(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103324495 A
(43)申请公布日 2013.09.25

(21)申请号 201210078662.1

(22)申请日 2012.03.23

(71)申请人 鸿富锦精密工业(深圳)有限公司
地址 518109 广东省深圳市宝安区龙华镇油松第十工业区东环二路 2 号
申请人 鸿海精密工业股份有限公司

(72)发明人 黄嘉庆

(51)Int.Cl.
G06F 9/445(2006.01)

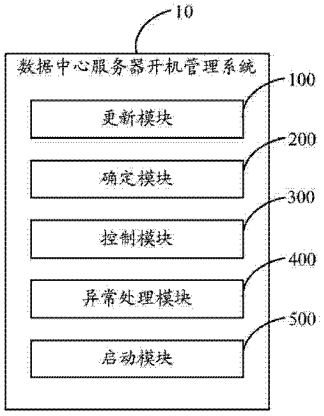
权利要求书2页 说明书5页 附图3页

(54)发明名称

数据中心服务器开机管理方法及系统

(57)摘要

一种数据中心服务器开机管理系统，该系统用于：从候选 BMC 中确定一个主 BMC；控制主 BMC 按照设定的启动顺序每隔预定时间依次向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备，并发送当前从 BMC 的信息给各候选 BMC；当主 BMC 出现故障时，从剩余候选 BMC 中重新确定新的主 BMC，并控制所述新的主 BMC 继续按照设定的启动顺序每隔预定时间依次发送指令至剩余的从 BMC，启动剩余从 BMC 对应的电源设备；当所有从 BMC 对应的电源设备均已启动后，依次启动当前的主 BMC 和所有候选 BMC 所对应的电源设备。本发明还提供一种数据中心服务器开机管理方法。本发明可以智能管理数据中心中所有服务器的开机顺序，且不受主 BMC 故障的影响。



CN 103324495 A

1. 一种数据中心服务器开机管理方法，其特征在于，该方法包括：

更新步骤：当候选基板管理控制器 BMC 接收到从 BMC 发送的数据包后，更新每个候选 BMC 中的开机管理链表；

确定步骤：从所述候选 BMC 中确定一个主 BMC；

控制步骤：控制主 BMC 按照设定的启动顺序每隔预定时间依次向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备，并发送当前从 BMC 的信息给各候选 BMC，该当前从 BMC 为当前正在主 BMC 的指令下启动对应电源设备的从 BMC；

异常处理步骤：当候选 BMC 在预先设定的等待时间内没有接收到主 BMC 发送的当前从 BMC 的信息时，从剩余候选 BMC 中重新确定新的主 BMC，并控制所述新的主 BMC 继续按照设定的启动顺序每隔预定时间依次发送指令至剩余的从 BMC，启动剩余从 BMC 对应的电源设备；及

启动步骤：当所有从 BMC 对应的电源设备均已启动后，依次启动当前的主 BMC 和所有候选 BMC 所对应的电源设备。

2. 如权利要求 1 所述的数据中心服务器开机管理方法，其特征在于，所述开机管理链表中包含多个节点，每个节点记录一个从 BMC 的信息，所述从 BMC 的信息包括从 BMC 的 IP 地址及历史平均开机功率，并设置一个指针 Index，指向该开机管理链表的当前节点，该当前节点用于记录当前正在主 BMC 的指令下启动对应电源设备的从 BMC 的信息。

3. 如权利要求 1 或 2 所述的数据中心服务器开机管理方法，其特征在于，所述设定的启动顺序是指主 BMC 按照各从 BMC 的历史平均开机功率从大到小或者从小到大的顺序来向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备。

4. 如权利要求 1 所述的数据中心服务器开机管理方法，其特征在于，每个候选 BMC 中创建有一个主 BMC 管理链表，所述主 BMC 管理链表中包含多个节点，每个节点记录一个候选 BMC 的信息，所述候选 BMC 的信息包括候选 BMC 的 IP 地址及预设的 ID，并设置一个指针 Master，指向该主 BMC 管理链表的当前节点，该当前节点用于记录当前的主 BMC 的信息；

在所述确定步骤及异常处理步骤中，按照主 BMC 管理链表中的 ID 号最小的原则确定主 BMC。

5. 如权利要求 4 所述的数据中心服务器开机管理方法，其特征在于，在所述启动步骤中，当前的主 BMC 和所有候选 BMC 按照主 BMC 管理链表中的 ID 号从小到大的顺序每隔预定时间依次启动对应的电源设备。

6. 一种数据中心服务器开机管理系统，其特征在于，该系统包括：

更新模块，用于当候选基板管理控制器 BMC 接收到从 BMC 发送的数据包后，更新每个候选 BMC 中的开机管理链表；

确定模块，用于从所述候选 BMC 中确定一个主 BMC；

控制模块，用于控制主 BMC 按照设定的启动顺序每隔预定时间依次向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备，并发送当前从 BMC 的信息给各候选 BMC，该当前从 BMC 为当前正在主 BMC 的指令下启动对应电源设备的从 BMC；

异常处理模块，用于当候选 BMC 在预先设定的等待时间内没有接收到主 BMC 发送的当前从 BMC 的信息时，从剩余候选 BMC 中重新确定新的主 BMC，并控制所述新的主 BMC 继续按照设定的启动顺序每隔预定时间依次发送指令至剩余的从 BMC，启动剩余从 BMC 对应的电

源设备；及

启动模块，用于当所有从 BMC 对应的电源设备均已启动后，依次启动当前的主 BMC 和所有候选 BMC 所对应的电源设备。

7. 如权利要求 6 所述的数据中心服务器开机管理系统，其特征在于，所述开机管理链表中包含多个节点，每个节点记录一个从 BMC 的信息，所述从 BMC 的信息包括从 BMC 的 IP 地址及历史平均开机功率，并设置一个指针 Index，指向该开机管理链表的当前节点，该当前节点用于记录当前正在主 BMC 的指令下启动对应电源设备的从 BMC 的信息。

8. 如权利要求 6 或 7 所述的数据中心服务器开机管理系统，其特征在于，所述设定的启动顺序是指主 BMC 按照各从 BMC 的历史平均开机功率从大到小或者从小到大的顺序来向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备。

9. 如权利要求 6 所述的数据中心服务器开机管理系统，其特征在于，每个候选 BMC 中创建有一个主 BMC 管理链表，所述主 BMC 管理链表中包含多个节点，每个节点记录一个候选 BMC 的信息，所述候选 BMC 的信息包括候选 BMC 的 IP 地址及预设的 ID，并设置一个指针 Master，指向该主 BMC 管理链表的当前节点，该当前节点用于记录当前的主 BMC 的信息；

所述确定模块及异常处理模块按照主 BMC 管理链表中的 ID 号最小的原则确定主 BMC。

10. 如权利要求 9 所述的数据中心服务器开机管理系统，其特征在于，当前的主 BMC 和所有候选 BMC 按照主 BMC 管理链表中的 ID 号从小到大的顺序每隔预定时间依次启动对应的电源设备。

数据中心服务器开机管理方法及系统

技术领域

[0001] 本发明涉及一种开机管理方法及系统，尤其是涉及一种数据中心服务器开机管理方法及系统。

背景技术

[0002] 数据中心（Data Center）通常包括几台乃至上万台服务器，为了减少电力负载，避免所有服务器同时开机，需要设定一个开机的先后顺序。目前业界普遍的做法是在BIOS（Basic Input Output System，基本输入输出系统）或BMC（Baseboard Management Controller，基板管理控制器）韧件中，给每一台服务器设定一个固定或随机时间T，服务器在延迟该时间T后会自动开机。这样就需要在每台服务器进行设置，过程繁琐，且容易出错，还存在随机时间T冲突的问题，即随机时间T相同，导致同时开机的情况出现。另外，如果采取一个主BMC控制所有的从BMC开机的策略，则当主BMC出现故障时，会导致剩余所有从BMC所在服务器无法开机。

发明内容

[0003] 鉴于以上内容，有必要提供一种数据中心服务器开机管理方法，可以智能管理数据中心中所有服务器的开机顺序，且不受主BMC故障的影响。

[0004] 鉴于以上内容，还有必要提供一种数据中心服务器开机管理系统，可以智能管理数据中心中所有服务器的开机顺序，且不受主BMC故障的影响。

[0005] 所述数据中心服务器开机管理方法包括：更新步骤：当候选基板管理控制器BMC接收到从BMC发送的数据包后，更新每个候选BMC中的开机管理链表；确定步骤：从所述候选BMC中确定一个主BMC；控制步骤：控制主BMC按照设定的启动顺序每隔预定时间依次向每个从BMC发送指令，启动该从BMC对应的电源设备，并发送当前从BMC的信息给各候选BMC，该当前从BMC为当前正在主BMC的指令下启动对应电源设备的从BMC；异常处理步骤：当候选BMC在预先设定的等待时间内没有接收到主BMC发送的当前从BMC的信息时，从剩余候选BMC中重新确定新的主BMC，并控制所述新的主BMC继续按照设定的启动顺序每隔预定时间依次发送指令至剩余的从BMC，启动剩余从BMC对应的电源设备；及启动步骤：当所有从BMC对应的电源设备均已启动后，依次启动当前的主BMC和所有候选BMC所对应的电源设备。

[0006] 所述数据中心服务器开机管理系统包括：更新模块，用于当候选基板管理控制器BMC接收到从BMC发送的数据包后，更新每个候选BMC中的开机管理链表；确定模块，用于从所述候选BMC中确定一个主BMC；控制模块，用于控制主BMC按照设定的启动顺序每隔预定时间依次向每个从BMC发送指令，启动该从BMC对应的电源设备，并发送当前从BMC的信息给各候选BMC，该当前从BMC为当前正在主BMC的指令下启动对应电源设备的从BMC；异常处理模块，用于当候选BMC在预先设定的等待时间内没有接收到主BMC发送的当前从BMC的信息时，从剩余候选BMC中重新确定新的主BMC，并控制所述新的主BMC继续按照设定的

启动顺序每隔预定时间依次发送指令至剩余的从 BMC，启动剩余从 BMC 对应的电源设备；及启动模块，用于当所有从 BMC 对应的电源设备均已启动后，依次启动当前的主 BMC 和所有候选 BMC 所对应的电源设备。

[0007] 相较于现有技术，所述的数据中心服务器开机管理方法及系统，可以通过主 BMC 每隔预定时间依次发送指令至每个从 BMC，启动该从 BMC 对应的电源设备，从而控制该从 BMC 所在的服务器开机。并且在该主 BMC 出现故障时，按照预定策略从剩余候选 BMC 中重新确定新的主 BMC，控制所述新的主 BMC 继续发送指令至剩余的从 BMC，启动剩余的从 BMC 对应的电源设备，确保数据中心中的所有服务器可以正常开机。

附图说明

[0008] 图 1 是本发明数据中心服务器开机管理系统较佳实施例的应用环境图。

[0009] 图 2 是本发明数据中心服务器开机管理系统较佳实施例的功能模块图。

[0010] 图 3 是本发明所用开机管理链表的示意图。

[0011] 图 4 是本发明所用主 BMC 管理链表的示意图。

[0012] 图 5 是本发明数据中心服务器开机管理方法较佳实施例的流程图。

[0013] 主要元件符号说明

控制电脑	1
数据中心	2
数据中心服务器开机管理系统	10
存储器	11
处理器	12
服务器	20
BMC	21
电源设备	22
更新模块	100
确定模块	200
控制模块	300
异常处理模块	400
启动模块	500

如下具体实施方式将结合上述附图进一步说明本发明。

具体实施方式

[0014] 参阅图 1 所示，是本发明数据中心服务器开机管理系统较佳实施例的应用环境图。所述数据中心服务器开机管理系统(以下简称为开机管理系统) 10 运行于控制电脑 1 中，所述控制电脑 1 通过网络与数据中心 2 连接。所述控制电脑 1 还包括通过数据总线相连的存储器 11 和处理器 12。所述数据中心 2 中包括多个服务器 20 (图中以四个为例)，每个服务器 20 中包括 BMC 21 及电源设备 22。可以理解，所述控制电脑 1 还应该包括其他必要的硬件系统与软件系统，如主板、操作系统等，由于这些设备都是本领域技术人员的习知常识，本实施例中不再一一描述。

[0015] 所述存储器 11 用于存储所述开机管理系统 10 的程序代码等资料。所述处理器 12 用于执行所述开机管理系统 10 的各功能模块，以完成本发明。

[0016] 其中，BMC 21 用于读取电源设备 22 上的信息(例如，电源设备 22 的功率、电压及电流等信息)，并可以控制该电源设备 22 在指定的时间启动。需要说明的是，BMC 21 不需

要电源设备 22 供电，服务器 20 通过电线接通外界的电源（图中未示出），BMC 21 就会启动。而启动电源设备 22 目的在于启动服务器 20 中的操作系统，使得该服务器 20 能够运行。

[0017] 所述 BMC 21 被分成多个候选 BMC 及从 BMC，开机管理系统 10 从所有候选 BMC 中选取一个作为主 BMC，该主 BMC 向所有从 BMC 发送相关指令，从 BMC 根据该指令启动对应的电源设备 22，从而控制该从 BMC 所在的服务器 20 开机。

[0018] 参阅图 2 所示，是本发明数据中心服务器开机管理系统较佳实施例的功能模块图。

[0019] 所述开机管理系统 10 包括更新模块 100、确定模块 200、控制模块 300、异常处理模块 400 及启动模块 500。

[0020] 所述更新模块 100 用于当候选 BMC 接收到从 BMC 发送的数据包后，更新每个候选 BMC 中的开机管理链表。

[0021] 每个从 BMC 在每次启动对应的电源设备 22，使该从 BMC 所在的服务器 20 开机后，都会动态记录历史平均开机功率 P_i，并向所有候选 BMC 发送包括历史平均开机功率 P_i 的数据包。

[0022] 每个候选 BMC 中管理一个开机管理链表（参阅图 3 所示），所述开机管理链表中包含多个节点，每个节点记录一个从 BMC 的信息，所述从 BMC 的信息包括从 BMC 的 IP 地址及历史平均开机功率 P_i。并设置一个指针 Index，指向该开机管理链表的当前节点，该当前节点用于记录当前正在主 BMC 的指令下启动对应电源设备 22 的从 BMC（以下简称为当前从 BMC）的信息。

[0023] 在本实施例中，当外部电源上电后，所有 BMC 21 启动，然后所有从 BMC 会发送包含历史平均开机功率 P_i 的数据包至所有候选 BMC，所有的候选 BMC 在接收到该数据包后更新开机管理链表。在本实施例中，开机管理链表中各节点按历史平均开机功率 P_i 由大到小依次排列，若 P_i 大小相等则按接收数据包的时间先后排列。

[0024] 所述确定模块 200 用于从所述候选 BMC 中确定一个主 BMC。在本实施例中，为了管理所有的候选 BMC，在每个候选 BMC 中创建一个主 BMC 管理链表（参阅图 4 所示）。所述主 BMC 管理链表中包含多个节点，每个节点记录一个候选 BMC（包括之后确定的主 BMC）的信息，所述候选 BMC 的信息包括候选 BMC 的 IP 地址及预设的 ID（比如 0 到 n）。并设置一个指针 Master，指向该主 BMC 管理链表的当前节点，该当前节点用于记录当前的主 BMC 的信息。在本实施例中，按照主 BMC 管理链表中的 ID 号最小的原则确定主 BMC，即初始时确定 ID 号为 0 的候选 BMC 为主 BMC。

[0025] 所述控制模块 300 用于控制主 BMC 按照设定的启动顺序每隔预定时间 T 依次向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备 22。所述主 BMC 还同时将开机管理链表中的指针 Index 移向记录当前从 BMC 的信息的节点，并发送当前从 BMC 的信息给各候选 BMC，候选 BMC 也将开机管理链表中的指针 Index 移向记录该当前从 BMC 的信息的节点。

[0026] 在本实施例中，所述设定的启动顺序是指主 BMC 按照各从 BMC 的历史平均开机功率 P_i 从大到小或者从小到大的顺序来向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备 22。在其他实施例中，还可以按照其他顺序启动电源设备 22，例如，按照各从 BMC 的编号大小来启动电源设备 22。

[0027] 所述异常处理模块 400 用于当候选 BMC 在预先设定的等待时间内没有接收到主

BMC 发送的当前从 BMC 的信息时，则判定此时主 BMC 出现故障不能工作，从剩余候选 BMC 中重新确定新的主 BMC，并控制所述新的主 BMC 继续按照设定的启动顺序每隔预定时间 T 依次发送指令至剩余的从 BMC，启动该剩余的从 BMC 对应的电源设备 22。当确定新的主 BMC 后，所有候选 BMC 的主 BMC 管理链表中的指针 Master 移向记录新的主 BMC 的信息的节点。

[0028] 在本实施例中，按照主 BMC 管理链表中的 ID 号最小的原则确定新的主 BMC。例如，之前的主 BMC 的 ID 号为 0，当该主 BMC 出现故障后，重新确定 ID 号为 1 的候选 BMC 为新的主 BMC。这样的情况下，即使主 BMC 出现故障，仍然可以由剩余的候选 BMC 顶替，确保了数据中心 2 中的所有服务器 20 可以正常开机。

[0029] 在本实施例中，设定所述等待时间为 3T（即上述预定时间 T 的三倍）。值得注意的是，在其他实施例中，新的主 BMC 也可以按照设定的启动顺序重新发送指令至所有从 BMC，重新控制所有从 BMC 依次启动对应的电源设备 22。

[0030] 所述启动模块 500 用于当所有从 BMC 对应的电源设备 22 均已启动后，即所有从 BMC 所在的服务器 20 均已开机后，依次启动当前的主 BMC 和所有候选 BMC 所对应的电源设备 22。在本实施例中，当前的主 BMC 和所有候选 BMC 按照主 BMC 管理链表中的 ID 号从小到大的顺序每隔预定时间 T 依次启动对应的电源设备 22。

[0031] 参阅图 5 所示，是本发明数据中心服务器开机管理方法较佳实施例的流程图。

[0032] 步骤 S10，当候选 BMC 接收到从 BMC 发送的数据包后，所述更新模块 100 更新每个候选 BMC 中的开机管理链表。

[0033] 步骤 S12，所述确定模块 200 从所述候选 BMC 中确定一个主 BMC。在本实施例中，按照主 BMC 管理链表中的 ID 号最小的原则确定主 BMC。

[0034] 步骤 S14，所述控制模块 300 控制主 BMC 按照设定的启动顺序每隔预定时间 T 依次向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备 22。在本实施例中，所述设定的启动顺序是指主 BMC 按照各从 BMC 的历史平均开机功率 P_i 从大到小或者从小到大的顺序来向每个从 BMC 发送指令，启动该从 BMC 对应的电源设备 22。

[0035] 步骤 S16，当候选 BMC 在预先设定的等待时间内没有接收到主 BMC 发送的当前从 BMC 的信息时，所述异常处理模块 400 判定此时主 BMC 出现故障不能工作，从剩余候选 BMC 中重新确定新的主 BMC，并控制所述新的主 BMC 继续按照设定的启动顺序每隔预定时间 T 依次发送指令至剩余的从 BMC，启动该剩余的从 BMC 对应的电源设备 22。在本实施例中，按照主 BMC 管理链表中的 ID 号最小的原则确定新的主 BMC，设定所述等待时间为 3T。

[0036] 步骤 S18，当所有从 BMC 对应的电源设备 22 均已启动后，即所有从 BMC 所在的服务器 20 均已开机后，所述启动模块 500 依次启动当前的主 BMC 和所有候选 BMC 所对应的电源设备 22。在本实施例中，当前的主 BMC 和所有候选 BMC 按照主 BMC 管理链表中的 ID 号从小到大的顺序每隔预定时间 T 依次启动对应的电源设备 22。

[0037] 综上所述，使用本发明数据中心服务器开机管理方法及系统，可以通过主 BMC 每隔预定时间 T 依次发送指令至每个从 BMC，启动该从 BMC 对应的电源设备 22，从而控制该从 BMC 所在的服务器 20 开机。并且在该主 BMC 出现故障时，按照预定策略从剩余候选 BMC 中重新确定新的主 BMC，控制所述新的主 BMC 继续发送指令至剩余的从 BMC，启动剩余的从 BMC 对应的电源设备 22，确保数据中心 2 中的所有服务器 20 可以正常开机。

[0038] 以上实施例仅用以说明本发明的技术方案而非限制，尽管参照较佳实施例对本发

明进行了详细说明，本领域的普通技术人员应当理解，可以对本发明的技术方案进行修改或等同替换，而不脱离本发明技术方案的精神和范围。

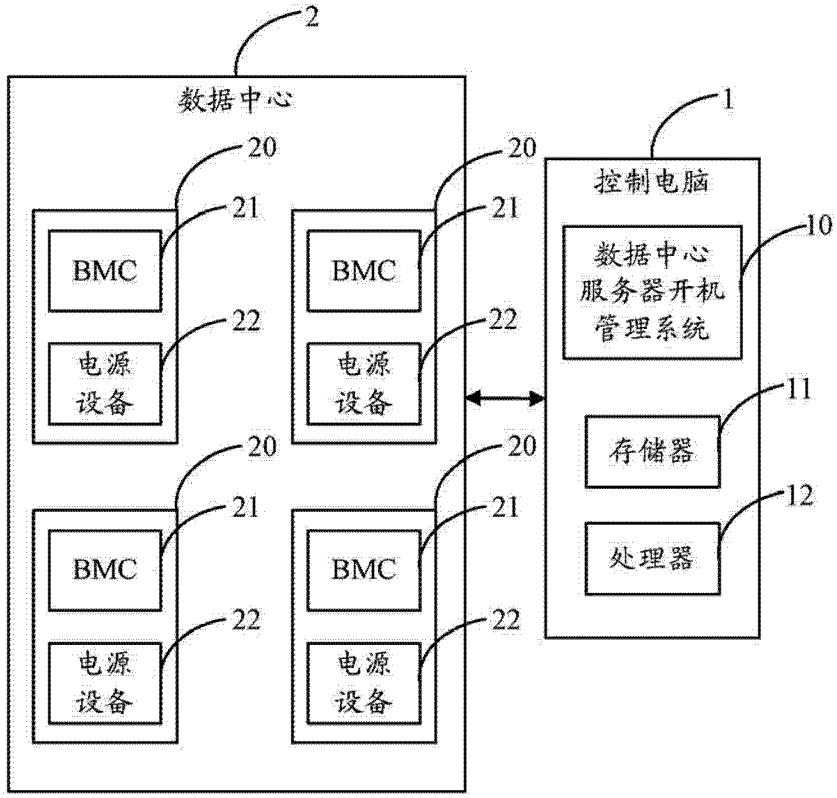


图 1

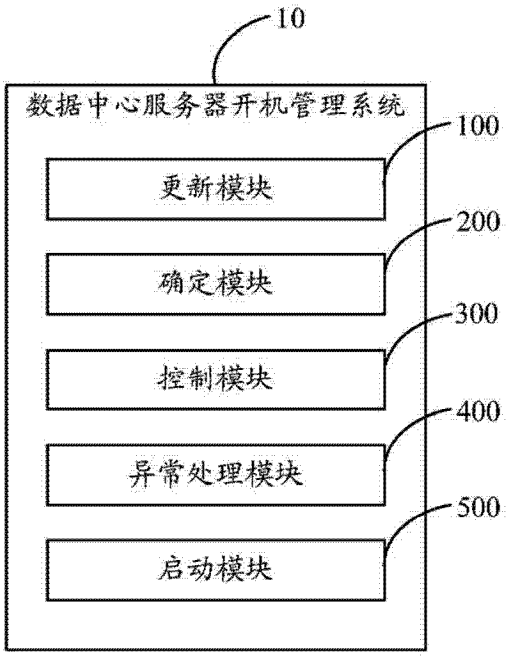


图 2

IP	历史平均开机功率
IP_0	P_0
IP_1	P_1
...	...
IP_n	P_n

Index

图 3

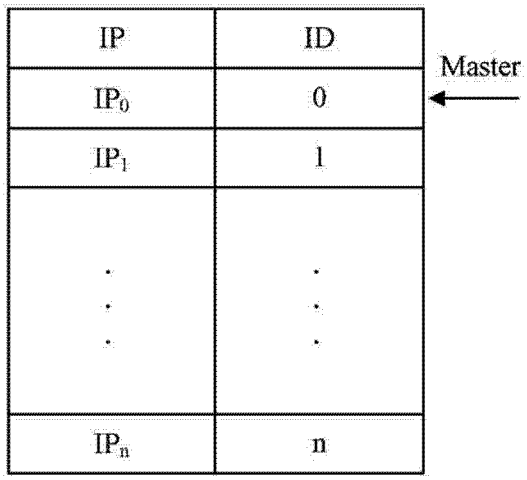


图 4

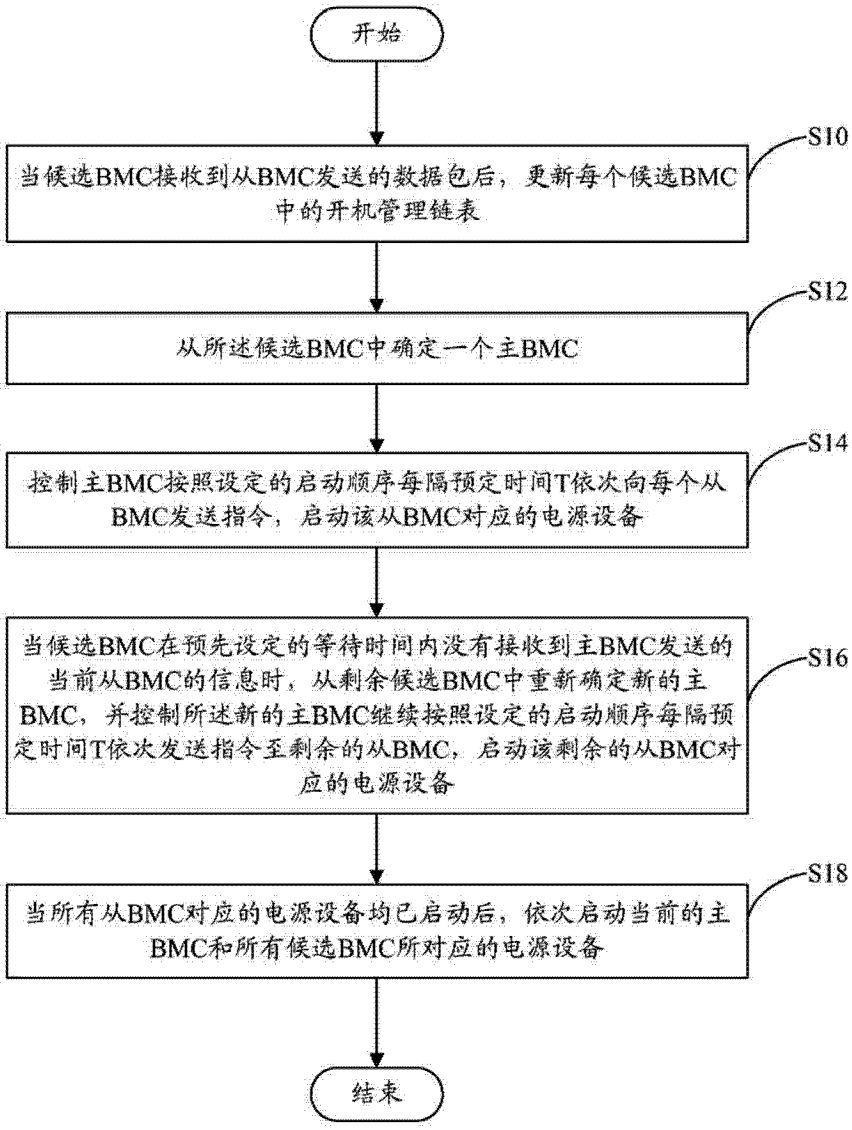

开始
当候选BMC接收到从BMC发送的数据包后，更新每个候选BMC中的开机管理链表
S10
从所述候选BMC中确定一个主BMC
S12
控制主BMC按照设定的启动顺序每隔预定时间T依次向每个从BMC发送指令，启动该从BMC对应的电源设备
S14
当候选BMC在预先设定的等待时间内没有接收到主BMC发送的当前从BMC的信息时，从剩余候选BMC中重新确定新的主BMC，并控制所述新的主BMC继续按照设定的启动顺序每隔预定时间T依次发送指令至剩余的从BMC，启动该剩余的从BMC对应的电源设备
S16
当所有从BMC对应的电源设备均已启动后，依次启动当前的主BMC和所有候选BMC所对应的电源设备
S18
结束

图 5