。

13. 根据权利要求 8 的系统，其中间接索引包括起始任务标签 ITT，并且其中直接索引包括转向标签或远程的转向标签 R_Key。

14. 一种在本地节点中执行以与远程节点通信的方法，包括：

使用第一通信协议与远程节点通信以建立第二通信协议的连接；

创建数据结构以允许与远程节点的通信，以建立第二通信协议的与远程节点的连接；

调用第二通信协议的扩展层；以及

传送数据结构到扩展层以用于使用第二通信协议与远程节点通信。

15. 根据权利要求 14 的方法，其中用以建立连接的与远程节点的通信通过套接字层来进行。

16. 根据权利要求 14 的方法，其中第一通信协议包括用来与远程节点通信的在 RNIC 适配器中实施的连网层，其中第二通信协议包括 RDMA 协议，并且其中扩展层包括使用 RDMA 协议处理与远程节点的通信的 iSER 层。

17. 根据权利要求 14 的方法，其中用以建立连接的与远程节点的通信使用利用第二通信协议通信的接口协议来进行，其中接口协议创建第二通信协议的数据结构，还包括：

通过扩展层协议发出调用以终止接口协议，其中扩展层使用由接口协议创建的数据结构与远程节点通信。

18. 根据权利要求 17 的方法，其中第一通信协议包括套接字层，其中第二通信协议包括 RDMA 协议，其中接口协议包括 SDP 层或分层于 IPoIB-RC 上的 TCP，其中扩展层包括使用 RDMA 协议处理与远程节点的通信的 iSER 层，并且其中无限带宽适配器用来与远程节点通信。

19. 根据权利要求 14 的方法，其中本地节点包括起始节点并且远程节点包括目标节点，其中当接收到来自目标节点的用以建立连接的最后应答时扩展层被调用，其中用以终止接口协议的调用响应于调用扩展层而发出。

20. 根据权利要求 14 的方法，其中除了响应于调用扩展层之外用以终止接口协议的调用响应于使用接口协议将最后应答消息发送到目标节点而发出。

21. 根据权利要求 14 的方法，其中本地节点包括起始节点并且远程节点包括目标节点，并且其中起始节点还执行：

通过起始节点适配器接收来自目标节点的能够包括使对存储位置的直接索引无效的消息，其中直接索引与第二通信协议兼容，并且其中所述起始节点适配器不允许索引的直接无效；

响应于确定直接索引不匹配映射中的一个索引而调用所述起始节点适配器以使直接索引无效；以及

响应于确定直接索引不匹配映射中的一个索引或没有提供直接索引使与包含于使无效信息中的间接索引关联的、在映射中指示的至少一个直接索引无效。

22. 根据权利要求 21 的方法，其中扩展层执行调用所述起始节点适配器以使直接索引

无效以及使无效消息中的间接索引关联的至少一个索引无效。

23. 根据权利要求 22 的方法，其中扩展层包括 iSER 层，其中目标节点包括 RNIC 或无限带宽适配器，并且其中被调用的所述起始节点适配器包括 RNIC 或无限带宽适配器。

24. 根据权利要求 23 的方法，其中扩展层响应于接收到来自起始节点的最后登录请求而被调用。

25. 根据权利要求 24 的方法，其中目标节点还执行 ：

使用第一通信协议发送最后登录应答给起始节点，其中数据结构在发送最后登录请求之后被传送到扩展层。

26. 根据权利要求 21 的方法，其中间接索引包括起始任务标签 ITT，并且其中直接索引包括转向标签 STag 或远程的转向标签 R_Key。

节点之间进行通信的方法及系统

技术领域

[0001] 本发明涉及一种允许节点之间通信的方法、系统及程序。

背景技术

[0002] 在存储环境中，数据存取命令从主机系统传送到管理对磁盘存取的存储控制器。存储控制器可以是主机系统内的卡或分开的装置。互联网小型计算机系统接口（iSCSI）协议用于利用包括以太网交换机和路由器的以太网连接的存储网络。这里所使用的术语“iSCSI”指的是由 IETF（互联网工程任务小组）标准机构定义的 iSCSI 协议的语法和语义，以及该协议的任何变体。在利用 iSCSI 的当前存储网络中，分组配置包括封装互联网协议（IP）和传输控制协议（TCP）包层的以太网包，包层进一步封装包括一个或多个 SCSI 命令的 iSCSI 包。当分组在任意网络段（链路）上从点到点流动时，以太网协议规定链路级差错检验以确定数据是否已在通过链路时被损坏。在网络数据传输操作中，起始设备在网络上传送数据或命令到目标设备。TCP/IP 包包括执行端到端检验的错误检测码，以确定在相对端所传送的分组是否已在传输过程中当分组通过交换机和路由器时被改变。检测错误的接收设备将发送否定应答给发送设备，以请求检测出错误的那些分组的重发。

[0003] 远程直接存储器存取（RMDA）协议允许一个网络节点以对存储器总线带宽及处理器开销的最小要求将信息直接放置在另一个网络节点的存储器中。TCP/IP 上的 RDMA（也称作 iWARP）定义可互操作协议以在标准 TCP/IP 网络上支持 RDMA 操作。RDMA 网络接口卡（RNIC）实施 RDMA 协议并执行 RDMA 操作以传送数据到本地及远程存储器。RDMA 协议的更多细节在以下规范中描述：由 RDMA 联盟（2003 年 4 月）公布的名称为“RDMA 协议动词规范（1.0 版）”；由 RDMA 联盟（2002 年 10 月）公布的“可靠传输上的直接数据放置（1.0 版）”；以及由 RDMA 联盟（2002 年 10 月）公布的“TCP 的标记 PDU 对齐组帧规范（1.0 版）”，并且这些规范在此引入作为参考。

[0004] 由 RDMA 联盟（2003 年 7 月）发布的 Michael Ko 等人的一个名称为“RDMA 的 iSCSI 扩展规范（1.0 版）”的规范定义通过在 RDMA 之上分层 iSCSI 而为 iSCSI 提供 RDMA 数据传送能力的协议，该规范在此整体引入作为参考。

[0005] 定义为 TCP/IP 上的 RDMA 也称作 iWARP 的一部分的许多特性以前定义为无限带宽网络中的操作。无限带宽适配器硬件支持 RDMA 操作。无限带宽也定义称作套接字字指引协议（SDP）的一组协议，其允许正常的 TCP/IP 套接字字应用程序以与它们在 TCP/IP 网络上操作时相同的方式跨越无限带宽网络发送消息。无限带宽及 SDP 协议的更多细节在出版物“无限带宽™体系结构，规范卷 1”，1.1 版（2002 年 11 月，版权无限带宽™同业公会）中描述，该出版物在此整体并入作为参考。

附图说明

[0006] 现在参考附图，其中相似参考数字自始至终表示一致的部分：

[0007] 图 1 说明实现实施方案的网络节点的例子；

[0008] 图 2 说明根据所描述的实施方案的计算体系结构的例子；

[0009] 图 3 说明分组格式；

[0010] 图 4 说明保存在映射上的信息；

[0011] 图 5a、5b、6a、6b、7 和 8 说明根据实施方案的执行以传送数据的操作；

[0012] 图 9，10、11a 和 11b 说明包括网关的实施方案；以及

[0013] 图 12、13 和 14 说明在网关中执行以处理并转发消息的操作。

发明内容

[0014] 本发明提供的是一种在本地节点处执行以与远程节点通信的方法、系统及程序。第一通信协议用来与远程节点通信以建立第二通信协议的连接。数据结构被创建来允许与远程节点的通信，以建立第二通信协议的与远程节点的连接。扩展层对于第二通信协议被调用。数据结构传送到扩展层以用于使用第二通信协议与远程节点通信。

具体实施方式

[0015] 在下面的描述中，参考构成其一部分并说明本发明的几种实施方案的附图。应当明白，其他实施方案也可以使用并且可以不背离本发明的范畴而作结构及操作上的改变。

[0016] 图 1 说明包括在网络 4 上通信的多个计算节点 2a，2b，...2n 的网络计算环境。网络可以包括局域网 (LAN)、广域网 (WAN)、存储域网 (SAN)。可选地，节点可以在总线例如 SCSI 总线等上通信。

[0017] 图 1 中的节点 2a，2b...2n 可以操作为起始或目标。图 2 说明包含于节点例如节点 2a，2b，2c 中以允许网络 4 上的通信的部件。节点 2 包括处理器 6 例如中央处理单元或组合体，以及操作系统 8。节点 2 还包含应用程序 14，包括用户应用程序例如数据库程序、服务器程序等。为了执行 I/O 操作，应用程序 14 将访问 SCSI 层 16 以产生 SCSI I/O 请求，SCSI 层 16 又将访问 iSCSI 层 18。iSCSI 层 18 通过执行 iSCSI 登录操作开始与目标节点的通信。为了登录，iSCSI 层与套接字层 20 例如 TCP 套接字层连系，以建立与远程目标节点的通信并登录。套接字层 20 包括用来在 iSCSI 层 18 和适配器 24 中的网络协议之间连系的编程接口。

[0018] 适配器 24 可以包括 RNIC 适配器或无限带宽 (IB) 适配器。适配器 24 包括 RDMA 层 26 和网络层 28 例如 TCP 层或无限带宽层，以为在网络 4 上的传输在传输层中包装分组或者将从网络 4 接收到的分组解包。

[0019] 在适配器 24 包括无限带宽适配器的情况下，节点 2 可以包括套接字指引协议 (SDP) 层 42，使得套接字层 20 与 SDP 层 42 连系并且 SDP 层 42 在套接字层 20 和 RDMA 层 26 之间连系。在无限带宽实施方案中，通过实施来自 iSCSI 层 18 的套接字层 20 访问至对 RDMA 层 26 的 RDMA 访问（直接地或者通过 IB 适配器驱动器 44），SDP 层 42 提供在使用套接字层 20 进行访问的应用程序 14 和适配器 24 中的 RDMA 层 26 之间的接口。在无限带宽和 RNIC 两种实施方案中，iSER 层 22 被提供，使得在登录之后，iSCSI 层 18 将访问 iSER 层 22 以访问 RNIC24。iSER 层 22 可以直接通过函数调用或通过包括 RDMA 动词 (verb) 层的 RNIC 驱动器 44 来访问 RNIC 24。在适配器 24 由 RNIC 驱动器构成的实施方案中，节点 2 可以不包括 SDP 层 42，但是在无限带宽适配器实施方案中，SDP 层 42 被包含。

[0020] RDMA 层 26 可以直接存取起始和目标节点中（本地或本地及远程地）以逻辑上相连方式的已登记存储位置。所定义的存储位置，例如存储区域或存储窗口，通过由 RDMA 层 26 所创建的并用来索引已登记的存储位置例如存储区域 32 的转向标签来标识。在 RNIC 实施中转向标签称为 STag，并且在无限带宽实施方案中转向标签称为远程的转向标签 R_Key 和本地的转向标签 L_Key（这里用于两者的统称术语为 #_Key）。在某些实施方案中，存储区域或称为存储窗口的存储区域的子集可以被登记，其中单独的 STag/#_Key 将与每个被登记的存储位置、区域或窗口关联。RDMA 层 26 使用 STag/#_key 来存取所索引的存储位置。在某些实施方案中，iSER 层 22 将通过访问 RDMA 动词层 44 来访问适配器 24 以登记存储区域。RDMA 动词层 44（RNIC/IB 适配器驱动器）包括将操作系统 8 与适配器 24 接口的设备驱动器。响应于来自 iSER 层 22 或 SDP 层 42 中的函数的访问以声明或登记存储位置例如存储区或窗口，适配器驱动器 44 将访问适配器 24。

[0021] RDMA 层 26 保存存储转换表 34，并且当登记存储区域时，将标识所登记存储区域和产生来索引该存储区域的 STag/#_key 以允许 RDMA 层 26 将 STag/#_key 与存储区域关联的条目添加到存储转换表 34 中。存储转换表 34 可以保存于适配器 24 的缓冲区中或存储器 30 中。Stag/#_Key 将返回给请求登记的 iSER 层 22 函数以用于 I/O 操作。

[0022] 在适配器 24 产生并返回 Stag/#_Key 到 iSER 层 22 之后，iSER 层 22 可以进行 I/O 操作。iSER 层 22 用头信息和从适配器 24 接收的 STag/R_Key 包装从 iSCSI 层 18 接收的分组，并将分组传递到适配器 24 以传送。

[0023] 为了管理 RDMA 数据传送，RDMA 层 26 保存发送队列 36、接收队列 38，以及完成队列 40。发送队列 36 和接收队列 38 包括 RDMA 层 26 用来管理 RDMA 数据传送请求的工作队列。完成队列 40 可以包括具有完成条目的包含一个或多个条目的可共享队列，以为多个工作队列提供单点的完成通知。队列 36，38 和 40 可以具有许多实例，可能对应每个逻辑连接，并且可以由适配器 24 分配于存储器 30 中或适配器 24 中的缓冲器内。

[0024] 图 3 说明与实施方案一起使用的传送包的格式。由 SCSI 层 16 产生的 SCSI 命令 50（例如读或写命令）由 iSCSI 层 18 封装于 iSCSI 协议数据单元 (PDU)52 中，协议数据单元 52 由 iSER 层 22 中的函数进一步封装于 iSER 头 54 中。包含 SCSI 命令 50 的 iSCSI PDU 52 还包括与根本的 SCSI 命令关联的、iSCSI 层 18 分配给每个已发出的 iSCSI 任务的起始任务标签 (ITT)56，以标识根本的 SCSI I/O 操作。ITT 56 唯一地标识任务会话宽。当目标节点应答来自起始的请求时，ITT 56 用来在 iSCSI 层 18 处将应答与原始请求关联。例如，当目标已完成操作并返回操作状态时发送的、在来自目标 iSCSI 层 18 的 iSCSI SCSI 应答 PDU 中的 ITT 56 由起始 iSCSI 层 18 用来将目标的 PDU 与原始 SCSI 写命令相关联。

[0025] iSER 头 54 将包括与 I/O 操作一起使用的 Stag/R_Key 以及指示接收广告 Stag/R_Key 的远程节点是否读或写由 Stag/R_Key 索引的存储区域（窗口）以及与请求相关的工作队列的信息。iSER 头 54 和 iSCSIPDU 52 进一步封装于一个或多个附加网络层 60 例如 TCP 层或无限带宽网络协议层中。在某些实施方案中，适配器 24 中的网络层 28 将 iSER 头 54 和 PDU 52 组装于网络层 60 例如 TCP、IB 等中。

[0026] iSER 层 22 还保存 ITT 到 Stag/#_Key 的映射 70（图 2）于存储器 30 中，映射 70 将代表 iSCSI 任务的 ITT 与用于传送该任务的数据的 STag/#_key 例如读或写 Stag/R_Key 关联，以使得目标节点通过 RDMA 通道读或写数据。图 4 说明映射 70 中的条目 72 的内容，包

括指示映射是对于本地 STag/L_key 还是远程 Stag/R_key 的本地 / 远程指示符 73。本地映射条目将索引本地存储窗口或区域的 Stag/L_Key 与用来存储（或取回）与 I/O 命令本地相关的数据的 ITT 关联，并且远程映射条目将远程节点上索引存储区域或窗口的 STag/R_key 与数据从中读出或传送到本地存储区域或窗口的 ITT 关联。映射条目 72 还包括代表 iSCSI 任务的相关 ITT 74、与该任务关联的 Stag/#_Key 76，以及关于任务的状态信息 78。

[0027] 图 5a，5b，6a，6b，7 和 8 说明根据所描述的实施方案处理 SCSI 写命令的操作。图 5a 和 5b 说明由起始节点 2 中的部件实施以连接包括 RNIC 或无限带宽适配器 24 的为起始节点 2 使用 RDMA 的目标节点的操作。参考图 5a，控制在方框 100 处开始，其中 iSCSI 层 18 连系套接字层 20 以与目标节点通信并“登录”新的 iSCSI 连接并与目标节点协商各种参数。起始和目标节点都可以包括图 2 中所示的体系结构并包括相配的 RNIC 和 / 或无限带宽适配器。然后在起始节点中执行的操作取决于起始节点 2 包括 RNIC 还是无限带宽适配器 24。如所讨论的，起始节点 2 可以基于安装于起始节点 2 中的适配器 24 的类型装载通信所需的层和代码。例如，如果节点包括无限带宽适配器，那么节点将装载 SDP 层 42，反之如果节点仅包括 RNIC 适配器，那么 SDP 层 42 可以不被装载。如果（在方框 102 处）iSCSI 层 18 通过 RNIC 适配器 24 连接，那么套接字层 20 通过 RNIC 适配器 24 中的 TCP 28 连网函数通信（在方框 104 处）以传递 iSCSI 消息到目标节点并通过套接字层 20 将接收到的应答传递回 iSCSI 层 18。否则，如果（在方框 102 处）iSCSI 层 18 通过无限带宽适配器 24 连接，套接字层 20 通过 SDP 层 42 通信（在方框 106 处）以传递 iSCSI 消息到目标节点并通过套接字层 20 将接收到的应答传递回 iSCSI 层 18。SDP 层 42 创建（在方框 108 处）RDMA 连接例如队列对 36，38、完成队列 40 等的数据结构，然后通过无限带宽适配器 24 和无限带宽网络 4 发送消息到目标或者从那里接收消息。

[0028] 如果（在方框 110 处）起始节点 2 不愿意通过 RNIC 或无限带宽适配器 24 中的连网层建立与远程节点的 RDMA 会话，那么起始节点 2 将断开协商连接并试图（在方框 112 处）定位其他 RDMA 可兼容目标节点。否则，如果（在方框 110 处）RDMA 会话是可接受的，那么（在方框 114 处）SDP 层 42（对于无限带宽适配器 24）或网络层 28（对于 RNIC 适配器）（在方框 114 处）返回来自目标节点的应答，并继续允许 iSCSI 层使用套接字 API 以触发 RNIC 连网层 28 或 SDP 层 42，以和目标一起发送和接收另外的登录请求和登录应答消息。

[0029] 参考图 5b，在建立登录参数之后，起始 iSCSI 层 18 通过对于 RNIC 适配器 24 由连网层（例如 TCP）或者对于无限带宽适配器由 SDP 层 42 处理的套接字调用发送（在方框 116 处）最后登录请求 PDU 给目标节点。当接收到（在方框 118 处）来自目标的最后登录应答 PDU 时，iSCSI 层 18 使用 iSER 原语调用 iSER 层 22，并传送其套接字信息句柄（指向 RNIC 适配器 24 的连网层 28 的控制结构或无限带宽适配器的 SDP RDMA 控制结构）。然后 iSER 层 22 接管（在方框 120 处）套接字控制结构并为 iSER/RDMA 操作配置及修改它们。在无限带宽适配器 24 的情况下，iSER 层 22 发出（在方框 122 处）特定调用以终止 SDP 会话并允许 iSER 层 22 之后拥有及使用原来由 SDP 使用的 RDMA 通信通道（队列对、完成队列等）。然后对于与目标节点的所有将来通信，iSCSI 层用 iSER 层通过 RDMA 通信（在方框 124 处）。

[0030] 图 6a 和 6b 说明由可具有节点 2 的体系结构（图 2）的目标节点中的部件实施以连接使用 RDMA 的起始节点的操作，其中目标节点 2 包括 RNIC 或无限带宽适配器 24。关于

图 2 描述的体系结构用来描述起始和目标节点两者中的部件。参考图 6a，目标节点 iSCSI 层 20 通过套接字调用允许（在方框 150 处）监听端口以等待来自起始节点的联系。套接字层 20 或者通过使用 RNIC 适配器 24 中的连网层 28（例如 TCP）或者通过使用在无限带宽网络上通信的无限带宽适配器的 SDP42 来允许该联系。当从起始节点 iSCSI 层 18 通过相同的套接字接口联系（在方框 152 处）时，目标节点与起始节点协商以获得一致，以进行 RDMA 连接。如果（在方框 154 处）RDMA 连接没被批准，那么目标节点断开协商连接并继续等待（在方框 156 处）直到建立另一个连接。否则，如果 RDMA 连接被批准，那么目标节点处的 RNIC 连网层 28 例如 TCP 或 SDP 层 42 通过套接字 API 将来自起始节点的应答返回（在方框 160 处）到 iSCSI 层 18，并继续允许 iSCSI 层使用套接字 API 来触发 RNIC 连网层或 SDP 层，以和起始节点一起接收和发送另外的登录请求和登录应答消息。当接收到来自起始的最后登录请求时，目标 iSCSI 层 18（在方框 162 处）使用 iSER 原语调用 iSER 层 22 并传送其套接字信息句柄（指向 RNIC 适配器的连网层 28 的控制结构或者 SDP 控制结构）和最后登录应答消息。

[0031] 参考图 6b，在 iSCSI 调用 iSER 层 22 之后，iSER 层 22（在方框 164 处）接管连接并建立 iSER RDMA 通信通道。对于 RNIC 适配器 24，连接信息是连网层 28 例如 TCP 先前所建立的，对于无限带宽适配器 24，连接信息包括由 SDP 层 42 创建及使用的 RDMA 连接和数据结构。在 RDMA RNIC 或无限带宽适配器任一种的情况下，数据 / 控制结构被修改以允许 iSER/RDMA 操作。对于无限带宽适配器 24，iSER 层 22 发出（在方框 166 处）特定调用以终止 SDP 会话并接管来自 SDP 层 42 的 RDMA 通信。在通信过程完全转换到 iSER RDMA 模式之前（在方框 168 处），由目标 iSCSI 发送给 iSER 的“最后”登录应答 PDU 将被发送给起始节点，作为在先前连接模式（例如 TCP/IP 或 SDP/IB）上发送的最后信息。然后对于起始节点的所有将来的交互，iSCSI 层通过使用 RDMA 的 iSER 层通信（在方框 170 处）。

[0032] 图 7 说明当 SCSI 命令完成并且接收到来自具有能够发送“无效发送消息”的 RDMA 适配器的目标节点的“无效发送”消息时由起始节点 2 执行的操作。当接收到（在方框 200 处）“无效发送”消息时，如果适配器 24 支持“无效发送”，那么适配器 24 使消息中指定的 STag/R_key 无效（在方框 202 处），其中 STag 可以在 RDMA 头中。如果起始节点不支持“无效发送”，这可能当无限带宽适配器 24 时发生，那么消息转发给 iSER 层 22 来处理。起始 iSER 层 22 接收（在方框 204 处）应答 PDU 并访问来自消息的 ITT 56（图 3），以确定与 ITT/STag 映射 70 中的所访问 ITT 关联的 STag 或 R_Key。如果（在方框 206 处）“无效发送”消息中的 STag/R_key 匹配 ITT/STag 映射 70 中的 STag/R_Key 并且起始的适配器支持自动无效，那么起始 iSER 层 22 使得与映射 70 中的 ITT 关联的没有被适配器宣告无效的任何另外的 STag/#_Key 无效（方框 208 处）。如果适配器不支持自动无效或者如果“无效发送”中的 STag/R_key 不匹配映射 70 中的一个 STag/R_key，那么 iSER 层 22 直接调用（在方框 210 处）适配器，使该 Stag/R_key 以及与映射 70 中的 ITT 关联的任何另外的 Stag/#Key 无效。然后 iSER 层 22 传送（在方框 212 处）SCSI 应答 PDU 给 iSCSI 层 18 并整理其 ITT/Stag/#_ Key 映射表 70。然后 iSCSI 层 118 通知 SCSI 层 16 操作的完成状态。

[0033] 图 8 说明当 SCSI 命令完成并且接收到来自具有无限带宽适配器 24 的目标节点的“立即数据发送”消息时由起始节点 2 执行的操作。当起始适配器 24（无限带宽或 RNIC）接收到（在方框 250 处）来自目标无限带宽适配器的“立即数据发送”消息或“发送”消息时，

消息由起始适配器 24 传送到 iSER 层 22，并且 iSER 层访问适配器 24 以使消息的“立即数据”部分中的 R_Key 无效，如果有的话。然后 iSER 层 22 促使适配器 24 使与映射 70 中的 ITT 关联的先前没有宣告无效的任何 #_Key/STag 无效（在方框 254 处）。iSER 层 22 传送（在方框 256 处）SCSI 应答 PDU 给 iSCSI 层 18 并整理其 ITT/Stag/#_Key 映射表 70。然后 iSCSI 层 18 通知 SCSI 层 16 操作的完成状态。

[0034] 协议网关

[0035] 图 9、10、11a 和 11b 说明网关 302，322，352 和 354 如何可以用来在节点之间传送消息。每个网关 302，322，352 和 354 可以包括目标和起始通过它传送消息的转发硬件设备例如交换机，路由器等。网关 302，322，352 和 354 包括协议转换器 314，334，362 和 364，以处理在起始和目标节点之间发送的、从一个协议发送到另一个的消息。可选地，网关 302 可以在目标节点或起始节点硬件中实施。网关 302，322，352 和 354 还包括提供以一个协议例如无限带宽到另一个协议例如 iWARP 的消息之间的协议映射 316，336，353 和 365。

[0036] 图 9 说明实施使用 iWARP 的 iSCSI/iSER 360 并包括 RNIC 310 以使用 iWARP 协议传送消息的目标节点 300。网关 302 接收从 iWARP 网络上的目标节点 300 指向实施使用无限带宽的 iSCSI/iSER 协议 308 并具有无限带宽适配器 312 的起始节点 304 的消息。协议转换器 314 将 iWARP 消息转换成符合无限带宽协议的与无限带宽适配器 312 可兼容的消息。

[0037] 图 10 说明实施使用无限带宽的 iSCSI/iSER 326 并包括无限带宽适配器 330 以使用无限带宽协议传送消息的目标节点 320。网关 322 接收从无限带宽中的目标节点 320 指向实施使用 iWARP 的 iSCSI/iSER 协议 328 并具有 RNIC 332 的起始节点 324 的消息。协议转换器 334 将无限带宽消息转换成符合 iWARP 协议的与使用 iWARP 协议操作的 RNIC 适配器 332 可兼容的消息。

[0038] 图 11a 说明其中网关 352 和 354 合作以将来自目标节点 350 的 iSER/IB 消息传送到起始节点 356 同时转换消息以供在中间 iWARP 网络上传送的实施方案。接收来自目标 350 的无限带宽消息的目标网关 352 将消息转换成与 iWARP 协议可兼容的格式并发送转换后的消息到 iWARP 网络上的起始网关 354。接收来自目标网关 352 的 iWARP 消息的起始网关 354 将消息转换成与起始节点 356 处所使用的无限带宽协议可兼容的格式，然后发送转换后的消息到起始节点 356。以这种方式，网关 352 和 354 用来转换消息，以便通过在使用无限带宽协议并包括使用无限带宽的 iSCSI/iSER 358，366 和无限带宽适配器 360，368 的两个节点 350 和 356 之间的 iWARP 网络传送。此外，在可选实施方案中，在起始和目标节点之间可以有执行许多消息转换以供不同的可能通信协议使用的许多网关。

[0039] 图 11b 说明其中两个网关 372 和 374 合作以将来自目标节点 370 的 iSER/iWARP 消息传送到起始节点 376 同时转换消息以供在中间无限带宽网络上传送的实施方案。接收来自目标节点 370 的 iWARP 消息的目标网关 372 将消息转换成与无限带宽协议可兼容的格式并发送转换后的消息到无限带宽网络上的起始网关 374。接收来自目标网关 372 的无限带宽消息的起始网关 374 将消息转换成与起始节点 376 处所使用的 iWARP 协议可兼容的格式，然后发送转换后的消息到 iWARP 网络上的起始 376。以这种方式，网关 372 和 374 用来转换消息，以便通过在使用 iWARP 协议并且包括使用 iWARP 的 iSCSI/iSER 378，390 和 iWARP 适配器 380，392 的两个节点 370 和 376 之间的无限带宽网络传送。此外，在可选实施方案中，在起始和目标节点之间可以有执行许多消息转换以供不同的可能通信协议使用的许多

网关。

[0040] 图 9、10、11a 和 11b 显示从目标节点流到起始节点的消息。但是，消息流可以从起始进行到目标节点，或者在任何两个类型的节点之间。例如，在这一点上图 11a 的网关 352 和 354 可以重复，并且图 11b 中的网关 372 和 374 可以重复。

[0041] 图 12、13 和 14 说明由协议转换器 314（图 9），334（图 10），362（图 11a），364，382（图 11b）和 386 执行的以将消息从传送消息的节点所使用的格式转换成实施不同于传送消息的节点所使用协议的、与接收消息的节点可兼容的格式的操作。参考图 12，协议转换器 314，334，362，364，382 和 386（网关）接收（在方框 400 处）来自目标节点（或通过另一网关节点来自目标节点）的 iSCSI/iSER 消息。如果（在方框 402 处）消息来自使用 iSER/iWARP 的目标节点（或目标的网关）并且如果（在方框 408 处）iWARP 消息类型映射到等价的无限带宽消息，那么协议转换器 314，364 或 382 将 iWARP 消息转换成（在方框 410 处）等价的无限带宽消息并将转换后的无限带宽消息转发到无限带宽网络上的起始节点例如 304 或 356（或随后的网关）。协议映射 316，336，353，365，384 和 388 可以为协议转换器 314，334，362，364，382 和 386 提供以不同格式例如无限带宽和 iWARP 的等价消息对，它们可以保存 iWARP 到无限带宽（以及无限带宽到 iWARP）消息类型的映射，使得协议映射指示以一个协议格式的消息如何可以被转换并映射其他协议格式。如果（在方框 408 处）iWARP 消息类型不映射到等价的无限带宽消息并且如果（在方框 412 处）消息不是具有 STag 的 iWARP 无效发送消息，那么协议转换器 314，364 或 382 丢弃（在方框 414 处）消息并抛出错误，因为这种转换不由协议转换器 314，364 或 382 处理。

[0042] 如果（在方框 412 处）消息是具有 STag 的 iWARP“无效发送消息”，那么协议转换器 314，364 或 382 创建（在方框 416 处）无限带宽“有（或无）请求事件发送”消息。协议转换器 314，364 或 382 将索引目标或起始中存储位置的直接参考的、来自 iWARP 消息的 STag 添加（在方框 418 处）到无限带宽消息中的立即数据域（可选地，丢弃 STag 并准备无任何立即数据发送）。协议转换器 314，364 或 382 传送（在方框 420 处）转换后的消息到无限带宽网络上的起始节点（或随后的网关）。从方框 410 或 420，控制进行到方框 422，其中如果有随后的网关，那么通过执行从图 13 中的方框 440 开始的操作，这种网关将 iSER/IB 消息转换成 iSER/iWARP。

[0043] 如果（在方框 402 处）来自目标节点的消息是以无限带宽协议的，那么控制进行到图 13 中的方框 440。如果（在方框 440 处）传输在到达起始节点之前在 iWARP 网络上继续到网关例如网关 352，并且如果（在方框 442 处）无限带宽消息类型映射到协议映射 353 中的等价 iWARP 消息，那么协议转换器 362 将无限带宽消息转换成（在方框 444 处）等价的 iWARP 消息并将转换后的消息在 iWARP 上转发到下一个网关 354。如果（在方框 442 处）无限带宽消息类型不映射到协议映射 353 中的等价 iWARP 消息，并且如果（在方框 446 处）消息不是无限带宽“立即数据发送”消息，那么错误被抛出（在方框 448 处）并且消息被丢弃。

[0044] 如果（在方框 446 处）消息是无限带宽“立即数据发送”消息，那么协议转换器 362 创建（在方框 450 处）iWARP 无效（或请求事件）发送消息并将来自无限带宽消息中的立即数据域的 R_Key 添加（在方框 452 处）到 iWARP“无效发送消息”中的 STag 域（可选地，丢弃 R_Key（立即数据）并建立无 STag 的发送消息）。协议转换器 362 将转换后的消息传

送（在方框 454 处）到 iWARP 网络上的网关 354，如图 11a 中所示。从方框 454 或 444，随后的网关可以通过执行从方框 400 开始的操作将 iSER/iWARP 消息转换成 iSER/ 无限带宽消息。

[0045] 如果（在方框 440 处）来自目标节点的无限带宽传输在到达起始节点之前不在 iWARP 网络上继续到网关（即无限带宽消息将穿过 iWARP 网络上的网关 322 或 374 直接继续到起始 324 或 376，如分别在图 10 和图 11b 中所示），那么控制进行到图 14 中的方框 480。如果（在方框 480 处）无限带宽消息类型映射到等价的 iWARP 消息，那么协议转换器 334 或 386 将无限带宽消息转换成（在方框 486 处）等价的 iWARP 消息并在 iWARP 上转发给起始节点。如果（在方框 480 处）无限带宽消息类型不映射到协议映射 336 或 388 中的等价 iWARP 消息，并且如果（在方框 488 处）消息不是无限带宽“立即数据发送”消息，那么错误被抛出（在方框 490 处），并且消息被丢弃。

[0046] 如果（在方框 488 处）消息是无限带宽立即数据发送消息，那么协议转换器 334 或 386 创建（在方框 492 处）iWARP 无效（请求事件）发送消息并将来自无限带宽消息中的立即数据域的 R_Key 添加（在方框 494 处）到 iWARP 无效发送消息中的 STag 域中（可选地，丢弃 R_Key（立即数据）并建立无 STag 的发送消息）。协议转换器 334 或 386 将转换后的消息传送（在方框 496 处）到 iWARP 网路上的起始节点，如图 10 或 11b 中所示。

[0047] 所描述的实施方案提供一种通过处理并且如果需要时将消息转换成与接收节点所使用的通信协议可兼容的格式来允许消息在使用不同传输协议的网络之间传送的技术。

[0048] 另外的实施方案细节

[0049] 这里所描述的实施方案可以使用制造软件、固件、硬件或其任意组合的标准编程和 / 或工程技术实施为一种方法、装置或制造产品。这里所使用的术语“制造产品”指的是在硬件逻辑（例如集成电路芯片、可编程门阵列 (PGA)、专用集成电路 (ASIC) 等）或计算机可读媒介，例如磁性存储媒介（例如硬盘驱动器、软盘、磁带等）、光存储器 (CD-ROM、光盘等）、易失性和非易失性存储设备（例如 EEPROM、ROM、PROM、RAM、DRAM、SRAM、固件、可编程逻辑等）中实施的代码或逻辑。计算机可读媒介中的代码由处理器访问并执行。实施优选实施方案的代码还可以通过传输媒介或从网络上的文件服务器访问。在这些情况下，实施代码的制造产品可以包括传输媒介，例如网络传输线、无限传输媒介、通过空间传播的信号、无限电波、红外信号等。因此，“制造产品”可以包括包含代码的媒介。另外，“制造产品”可以包括其中代码被包含、处理和执行的硬件及软件部件的组合。当然，本领域技术人员应答认识到，可以不背离本发明的范畴对该配置作许多修改，并且制造产品可以包括本领域中已知的任何信息承载媒介。

[0050] 所描述的操作可以由电路系统执行，其中“电路系统”指的是硬件或软件或其组合。用于执行所描述实施方案的操作的电路系统可以包括硬件设备，例如集成电路芯片、可编程门阵列 (PGA)、专用集成电路 (ASIC) 等。电路系统也可以包括处理器部件例如集成电路，以及计算机可读媒介例如存储器中的代码，其中代码由处理器执行，以进行所描述实施方案的操作。

[0051] 在所描述的实施中，物理层使用以太网协议。在可选实施中，提供分组的链路到链路校验求和 /CRC（或其他数据检测方案）的可选协议可以用来代替以太网，例如串行高级技术附加 (SATA)、无限带宽、串行附加 SCSI 电缆等。

[0052] 在所描述的实施中，传输层包括 iSCSI 协议。在可选实施中，本领域中已知的用于以分组传送 I/O 命令并提供端到端校验求和 /CRC（或其他数据检测方案）的其他协议也可以使用。

[0053] 在所描述的实施中，被包装的 I/O 命令包括 SCSI 命令。在可选实施中，命令可以是与 SCSI 不同的 I/O 命令格式，例如高级技术附加 (ATA)。

[0054] 在所描述的实施方案中，iSCSI 层调用 iSER 层以访问 RDMA 数据传送能力。在另外的实施方案中，不同于 iSCSI 的数据传输协议层例如应用或其他数据传输协议可以调用 iSER 层以访问 RDMA 数据传送能力。

[0055] 在可选实施方案中，无限带宽上 IP 协议（具有可靠连接 --RC) 可以代替 SDP 用来跨越无限带宽网络传送使用协议例如 TCP 编码的分组。关于无限带宽上 IP 协议（具有可靠连接 --RC) 的更多细节在由 IETF 出版为“draft-kashyap-ipoib-connected-mode-01.txt”(2003 年 9 月）的出版物“无限带宽上 IP :连接模式”中描述，该出版物在此整体引入作为参考。例如，SDP 层可以代替分层于 IPoIB(RC) 实施上方的 TCP 栈，并且该 TCP/IPoIB 组合的任何部分可以放置于结点 2 软件或适配器 24 中。在这种实施方案中，IPoIB(RC) 函数可以根据 IPoIB(RC) 规范调用所需的 RDMA 层 26。

[0056] 在另外的实施方案中，不同于 TCP 的协议可以用来在允许 IP 的网络上传送分组，例如流控制传输协议 (SCTP)，该协议在出版物“流控制传输协议”，RFC 2960（以太网协会，2000) 中定义，该出版物在此整体引入作为参考。

[0057] 图 5a、5b、6a、6b、7 和 8 描述以特定顺序发生的特定操作。在可选实施中，某些操作可以以不同的顺序执行、修改或移除。此外，步骤可以添加到上述逻辑，而仍然符合所描述的实施。此外，这里所描述的操作可以顺序地发生或者某些操作可以并行处理。此外，操作可以由单个处理单元或者由分布式处理单元执行。

[0058] 前面实施的描述已为了说明及描述的目的而提供。这不打算是排他性的或将本发明限制于所描述的精确形式。根据上面的讲授许多修改和变化是可能的。打算是本发明的范畴不是由本详细说明书而是由这里附加的权利要求所限定。上面的说明书、例子和数据提供本发明的组成部分的制造及使用的完整描述。因为可以不背离本发明的本质和范畴而进行本发明的许多实施，本发明属于下文所附加的权利要求。

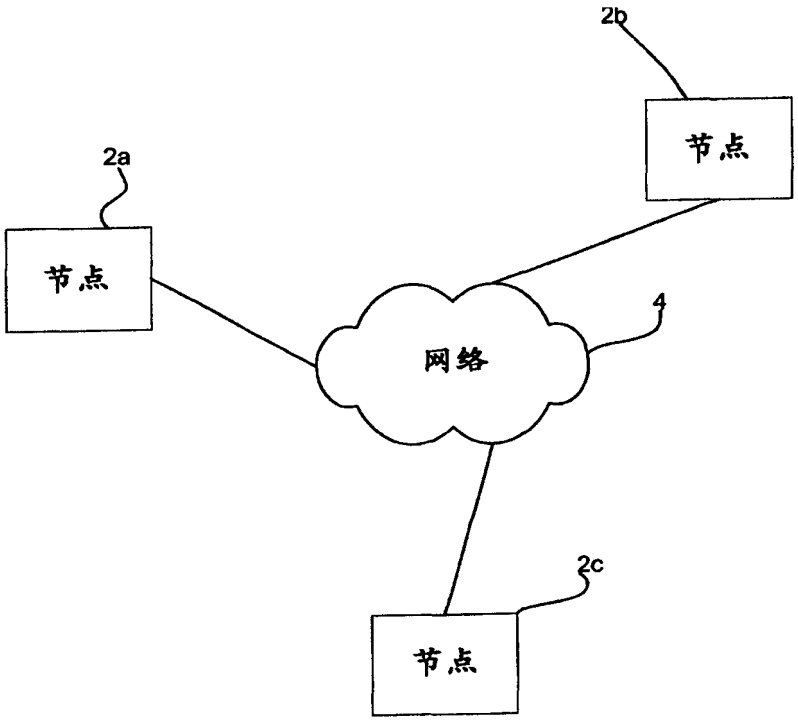


图 1

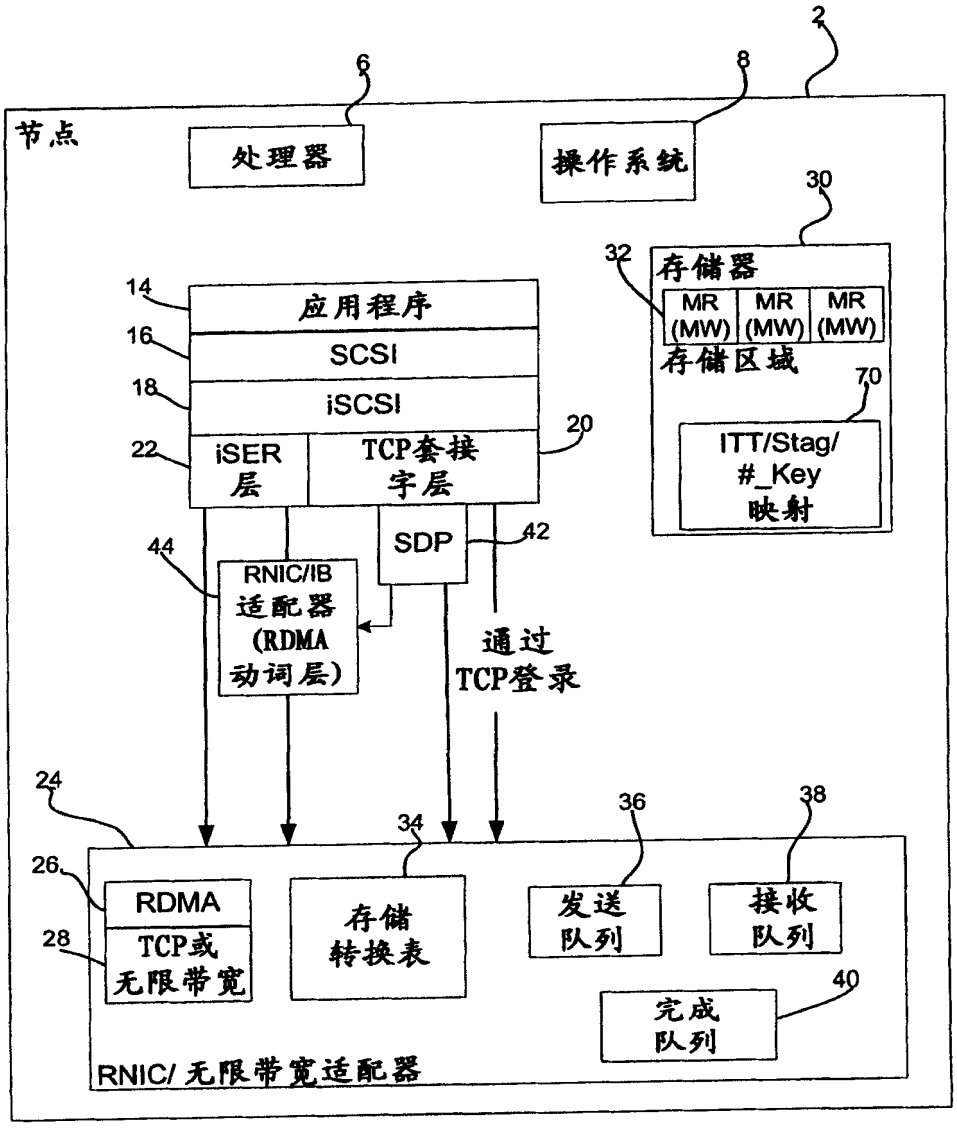

2
6
8
节点
处理器
操作系统
30
32
存储器
MR (MW)
MR (MW)
MR (MW)
存储区域
14
应用程序
16
SCSI
18
iSCSI
20
22
iSER 层
TCP套接字层
70
ITT/Stag/ #_Key 映射
SDP
42
44
RNIC/IB 适配器 (RDMA 动词层)
通过 TCP登录
24
34
36
38
26
RDMA
28
TCP或 无限带宽
存储 转换表
发送 队列
接收 队列
40
完成 队列
RNIC/ 无限带宽适配器

图 2

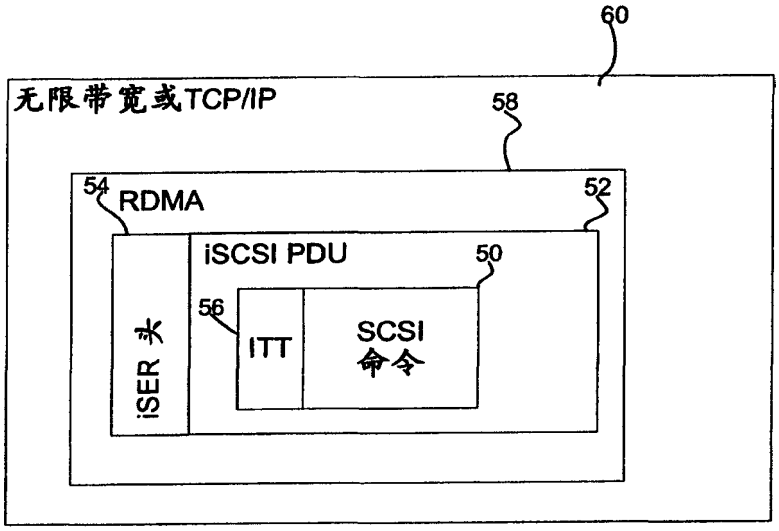

60
无限带宽或TCP/IP
58
54
RDMA
52
iSCSI PDU
50
56
ITT
SCSI
命令
iSER 头

图 3

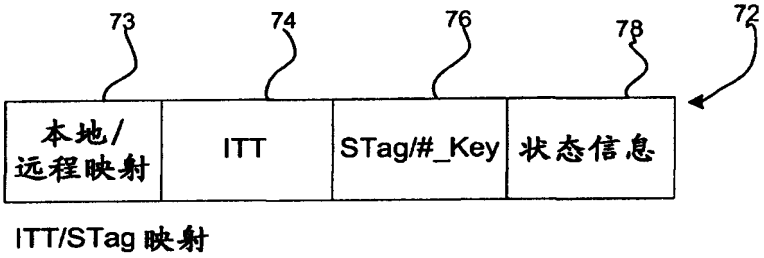

73
74
76
78
72
本地/
远程映射
ITT
STag/#_Key
状态信息
ITT/STag 映射

图 4

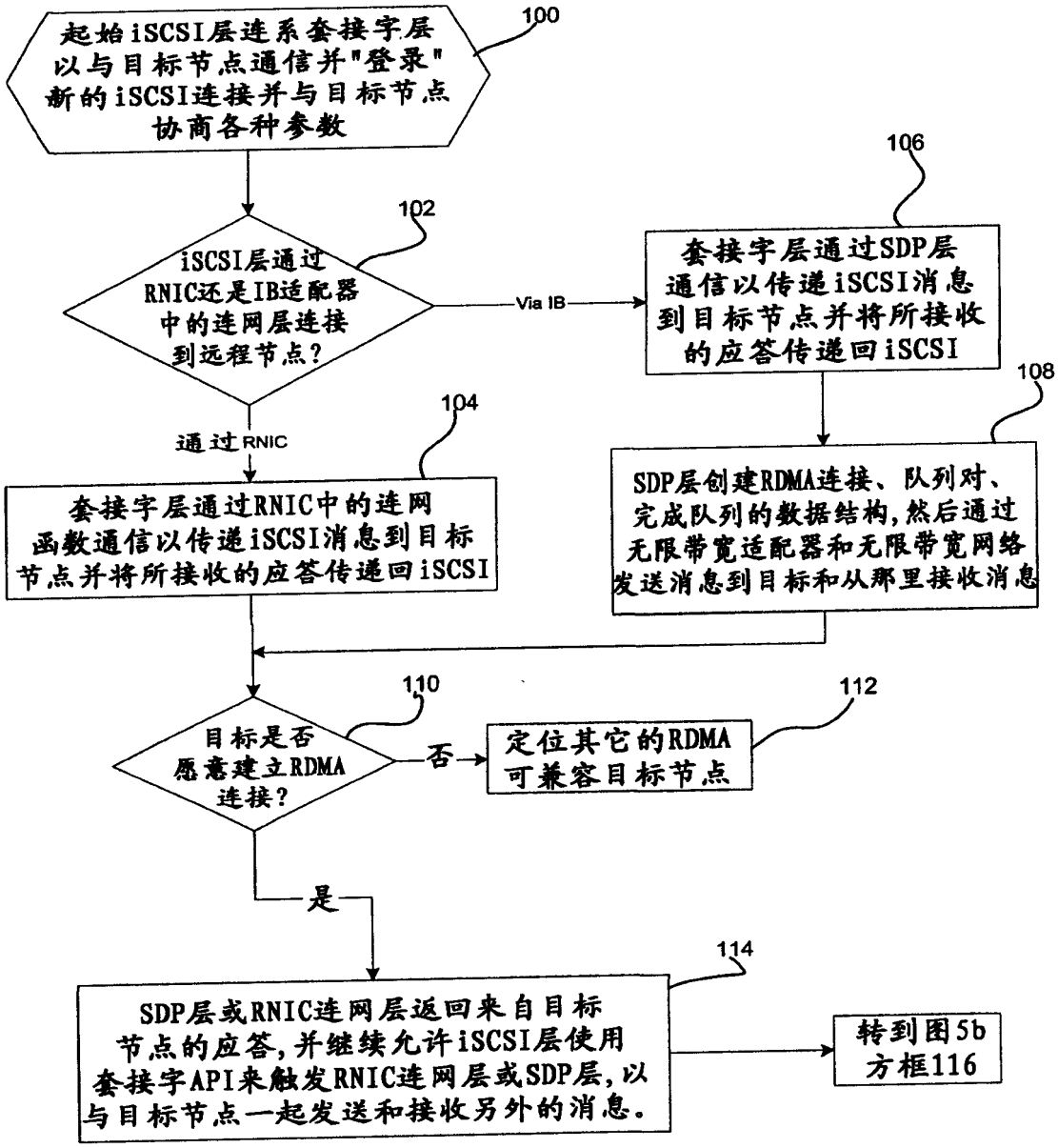

起始iSCSI层连系套接字层以与目标节点通信并"登录"新的iSCSI连接并与目标节点协商各种参数
100
iSCSI层通过RNIC还是IB适配器中的连网层连接到远程节点?
102
Via IB
套接字层通过SDP层通信以传递iSCSI消息到目标节点并将所接收的应答传递回iSCSI
106
通过RNIC
104
套接字层通过RNIC中的连网函数通信以传递iSCSI消息到目标节点并将所接收的应答传递回iSCSI
SDP层创建RDMA连接、队列对、完成队列的数据结构,然后通过无限带宽适配器和无限带宽网络发送消息到目标和从那里接收消息
108
目标是否愿意建立RDMA连接?
110
否
定位其它的RDMA可兼容目标节点
112
是
SDP层或RNIC连网层返回来自目标节点的应答,并继续允许iSCSI层使用套接字API来触发RNIC连网层或SDP层,以与目标节点一起发送和接收另外的消息。
114
转到图5b方框116

图 5a

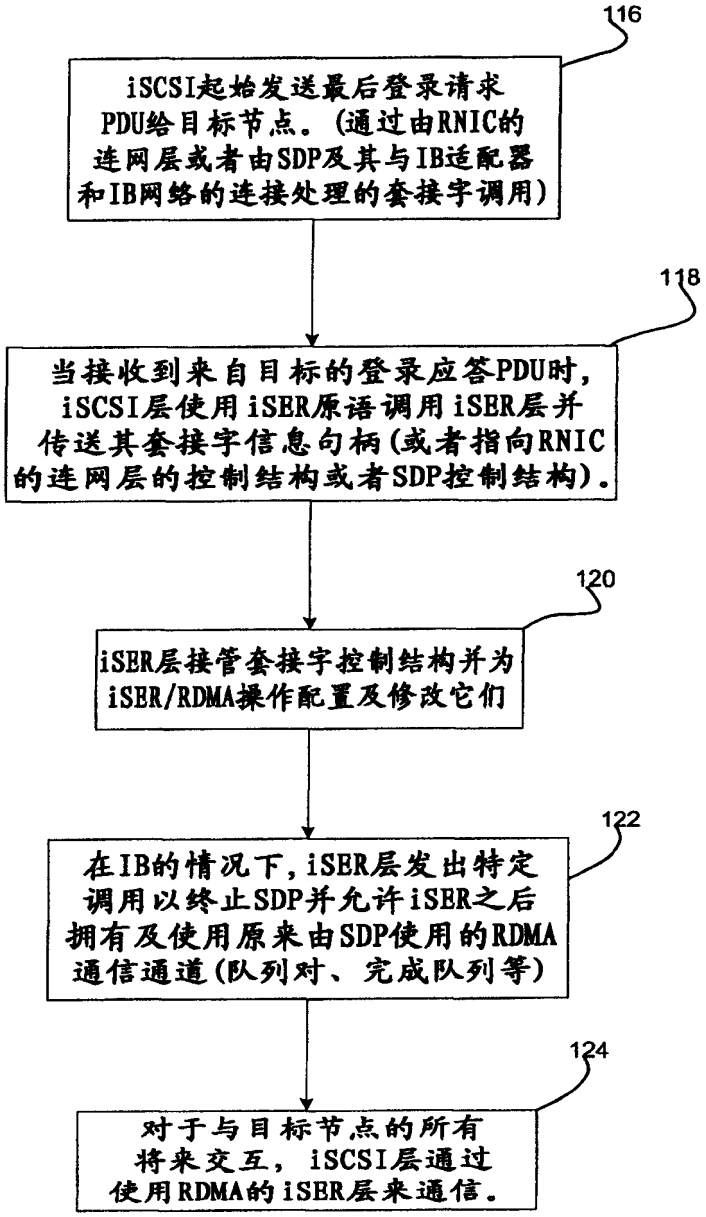

116
iSCSI起始发送最后登录请求PDU给目标节点。(通过由RNIC的连网层或者由SDP及其与IB适配器和IB网络的连接处理的套接字调用)
118
当接收到来自目标的登录应答PDU时，iSCSI层使用iSER原语调用iSER层并传送其套接字信息句柄(或者指向RNIC的连网层的控制结构或者SDP控制结构)。
120
iSER层接管套接字控制结构并为iSER/RDMA操作配置及修改它们
122
在IB的情况下，iSER层发出特定调用以终止SDP并允许iSER之后拥有及使用原来由SDP使用的RDMA通信通道(队列对、完成队列等)
124
对于与目标节点的所有将来交互，iSCSI层通过使用RDMA的iSER层来通信。

图 5b

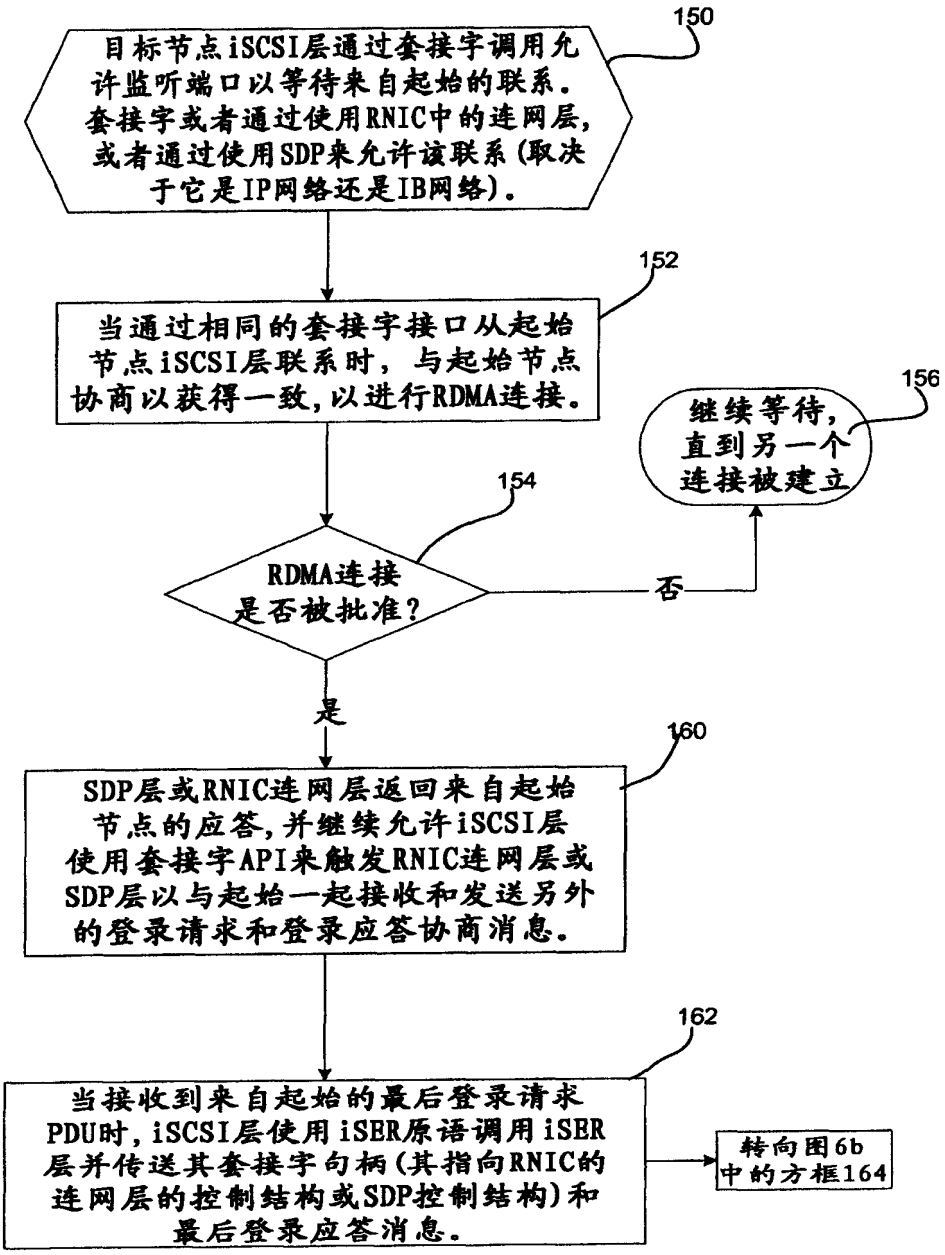

150
目标节点iSCSI层通过套接字调用允许监听端口以等待来自起始的联系。套接字或者通过使用RNIC中的连网层，或者通过使用SDP来允许该联系(取决于它是IP网络还是IB网络)。
152
当通过相同的套接字接口从起始节点iSCSI层联系时，与起始节点协商以获得一致，以进行RDMA连接。
156
继续等待，直到另一个连接被建立
154
RDMA连接是否被批准？
否
是
160
SDP层或RNIC连网层返回来自起始节点的应答，并继续允许iSCSI层使用套接字API来触发RNIC连网层或SDP层以与起始一起接收和发送另外的登录请求和登录应答协商消息。
162
当接收到来自起始的最后登录请求PDU时，iSCSI层使用iSER原语调用iSER层并传送其套接字句柄(其指向RNIC的连网层的控制结构或SDP控制结构)和最后登录应答消息。
转向图6b中的方框164

图 6a

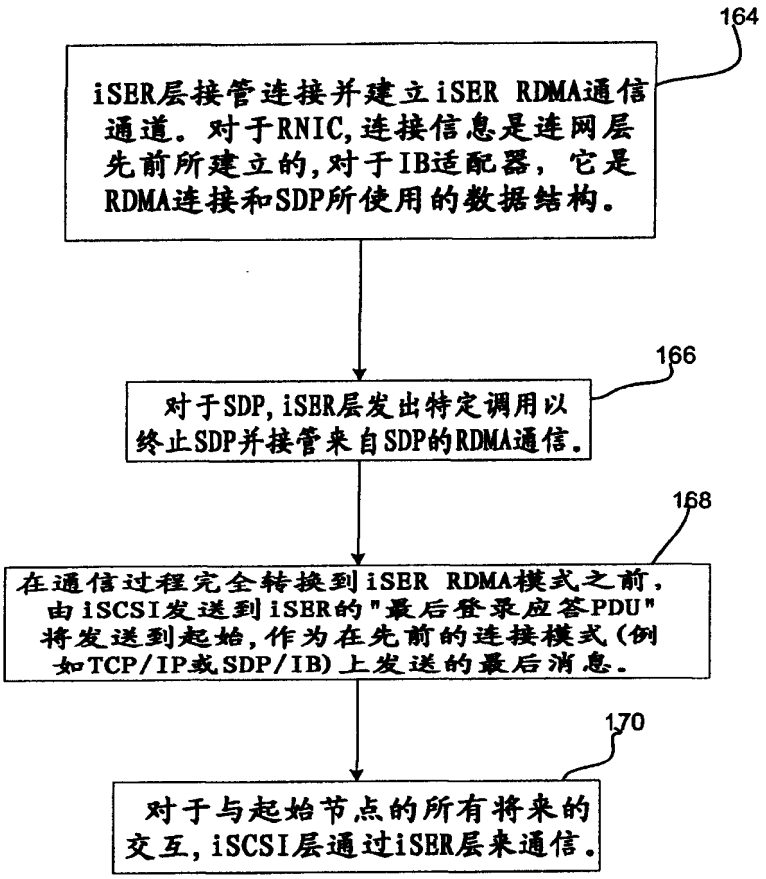

164
iSER层接管连接并建立iSER RDMA通信通道。对于RNIC,连接信息是连网层先前所建立的,对于IB适配器，它是RDMA连接和SDP所使用的数据结构。
166
对于SDP,iSER层发出特定调用以终止SDP并接管来自SDP的RDMA通信。
168
在通信过程完全转换到iSER RDMA模式之前，由iSCSI发送到iSER的"最后登录应答PDU"将发送到起始,作为在先前的连接模式(例如TCP/IP或SDP/IB)上发送的最后消息。
170
对于与起始节点的所有将来的交互,iSCSI层通过iSER层来通信。

图 6b

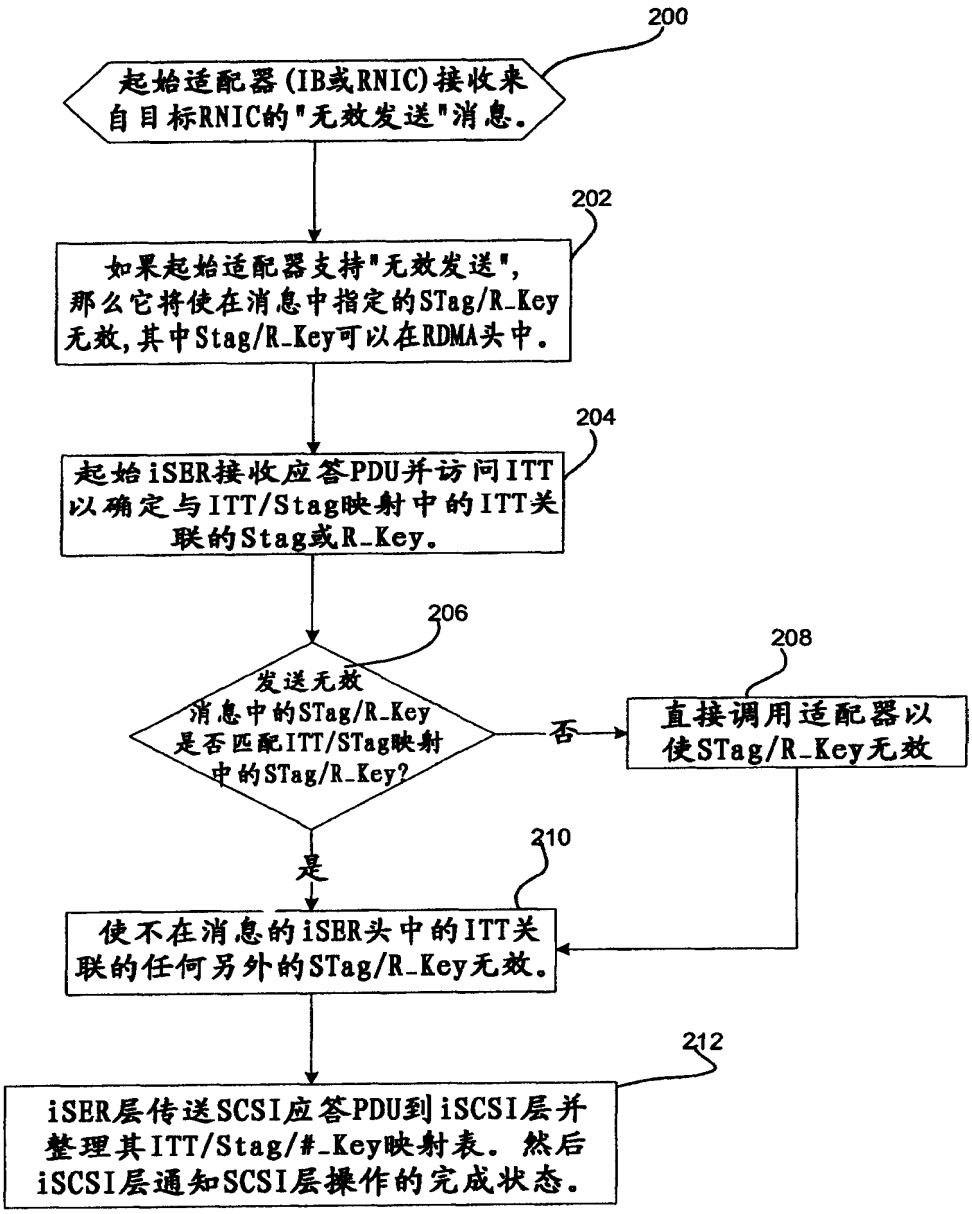

200
起始适配器(IB或RNIC)接收来自目标RNIC的"无效发送"消息。
202
如果起始适配器支持"无效发送",那么它将使在消息中指定的STag/R_Key无效,其中Stag/R_Key可以在RDMA头中。
204
起始iSER接收应答PDU并访问ITT以确定与ITT/Stag映射中的ITT关联的Stag或R_Key。
206
发送无效消息中的STag/R_Key是否匹配ITT/STag映射中的STag/R_Key?
否
208
直接调用适配器以使STag/R_Key无效
是
210
使不在消息的iSER头中的ITT关联的任何另外的STag/R_Key无效。
212
iSER层传送SCSI应答PDU到iSCSI层并整理其ITT/Stag/#_Key映射表。然后iSCSI层通知SCSI层操作的完成状态。

图 7

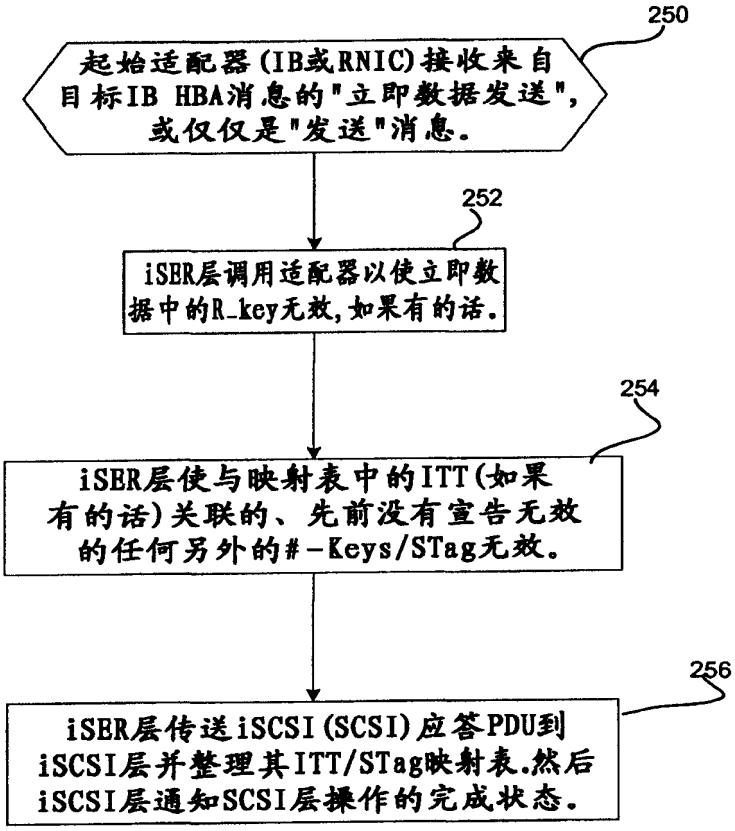

250
起始适配器(IB或RNIC)接收来自目标IB HBA消息的"立即数据发送",或仅仅是"发送"消息。
252
iSER层调用适配器以使立即数据中的R_key无效,如果有的话。
254
iSER层使与映射表中的ITT(如果有的话)关联的、先前没有宣告无效的任何另外的#-Keys/STag无效。
256
iSER层传送iSCSI(SCSI)应答PDU到iSCSI层并整理其ITT/STag映射表.然后iSCSI层通知SCSI层操作的完成状态。

图 8

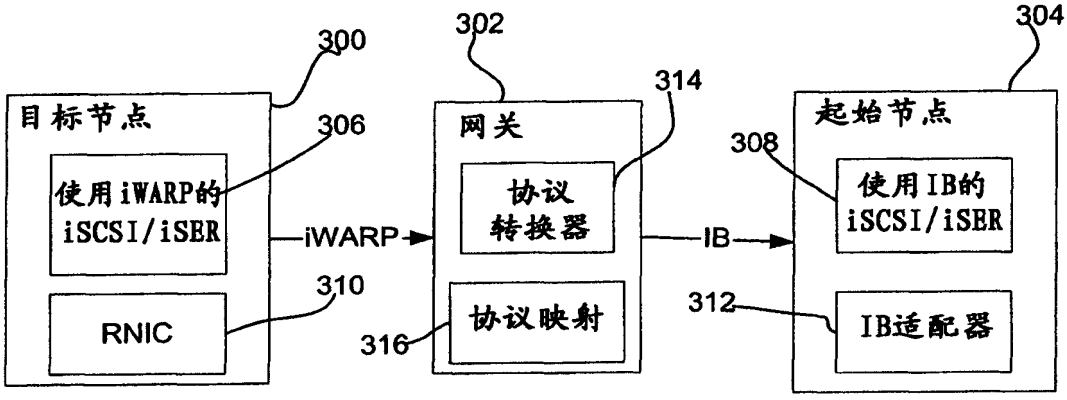

300
302
304
目标节点
网关
起始节点
306
314
308
使用iWARP的iSCSI/iSER
协议转换器
使用IB的iSCSI/iSER
iWARP
IB
310
312
RNIC
协议映射
IB适配器
316

图 9

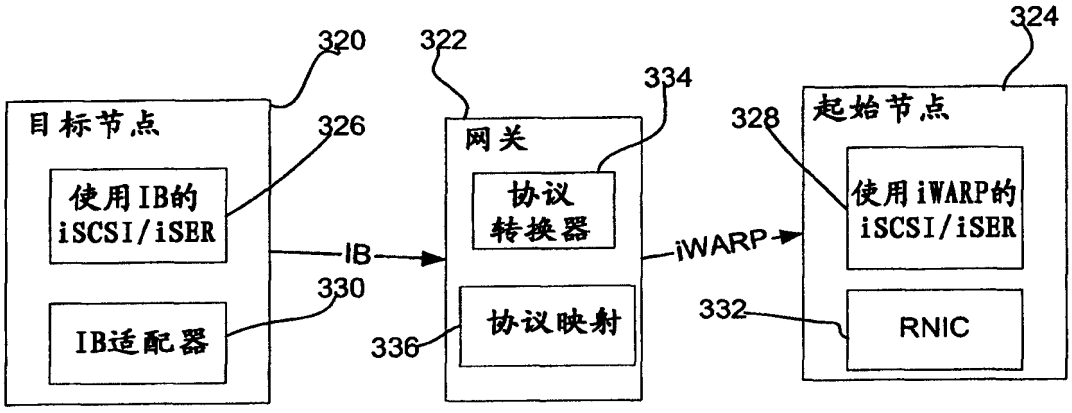


图 10

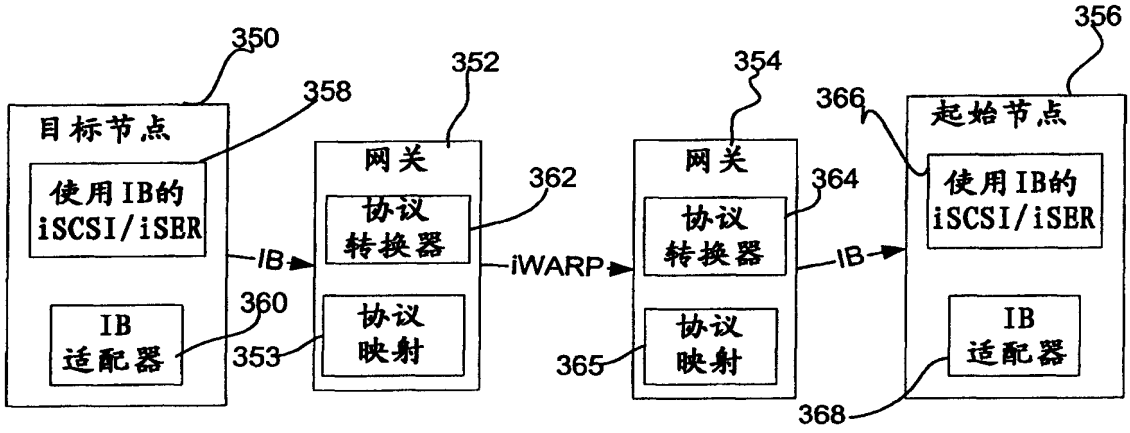


图 11a

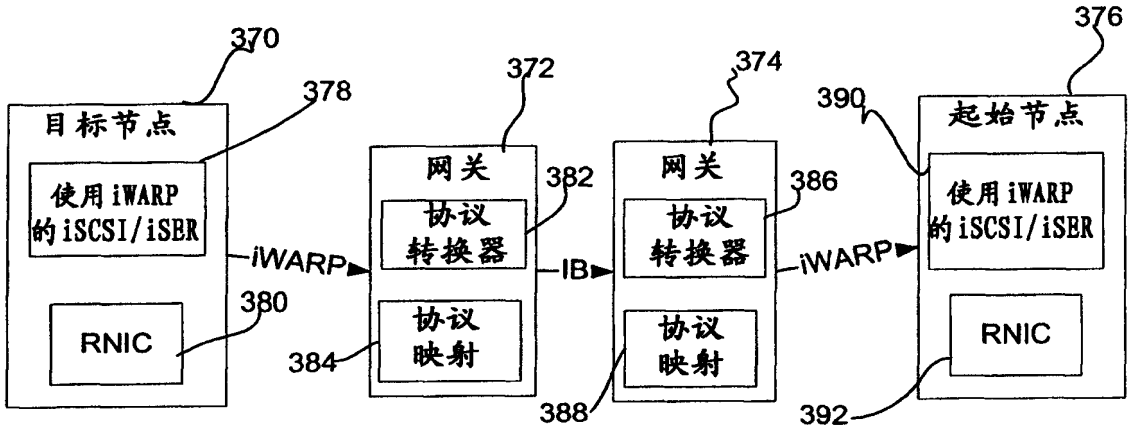


图 11b

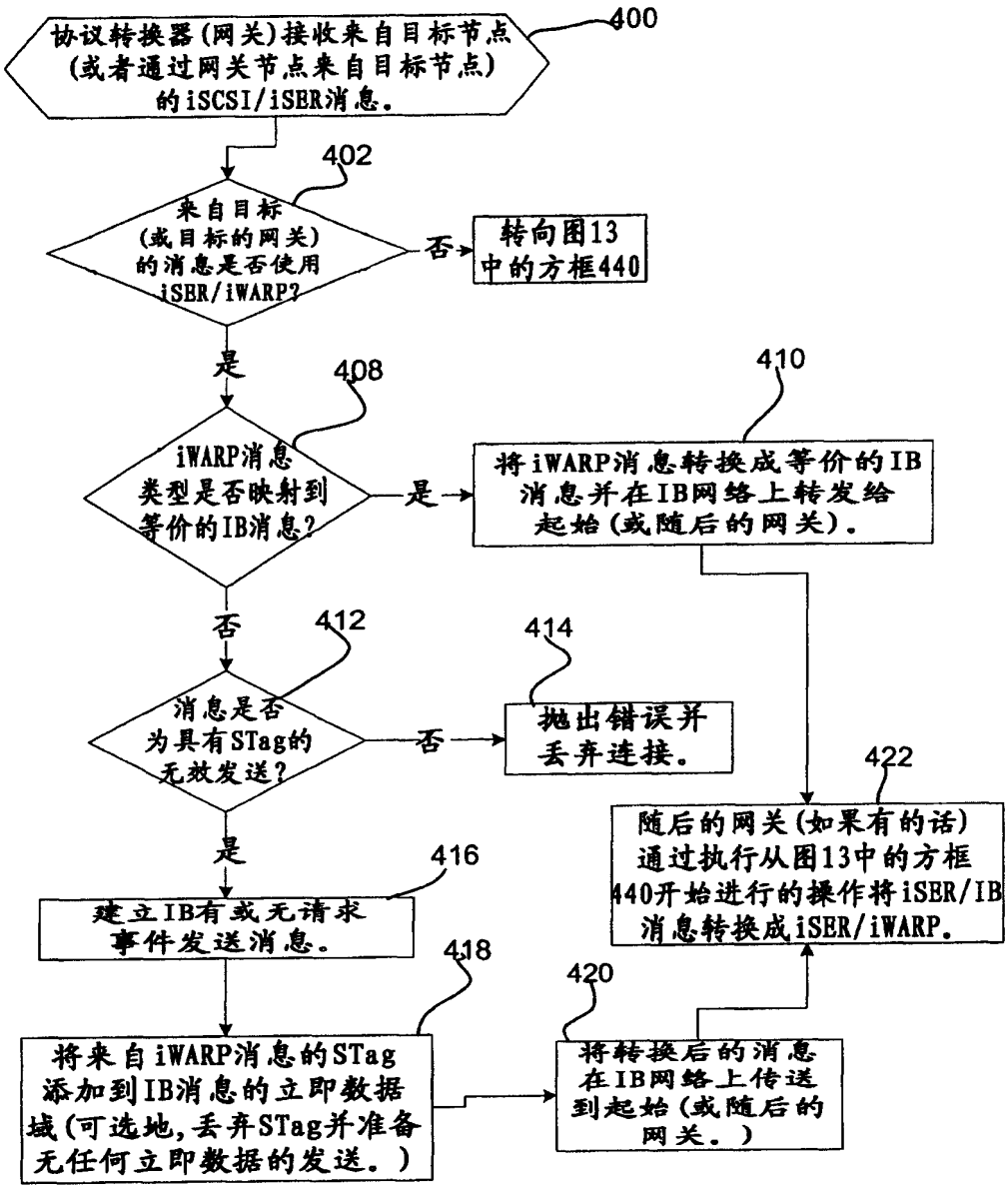

协议转换器(网关)接收来自目标节点(或者通过网关节点来自目标节点)的iSCSI/iSER消息。
400
402
来自目标(或目标的网关)的消息是否使用iSER/iWARP?
否
转向图13中的方框440
是
408
iWARP消息类型是否映射到等价的IB消息?
是
410
将iWARP消息转换成等价的IB消息并在IB网络上转发给起始(或随后的网关)。
否
412
消息是否为具有STag的无效发送?
否
414
抛出错误并丢弃连接。
是
416
建立IB有或无请求事件发送消息。
418
将来自iWARP消息的STag添加到IB消息的立即数据域(可选地,丢弃STag并准备无任何立即数据的发送。)
420
将转换后的消息在IB网络上传送到起始(或随后的网关。)
422
随后的网关(如果有的话)通过执行从图13中的方框440开始进行的操作将iSER/IB消息转换成iSER/iWARP。

图 12

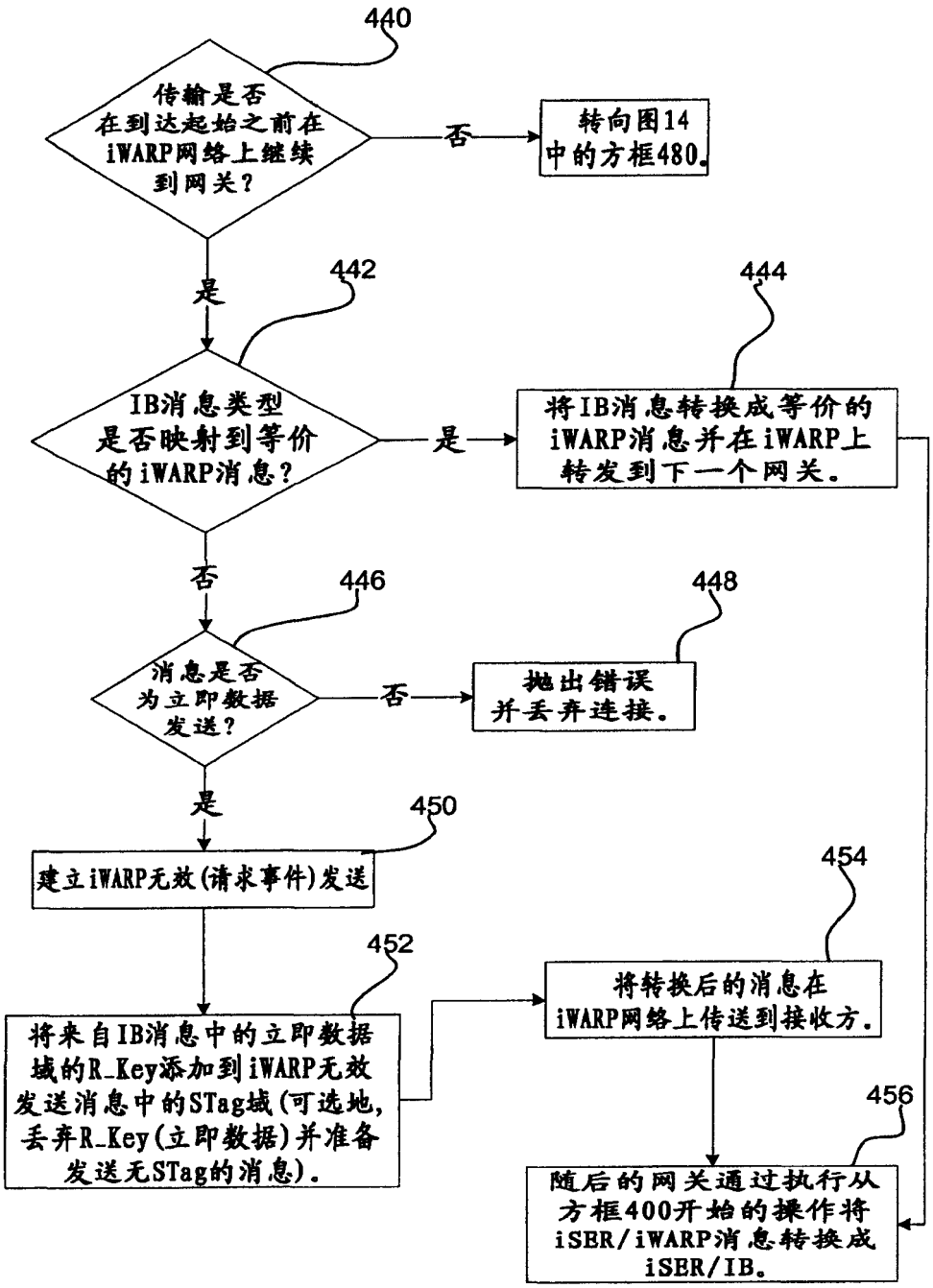

440
传输是否在到达起始之前在iWARP网络上继续到网关？
否
转向图14中的方框480。
是
442
IB消息类型是否映射到等价的iWARP消息？
是
444
将IB消息转换成等价的iWARP消息并在iWARP上转发到下一个网关。
否
446
消息是否为立即数据发送？
否
448
抛出错误并丢弃连接。
是
450
建立iWARP无效(请求事件)发送
452
将来自IB消息中的立即数据域的R_Key添加到iWARP无效发送消息中的STag域(可选地，丢弃R_Key(立即数据)并准备发送无STag的消息)。
454
将转换后的消息在iWARP网络上传送到接收方。
456
随后的网关通过执行从方框400开始的操作将iSER/iWARP消息转换成iSER/IB。

图 13

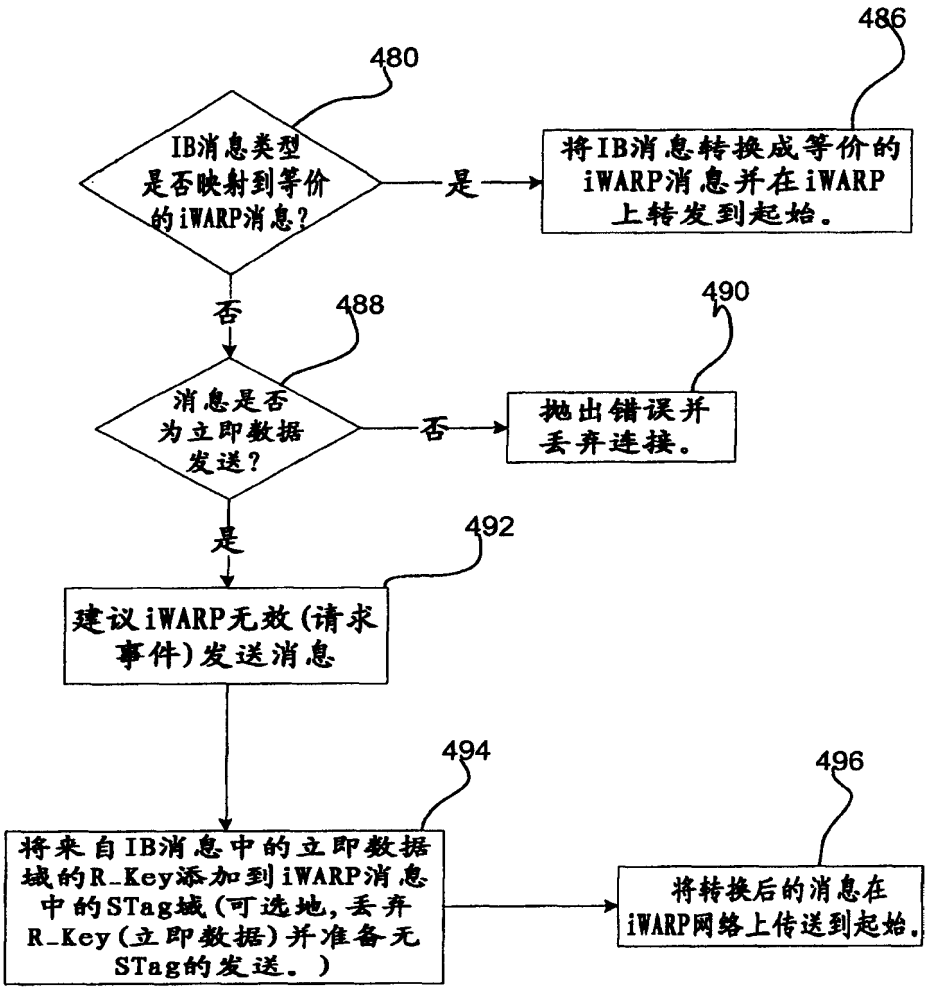

480
IB消息类型是否映射到等价的iWARP消息?
是
486
将IB消息转换成等价的iWARP消息并在iWARP上转发到起始。
否
488
消息是否为立即数据发送?
否
490
抛出错误并丢弃连接。
是
492
建议iWARP无效(请求事件)发送消息
494
将来自IB消息中的立即数据域的R_Key添加到iWARP消息中的STag域(可选地,丢弃R_Key(立即数据)并准备无STag的发送。)
496
将转换后的消息在iWARP网络上传送到起始。

图 14