



[12] 发明专利说明书

专利号 ZL 03810006.1

[45] 授权公告日 2007 年 10 月 10 日

[11] 授权公告号 CN 100342333C

[22] 申请日 2003.2.20 [21] 申请号 03810006.1

[30] 优先权

[32] 2002. 3. 8 [33] US [31] 10/092,793

[86] 国际申请 PCT/US2003/005085 2003. 2. 20

[87] 国际公布 WO2003/077116 英 2003. 9. 18

[85] 进入国家阶段日期 2004. 11. 2

[73] 专利权人 英特尔公司

地址 美国加利福尼亚州

[72] 发明人 P·豪金斯

[56] 参考文献

US4141066A 1979. 2. 20

US5644700A 1997. 7. 1

US6085333A 2000. 7. 4

审查员 王学睿

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 杨 凯 王 勇

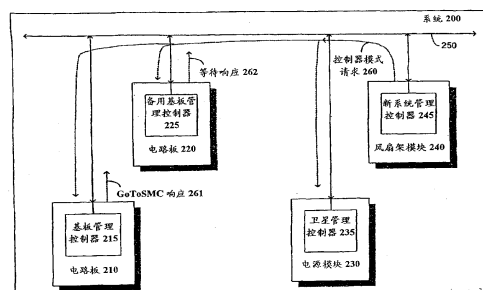
权利要求书 2 页 说明书 16 页 附图 6 页

[54] 发明名称

系统管理控制器协商协议

[57] 摘要

一种计算机系统模块包括一个系统管理控制器，以便与其它系统管理控制器协商以确定该控制器的初始操作模式。在实施例中，与其它系统管理控制器进行的协商至少部分基于控制器能力、用户配置的首选项、模块类型和地理地址之一。



1. 一种用于确定基板管理控制器模式的方法，所述方法包括：
从计算机系统多个系统管理控制器发送模式请求到所述计
5 算系统中的其它系统管理控制器；以及
根据不存在对所述系统管理控制器之一发送的模式请求的响
应，确定该系统管理控制器为基板管理控制器，
其中，所述方法还包括从所述系统管理控制器中的第一控制器
发送对从所述系统管理控制器中的第二控制器收到的模式请求的响
10 应，其中，所述第一管理控制器的所述响应至少部分基于所述第一
系统管理控制器的协商协议状态，
所述模式请求和响应作为符合智能平台管理接口规范的消息发
送。
2. 如权利要求 1 所述的方法，其特征在于：所述第一管理控制
15 器的所述响应至少部分基于所述第一系统管理控制器的所述相对优
先级。
3. 如权利要求 2 所述的方法，其特征在于：所述优先级基于控
制器模式能力、用户首选项、控制器的模块类型或控制器地理地址
中的至少一项。
- 20 4. 如权利要求 3 所述的方法，其特征在于：所述控制器地理地
址是确定优先级中最后考虑的因素。
5. 如权利要求 1 所述的方法，其特征在于：所述方法还包括至
少部分基于所述响应内容确定所述第二系统管理控制器的所述模
式。
- 25 6. 如权利要求 1 所述的方法，其特征在于：所述方法作为控制
器启动过程的一部分来执行。
7. 如权利要求 1 所述的方法，其特征在于：如果所述控制器发
送了极限数量的请求而未在超时期内收到响应，则不存在对模式请

求的响应。

8. 如权利要求 7 所述的方法，其特征在于：在确定所述系统管理控制器为所述基板管理控制器前，未在超时期内收到对控制器模式请求的响应的系统管理控制器至少重试所述请求一次。

- 5 9. 如权利要求 1 所述的方法，其特征在于：所述模式请求包含与所述操作模式相关的信息，发送所述请求的所述控制器可以在所述模式下操作。

系统管理控制器协商协议

5 技术领域

本发明的实施例涉及系统管理。具体地说，本发明的实施例涉及用于确定系统管理控制器操作模式的协商协议。

背景技术

10 计算机和其它电子系统包含在系统使用期内可能发生故障的多种组件。为了减少和/或纠正此类故障，一些系统包括了内置特性，如监视和控制系统硬件“健康状态”或性能的能力。此类功能有时称为系统管理，也可以以其它名称引用，如管理、硬件管理、平台管理等等。系统管理特性可包括例如对诸如温度、压力、风扇、电
15 源、总线错误、系统物理安全等要素的监控。另外，系统管理特性还可包括确定帮助识别故障硬件组件的信息及发出告警以指出组件有故障。可用于处理系统管理功能的组件之一是系统管理控制器（本文中也称为“控制器”）。系统管理控制器可以是微处理器、微控制器、专用集成电路（ASIC）或控制系统管理任务的其它类型的处理
20 单元。系统管理控制器可执行诸如接收系统管理信息，发送消息以控制系统性能，记录系统管理信息等任务。例如，管理控制器可从温度传感器接收表示系统温度在上升的指示，可发送命令以加快风扇速度，并可记录温度读数。

25 系统中的一个管理控制器可发挥中央系统管理控制器的作用，并执行诸如记录事件、收集现场可更换单元（FRU）库存信息、用户接口、主机 CPU 接口等中央管理功能。系统的中央管理控制器可称为系统的基板管理控制器（BMC）。其它非中央管理控制器可称为卫星管理控制器（SMC）。SMC 可执行针对系统特殊部分或功能的系统管理。例如，计算机系统可包含由总线连接的多个电路板和其

它组件，其中一块电路板包含了该系统的 BMC，而其它电路板包含了执行其它系统管理功能的 SMC。

5 一些系统管理控制器能够在 BMC 模式或 SMC 模式下操作（即，发挥 BMC 或 SMC 的作用）。在一些以前的系统中，连接到电路板的系统管理控制器可根据该板插入的插槽适配其功能。在此类系统中，系统机箱中的指定插槽可保留用于为该系统执行 BMC 功能的插板，并且可具有为驻留模块提供此类指示的引线。这种情况下，系统管理控制器可在重置时确定其是否在 BMC 插槽中，并且如果在 BMC 插槽中，则可将自身设为充当 BMC（即，将自身设为 BMC 模式）。在此类系统中，组装系统或改换电路板的人员可能需要确定哪个插槽是 BMC 插槽，并确保具有所需 BMC 能力的插板放置在适当的 BMC 插槽中。

发明内容

15 本发明提供一种用于确定基板管理控制器模式的方法，所述方法包括：从计算机系统多个系统管理控制器发送模式请求到所述计算机系统其它系统管理控制器；以及根据不存在对所述系统管理控制器之一发送的模式请求的响应，确定该系统管理控制器为基板管理控制器，其中，所述方法还包括从所述系统管理控制器中的第一控制器发送对从所述系统管理控制器中的第二控制器收到的模式请求的响应，其中，所述第一管理控制器的所述响应至少部分基于所述第一系统管理控制器的协商协议状态，所述模式请求和响应作为符合智能平台管理接口规范的消息发送。

25 附图说明

图 1 是根据本发明实施例，带有用于执行模式协商协议的管理控制器的系统方框图；

图 2 是根据本发明实施例，带有在执行模式协商协议的管理控

制器的系统方框图；

图 3 是根据本发明实施例，确定管理控制器初始操作模式的方法流程图；

图 4 是根据本发明实施例，响应控制器模式请求方法的流程图；

5 图 5 是根据本发明实施例的控制器模式请求和响应方框图；

图 6 是根据本发明实施例，说明管理控制器协商协议的状态和状态转换的状态图。

具体实施方式

10 根据本发明的实施例，系统管理控制器与其它系统管理控制器协商，以确定初始操作模式（例如，在重置或其它初始操作之后的模式）。此类协商例如可通过在系统管理控制器之间发送消息而实现。在一个实施例中，一个系统管理控制器在重置后根据不存在对该系统管理控制器发送的一个或多个控制器模式请求的响应，确定
15 其初始操作模式为中央管理控制器模式（例如，BMC 模式）。在另一实施例中，系统管理控制器的初始模式可基于由该系统管理控制器接收的响应内容。

本发明的实施例提供控制器模式协商协议。在一个实施例中，系统中的每个系统管理控制器用于执行协商协议。协商协议可针对
20 诸如系统启动等事件执行，或者在单个系统管理控制器重置时执行。例如，在系统开机时，系统中的每个系统管理控制器可根据协商协议向其它系统管理控制器发送控制器模式请求，并且可根据对控制器模式请求的响应转换到初始模式。协商协议还可定义系统管理控制器响应其接收的模式请求的协议。在本发明的实施例中，控制器
25 通过一系列协商状态切换状态，这些状态可包括：请求、等待、SMC、备用 BMC 和活动 BMC。在实施例中，协商可至少部分基于诸如控制器能力、用户配置首选项、模块类型和地理（物理）地址等标准。

图 1 是根据本发明实施例，带有适于执行模式协商协议的管理

控制器的系统方框图。图 1 显示的系统 100 可以是任何类型的电子系统，如通用计算机系统，特殊用途的计算机系统等。系统 100 包含四个模块 110、120、130 和 140，它们可以是例如插入系统机箱插槽中的电路板。当然，在其它实施例中，系统可包含更多或更少的模块。每个模块 110、120、130 和 140 可以是电源、风扇架、CPU 板或任何其它类型的组件。系统 100 中的各个控制器可通过输入/输出端口耦合到系统管理总线 150，系统管理总线可以是承载管理信息的任何类型总线。系统管理总线 150 的示例有符合飞利浦半导体公司 (Philips Semiconductor Corporation) 开发的 I²C 总线规范的 Inter-IC 总线(I²C)、符合 SBS 实施者论坛(SBS Implementers Forum)的 SMBus 规范(2000 年 8 月 2.0 版)的系统管理总线(SMBus)或符合智能平台管理总线通信协议规范(Intel 公司等公司 v1.0)的智能平台管理总线(IPMB)。系统管理总线可配置为任何类型的拓扑，如单总线、星形、双总线或混合拓扑。如果使用双总线拓扑，则系统管理控制器可具有第二输入/输出端口，以发送完全相同的一份系统管理消息到其它系统管理控制器。系统管理控制器可使用各种类型的消息格式与其它系统组件进行通信，这些消息格式如智能平台管理接口规范(Intel 公司等公司，v1.5, rev.1, 2001 年 2 月 21 日)(本文称为 IPMI)中定义的格式。

系统 100 中所示的每个模块包含系统管理控制器(113、123、133、143)和计算机可读介质(115、125、135、145)。每个系统管理控制器可以是能够执行如上所述系统管理功能的处理器。每个计算机可读介质可以是能够存储指令的任何类型的介质，如只读存储器(ROM)、可编程只读存储器(PROM)或可擦可编程只读存储器(EPROM)。在一个实施例中，计算机可读介质是非易失性存储器。图 1 中显示的每个计算机可读介质存储了模块协商协议指令(117、127、137、147)。模式协商协议指令可以是例如软件指令、固件指令、微码或任何其它类型的指令，这些指令可由关联的系统管理控

制器执行，以便与其它系统管理控制器协商，从而确定系统管理控制器的初始操作模式。诸如执行系统管理功能的指令等其它指令也可存储在一个或多个计算机可读介质上，并且可由系统管理控制器执行。在其它实施例中，系统管理控制器和模块协商协议指令可实现为 ASIC、可编程逻辑阵列（PLA）或执行系统管理功能的任何其它类型的处理装置。

在本发明实施例操作的一个示例中，系统管理控制器 113 可执行模式协商协议指令 117 以便与系统管理控制器 123、133 和 143 协商，从而确定一个或多个系统管理控制器 113、123、133 和 143 的初始系统管理模式。在一个实施例中，控制器的可能系统管理模式可以为活动 BMC 模式、备用 BMC 模式和 SMC 模式。在此实施例中，活动 BMC 可执行系统的 BMC 功能，而备用 BMC 可在当前活动 BMC 发生故障时变为活动 BMC（例如，可与活动 BMC 一样接收并记录相同的管理信息）。在其它实施例中，可存在更多或更少的可能的管理模式。

在参照图 2 所述的最简单实施例中，除一个系统管理控制器以外，其它所有系统管理控制均已采用了一种操作模式。这种情况下，未初始化的管理控制器可根据协商协议发送模式请求，并且其它管理控制器可根据协商协议作出响应。在另一实施例中，整个系统进行了初始化或重置，并且所有系统管理控制器可在大致相同的时间向彼此发送模式请求。

图 2 是根据本发明实施例，带有在执行模式协商协议的管理控制器的计算机系统 200 方框图。计算机系统 200 显示为包括电路板 210、电路板 220、电源模块 230、风扇架模块 240 及管理总线 250，它们可以与图 1 中的模块 110-140 和管理总线 150 相同。在图 2 所示实施例中，电路板 210 包括基板管理控制器 215，电路板 220 包括备用管理控制器 225，并且电源模块 230 包含卫星管理控制器 235。为便于说明，在此实施例中，管理控制器 215、225 和 235 已经采用了

一种操作模式。例如，BMC 215 可以是系统 200 的活动中央管理控制器，备用 BMC 225 已经准备在 BMC 215 发生故障时变为系统 200 的中央管理控制器，并且 SMC 235 可监视并控制电源模块 230 的操作。

- 5 如系统 200 中所示的那些管理控制器可以在 BMC 模式、备用 BMC 模式或 SMC 模式中的一种、一些或所有模式下操作。例如，BMC 215 可能还能够作为备用 BMC 或 SMC 操作，备用 BMC 225 可能还能够作为 BMC 或 SMC 操作，并且 SMC 235 可能只能够作为 SMC 操作。例如，在其它实施例中，SMC 225 可能能够作为 BMC 操作，
- 10 和/或 BMC 215 可能不能够作为 SMC 操作。

 风扇架模块 240 显示为包括新的系统管理控制器 245。为便于说明，此控制器标记为“新的”以显示一种情况，即一个系统管理控制器在进行初始化，而其它系统管理控制器已经采用了一种操作模式。例如，管理控制器可在整个系统开启或重置时，或者（在图 2

15 所示情况下）在包括管理控制器的特殊模块开启或重置时，进行初始化。例如，图 2 可说明这种情况，其中风扇架模块 240 作为热交换操作的一部分安装到了系统 200 中。

 如图 2 所示，在重置期间或重置之后，新的系统管理控制器 245 可通过系统管理总线 250 向其它系统管理控制器 215、225 和 235 发送控制器模式请求 260。控制器模式请求 260 可以是例如符合 IPMI

20 规范的命令。在此实施例中，活动 BMC 215 通过发送 GoToSMC 响应 261 来响应控制器模式请求；并且备用 BMC 225 通过发送等待响应予以响应。在一个实施例中，SMC 235 不发送对控制器模式请求 260 的响应。在一个实施例中，如下所述，所发送的对模式请求消息的

25 响应可由模式协商协议确定。例如，协议可规定，如果活动 BMC 收到来自具有相等或更低优先级的管理控制器的模式请求，则活动 BMC 将以 GoToSMC 响应作为响应。协议还可规定，如果备用 BMC 从具有更高优先级的管理控制器收到模式请求，则备用 BMC 的响应将为

等待响应。另外，协议可规定，无论相对优先级如何，SMC 不响应模式请求。如下所述，用于确定两个管理控制器相对优先级的因素可包括控制器能力、用户配置的首选项、模块类型和地理地址。

5 协商协议还可定义收到模式响应（或未能收到模式响应）时采取的动作。例如，协议可规定，在收到 GoToSMC 命令后控制器转换到 SMC 状态。又如参照图 3 所述，协议可规定，管理控制器如果发送了模式请求但未在超时期（即，门限时间量）内收到响应，则它转换到主 BMC 状态。

10 图 3 是根据本发明实施例确定管理控制器初始操作模式的方法流程图。此方法例如可在其它系统管理控制器重置或启动时由一个系统管理控制器（如图 2 的新系统管理控制器）执行。例如，新系统管理控制器 245 可将图 3 所示方法作为部分启动例程或其启动例程执行。在图 3（和图 6）所示实施例中，系统管理控制器具有除上述 BMC、备用 BMC 和 SMC 外的多个非操作状态。具体地说，图 3
15 显示了进入请求状态（302）的控制器。在一个实施例中，图 3 所示方法是控制器模式协商协议的一部分，并且该方法例如可由图 1 所示的模式协商协议指令执行。

在图 3 所示实施例中，先对系统管理控制器进行了重置（301）。随后，系统进入请求状态（302）。例如，新系统管理控制器 245 可
20 执行控制器启动例程，该例程以重置控制器开始，然后进入请求状态。在请求状态，控制器发送控制器模式请求消息（如图 2 的控制器模式请求 260）（303）。在一个实施例中，请求状态的唯一特性可以是发送控制器模式请求。接着，控制器可等待模式请求的响应（304）。如果在超时期（例如，100 毫秒）内收到响应，则控制器
25 可进入响应中指定的状态（305）。例如，新系统管理控制器 245 可从系统中的另一管理控制器接收 GoToSMC 响应。在采用指定状态后，除执行所采用的操作模式功能外，系统管理控制器随后可处理其它控制器的请求（309）。

如果在超时期（例如，100 毫秒）内未收到响应，则控制器可确定是否达到了重试限制（307）。如果尚未达到重试限制，则如上所述，控制器可转换回到请求状态，可发送另一控制器模式请求，以及可等待。在一个实施例中，重试限制为三次重试。当然，可使用
5 其它超时期和重试限制。如果达到了重试限制，则控制器可将自身设为活动 BMC 模式（308）。在采用 BMC 模式后，除执行 BMC 功能外，控制器随后可处理其它控制器的请求（309）。因此，在此实施例中，如果控制器未收到对控制器模式请求的响应，则它可采用 BMC 模式。优先级可基于任何不同的因素，例如，如下参照图 5 所述的那些因素。在其它实施例中，控制器可能没有重置状态。另外，
10 如上所述，状态的采用可不需要控制器采取任何肯定动作。

在上述示例中，除一个控制器外的所有其它控制器事先采用了一种操作模式。然而，图 3 所示的方法也适用于其它情况，如在整个系统重置且所有控制器在大致相同的时间初始化时。根据实施例，
15 在所有控制器初始化时，它们可在大致相同的时间各自执行如图 3 所示的方法。因此，系统中的多个管理控制器各自可向其它管理控制器发送模式请求。这些控制器中的每个控制器随后可对收到的每个模式请求作出响应（或不响应），例如，通过发送响应消息到发送请求的控制器。在一个实施例中，根据不存在对控制器发送的模式请求的响应，确定控制器之一最初是系统的中央管理控制器。
20

在又一实施例中，如果控制器发送的请求数量超过极限值而未在超时期内收到响应，则不存在模式请求的响应。

图 4 是根据本发明实施例响应控制器模式请求的方法流程图。此方法例如可由系统管理控制器在收到控制器模式请求（如图 2 所示的控制器模式请求 260 或图 5 所示的控制器模式请求 510）时执行。
25 在一个实施例中，图 4 所示方法是控制器模式协商协议的一部分，并且该方法可通过例如图 1 所示的协商协议指令执行。该方法可在某个控制器发送模式请求之前或之后或者同时由该控制器执行（如

图3方法所示)。

根据图4所示实施例,第一系统管理控制器(可称为接收器)从第二系统管理控制器(可称为请求者)接收控制器模式请求(401)。接收器随后可决定其当前协商协议状态(402)。例如,接收器可确定其当前处于SMC模式。接收器随后可确定模式协商协议是否规定了在这种情况下响应应基于相对优先级(403),并且倘若如此,则可确定请求者与接收器的相对优先级(404)。随后,接收器可确定模式协商协议是否规定了这种情况下应发送响应(405),并且倘若如此,则可将响应发送到第二系统管理控制器(406)。响应可指定第二系统管理控制器要进入的状态。如果不要发送响应,则接收器随后可确定模式协商协议是否规定了这种情况下接收器应转换到等待状态(407),并且倘若如此,则接收器自己切换为等待状态(408)。优先级可如下参照图5所述那样确定,并且图6和下表1中显示了针对不同情况的模式协商协议实施例示例。

因此,根据本发明的实施例,发送回控制器模式请求发送方的响应可至少部分基于接收器的当前状态。响应可至少部分基于接收器的控制器模式能力,并可至少部分基于用户配置的模式首选项。图3和图4中所示的方法可作为控制器启动过程的一部分执行。虽然按所示顺序讨论图3和图4所示的方法步骤,但在其它实施例中,可按不同顺序执行一些步骤。例如,系统管理控制器可发送控制器模式请求,并且在大致相同的时间响应收到的一个或多个请求。

图5是根据本发明实施例的控制器模式请求510和响应520的方框图。控制器模式请求510可以是例如图2的控制器模式请求260,并且响应520例如可以是图2的GoToSMC响应261或等待响应262。在一个实施例中,控制器模式请求510和响应520是符合IPMI规范的消息,在这种情况下,它们可称为命令。在一个实施例中,控制器模式请求510和响应520可使用IPMI规范中定义的组扩展网络函数(如网络函数=2Ch/2Dh)。如上所述,控制器模式请求510和响

应 520 可用于模式协商。在一个实施例中，控制器可在处于请求状态时广播控制器模式请求 510，并且可在发送控制器模式请求后等待接收响应。

5 如图 5 所示，控制器模式请求 510 包括首部 511、能力 512、用户首选项 513、模块类型 514 和地理地址 515 字段。在一个实施例中，控制器模式请求 510 中的信息可用于确定相对控制器优先级。在一个实施例中，能力集是第一级优先检查。在又一实施例中，如果能力集相同，则将用户配置首选项用作第二级优先检查。在又一实施例中，如果能力集和用户首选项均相同，则模块类型用作第三级优先检查。在又一实施例中，在其它标准相同时，使用地理地址。当然，控制器模式请求 510 和响应 520 是唯一的可能格式示例。可为
10 这些字段分配确定相对控制器优先级的其它优先级顺序，并且可使用其它字段。

15 在一个实施例中，能力字段 512 可指示发送控制器模式请求 510 的控制器系统管理模式能力。在一个实施例中，可用的能力集为纯 BMC (BMC-only)、BMC/SMC 和纯 SMC (SMC-only)。在又一实施例中，纯 BMC 具有最高优先级，并且纯 SMC 具有最低优先级。在一个实施例中，可以为纯 BMC 的唯一模块是专用于机箱中央管理代理的模块，它可称为“机箱管理模块” (CMM)，设计用于
20 星形或混合拓扑。在一个实施例中，具有 BMC/SMC 能力集的控制器 (即，可充当 BMC 或 SMC 的控制器) 可选择实施用户配置特性，以使用户指定首选项 BMC、SMC，或无首选项。例如，用户可使用 BIOS 设置选项、软件设置、DIP 开关、跳线设置或运行或加载软件输入此类首选项。此信息可包括在控制器模式请求 510 的用户首选项
25 项字段 513 中。在一个实施例中，包括纯 BMC 和纯 SMC 模块在内不实施用户配置首选项特性的模块可报告无首选项。在一个实施例中，纯 BMC 是最高优先级，无中间优先级首选项，并且纯 SMC 是最低优先级。由于不同的模块类型可具有不同的地理地址域，因此，

在实施例中，模块类型可用于确定优先级。在一个实施例中，用于模块类型字段 514 的按从低到高优先级顺序的不同可用值为电源模块、其它机箱特定类型、风扇架、节点板、开关板和专用 CMM。当然，可使用其它模块类型和其它顺序或优先级。

5 对于控制器为其组成部分的模块，地理地址字段 515 可包含其地理地址（例如，插槽地址）。在一个实施例中，其它标准比较的结果等同时，带有较低地理地址的控制器确定为具有更高的优先级。在又一实施例中，处于 BMC 状态的控制器还可使用地理地址来确定如何响应。例如，BMC 可使用地理地址来确定哪些模块在初始上电后应

10 为活动状态。

 如上所述，接收控制器模式请求的控制器可根据其当前状态和请求者相对于自己的优先级作出响应。图 5 显示了包含首部字段 521 和数据字段 523 的响应 520。在一个实施例中，首部 520 包含完成码。在一个实施例中，数据字段 521 可包含等待响应（该响应指示响应接收器应转到等待状态）或 GoToSMC 响应（该响应指示响应接收器应转到 SMC 状态）。在又一实施例中，数据字段还可包含诸如备用 BMC 等其它响应。当然，在其它实施例中，消息的类型和格式可不同于图 5 所示的类型。例如，其它消息也可由模式协商协议使用。

15

 图 6 是根据本发明实施例，说明管理控制器协商协议的状态和状态转换的状态图。根据一个实施例，在执行协商协议的系统管理控制器可以处于如下图 6 所示的协商状态之一：请求 620、等待 630、SMC 640、备用 SMC 650 和活动 BMC 660。在此实施例中，处于 SMC 状态的控制器可以处于 SMC 模式，处于备用 BMC 状态的控制器可以处于备用 BMC 模式，并且处于活动 BMC 状态的控制器可以处于活动 BMC 模式。在一个实施例中，一些控制器将支持所有这 5 种状态，而其它控制器可能只支持其中一部分状态。纯 SMC 状态可能是只可在纯 SMC 模式下操作的控制器唯一可用的状态；纯 BMC 状态可能是只可在纯 BMC 模式下操作的控制器唯一可用的状态，并且所

20

25

有 5 种状态均可用于能够在 SMC 或 BMC 模式下操作的控制器。在图 6 中，重置 610 可指示控制器已经历了重置条件（即，控制器重置或启动的条件）。

在一个实施例中，控制器从重置中恢复时，进入请求状态 620。

- 5 在请求状态中，控制器可广播控制器模式请求并等待响应。接收控制器模式请求的其它控制器可根据其当前状态和相对于请求者的优先级作出响应。根据实施例，协商协议支持优先分级，因此，不能够充当 SMC 的模块与能够充当 SMC 的模块相比，可优先用作 BMC。BMC 优先级可基于能力、首选项设置、模块类型和地理地址。在一个实施例中，如果未收到控制器模式请求的响应（重试后），则请求者可
- 10 可将自身设为活动 BMC 模式。否则，请求者会经任一 GoToSMC 响应得知要运行的模式。在一个实施例中，控制器还可接收未经请求的消息，该消息不是为响应某个控制器模式请求而发送的，其请求控制器采用一定模式。此类未经请求的消息可称为设置模式命令。
- 15 在一个实施例中，设置模式命令由 BMC 在系统操作期间发送，以便在采用初始模式后改变控制器模式。

- 下面将更详细地描述根据本发明实施例的各种状态转换。在处于请求状态的控制器广播控制器模式请求后，它可接收一个或多个响应，如 GoToSMC 响应(622)或等待响应(621)。如果收到 GoToSMC
- 20 响应，则控制器可转换为 SMC 状态(640)。如果收到等待响应，则控制器可转换为等待状态(630)。如果在超时和重试后仍未收到响应，则控制器可转换为活动 BMC 状态(624)。另外，处于请求状态的控制器可接收一个或多个设置模式命令，这些命令可指示控制器进入备用 BMC 模式(623)或者可指示控制器进入 SMC 模式
- 25 (622)。

在所示实施例中，处于等待状态 630 的控制器可等待接收设置模式命令或 GoToSMC 响应。如果收到 GoToSMC 响应，则控制器转换为 SMC 状态(632)。如果收到设置模式命令，则控制器可转换

为设置模式命令中指定的适当状态(例如,转换为备用 BMC 状态 633 或转换为 SMC 状态 632)。在此实施例中,如果在超时期内未收到 GoToSMC 响应或设置模式命令,则控制器可转换回请求状态(631),在该状态它可重新广播控制器模式请求。

5 处于 SMC 状态的控制器可充当卫星管理控制器。如图 6 所示,如果处于 SMC 状态的控制器收到设置模式命令,则控制器可转换为设置模式命令中指定的适当状态(例如,转换为备用 SMC 状态 641)。

 处于备用 BMC 状态的控制器可充当备用 BMC。如上所述,在一个实施例中,处于备用 BMC 状态的控制器可保持与活动 BMC 的同步状态信息,并可为活动 BMC 执行看门狗功能。在又一实施例中,如果活动 BMC 发生故障,则备用 BMC 将转换为活动 BMC 状态(652)。视管理拓朴和安装的模块而定,可在活动 BMC 发生故障时选择新的备用 BMC。如图 6 所示,如果处于备用 BMC 状态的控制器收到设置模式命令,则控制器可转换为设置模式命令中指定的适当状态(例如,转换为 SMC 状态 651 或转换为活动 BMC 状态 652)。

 在活动 BMC 状态中,控制器可执行正常的 BMC 功能。在一个实施例中,活动 BMC 可选择适合于拓朴的备用 BMC,并且可与备用 BMC 同步状态信息。在上述实施例中,BMC 最后负责通知其它协商控制器进入 SMC 状态。例如,在一个实施例中,在使用双总线拓朴时,只有在建立备用 BMC 后 BMC 才会通知控制器进入 SMC 模式。在此实施例中,不允许控制器在建立备用 BMC 前进入 SMC 模式。如果活动 BMC 在建立备用 BMC 前将要发生故障并且所有其它控制器已进入 SMC 状态,则系统可不设 BMC。在一个实施例中,为星形或混合拓朴专门设计的 CMM 可在建立备用 BMC 前通知其它非 CMM 模块进入 SMC 状态,这是因为只有另一星形或混合 CMM 才可作为备用 BMC。在一个实施例中,如果热交换使用了比备用 BMC 优先级更高的控制器,或者由于其它原因,则活动 BMC 可在收到备用设置模式命令时转换为备用 BMC 状态(662),例如,这可在用

户触发了从备用 BMC 切换到活动 BMC 模式（这可称为“故障转移（failover）”）时发生。

5 根据一个实施例，接收控制器模式请求的控制器（即接收器）可如下表 1 所示响应请求者（即发送请求的控制器）。此表显示了 15 种不同的情况。如下面所示，响应可基于接收器的状态和请求者的相对优先级。在一些情况下，响应要取决于请求者的相对地理优先级，而在一些情况下，响应要取决于是否已经建立了备用 BMC。在表 1 中，标记“X”表示，对于此情况，响应内容不基于此标准。例如，如上参照图 5 所述，相对的控制器的优先级可依据能力、用户首选项、模块类型和地理地址来确定。

10

情况	接收器状态	请求者的优先级 w/o GA	请求者的 GA 优先级	设立备用	响应
1	活动 BMC	更高	X	X	等待
2	活动 BMC	等于或更低	X	是	等待 或 GoToSMC
3	活动 BMC	等于或更低	X	否	等待（或 GoToSMC） ^[1]
4	备用 BMC	更高	X	是	等待
5	备用 BMC	等同	X	是	等待 或 GoToSMC
6	备用 BMC	更低	X	是	GoToSMC
7	SMC	X	X	X	无响应
8	等待	更高	X	X	无响应
9	等待	等同	更高	X	无响应
10	等待	等同	更低	X	等待
11	等待	更低	X	X	等待

12	请求	更高	X	X	无响应并将自身设为等待
13	请求	等同	更高	X	无响应并将自身设为等待
14	请求	等同	更低	X	等待
15	请求	更低	X	X	等待

表 1

在表 1 的前三种情况中，控制器模式请求的接收器为活动 BMC。如果请求者的优先级（无地理地址）高于接收器的优先级，则可发送等待响应。请求者可具有比活动 BMC 更高优先级的情形例如：请求者以热交换方式换入的情形，或者请求者花费较长时间从休息状态恢复的情形。具有比活动 BMC 更高优先级的请求者可进入等待状态而不是直接成为活动 BMC，这样，它可在切换为活动 BMC 前实现同步。如果请求者的优先级（无地理地址）等于或低于接收器的优先级，并且已建立备用 BMC，则可发送等待响应或 GoToSMC 响应。如果请求者要成为新的备用 BMC，则可发出等待响应。如果尚未建立备用 BMC，则在一个实施例中，如果活动 BMC 是专门为星形或混合拓扑设计的 CMM，而请求者不是 CMM，则接收器可只发出 GoToSMC 响应；否则，接收器可发出等待响应。

在表 1 的第 4 到第 6 种情况中，接收器处于备用 BMC 状态（根据定义这表示备用 BMC 已建立）。在所示实施例中，如果请求者的优先级（无地理地址）高于接收器的优先级，则可发送等待响应；如果请求者的优先级（无地理地址）低于接收器的优先级，则可发送 GoToSMC 响应。通常，如果请求者的优先级（无地理地址）等于接收器的优先级，则将发送 GoToSMC 响应；但在这种情况下，例如在决定要改变备用 BMC 时，可发送等待响应。

在表 1 中其余情况 7-15 中，响应不取决于是否建立了备用。在情况 7 中，接收器处于 SMC 状态，并且无论相对优先级如何均不发

送响应。因此，在此实施例中，处于 SMC 状态的控制器不响应控制器模式请求。在情况 8-11 中，接收器处于等待状态，并且地理优先级用于打破平衡。在这些情况中，如果请求者具有更高的优先级，则不发送响应；如果请求者具有更低的优先级，则发送等待响应。

5 最后，在情况 12-15 中，接收器处于请求状态，并且地理优先级用于打破平衡。在这些情况中，如果请求者具有更高的优先级，则不发送响应并且接收器将自身设为等待状态；如果接收器处于请求状态并且请求者具有更低的优先级，则发送等待响应。

10 表 1 只表示根据本发明实施例的协商协议的一个实施例。在其它实施例中，可能有其它接收器状态，并且在一种或多种情况下的响应可能不同。

在以上公开的实施例中，不接收针对控制器模式请求和重试的任何响应（并且不将自身设为等待状态）的控制器可将自身设为活动 BMC 状态。在所示实施例中，已建立备用 BMC 以防止控制器在可以建立备用 BMC 前进入 SMC 状态时，只可由 BMC 通知控制器进入 SMC 模式。如果活动 BMC 在建立备用 BMC 前将要发生故障，并且所有其它控制器已进入 SMC 状态，则系统可不建立 BMC。使用本发明实施例中公开的模式协商协议可自动确定哪些控制器将作为活动和备用 BMC，同时避免控制器之间的冲突。

20 本文具体示意和/或描述了本发明实施例的几个示例。然而，应理解，在不脱离本发明精神和预期范围的情况下对本发明所作的修改和变化由以上教导所涵盖，并且在所附权利要求书范围内。例如，用于响应请求的优先级确定和协议可与如上所示有所不同。又如，系统管理协商协议可用硬件或软件来实现。

25

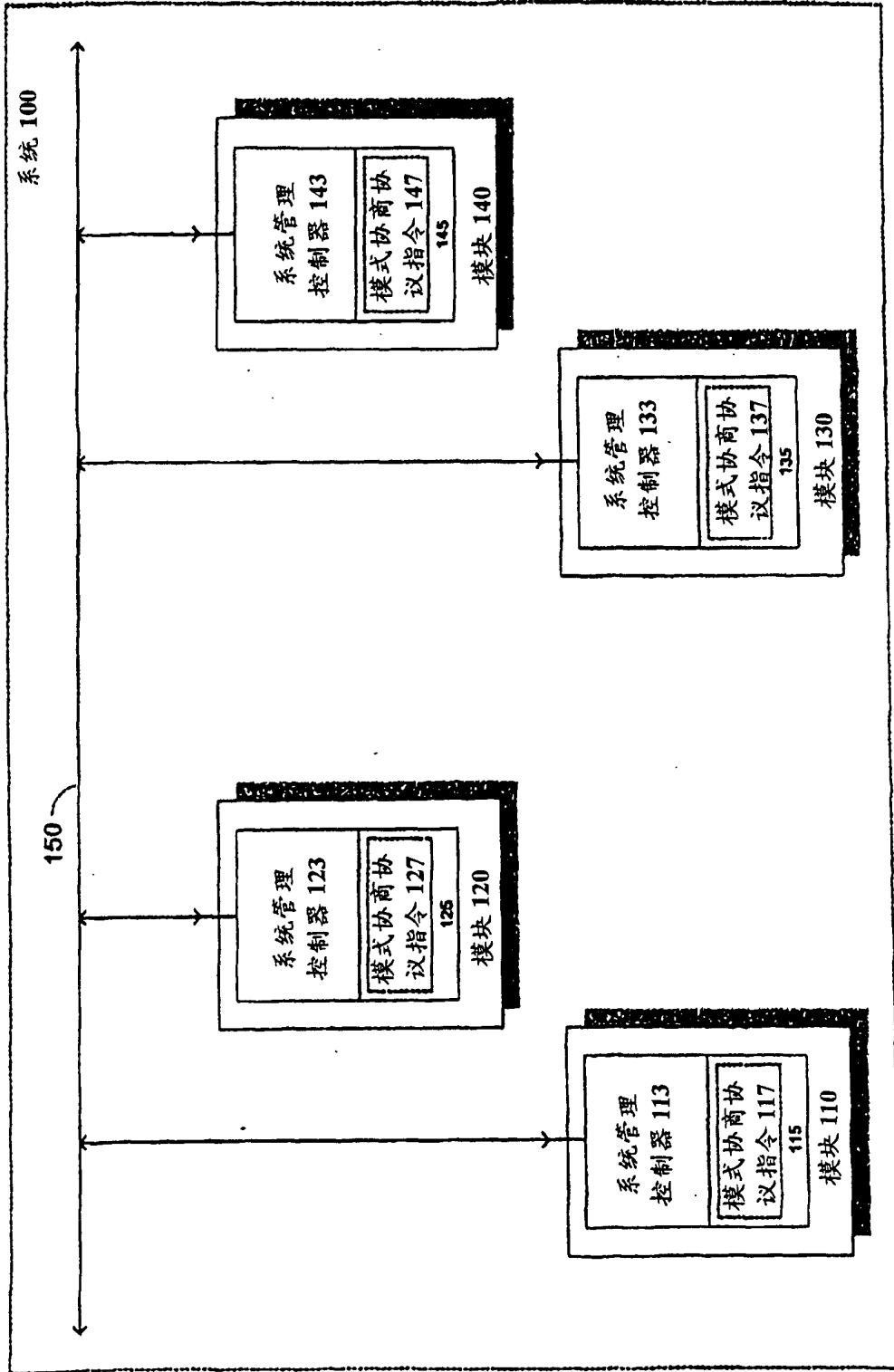


图 1

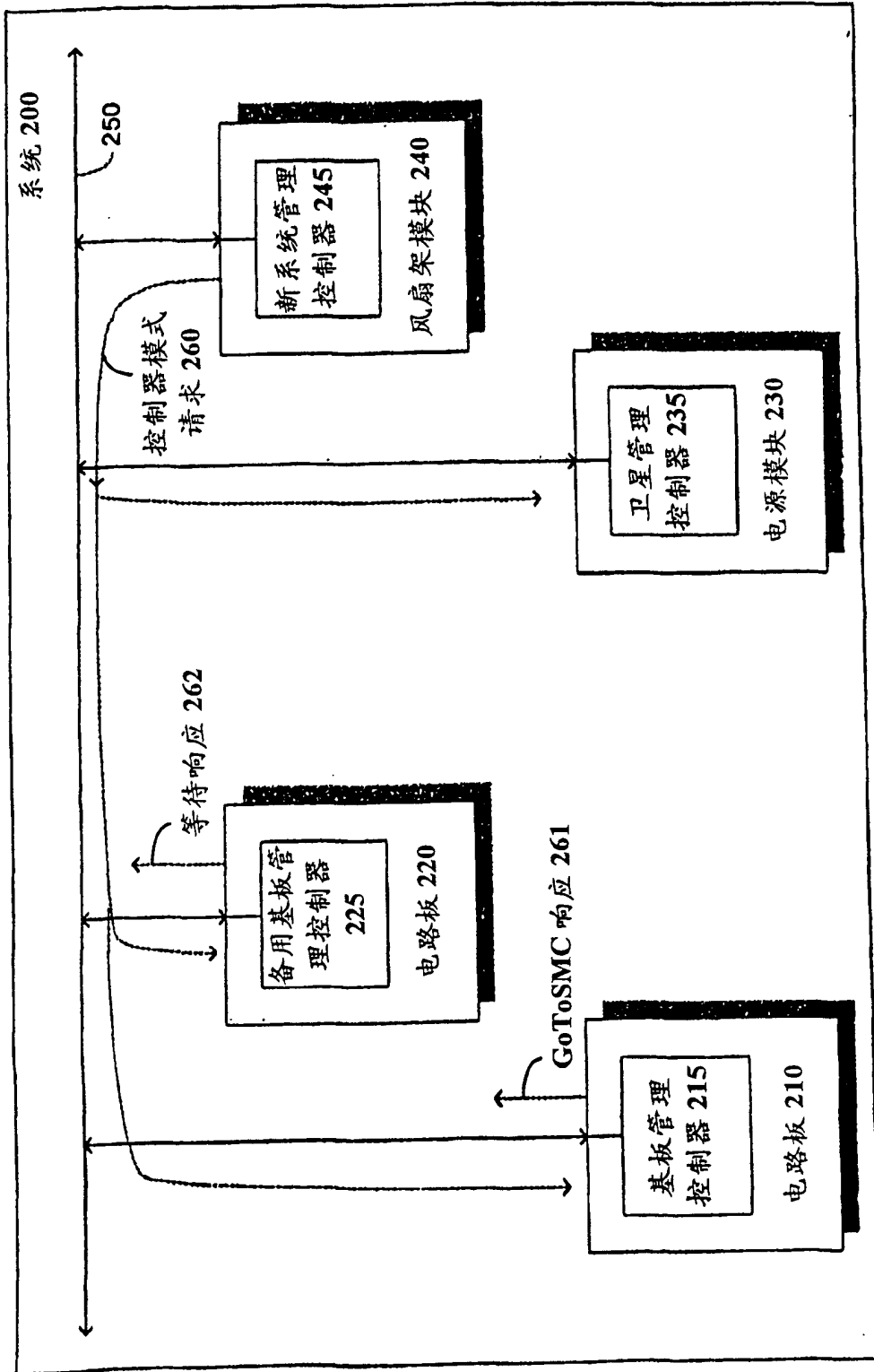


图 2

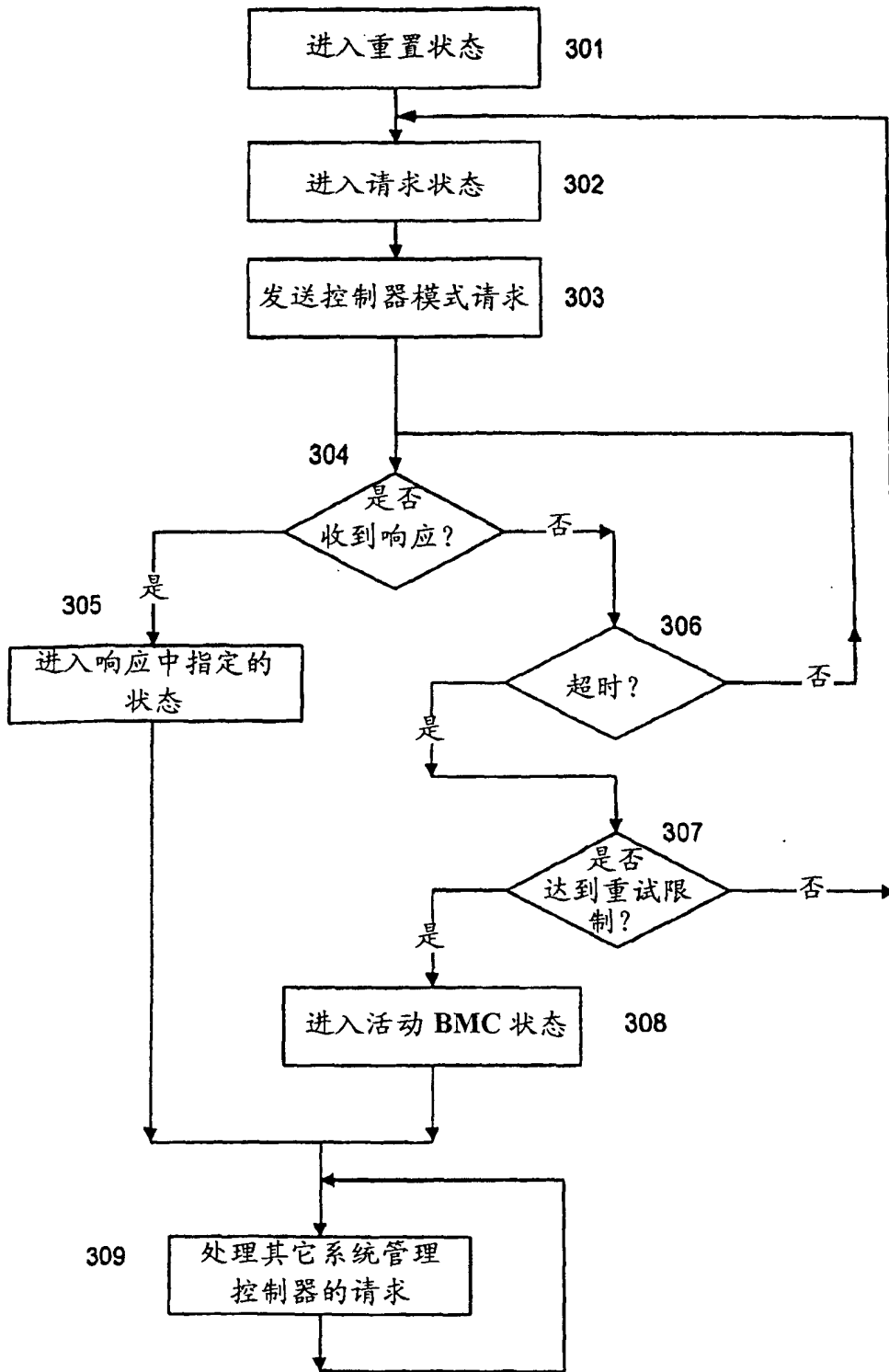


图 3

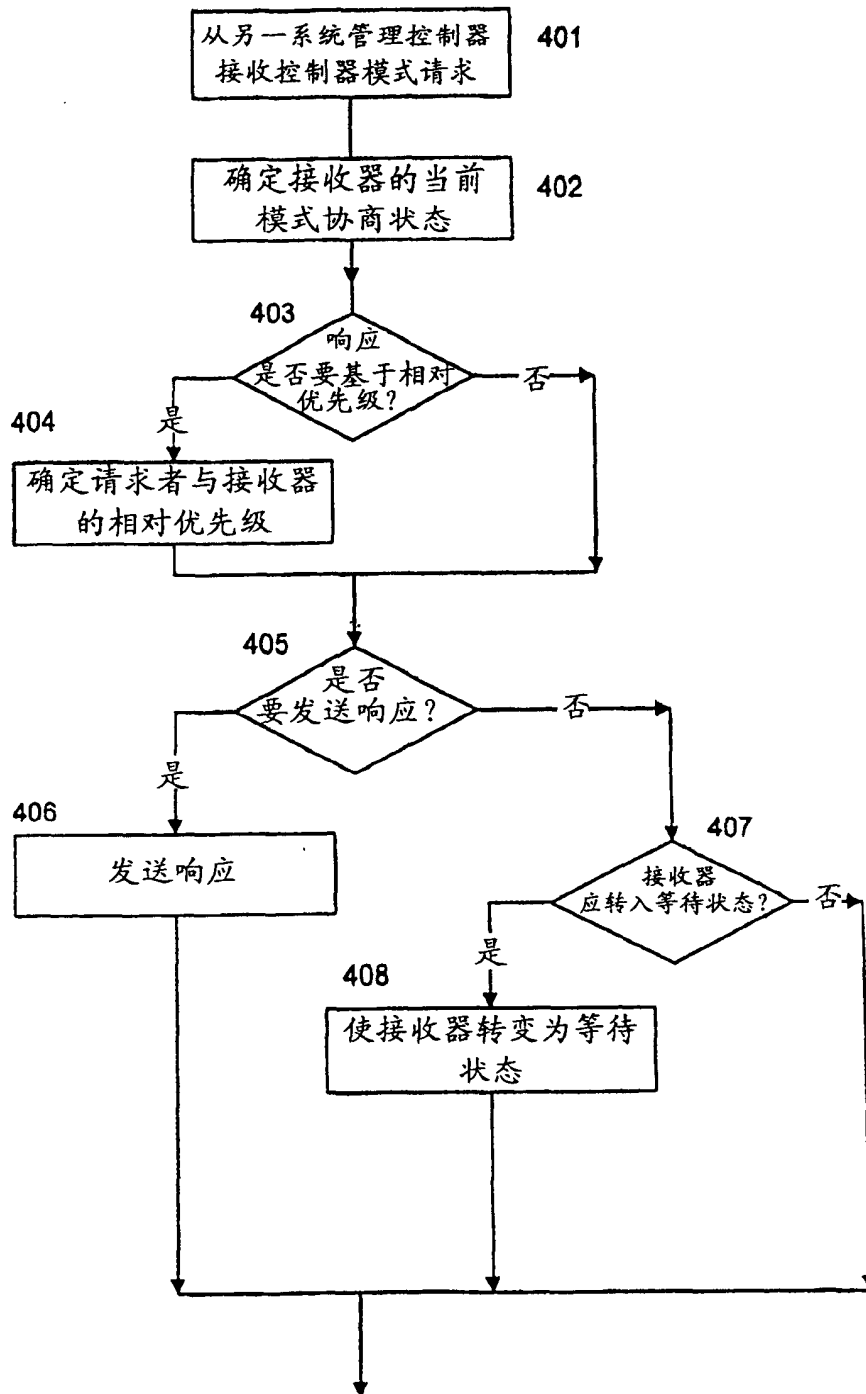


图 4

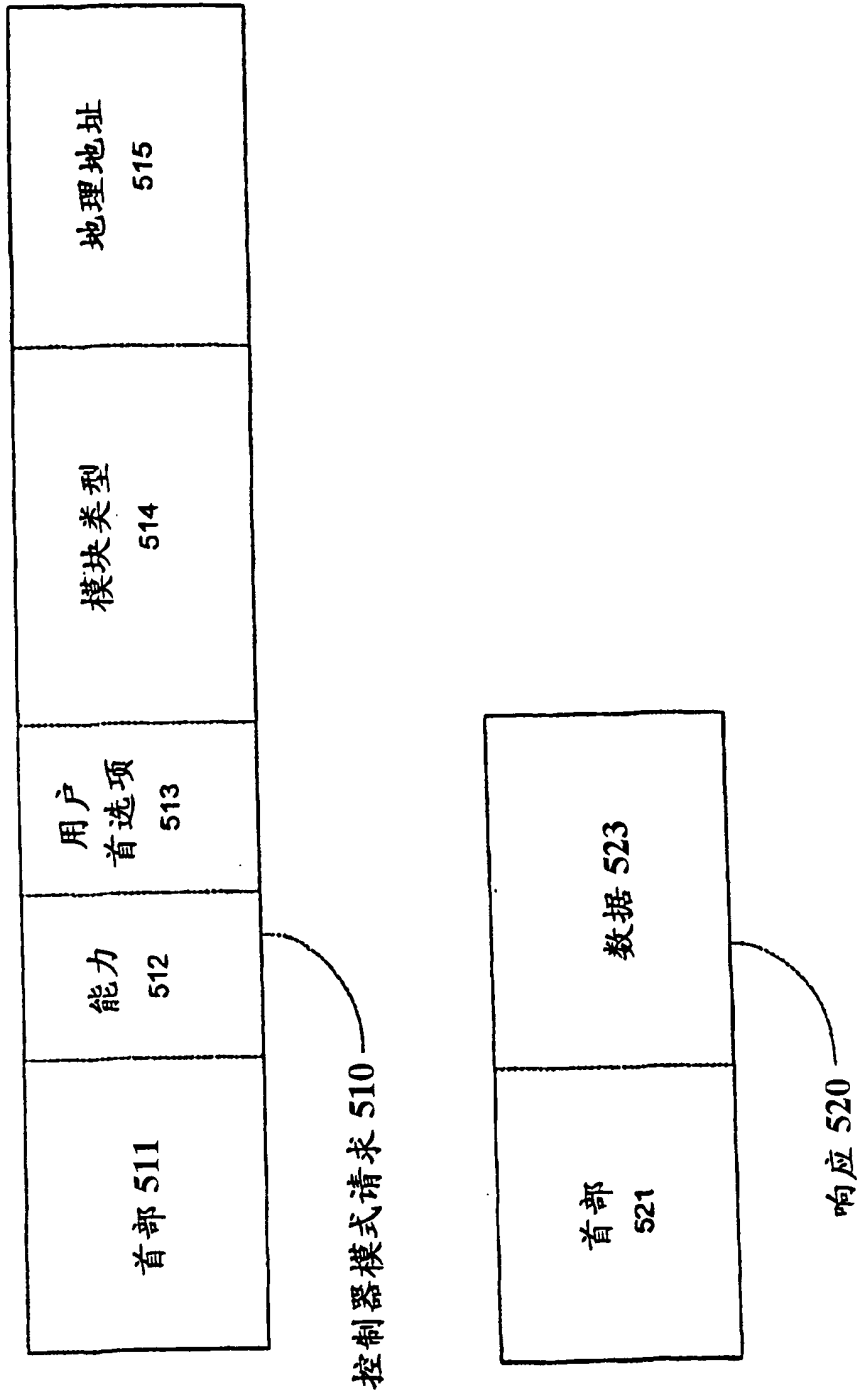


图 5

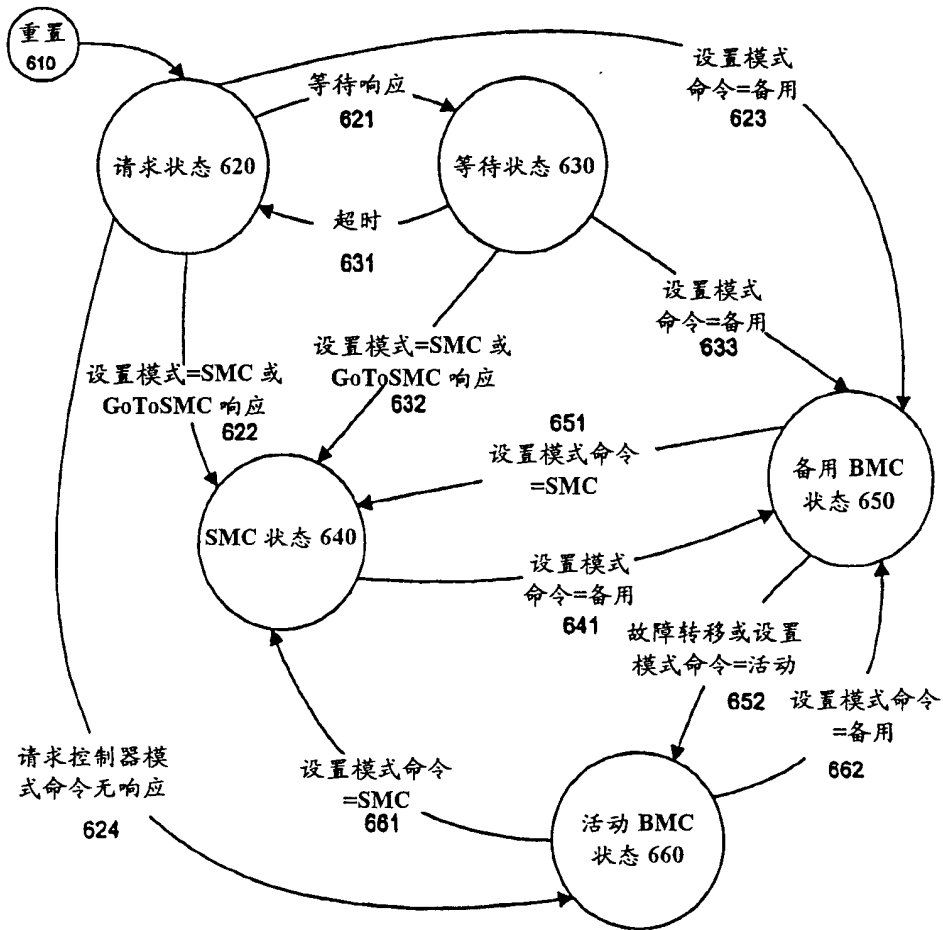


图 6