



(12)发明专利

(10)授权公告号 CN 103716175 B

(45)授权公告日 2018.09.28

(21)申请号 201310265220.2

(22)申请日 2013.06.28

(65)同一申请的已公布的文献号

申请公布号 CN 103716175 A

(43)申请公布日 2014.04.09

(30)优先权数据

13/629,864 2012.09.28 US

(73)专利权人 阿瓦亚公司

地址 美国新泽西州

(72)发明人 N·奥康纳 D·伦纳德

(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 宋海宁

(51)Int.Cl.

H04L 12/24(2006.01)

H04L 29/06(2006.01)

(56)对比文件

CN 101517555 A, 2009.08.26,

US 2011173441 A1, 2011.07.14,

审查员 王志豪

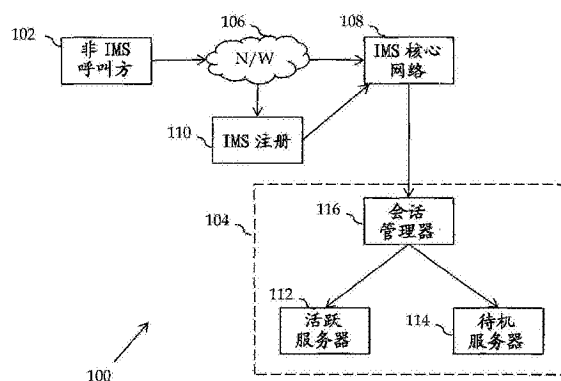
权利要求书4页 说明书11页 附图7页

(54)发明名称

用于确保企业IMS网络中高可用性的系统和
方法

(57)摘要

本发明通过高度可用的应用服务器架构,使在IMS网络的应用层部署的应用能够频繁地执行关键业务功能。这可以通过使用会话管理器实现,会话管理器管理高度可用的应用架构的活跃服务器和(一个或多个)待机服务器。此外,会话管理器通过以串行或并行方式将SIP消息路由到应用服务器架构来管理活跃服务器和(一个或多个)待机服务器的角色。此外,活跃服务器和待机服务器通过它们在会话管理器所确定的信令路径中的位置自动保持同步。会话管理器确保如果活跃服务器发生故障,待机服务器必须能够以无缝方式接管活跃服务器的角色。



1. 一种高可用性企业IMS系统,包括:

会话管理器,该会话管理器对被管理的服务器的活跃和待机角色进行管理以定义活跃应用服务器和待机应用服务器,其中所述活跃应用服务器的堆栈单元和所述待机应用服务器的堆栈单元独立地工作;其中:

所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元处理与和所述会话管理器的通信相关联的事件;

所述会话管理器通过将数据排序到活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元两者来管理所述活跃应用服务器和待机应用服务器的同步;并且

所述会话管理器在检测到活跃应用服务器不可用时指示待机应用服务器接管所述活跃应用服务器的角色。

2. 根据权利要求1所述的系统,其中所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元各自包括SIP堆栈单元;其中:

所述活跃应用服务器和待机应用服务器各自还包括业务逻辑单元;

所述活跃应用服务器的SIP堆栈单元处于活跃状态;

所述待机应用服务器的SIP堆栈单元处于活跃状态;

所述活跃应用服务器的业务逻辑单元处于产生事件的活跃状态;

所述待机应用服务器的业务逻辑单元处于非活跃状态;并且

所述会话管理器处理从活跃的业务逻辑单元接收的事件。

3. 根据权利要求1所述的系统,其中所述会话管理器以并行方式将SIP消息路由到所述活跃应用服务器和待机应用服务器,其中所述待机应用服务器和活跃应用服务器两者均配置为处理接收的SIP事件。

4. 根据权利要求1所述的系统,其中所述会话管理器以串行方式将SIP消息路由到所述活跃应用服务器和待机应用服务器,其中所述待机应用服务器和活跃应用服务器两者均配置为处理接收的SIP事件。

5. 根据权利要求1所述的系统,其中所述会话管理器在活跃应用服务器和待机应用服务器之间没有直接轮询的情况下控制所述活跃应用服务器和待机应用服务器的状态,使得活跃服务器和待机服务器彼此的状态独立。

6. 根据权利要求1所述的系统,其中所述会话管理器还配置为基于所述活跃应用服务器中的故障判定,将所述活跃应用服务器移出活跃角色。

7. 根据权利要求1所述的系统,其中所述会话管理器通过使用健康检查机制来监视所述活跃应用服务器和待机应用服务器的状态。

8. 根据权利要求1所述的系统,其中所述会话管理器配置为确保排序的事件同时到达所述活跃应用服务器和待机应用服务器。

9. 根据权利要求1所述的系统,所述活跃应用服务器预先向会话管理器传送失效转移请求,以便在活跃应用服务器变为不可用之前引起待机应用服务器向活跃角色的转换。

10. 根据权利要求1所述的系统,其中所述会话管理器通过检测活跃应用服务器对请求缺乏响应来确定所述活跃应用服务器不可用。

11. 根据权利要求1所述的系统,其中:

所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元各自包括SIP堆栈单

元;

所述活跃应用服务器和待机应用服务器各自还包括业务逻辑单元;

所述活跃应用服务器的SIP堆栈单元处于活跃状态;

所述待机应用服务器的SIP堆栈单元处于活跃状态;

所述活跃应用服务器的业务逻辑单元处于产生事件的活跃状态;

所述待机应用服务器的业务逻辑单元处于非活跃状态;并且

所述会话管理器通过向待机应用服务器发送SIP消息以指示所述待机应用服务器激活它的业务逻辑单元来触发失效转移。

12. 根据权利要求1所述的系统,其中所述会话管理器通过在原子网络事务的上下文中将数据排序到活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元两者来管理所述活跃应用服务器和待机应用服务器的同步,由此使得待机应用服务器能够在任何时间戳接管活跃角色。

13. 根据权利要求1所述的系统,其中:

所述会话管理器通过将邀请序列分叉到活跃应用服务器和待机应用服务器两者来管理所述活跃应用服务器和待机应用服务器的同步;

所述会话管理器从活跃应用服务器和待机应用服务器两者接收回序列;并且

所述会话管理器向下一hop发送活跃应用服务器序列,但不发送待机应用服务器序列。

14. 根据权利要求13所述的系统,其中:

所述会话管理器还在向下一hop发送活跃服务器序列之前进行等待,直到从活跃应用服务器和待机应用服务器两者接收到会话。

15. 根据权利要求1所述的系统,其中:

所述会话管理器执行序列到活跃应用服务器和待机应用服务器的并行分叉;

所述会话管理器在从活跃服务器接收到邀请阶段完成消息并且没有从待机服务器接收到邀请阶段完成消息时继续进行下一hop,由此策略性地企业IMS内分叉和终止序列对话以使得待机服务器遮蔽活跃应用服务器并保持与活跃应用服务器完全同步。

16. 根据权利要求1所述的系统,其中处理器还执行存储在存储器中的代码以实现显示SIP实体配置输入的图形用户接口,该SIP实体配置输入具有控制对于排序是期望串行模式还是并行模式的字段。

17. 根据权利要求1所述的系统,其中:

所述处理器还执行存储在存储器中的代码以实现图形用户接口,该图形用户接口允许用户选择确定活跃应用服务器和待机应用服务器是否以代理服务器模式或者B2B模式运行的控制,使得:

在代理服务器模式下,与B2B模式相比实现较快的传播,但是不具有建立的排序对话的状态和控制;并且

在B2B模式下,与代理服务器模式相比实现较慢的传播,但是具有建立的排序对话的状态和控制;以及

由此便于选择用速度来换取期望的功能恢复水平。

18. 一种在企业IMS网络中提供高可用性服务的计算机实现的方法,包括:

通过以下处理来由处理器对活跃应用服务器和待机应用服务器执行排序和同步,其中

所述活跃应用服务器的堆栈单元和所述待机应用服务器的堆栈单元独立地工作,并且所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元处理与和会话管理器的通信相关联的事件:

通过处理器将SIP消息路由到所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元两者,使得所述活跃应用服务器和待机应用服务器被同步;

通过处理器基于所述活跃应用服务器的状态确定所述活跃应用服务器和待机应用服务器的活跃和待机角色;以及

通过处理器在检测到活跃应用服务器不可用时指示待机应用服务器接管所述活跃应用服务器的角色。

19. 根据权利要求18所述的方法,其中所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元各自包括SIP堆栈单元,所述活跃应用服务器和待机应用服务器各自还包括业务逻辑单元,所述方法进一步包括:

通过处理器指定所述活跃应用服务器的SIP堆栈单元处于活跃状态;

通过处理器指定所述待机应用服务器的SIP堆栈单元处于活跃状态;

通过处理器指定所述活跃应用服务器的业务逻辑单元处于产生事件的活跃状态;

通过处理器指定所述待机应用服务器的业务逻辑单元处于非活跃状态;

所述方法进一步包括基于所述活跃应用服务器中有故障的判定并且通过将所述待机应用服务器的业务逻辑单元改变为活跃状态来将所述活跃应用服务器移出活跃角色。

20. 根据权利要求18所述的方法,其中:

通过处理器将SIP消息路由到所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元两者包括:

通过处理器以串行方式将SIP消息路由到所述活跃应用服务器和待机应用服务器,其中所述待机应用服务器和活跃应用服务器两者均配置为处理接收的SIP事件。

21. 根据权利要求18所述的方法,其中:

通过处理器将SIP消息路由到所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元两者包括:

通过处理器以并行方式将SIP消息路由到所述活跃应用服务器和待机应用服务器,其中所述待机应用服务器和活跃应用服务器两者均配置为处理接收的SIP事件。

22. 根据权利要求18所述的方法,还包括:

通过处理器在活跃应用服务器和待机应用服务器之间没有直接轮询的情况下控制所述活跃应用服务器和待机应用服务器的状态,使得活跃服务器和待机服务器彼此的状态独立。

23. 根据权利要求18所述的方法,还包括:

由处理器通过使用健康检查机制来监视所述活跃应用服务器和待机应用服务器的状态。

24. 根据权利要求18所述的方法,其中:

通过处理器在检测到活跃应用服务器不可用时指示待机应用服务器接管所述活跃应用服务器的角色包括:

通过处理器从活跃应用服务器接收预先传送的失效转移请求,并且指示待机应用服务

器在所述活跃应用服务器变为不可用之前接管该活跃应用服务器的角色;或者

通过检测活跃应用服务器对请求缺乏响应来确定所述活跃应用服务器不可用,并然后指示待机应用服务器接管该活跃应用服务器的角色。

25. 根据权利要求18所述的方法,其中:

通过处理器将SIP消息路由到所述活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元两者包括:

在原子网络事务的上下文中将数据排序到活跃应用服务器的堆栈单元和待机应用服务器的堆栈单元两者,由此使得待机应用服务器能够在任何时间戳接管活跃角色。

用于确保企业IMS网络中高可用性的系统和方法

技术领域

[0001] 本发明总体上涉及通信网络。更具体地,本发明涉及企业IMS网络中高可用性的维护。

背景技术

[0002] 一般来说,在IMS网络的应用层部署的应用经常执行关键功能。因此,对于这些应用,高可用性和零服务中断是重要的考虑。常规地,很多企业IMS应用通过由活跃的服务器使用频带外(out-of-band)机制向其指定的待机服务器发送信息,来实现高可用性。常常,专业应用的供应商被迫使用这种复杂的方法,因为他们在IMS会话管理领域没有影响或产品能力。

[0003] 另一种实现维护高可用性和零服务中断的目标的方法可以通过冗余,即,具有能够在一个或多个服务器完全发生故障时能够接管的多个备用服务器(backup server)。实现该目标的困难在于(一个或多个)冗余部件不能正常地且自然地用作待机服务器、同时仍被包含在信令路径中,即,冗余部件不能正常地且自然地接收与原部件接收的信息相同的信息。另外,冗余部件的维护很昂贵和困难。

[0004] 因此,需要改进常规技术来提供应用服务器的高可用性,以更容易和便宜的方式实现高度可用的服务器。此外,需要改善应用服务器的容错能力可靠性的技术,以确保在发生故障时的零停机时间。

发明内容

[0005] 根据本发明的实施例提供包括了活跃应用服务器的高可用性企业IMS系统。活跃应用服务器可以向其指定客户提供服务。此外,该IMS系统包括待机应用服务器,该待机应用服务器通过与活跃应用服务器完全同步的状态、并且能够被用作活跃应用服务器来确保高可用性。此外,该IMS系统包括会话管理器,该会话管理器通过向应用服务器通知其活跃/待机角色以及通过至少以串行方式将SIP消息路由到应用服务器,以对活跃和待机应用服务器进行排序、管理和同步。此处使用的会话管理器可以指现有技术中已知的、具有可选的高可用性(HA)部件的IP多媒体子系统(IMS)服务呼叫状态控制功能((S-CSCF))。

[0006] 在本发明另一实施例中,提供高可用性的企业IMS系统。该IMS系统包括活跃应用服务器。该活跃应用服务器可以向其指定顾客提供服务。此外,该IMS系统包括待机应用服务器,该待机应用服务器通过与活跃应用服务器完全同步的状态、并且能够用作活跃应用服务器来确保高可用性。此外,该IMS系统包括会话管理器,该会话管理器通过向应用服务器通知其活跃/待机角色以及通过至少以并行方式将SIP消息路由到应用服务器,以对活跃和待机应用服务器进行排序、管理和同步。

[0007] 在串行或并行配置中,待机应用服务器的实施例可以包括多个向其客户提供高可用性服务的备用服务器。多个备用服务器由会话管理器管理。

[0008] 在本发明又一实施例中,公开了提供高可用性的企业IMS系统的方法。该方法包括

同步和管理一对高可用性的活跃和待机应用服务器,这是通过会话管理器向这些服务器通知它们的活跃/待机角色以及以串行或并行方式向这些应用服务器路由SIP消息实现的。此外,该方法包括,基于从会话管理器接收的指令,由待机应用服务器无缝接管以作为活跃应用服务器。

[0009] 根据本发明的第一方面,提供了一种高可用性企业IMS系统,包括:活跃应用服务器;一个或多个待机应用服务器,配置为与所述活跃应用服务器基本同步;以及会话管理器,配置为基于所述活跃应用服务器的状态来管理所述活跃应用服务器和待机应用服务器的活跃和待机角色。

[0010] 根据第一方面的系统,其中所述活跃应用服务器和待机应用服务器还包括业务逻辑单元和SIP堆栈单元。

[0011] 根据第一方面的系统,其中所述会话管理器还配置为,基于所述活跃应用服务器的状态,指示所述活跃应用服务器的业务逻辑单元为活跃的。

[0012] 根据第一方面的系统,其中所述会话管理器配置为以串行方式将SIP消息路由到所述活跃应用服务器和待机应用服务器。

[0013] 根据第一方面的系统,其中所述待机应用服务器和活跃应用服务器均配置为处理SIP事件。

[0014] 根据第一方面的系统,其中所述会话管理器还配置为,基于所述活跃应用服务器中的故障判定,将所述活跃应用服务器移出活跃角色。

[0015] 根据第一方面的系统,其中所述会话管理器还配置为指示所述待机服务器将其角色改变为活跃。

[0016] 根据第一方面的系统,其中所述会话管理器配置为,确保排序的事件同时到达所述活跃应用服务器和待机应用服务器。

[0017] 根据本发明的第二方面,提供了一种高可用性企业IMS系统,包括:活跃应用服务器;待机应用服务器,配置为与所述活跃应用服务器基本同步;以及会话管理器,配置为以并行方式将SIP消息路由到所述活跃应用服务器和待机应用服务器,该会话管理器还配置为基于所述活跃应用服务器的状态来管理所述活跃应用服务器和待机应用服务器的活跃和待机角色。

[0018] 根据第二方面的系统,其中所述活跃应用服务器和待机应用服务器还包括业务逻辑单元和SIP堆栈单元。

[0019] 根据第二方面的系统,其中所述会话管理器配置为,基于所述活跃应用服务器的状态,指示所述活跃应用服务器的业务逻辑单元为活跃的。

[0020] 根据第二方面的系统,其中所述会话管理器配置为,基于所述活跃应用服务器的状态,指示所述待机应用服务器的业务逻辑单元。

[0021] 根据第二方面的系统,其中所述待机应用服务器和活跃应用服务器均配置为处理SIP事件。

[0022] 根据第二方面的系统,其中所述会话管理器还配置为,基于所述活跃应用服务器中故障的判定,将所述活跃应用服务器移出活跃角色。

[0023] 根据第二方面的系统,其中所述会话管理器还配置为,指示所述待机服务器将其角色改为活跃。

[0024] 根据第二方面的系统,其中所述会话管理器配置为,确保排序的事件同时到达所述活跃应用服务器和待机应用服务器。

[0025] 根据第二方面的系统,其中所述活跃和待机应用服务器位于地理上远离会话管理器的位置。

[0026] 根据本发明的第三方面,提供了一种在企业IMS网络中提供高可用性服务的方法,包括:对活跃应用服务器和待机应用服务器进行排序和同步;以并行方式将SIP消息路由到所述活跃应用服务器和待机应用服务器;以及基于所述活跃应用服务器的状态,确定所述活跃应用服务器和待机应用服务器的活跃和待机角色。

[0027] 根据本发明的第三方面,进一步包括基于所述活跃应用服务器中有故障的判定,将所述活跃应用服务器移出活跃角色。

[0028] 根据本发明的第三方面,指示所述待机应用服务器将其角色改变为活跃。

[0029] 根据本发明的第四方面,提供了一种高可用性企业IMS系统,包括:多个冗余服务器;以及与应用客户的接口,该接口被配置为允许应用客户选择要与之通信的应用服务器。

[0030] 根据本发明的第四方面,其中每个所述冗余服务器配置为以备用角色运行、并且配置为与应用服务器实质上同步,其中所有所述的冗余服务器彼此实质上同步。

[0031] 根据本发明的第四方面,其中所述冗余服务器中的任何一个都不配置为在待机角色下运行,并且每一个所述冗余服务都配置为产生为客户产生事件。

[0032] 根据本发明的第四方面,其中所述应用客户的接口被配置为从每一个所述冗余服务器提供馈送。

[0033] 根据本发明的第四方面,其中预定的冗余服务器被配置为可被选择为所述应用服务器。

附图说明

[0034] 考虑了对本发明实施例的下列详细描述,尤其当结合附图时,本发明的上述及另外的特征和优点将变得显而易见,在附图中各个图中相似的附图标记用于表示相似的部件,其中:

[0035] 图1是本发明的各种实施例可以操作的环境的简化框图;

[0036] 图2是根据本发明的说明性实施例大致示出的容错企业IMS系统的简化框图;

[0037] 图3是本发明的各种实施例可以操作的另一环境的简化框图;

[0038] 图4是根据本发明的说明性实施例为产品提供高可用性馈送(feeds)的企业IMS环境的简化框图。

[0039] 图5是IMS架构中高水平的呼出呼叫流的图示。

[0040] 图6A-B示出了根据本发明的实施例、在企业IMS网络中提供一对高可用性应用服务器的方法的流程图。

[0041] 本申请使用的标题仅用于组织的目的,并不意在限制说明书和权利要求的范围。贯穿本申请中使用的词语“可以”是在允许的意义(即,意味着有可能……)上而不是在强制的意义(即,意思是必须)上使用的。类似地,词语“包括”及其各种形式意思是包括而不限干。为了有利于理解,在可能的地方使用了相似的附图标记以指示这些图所共有的相似的元件。图的可选部分可能使用虚线或点线示出,除非使用的上下文另外指出。

具体实施方式

[0042] 将参考附图在下文中更完整地描述本发明的说明性实施例,在附图中示出了本发明的一些而不是全部实施例。实际上,本发明可以以很多不同形式实现并且不应当认为是限于本文中阐述的实施例,更确切地,提供这些实施例以便本公开满足适用的法律要求。贯穿说明书相似的附图表示表示相似的元件。

[0043] 图1是本发明的各种实施例可以操作的环境100的简化框图。如图所示,非IMS呼叫者102可以经由网络106连接到组织104,组织104向其客户提供在线服务。网络106可以是任何网络,例如但不限于GSM网络、IP网络、PSTN等。此外,组织104可以连接到IMS核心网络108。因此,任何呼叫者必须向IMS网络登记以便与组织104通信。在本发明的示例性实施例中,组织104可以向其客户提供高可用性网络服务,即,来自该组织的服务故障的概率可以忽略。

[0044] 在一个实施例中,非IMS呼叫者可能正在使用GSM电话来连接仅能够经由IMS网络访问的组织104。此时,呼叫者102可以连接到GAM网络106以便在IMS注册中心110注册其自己。IMS注册中心110可以允许访问呼叫者102以与IMS核心网络108连接。之后,呼叫者102可以经由IMS核心网络108连接到其所需的组织104。在另一实施例中,非IMS呼叫者可能正在使用SIP终端来连接到组织104。因此,呼叫者102可以连接到IP网络106以便在IMS注册中心110注册其自己,在这种情况下IMS注册中心110可以是SIP终端。因此,SIP终端110可以允许接入呼叫者102以与IMS核心网络108连接。之后,呼叫者102可以经由IMS核心网络108连接到其所需的组织104。

[0045] 此外,在IMS注册中心110的注册过程期间,呼叫者102必须在S-CSCF(服务呼叫会话控制功能)中注册。注册过程通常是在将呼叫传递到其目的地(该目的地的联系地址通常在S-CSCF中注册)之前的预备过程。此外,在一般的情形中,SIP电话独立地发起注册。然而,在GSM电话的情况下不是如此。因此,需要代表GSM电话处理IMS注册的注册功能,诸如注册110。注册功能可以实现为独立实体或者也可以集成到IMS核心节点108中。

[0046] 此外,组织104可以包括多个服务器来向其客户提供高可用性服务,其中有一个或多个备用服务器,例如包括活跃服务器112和待机服务器114。服务器112和114由会话管理器116管理。会话管理器116可以配置为管理哪个服务要提供给哪个客户。会话管理器116配置为确保活跃服务器112保持为其客户服务。如果活跃服务器112出故障,会话管理器116就可以用活跃服务器112的不反应状态来跟踪它,并且可以命令待机服务器114用作活跃服务器112。同时,会话管理器116可以不停地轮询服务器112以便知晓服务器112是否恢复服务。这样,会话管理器116配置为确保如果活跃服务器故障,则待机服务器无缝接管,组织104提供的服务零中断。

[0047] 在本发明的一个示例性实施例中,待机服务器114保持与活跃服务器112实时同步。本领域技术人员将理解,待机服务器可以简单地且无缝地接管。会话管理器116配置为管理待机服务器114与活跃服务器112的同步。在一个实施例中,会话管理器116可以配置为使用排序(sequencing)来在企业IMS网络中提供高可用性应用的机制。会话管理器116可以以串行(串行排序)或以并行方式将必要的呼叫数据组织到活跃服务器112和待机服务器114中。活跃服务器112和待机服务器114的同步可以实现,其中会话管理器116可以在CSCF

(呼叫会话控制功能)实体中扮演其通常的角色。结合本发明的图2进一步解释所述同步和排序技术。

[0048] 图2是根据本发明的说明性实施例大致示出的容错企业IMS系统的简化框图。在本发明的一个示例性实施例中,系统200能够提供增强的高可用性服务,同时仍保持企业IMS网络的会话和应用层之间的宽松耦合。

[0049] 如图所示,系统200包括会话管理器116(之前结合本发明的图1描述)以及高可用性服务提供节点202,节点202向其客户提供零中断服务。这一对高可用性服务提供设备包括服务器112和114(之前结合本发明的图1描述)。此外,如图所示,两个服务器112和114都包括业务逻辑单元(204和206)以及SIP堆栈(208和210)。

[0050] 在一个实施例中,业务逻辑单元204和206负责基于从会话管理器116依次来的SIP消息向其业务逻辑单元客户214提供服务。此外,SIP堆栈单元208和210可以负责处理需要与会话管理器116通信的SIP事件。在一个实施例中,业务逻辑单元和SIP堆栈单元可能能够独立地操作。

[0051] 此外,在活跃服务器112中,其业务逻辑单元204和SIP堆栈单元208可以处于活跃状态。然而,在待机服务器114中,SIP堆栈210可以处于活跃状态并且业务逻辑单元206可以处于非活跃状态。如图所示,活跃服务器112的SIP堆栈单元208和待机服务器114的SIP堆栈单元210都与会话管理器116通信。

[0052] 此外,会话管理器116(如图2所示)可以配置为提供可配置的高可用性服务选项。高可用性服务选项可以配置为在完成活跃服务器112的排序之后,产生待机服务器114的串行排序。此外,可以由会话管理器116使用基于SIP的机制来管理活跃服务器112和待机服务器114的角色。此外,会话管理器116可以确保待机服务器114知道其处于仅同步模式的角色。类似地,会话管理器116也可以确保活跃服务器112知道其处于事件处理的角色。所有服务器都将活跃地处理来自会话管理器116的SIP消息。然而,仅活跃的业务逻辑单元(BLU)需要为其客户产生事件。在一些实施例中,一些逻辑客户可能接收双馈送,因此两个都将产生事件,这意味着活跃和待机之间没有任何区别。

[0053] 此外,在本发明的一个示例性实施例中,待机服务器114的数据和状态可以与活跃服务器112的数据和状态相同。通过活跃服务器112和待机服务器114之间的同步,这是可能的。此外,同步可以基于会话管理器116使用的排序技术。然而,业务逻辑单元和SIP堆栈单元的功能性在活跃服务器112和待机服务器114中可以不同。

[0054] 所述功能性变化可能是由于待机服务器114的业务逻辑单元206的待机状态。然而,活跃服务器的逻辑状态204可以处于活跃状态。因此,待机服务器总是以与活跃服务器112相同的方式处理SIP事件。如果活跃服务器112发生故障并且会话管理器116指示待机服务器114无缝接管活跃服务器112的角色,待机服务器114可以简单地通知业务逻辑单元206来为其指定客户服务。

[0055] 在本发明的一个示例性实施例中,同步处理由会话管理器116操作。此外,在本发明的一个示例性实施例中,会话管理器116的主要角色是定义活跃服务器112和待机服务器114的角色。因此,不需要活跃服务器112和待机服务器114彼此通信和协调来确定在特定时间段谁将扮演哪个角色。

[0056] 例如,如果在一个时刻活跃服务器112发生故障,则待机服务器114将不依赖于活

跃服务器112得到该信息,待机服务器也不需要定期地直接轮询活跃服务器112来确定活跃服务器是否正常工作。直接轮询给服务器强加了额外的工作负担并且可能不是很可靠的技术。因此,这个任务由会话管理器116实现以保持活跃服务器112和待机服务器114不依赖于彼此的状态。这增强了系统200的容错能力并且确保向客户提供零中断服务。

[0057] 此外,会话管理器116可以配置为使用SIP“OPTIONS(选项)”或其它基于SIP的健康检查机制来确定服务器112和114中各自的状态。服务器112和114的状态检查可以向会话管理器116确保服务被无缝地提供给客户,并且在有任何故障的情况下,会话管理器116可以认识到故障并且可以更新待机服务器114来接管,甚至在向客户提供的服务中断之前也是如此。BLU204可以通知会话管理器116其将要发生故障,有效地请求失效转移(fail-over)。例如,可以提前检测到并且触发所请求的失效转移的条件可以包括:内存不足或者管理员已经指示服务器112和/或BLU204关闭等等。

[0058] 此外,会话管理器116可以配置为检测对活跃服务器112的任何SIP请求的缺乏响应,以判定活跃服务器112是否正发生故障。因此,无论何时活跃服务器发生故障,会话管理器116都可以通过向SIP堆栈210发送SIP OPTIONS消息,按时指示待机服务器114接管活跃服务器的角色,所述SIP OPTIONS消息然后被传送到待机BLU206,BLU206于是接管。

[0059] 此外,在一个实施例中,会话管理器116甚至能够在活跃服务器不处于故障阶段的情况下切换活跃服务器和待机服务器的角色。在会话管理器116确定可能导致活跃服务器112将来发生故障的活跃服务器112的任何问题时,这可能是需要的。此外,会话管理器116可以在系统200需要的任何维护的情况下切换服务器的角色。

[0060] 如果会话管理器116确定系统中的失效转移是必要的,会话管理器116可以通过向待机服务器114发送SIP消息来触发失效转移。待机服务器114仅需要激活其业务逻辑单元206,因为其SIP堆栈单元210已经是活跃的。此外,只有当会话管理器116确定先前活跃的服务器112发生故障(不响应、或者不能处理SIP消息),会话管理器116才可以从后续消息的信令路径删除所述服务器112。

[0061] 此外,在检测到活跃服务器112的不响应状态后,会话管理器116可以自动从序列中删除获取服务器112。会话管理器116也通知待机服务器114将其角色改变成活跃的(使用OPTIONS或INFO命令传递),并且仅继续排序新的活跃服务器112。以这种方式,待机服务器114无缝接管并且服务没有中断。此外,会话管理器116可以继续轮询不响应的服务器112(用SIP OPTIONS命令)以便发现活跃服务器112的服务的任何恢复。

[0062] 一些应用要求呼叫建立必须是高度可用的,否则不进行呼叫建立。结果,在本发明的另一示例性实施例中,活跃服务器112和待机服务器114的排序可选地在原子网络事务(atomic network transaction)的环境中进行。原子网络事务保证事务可以完全退回重来、或者可以完全成功。这可以确保被排序的事件同时到达两个服务器,这使得待机服务器114能够在任何时间戳接管活跃服务器112,因为待机服务器114和活跃服务器112被会话管理器116完全排序和同步。

[0063] 图3是本发明的各种实施例可以操作的另一环境300的简化框图。如图所示,IMS呼叫者302经由网络306连接到向客户提供在线服务的组织304。网络306可以是IMS网络。此外,在本发明的一个示例性实施例中,组织304可以通过将其活跃服务器和待机服务器放置在彼此分开并且与它们的会话管理器(诸如会话管理器116)分开的地理位置,以向其客户

提供高可用性网络服务。这可以为组织304所提供的服务提供额外的可用度和/或可靠性，也为服务增加可靠性。

[0064] 此外，如图3所示，会话管理器116、活跃服务器112和待机服务器114均经由网络308连接。网络308可以是任何网络，诸如但不限于：IP网络、IMS网络、GSM网络、CDMA网络、或者PSTN等。如果网络308是IMS网络，则会话管理器116、活跃服务器112和待机服务器114可能能够彼此直接通信，因为它们均经由同一网络连接。另一方面，如果网络308是非IMS网络的任何其它网络，则可能需要会话管理器116、活跃服务器112和待机服务器114传递额外的安全层。在一个实施例中，会话管理器116、活跃服务器112和待机服务器114可能需要在IMS网络(如先前在图1所示)注册其自己以便彼此通信。此外，为了在会话管理器116中配置该特征，可以使被排序的服务器的SIP Entity(实体)配置内的新可选字段对于其管理员是可见的。管理员可以输入用于该应用(SIP Entity)的待机服务器的完全合格的名字，它用作初始活跃服务器的待机服务器。

[0065] 在环境300中，通过配置组织304的会话管理器116来以并行方式而不是串行方式排序活跃服务器112和待机服务器114，减少呼叫者302向组织304进行呼叫的建立时间。此外，会话管理器116可以配置为使用其向其客户提供高可用性应用支持的“并行模式”，因为活跃服务器112和待机服务器114均位于与会话管理器116地理上远离的位置。此处，会话管理器116可以将序列(例如“邀请”序列)分叉到活跃服务器112和待机服务器114。邀请序列的分叉可以产生该序列的副本。因此，相同的序列可以经由IMS网络以并行方式发送到这两个服务器。

[0066] 此外，在待机服务器114完成该序列之后，待机服务器114将其序列发送回会话管理器116。在这种情况下会话管理器116并不将待机服务器114的序列传播到其下一个SIP实体。此外，配置会话管理器116以确保服务器112和114在继续下一阶段的处理之前已经处理了该序列，可以实现事务隔离。

[0067] 或者，在并行分叉的情形下，驻留在会话管理器116上的HA单元117可以在接收到来自活跃服务器112的INVITE(完成阶段)、但是没有从(一个或多个)待机服务器接收到INVITE时，继续进行下一hop(例如图4的消息460)。稍后，当从下游hop接收到SIP响应消息时，遵守RFC-3261将确保活跃服务器112被包含在信令路径中。由于待机服务器114不在SIP信令路径中，会话管理器116必须将该响应(在适当的对话中)抄送给待机服务器114。有效的是，会话管理器116有目标地、策略地在企业IMS网络中分叉和终止序列对话。这使得待机服务器114遮蔽活跃服务器112。结果，待机服务器114保持完全同步，一旦会话管理器116检测到活跃服务器112不响应就能够接管处理。由此，从呼叫信令的角度看，战略上，在发生故障时经历零停机时间。此外，为了在会话管理器116中配置该特征，可以使被排序的服务器的SIP Entity配置内的新字段对于其管理员是可见的。管理员可以使用该新字段来控制是期望串行还是并行模式来排序。

[0068] 此外，在一个示例性实施例中，活跃和待机服务器可以配置为以代理服务器或B2B模式运行。代理服务器模式的优点是更快传播(在串行高可用性模式中特别有用)，速度的代价是没有B2B，在活跃服务器断电时，已建立的排序对话的状态和不控制是可能的。在这种情况下顾客可以选择用速度来换取期望的功能恢复水平。

[0069] 图4是根据本发明的说明性实施例用于为产品提供高可用性馈送(feeds)的企业

IMS环境的简化框图。在这种情况下,产品可以包括呼叫中心等提供的特征或能力。如图所示,顾客402呼叫了代理404。在一个实施例中,顾客402和代理404都是非IMS网络的一部分。此外,代理404可以代表为一种或多种产品提供高可用性在线馈送的组织的执行官或代表。

[0070] 如图所示,顾客402连接到该组织的家庭网络的信令层面的IMS S-CSCF节点406(服务-呼叫会话控制功能节点)。IMS S-CSCF406基本是能够执行会话控制的SIP服务器。在一个实施例中,节点406可以用作会话管理器(正如先前结合本发明的图1所解释的)。节点406还连接到应用服务器408和另一应用服务器410。在一个实施例中,应用服务器408和410针对不同的功能(需要或不需要高可用性馈送)。如图所示,应用服务器408是向业务逻辑单元客户412提供高可用性馈送的‘HA对’。另一方面,服务器410的排序应用被示为不为产品提供高可用性馈送。

[0071] 此外,服务器408可以包括用于向业务逻辑单元客户412提供高可用性馈送的一对服务器。该对服务器包括先前描述为服务器112(如先前在本发明的图1、2、3中所描述的)的‘Sequence Appl-Active’。此外,该对服务器包括先前描述为服务器114(如先前在本发明的图1、2、3中所描述的)的‘Sequence Appl-Backup’。在不可用的情况下,‘Sequence Appl-Backup’能够接管‘Sequence Appl-Active’而没有任何服务丢失,以确保呼叫处理像往常一样继续,任何事件都不中断。

[0072] 此外,如图所示,顾客402首先通过从顾客402发送到节点406的消息450与代理404通信。节点406通过将不同的应用以及相关消息排序来与高可用性服务器408通信。Sequence Appl-Active从节点406接收INVITE消息452并且向节点406回应消息454。类似地,Sequence Appl-Backup接收IVITE消息456并且回应以消息458,表明服务器408已经完成其处理。如果节点406未从Sequence Appl-Active接收到消息454,则Sequence Appl-Backup可以被指示来接管Sequence Appl-Active。这样,HA对408保持了业务逻辑单元客户412的高可用性。

[0073] 相反,不需要高可靠性的排序的应用处理通过与服务器410的通信示出。节点406向节点410发送消息460。服务器410执行其处理并且通过消息462向节点406返回结果,消息462表明服务器410完成了其处理。

[0074] 此外,每个IMS呼叫经过两个不同阶段;发起和终止。对于任何一个呼叫,S-CSCF确定已经被配置为该呼叫者发起的呼叫排序的所有应用的列表,执行发起处理。在图4中,没有这些,因为呼叫者在IMS网络之外。此外,当所有应用已经被发起排序(originating sequenced)时,S-CSCF确定需要被定期排序(term sequenced)的所有应用的列表。在图4中,有三个这种需要定期排序的应用。当所有应用已经被定期排序时,IMS处理完成,并且呼叫按照标准SIP被路由。此外,有四个感兴趣的IMS阶段。第一阶段是“orig(发起)”,其中SIP INVITE消息被发送给序列应用用来发起处理。第二阶段是“orig-done(发起-结束)”,其中Sequence app将SIP INVITE消息发送回S-CSCF以表明发起处理完成。类似地,第三和第四阶段是“term(定期)”和“term-done(定期-结束)”。用户被分别配置用于“orig”和“term”排序。例如,呼叫授权应用可能仅对正在本地发起的呼叫感兴趣。因此,对该用户的呼入呼叫,不需要排序。

[0075] 此外,图4示出了从未知用户到达CSSF的呼叫,对于该未知用户没有配置IMS排序。该图也示出了第二常规排序应用。作为本发明的一部分,第一排序应用具有在活跃服务器

发生故障时能够接管的待机服务器。因此前两个排序应用是HA对,并且不应当需要明确地配置为不同的序列。

[0076] 图5是IMS架构中高水平呼出呼叫流的图示,这是现有技术中已知的。如图所示,用户Alice502是代理并且是IMS网络的一部分。用户Alice已经呼叫了也是IMS网络的一部分的顾客用户Bob504。此外,框506代表为其客户管理高可用性的会话管理器(先前已经结合本发明的图1、2、3描述)。

[0077] 此外,如图4所示,框508和510是被排序的应用。用户Alice502是IMS系统已知的,并且两个应用508和510被排序用于发起来自用户Alice502的呼叫。此外,没有应用被排序用于在用户Bob504终止的呼叫。由于用户Alice502呼叫用户Bob504,SM506作为第一排序的应用路由到CM-FS508。此外,CM-FS508将新INVITE(IMS阶段origDone)发送回SM506。之后,SM506作为呼叫的“发起”侧的第二排序应用路由到AACC510。然后,AACC510知晓用户Alice502正在发起呼叫。于是AACC510产生B2BUA并且将新INVITE(阶段origDone)发送回SM506。然后,AACC510也在其业务逻辑中模拟该呼叫。由于用户Bob504未被排序,SM506不进行IMS定期处理并且简单地将该呼叫路由到用户Bob504。此外,当用户Bob504返回180铃声时,消息穿过AACC510B2BUA并且用户Bob504的呼叫模拟状态被更新为报警。换言之,该呼叫并未到达用户Bob504,直到活跃服务器和待机服务器的完全状态都已经被模拟。因此,为了被同步,活跃服务器应当在待机服务器之前看到和处理该呼叫。

[0078] 图6A-B示出了根据本发明的实施例、在企业IMS网络中提供一对高可用性应用服务器的方法流程图。当阅读了结合本发明的图1、2、3的描述时,可以更清楚地理解该方法。该方法执行顺序不意在理解为限制,进一步,任何数量的方法步骤可以组合以便实现该方法或者备选方法,而不脱离本发明的范围。

[0079] 在步骤602,呼叫方,可以是向其客户提供某些在线服务的组织的顾客,呼叫该组织以获得特定服务的高可用性馈送。在步骤604,呼叫方连接到管理高可用性的应用服务器对的会话管理器,所述应用服务器提供呼叫方期望的服务。同样,在步骤606,会话管理器对所述应用服务器对进行排序,以便使得这一对应用服务器彼此同步。此外,在步骤606,会话管理器不停检查应用服务器扮演的角色是活跃服务器还是待机服务器。

[0080] 此外,会话管理器提供可配置的高可用性选项,以在活跃序列完成之后自动地产生待机应用服务器的串行排序。此外,会话管理器使用基于SIP的机制来管理活跃应用角色和待机应用角色,以便活跃和待机应用可以知道它们的角色。待机服务器可以以与活跃服务器112相同方式正常地处理SIP事件。然而,待机服务器可以不为其客户产生业务逻辑事件。

[0081] 此外,会话管理器通知待机服务器其角色,而不是应用服务器需要自己协调这些角色。会话管理器可以使用基于SIP的健康检查机制来确定每个服务器的状态。在一个实施例中,缺乏对任何SIP请求的响应,可以被会话管理器用来指示角色变化要发生。

[0082] 此外,在步骤608,如果会话管理器确定活跃服务器发生故障,则在步骤610,会话管理器可以向待机服务器发送SIP消息以进行接管。待机服务器仅需要激活其业务逻辑,因为其SIP堆栈已经是活跃的。之后,只有当先前活跃的服务器发生故障(不响应、或者不能处理SIP消息),会话管理器才从随后消息的信令路径删除所述活跃的服务器。否则,会话管理器保持追踪活跃服务器(如步骤606所示)以确定是否在任何时间活跃服务器发生故障。

[0083] 此外,如果会话服务器检测到活跃应用服务器不响应,会话管理器自动从序列中删除活跃服务器。此外,会话管理器通知待机服务器将其角色改变为活跃的(如步骤610所示)。另一方面,会话管理器继续仅对新的活动应用进行排序。以这种方式,待机服务器无缝接管并且服务没有中断。此外,在步骤612,会话管理器继续用SIP指令轮询不响应的机器,以便会话管理器能够发现服务的任何恢复。

[0084] 在另一实施例,活跃应用和待机应用的排序可选地在原子网络事务的上下文中进行。这确保了排序的事件被同时发送到两个服务器。在另一实施例,通过配置会话管理器以并行而不是串行地对活跃和待机应用进行排序,可以减少呼叫建立时间。在本发明的该实施例中,会话管理器可以配置为使用其高度可用的应用支持的“并行模式”,通常其中活跃和待机都在地理上远离会话管理器的位置。会话管理器可以将该序列复制到活跃和待机二者。当待机将其序列完成回到会话管理器时,这种情况下的会话管理器不将待机服务器传播到下一SIP实体。通过配置会话管理器以确保两个应用都在继续下一阶段的处理之前已经处理了该序列,可以实现事务隔离。

[0085] 或者,会话管理器可以开始要将INVITE消息传播到下游的下一阶段。之后的阶段包括响应,当会话管理器已经接收到其从活跃服务器需要的东西时。当发生该对话的确认消息时,会话管理器将这些发送回待机应用。有效的,会话管理器有目的的、在企业IMS网络中策略地分叉和终止序列对话,使得待机服务器遮蔽活跃服务器。这意味着待机保持与对话完全同步,并且会话管理器一检测到活动应用不响应,就能够接管该对话的处理。从呼叫信令的角度看,在待机应用发生故障时经历零停机时间。

[0086] 此外,该方法不限于此处提及的以上信息。在图1、2、3、4和5解释的各种实施例可以由此处上面解释的每个方法使用。此外,本发明不限于上面提及的实施例和例子,可以在不脱离本发明的范围的情况下可以根据本发明实现很多其它实施例和例子。

[0087] 本领域技术人员可以理解本发明不限于上面提及的实施例。此外,通过该系统提供的特征也可以实现各种其它实施例。此外,诸如‘会话管理’、‘活跃服务器’‘待机服务器’的术语的使用不能理解为限制本发明的方面,因为这些术语仅用于更好的解释的目的。

[0088] 有利的是,本发明提供了提供一对高可用性应用服务器的方法和系统,该方法和系统实现了服务部署的零中断。同样,从呼叫信令的角度看,在发生服务器故障时经历零停机时间。此外,本发明的实施例通过支持待机服务器维护成本降低来实现应用服务器的高可用性,可能能够减少系统的资金和运行成本。

[0089] 本领域技术人员可以理解本发明不限于上面提及的优点。此外,在不脱离本发明的范围的情况下,按照上文中给出的描述,可以理解很多其它优点。

[0090] 参考根据本发明实施例的方法和系统的框图和示意性图示,在上文中描述了本发明的实施例。应当理解,图中的每个方框以及图中方框的组合,都可以由计算机程序指令实现。这些计算机程序指令可以装载到一个或多个通用计算机、专用计算机或其它可编程数据处理转换器上,从而生产出一种机器,使得在计算机或其它可编程数据处理转换器上执行的这些指令产生实现一个或多个方框中规定的功能的装置。也可以把这些计算机程序指令存储在计算机可读存储器中,这些指令使得计算机或其它可编程数据处理装置以特定方式工作,从而,存储在计算机可读存储器中的指令就产生出包括实现一个或多个方框中规定的功能的指令装置的制造品。

[0091] 尽管已经结合当前被认为是最实际的内容和各种实施例描述了本发明,应当理解,本发明不限于所公开的实施例,相反,本发明意图覆盖包含在所附权利要求的精神和范围内的各种修改和等效布置。已经在计算装置、电话和被计算机执行的、诸如程序模块的计算机可执行指令的一般上下文中描述了本发明。通常,程序模块包括执行特定任务或者实现特定摘要数据类型的例程、程序、字符、要素、数据结构等。本领域技术人员将理解本发明可以以其它计算机系统配置实践,所述其它计算机系统配置包括手持装置、多处理器系统、基于微处理器的或者可编程的消费电子装置、网络PC、微型计算机、大型计算机等等。此外,本发明也可以在分布式计算世界中实践,在分布式计算世界中任务由通过通信网络链接的远程处理装置执行。在分布式计算世界中,程序模块可以位于本地和远程存储器存储装置中。

[0092] 本文档描述使用例子公开了本发明,包括最佳模式,并且也使得本领域任何人员能够实践本发明,包括制作和使用任何装置或系统以及执行任何引入的方法。本发明的可专利范围在权利要求中定义,并且可以包括本领域技术人员想到的其它例子。如果这些其它例子的结构元素与权利要求的文字语言没有不同,或者如果它们包括与权利要求的文字语言没有实质区别的等效结构元素,则这些其它例子处于权利要求的范围内。

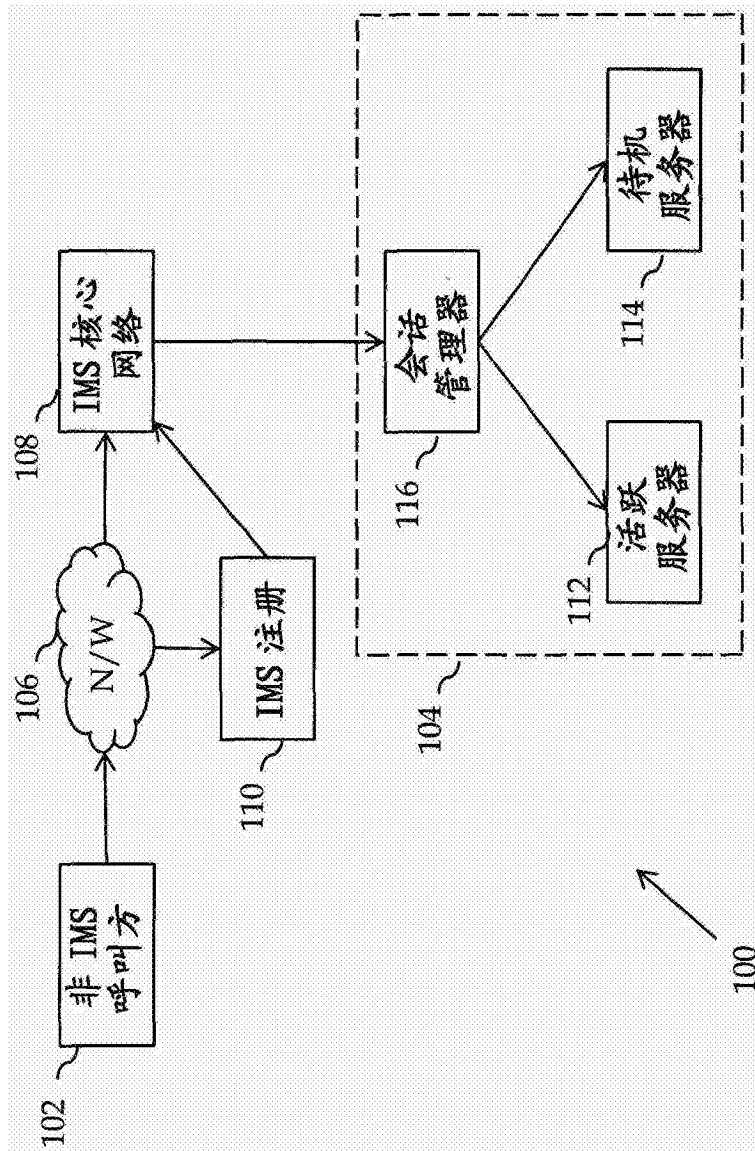


图1

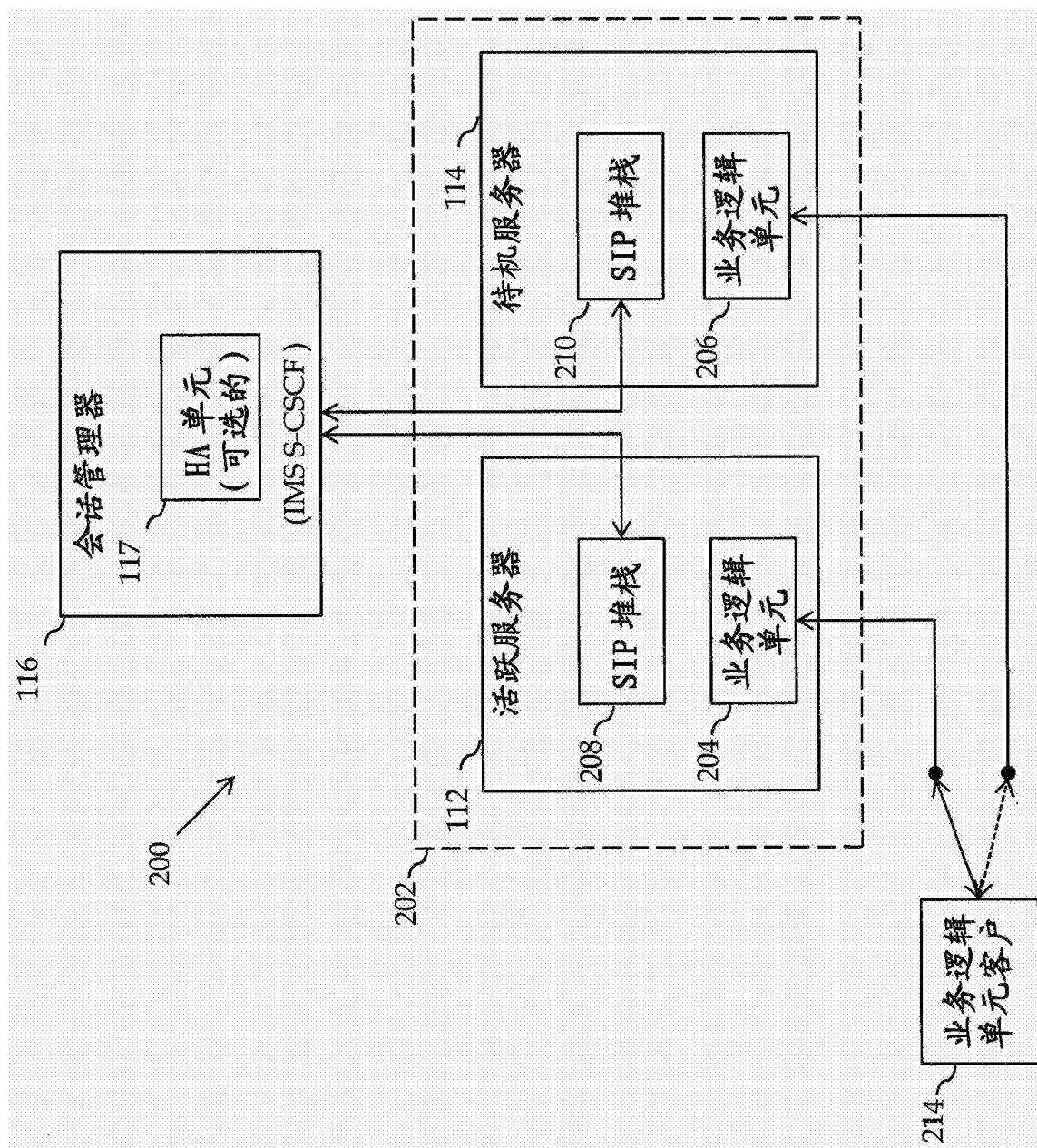


图2

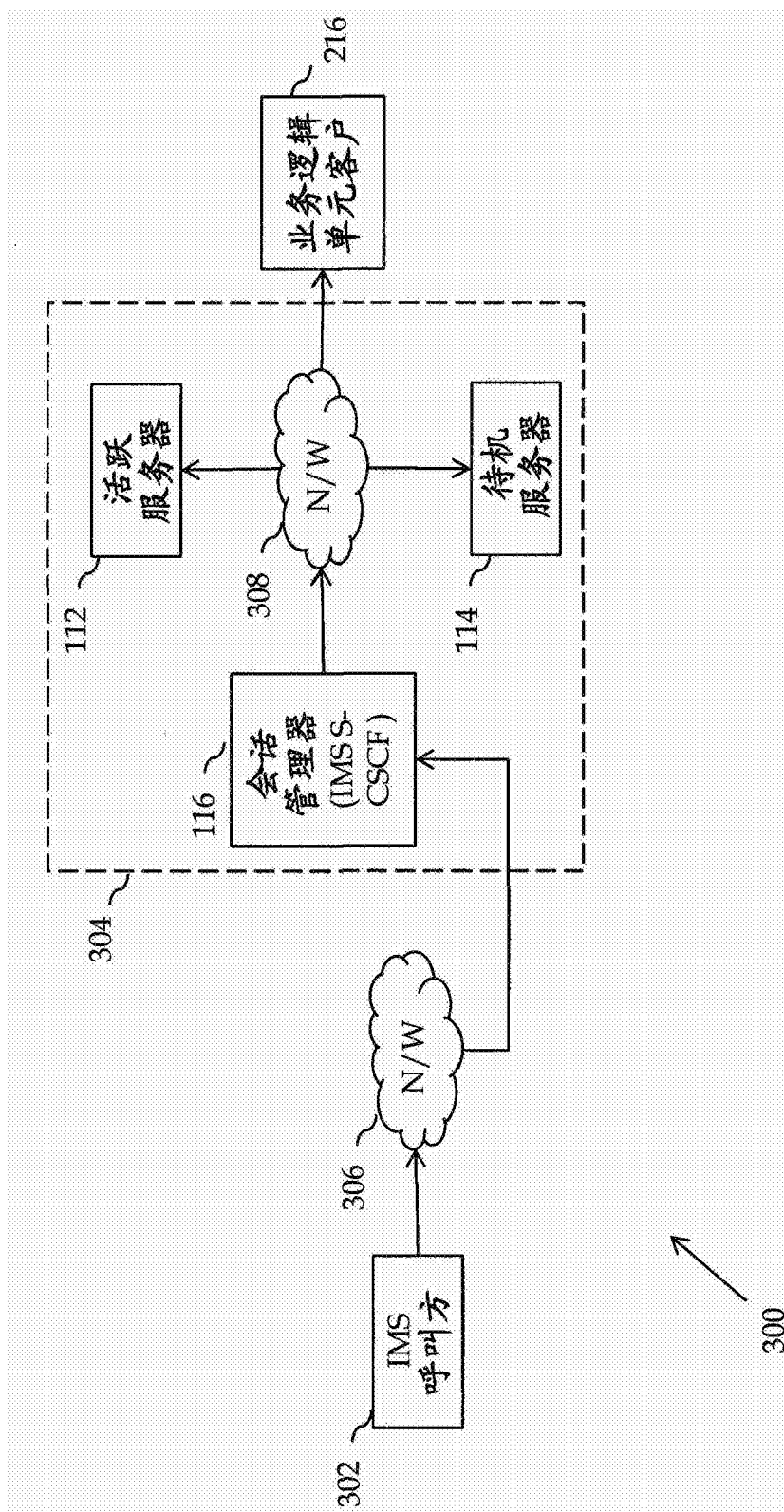


图3

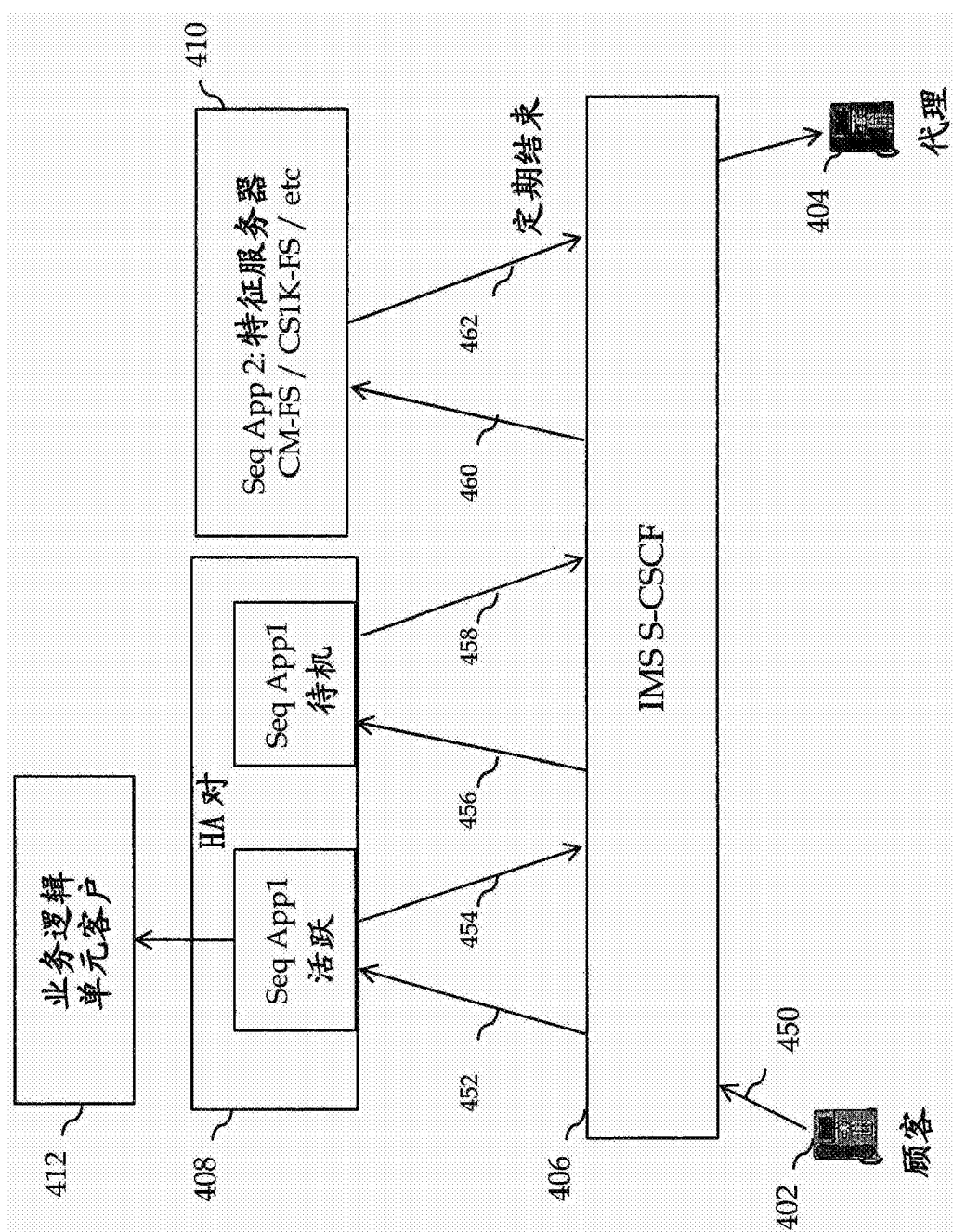


图 4

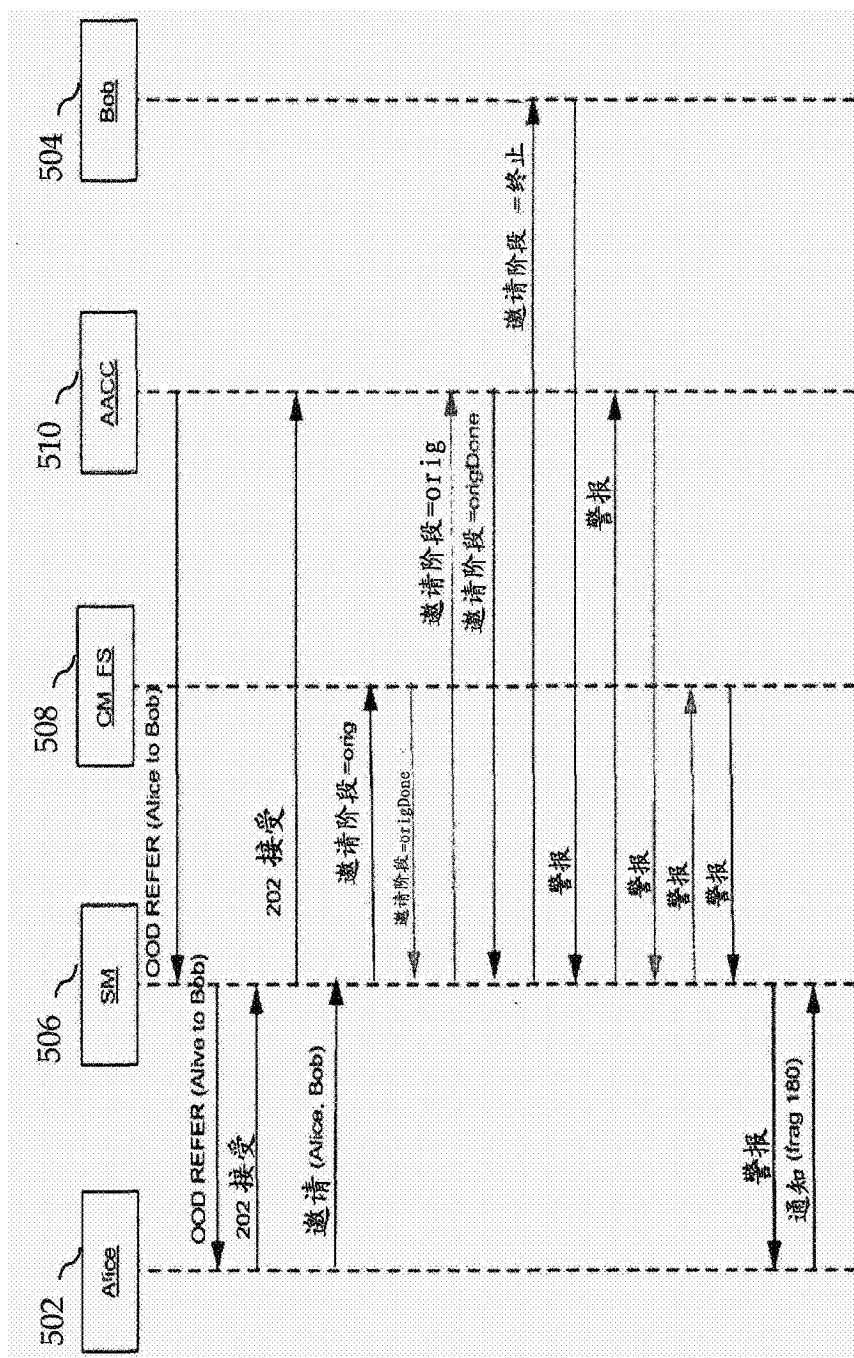


图5

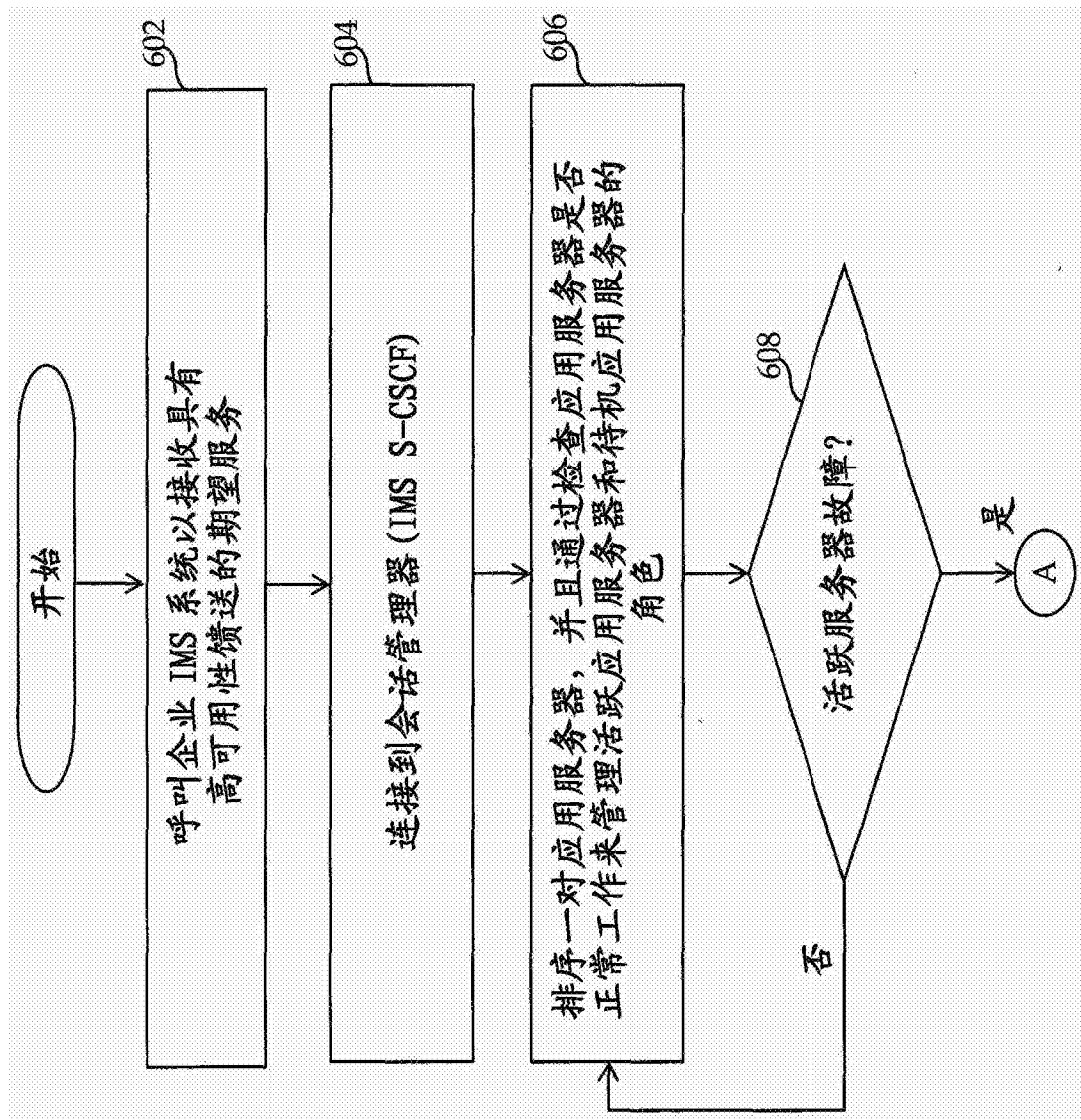


图6A

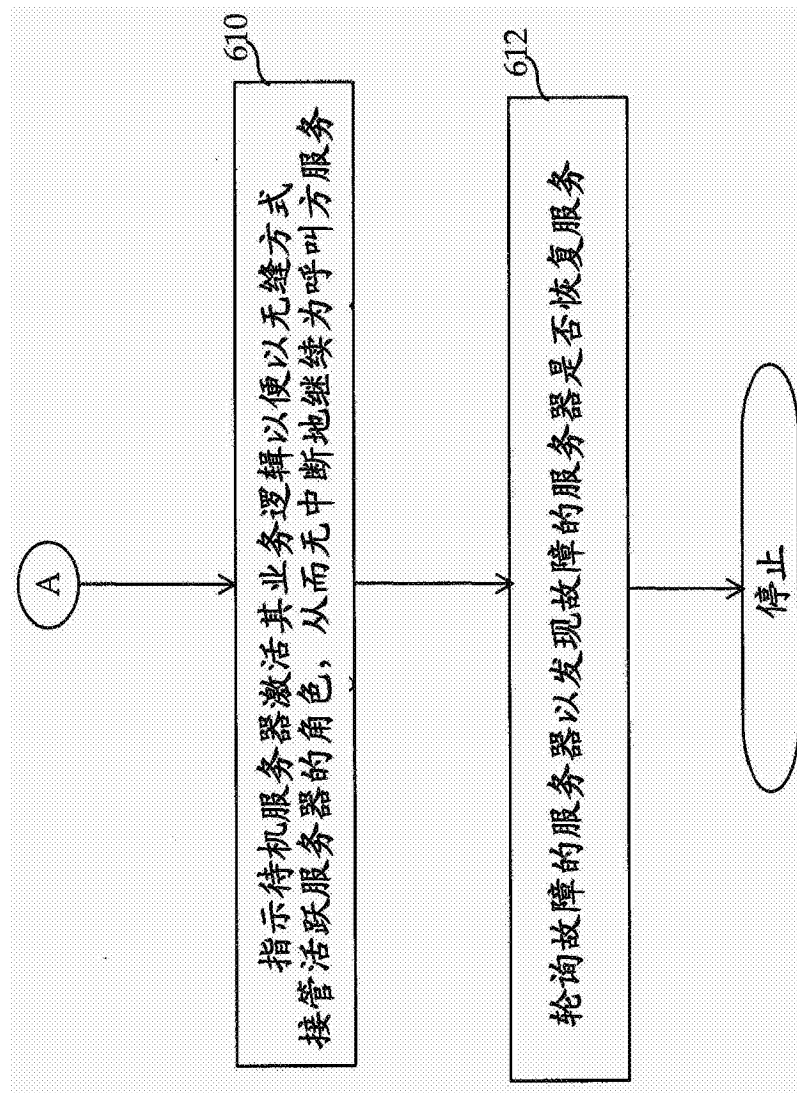


图6B