

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局(43) 国际公布日
2016 年 4 月 21 日 (21.04.2016) WIPO | PCT(10) 国际公布号
WO 2016/058318 A1(51) 国际专利分类号:
H04L 12/24 (2006.01)

春路甲 48 号盈都大厦 A 座 16 层, Beijing 100098 (CN)。

(21) 国际申请号: PCT/CN2015/073964

(22) 国际申请日: 2015 年 3 月 10 日 (10.03.2015)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权: 201410555238.0 2014 年 10 月 17 日 (17.10.2014) CN

(71) 申请人: 中兴通讯股份有限公司 (ZTE CORPORATION) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。

(72) 发明人: 陈露静 (CHEN, Lujing); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。 周智伟 (ZHOU, Zhiwei); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。 刘磊 (LIU, Lei); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。

(74) 代理人: 北京康信知识产权代理有限责任公司 (KANGXIN PARTNERS,P.C.); 中国北京市海淀区知

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国国际公布:

- 包括国际检索报告(条约第 21 条(3))。

(54) Title: ELASTIC VIRTUAL MACHINE (VM) RESOURCE SCALING METHOD, APPARATUS AND SYSTEM

(54) 发明名称: 虚拟机 VM 资源弹性伸缩处理方法、装置及系统

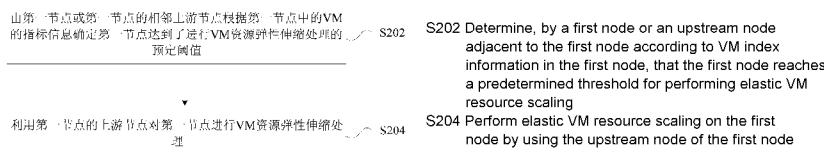


图 2 / Fig. 2

(57) **Abstract:** The present invention provides an elastic virtual machine (VM) resource scaling method, apparatus and system. The method comprises: determining, by a first node or an upstream node adjacent to the first node according to VM index information in the first node, that the first node reaches a predetermined threshold for performing elastic VM resource scaling; and performing elastic VM resource scaling on the first node by using the upstream node of the first node. The problem in the related art of excessive information exchange with an NFVO and high impact on the NFVO is solved, thereby achieving the effect of reducing the information exchange with the NFVO and lowering the impact on the NFVO.

(57) **摘要:** 本发明提供了一种虚拟机 VM 资源弹性伸缩处理方法、装置及系统, 其中, 该方法包括: 由第一节点或第一节点的相邻上游节点根据第一节点中的 VM 的指标信息确定第一节点是达到了进行 VM 资源弹性伸缩处理的预定阈值; 利用第一节点的上游节点对第一节点进行 VM 资源弹性伸缩处理。解决了相关技术中存在的对 NFVO 信息交互过多, 冲击大的问题, 进而达到了减少与 NFVO 的信息交互, 降低对该 NFVO 的冲击的效果。

虚拟机 VM 资源弹性伸缩处理方法、装置及系统

技术领域

本发明涉及通信领域，具体而言，涉及一种虚拟机 VM 资源弹性伸缩处理方法、装置及系统。

5 背景技术

在当前通讯领域，如何在电信设备中构建高效可靠的弹性伸缩服务是业内面临的重要问题和核心难题。

目前大多数的解决方案在安全性和可靠性上均不是很完善，特别是虚拟机（Virtual Machine，简称 VM）的弹性伸缩指标监控部分，目前对虚拟机指标信息的监控多以集中式的单点监控方式为主，其中监控节点多放在上游，多以网络功能虚拟化编排（Network Functions Virtualisation Orchestrator，简称 NFVO）为主，这种情况对 NFVO 的冲击过大，要是上游节点出现问题，就很容易会引发单点故障。

图 1 是相关技术中的资源控制结构图，并且，目前的监控方式一般都是虚拟机将本机的指标信息定时采集上报给上游节点，然后由上游节点判断是否达到了伸缩规则的阀值，此种方式的优点是对集中式的批量管理等比较方便，但是缺点也很明显，比如：

1、消息在传送过程中出现丢包或者网络阻塞时将无法进行实时监控。

2、将本机采集的指标信息交由上游节点进行伸缩判断的话，在精确性上也会有问题。

针对相关技术中存在的对 NFVO 信息交互过多，冲击大的问题，目前尚未提出有效的解决方案。

发明内容

本发明提供了一种虚拟机 VM 资源弹性伸缩处理方法、装置及系统，以至少解决相关技术中存在的对 NFVO 信息交互过多，冲击大的问题。

根据本发明的一个方面，提供了一种虚拟机 VM 资源弹性伸缩处理方法，包括：由第一节点或第一节点的相邻上游节点根据所述第一节点中的 VM 的指标信息确定所述第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理。

5 优选地，所述 VM 的指标信息包括以下信息至少之一：内存信息、中央处理器 CPU 信息、硬盘信息、宽带信息、所述 VM 所承载的业务指标信息、所述 VM 的处理能力。

优选地，当所述第一节点为 VM 时，利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理包括：所述 VM 向虚拟化网络功能 VNF 发送用于请求进行所述 VM 资源弹性伸缩处理的第一请求消息；所述 VNF 根据所述第一请求消息向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；所述 VNFM 根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；所述 NFVO 根据所述 VM 资源变更请求对所述 VM 进行所述 VM 资源弹性伸缩处理。

15 优选地，当所述第一节点为虚拟化网络功能 VNF 时，利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理包括：所述 VNF 向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；所述 VNFM 根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；所述 NFVO 根据所述 VM 资源变更请求对所述 VNF 进行所述 VM 资源弹性伸缩处理。

20 优选地，当所述第一节点为虚拟化网络功能管理 VNFM 时，利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理包括：所述 VNFM 向网络功能虚拟化编排 NFVO 发送用于请求全局的资源访问接口进行预申请资源的第三请求消息；在进行所述预申请资源申请成功的情况下，所述 VNFM 向所述 NFVO 发送 VM 资源变更请求；所述 NFVO 根据所述 VM 资源变更请求对所述 VNFM 进行所述 VM 资源弹性伸缩处理。

根据本发明的另一方面，提供了一种虚拟机 VM 资源弹性伸缩处理装置，包括：确定模块，位于第一节点或第一节点的相邻上游节点中，设置为根据所述第一节点中的 VM 的指标信息确定所述第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；处理模块，位于所述第一节点的各个上游节点中，设置为对所述第一节点进行所述 VM 资源弹性伸缩处理。

优选地，所述处理模块包括：第一发送单元，位于 VM 中，设置为当所述第一节点为所述 VM 时，向虚拟化网络功能 VNF 发送用于请求进行所述 VM 资源弹性伸缩处理的第一请求消息；第二发送单元，位于所述 VNF 中，设置为根据所述第一请求消息向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的消息；第三发送单元，位于所述 VNFM 中，设置为根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；第一处理单元，位于所述 NFVO 中，设置为根据所述 VM 资源变更请求对所述 VM 进行所述 VM 资源弹性伸缩处理。

优选地，所述处理模块包括：第四发送单元，位于虚拟化网络功能 VNF 中，设置为当所述第一节点为所述 VNF 时，利用所述 VNF 向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；第五发送单元，位于所述 VNFM 中，设置为根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；第二处理单元，位于所述 NFVO 中，设置为根据所述 VM 资源变更请求对所述 VNF 进行所述 VM 资源弹性伸缩处理。

优选地，所述处理模块包括：第六发送单元，位于虚拟化网络功能管理 VNFM 中，设置为当所述第一节点为所述 VNFM 时，利用所述 VNFM 向网络功能虚拟化编排 NFVO 发送用于请求全局的资源访问接口进行预申请资源的第三请求消息；第七发送单元，位于所述 VNFM 中，设置为在进行所述预申请资源申请成功的情况下，向所述 NFVO 发送 VM 资源变更请求；第三处理单元，位于所述 NFVO 中，设置为根据所述 VM 资源变更请求对所述 VNFM 进行所述 VM 资源弹性伸缩处理。

根据本发明的再一方面，提供了一种系统，包括上述任一项所述的虚拟机 VM 资源弹性伸缩处理装置。

通过本发明，采用由第一节点或第一节点的相邻上游节点根据所述第一节点中的 VM 的指标信息确定所述第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理，解决了相关技术中存在的对 NFVO 信息交互过多，冲击大的问题，进而达到了减少与 NFVO 的信息交互，降低对该 NFVO 的冲击的效果。

附图说明

此处所说明的附图用来提供对本发明的进一步理解，构成本申请的一部分，本发明的示意性实施例及其说明用于解释本发明，并不构成对本发明的不当限定。在附图中：

图 1 是相关技术中的资源控制结构图；

图 2 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理方法的流程图；

图 3 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置的结构框图；

图 4 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置中处理模块 34 的
5 结构框图一；

图 5 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置中处理模块 34 的
结构框图二；

图 6 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置中处理模块 34 的
结构框图三；

10 图 7 是根据本发明实施例的系统的结构框图；

图 8 是根据本发明实施例的 VM 弹性伸缩过程图；

图 9 是根据本发明实施例的 VNF 弹性收缩过程图；

图 10 是根据本发明实施例的 VNFM 弹性伸缩过程图。

具体实施方式

15 下文中将参考附图并结合实施例来详细说明本发明。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

在本实施例中提供了一种虚拟机 VM 资源弹性伸缩处理方法，图 2 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理方法的流程图，如图 2 所示，该流程包括如下步骤：

20 步骤 S202，由第一节点或第一节点的相邻上游节点根据第一节点中的 VM 的指标信息确定第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；

步骤 S204，利用第一节点的上游节点对第一节点进行 VM 资源弹性伸缩处理。

通过上述步骤，由第一节点或第一节点的相邻上游节点根据第一节点中的 VM 的指标信息确定第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；利用第一节点的上游节点对第一节点进行 VM 资源弹性伸缩处理，实现了由节点本身或节点的相邻

上游节点对节点进行状态监控，减少了对 NFVO 节点的冲击，解决了相关技术中存在的对 NFVO 信息交互过多，冲击大的问题，进而达到了减少与 NFVO 的信息交互，降低对该 NFVO 的冲击的效果。

其中，上述 VM 的指标信息包括以下信息至少之一：内存信息、中央处理器 CPU 5 信息、硬盘信息、宽带信息、VM 所承载的业务指标信息、VM 的处理能力。

在一个优选的实施例中，当第一节点为 VM 时，利用第一节点的上游节点对第一节点进行 VM 资源弹性伸缩处理包括：VM 向虚拟化网络功能 VNF 发送用于请求进行 VM 资源弹性伸缩处理的第一请求消息；VNF 根据第一请求消息向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；VNFM 10 根据第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；NFVO 根据 VM 资源变更请求对 VM 进行 VM 资源弹性伸缩处理。从而实现了在对 VM 进行 VM 资源弹性伸缩处理时，极大的减少了和 NFVO 之间的信息交互，降低了对 NFVO 的冲击。

在一个优选的实施例中，当第一节点为虚拟化网络功能 VNF 时，利用第一节点的 15 上游节点对第一节点进行 VM 资源弹性伸缩处理包括：VNF 向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；VNFM 根据第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；NFVO 根据 VM 资源变更请求对 VNF 进行 VM 资源弹性伸缩处理。从而实现了在对 VNF 进行 VM 资源弹性伸缩处理时，极大的减少了和 NFVO 之间的信息交互，降低了对 NFVO 的冲 20 击。

在一个优选的实施例中，当第一节点为虚拟化网络功能管理 VNFM 时，利用第一节点的上游节点对第一节点进行 VM 资源弹性伸缩处理包括：VNFM 向网络功能虚拟化编排 NFVO 发送用于请求全局的资源访问接口进行预申请资源的第三请求消息；在 25 进行预申请资源申请成功的情况下，VNFM 向 NFVO 发送 VM 资源变更请求；NFVO 根据 VM 资源变更请求对 VNFM 进行 VM 资源弹性伸缩处理。从而实现了在对 VNFM 进行 VM 资源弹性伸缩处理时，极大的减少了和 NFVO 之间的信息交互，降低了对 NFVO 的冲击。

在本实施例中还提供了一种虚拟机 VM 资源弹性伸缩处理装置，该装置用于实现上述实施例及优选实施方式，已经进行过说明的不再赘述。如以下所使用的，术语“模块”可以实现预定功能的软件和/或硬件的组合。尽管以下实施例所描述的装置较佳地 30 以软件来实现，但是硬件，或者软件和硬件的组合的实现也是可能并被构想的。

图 3 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置的结构框图，如图 3 所示，该装置包括确定模块 32 和处理模块 34。下面对该装置进行说明。

确定模块 32，位于第一节点或第一节点的相邻上游节点中，设置为根据第一节点中的 VM 的指标信息确定第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；
5 处理模块 34，位于第一节点的各个上游节点中，连接至上述确定模块 32，设置为对第一节点进行 VM 资源弹性伸缩处理。其中，上述 VM 的指标信息可以包括以下信息至少之一：内存信息、中央处理器 CPU 信息、硬盘信息、宽带信息、VM 所承载的业务指标信息、VM 的处理能力。

图 4 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置中处理模块 34 的
10 结构框图一，如图 4 所示，该处理模块 34 包括第一发送单元 42、第二发送单元 44、第三发送单元 46 和第一处理单元 48。下面对该处理模块 34 进行说明。

第一发送单元 42，位于 VM 中，设置为当第一节点为 VM 时，向虚拟化网络功能 VNF 发送用于请求进行 VM 资源弹性伸缩处理的第一请求消息；第二发送单元 44，位于 VNF 中，连接至上述第一发送单元 42，设置为根据第一请求消息向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；
15 第三发送单元 46，位于 VNFM 中，连接至上述第二发送单元 44，设置为根据第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；第一处理单元 48，位于 NFVO 中，连接至上述第三发送单元 46，设置为根据 VM 资源变更请求对 VM 进行 VM 资源弹性伸缩处理。

20 图 5 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置中处理模块 34 的结构框图二，如图 5 所示，该处理模块 34 包括第四发送单元 52、第五发送单元 54、和第二处理单元 56。下面对该处理模块 34 进行说明。

第四发送单元 52，位于虚拟化网络功能 VNF 中，设置为当第一节点为 VNF 时，利用 VNF 向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；第五发送单元 54，位于 VNFM 中，连接至上述第四发送单元 52，设置为根据第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；第二处理单元 56，位于 NFVO 中，连接至上述第五发送单元 54，设置为根据 VM 资源变更请求对 VM 进行 VM 资源弹性伸缩处理。
25

图 6 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理装置中处理模块 34 的
30 结构框图三，如图 6 所示，该处理模块 34 包括第六发送单元 62、第七发送单元 64、和第三处理单元 66。下面对该处理模块 34 进行说明。

第六发送单元 62，位于虚拟化网络功能管理 VNFM 中，设置为当第一节点为 VNFM 时，利用 VNFM 向网络功能虚拟化编排 NFVO 发送用于请求全局的资源访问接口进行预申请资源的第三请求消息；第七发送单元 64，位于 VNFM 中，连接至上述第六发送单元 62，设置为在进行预申请资源申请成功的情况下，向 NFVO 发送 VM
5 资源变更请求；第三处理单元 66，位于 NFVO 中，连接至上述第七发送单元 64，设置为根据 VM 资源变更请求对 VM 进行 VM 资源弹性伸缩处理。

图 7 是根据本发明实施例的虚拟机 VM 资源弹性伸缩处理系统的结构框图，如图 7 所示，该虚拟机 VM 资源弹性伸缩处理系统 72 包括上述任一项的虚拟机 VM 资源弹性伸缩处理装置 74。

10 为了解决相关技术中存在的对 NFVO 信息交互过多，冲击大的问题，在本发明实
施例中还提供了一种资源按需供给和动态管理的方法，通过本发明实施例中的方法可
以减轻对上游节点 NFVO 的冲击，有效消除单点故障，并减少消息交互，同时让采集
的指标信息更加精确，从而实现了构建高效、安全可靠的弹性伸缩服务的目的。

15 在该实施例中通过使原来统一管理虚拟资源的 NFVO 节点下沉，通过权限下放分
层管理的方式，来减小对 NFVO 的冲击，并提供资源管理接口给 VNFM，使其也能访
问预占资源、VM 状态、VM 已使用资源等全局虚拟资源信息，从而不用每次都通过
NFVO 访问。

优选实施例一 当上述第一节点为 VM 时的 VM 资源弹性伸缩处理：

最下游的 VM 资源创建好弹性伸缩规则后，各个 VM 是通过服务器负载均衡
20 (Server Load Balancing，简称为 SLB)的方式绑定在一起组成的虚拟化网络功能
(Virtualised Network Function，简称为 VNF)，多指网元侧，相对于监控 VNF 层级的
指标信息而言，监控 VNF 下每个 VM 的指标信息显得更为精确和及时，这里的 VM
监控指标主要由内存、CPU、硬盘、带宽等系统信息组成，通过每个 VM 自身进行实
时状态监控，并由自身判断是否达到弹性伸缩条件，若是 VM 自身无法进行判断是否
25 达到了弹性伸缩规则的条件，则将监控的信息数据交由上游的 VNF 节点进行判断。若
是达到伸缩条件，则发送第一请求消息给 VNF 进行弹性伸缩处理，并发送请求给虚拟
化网络功能管理 (Virtualised Network Function Manager，简称为 VNFM)，当系统整体
负荷较低的情况下，如果达到了缩容条件，则由 VNF 发送第二请求消息给 VNFM，
并触发弹性伸缩请求，VNFM 发送 VM 资源请求变更消息给 NFVO，通过对 VM 进行
30 一系列状态迁移操作，释放 VM 资源，然后对 VNF 进行 SLB 处理，降低系统总体拥
有成本 (Total Cost of Ownership，简称为 TCO)；当系统整体负荷较高的情况下，如

果达到了扩容条件，则由 VNF 发送第二请求消息给 VNFM，VNFM 发送 VM 资源请求变更消息给 NFVO，并根据 VNF 绑定的 VM 配置，通过对 VM 进行一系列状态迁移操作，启动相应的 VM，然后对 VNF 进行 SLB 处理，以此扩展、扩大系统处理能力。

5 综上可知，NFVO 配合 VNFM、VNF 完成 VM 的弹性伸缩。该过程由 VM 发起，VNFM 收到 VNF 的创建请求后，计算所需修改的资源列表，向 NFVO 申请分配资源，完成资源的预占和分配，VNFM 启动新增 VM 并向 VNF 通知完成。图 8 是根据本发明实施例的 VM 弹性伸缩过程图，如图 8 所示，该过程包括如下步骤：

10 步骤 S802：云数据中心平台部署完成，VM 资源创建好弹性伸缩规则，虚拟网络部署完成，VM 信息采集完成；

 步骤 S804：VM 自身进行状态监控以及弹性判断，若 VM 自身无法进行判断是否达到了弹性伸缩规则的条件，则将监控的信息数据交由上游的 VNF 节点进行判断；

 步骤 S806：若达到伸缩条件，则由 VM 发送第一请求消息给 VNF 进行弹性伸缩处理；

15 步骤 S808：VNF 发送请求对 VM 资源进行弹性伸缩处理的弹性伸缩请求给 VNFM；

 步骤 S810：VNFM 向 NFVO 发送 VM 资源变更请求，其中，VNFM 解析 VNF 发起的创建 VM 请求，请求中携带 VM 名称和操作、以及创建 VM 的个数：

 如果 VM 已经存在，当 VM 处于异常状态时，直接上报错误；

20 如果 VM 已经存在，当 VM 处于运行状态时，直接上报成功；

 如果 VM 已经存在，VM 处于其他正常状态，则 VNFM 启动 VM 到正常运行状态，然后上报成功；

 如果 VM 不存在，则进入下一步；

25 VNFM 向 NFVO 请求全局的资源访问接口进行预申请资源，资源申请成功上报消息并进入下一步。若 NFVO 无应答超时，则流程终止并上报错误；如果预申请资源失败，则上报失败，流程终止；VNFM 发送资源分配请求给 NFVO，资源分配成功上报消息并进入下一步；若 NFVO 无应答超时，则流程终止并上报错误；如果资源分配失败，则上报失败，流程终止；

步骤 S812：NFVO 成功创建 VM，最终使 VM 达到初始状态，VNFM 生命周期业务向 VNF 发送启动创建的 VM；若启动 VM 失败，则上报错误；VNFM 上报消息返回 VNF 创建结果，携带操作动作和 VM 名称，以及每个 VM 的操作结果；VNF 对新增 VM 进行 SLB 处理。

5 优选实施例二 当上述第一节点为 VNF 时的 VM 资源弹性伸缩处理

VNF 节点除了有可能需要判断下游的 VM 指标信息外，还需要同时判断并监控自身的指标信息，这里的监控指标主要涉及内存、CPU、硬盘、带宽等系统信息以及网元侧所承载的业务指标信息，比如每秒事务处理次数（Transactions Per Second，简称 TPS），某种业务的负荷情况等，若是 VNF 自身无法进行判断是否达到了弹性伸缩规则的条件，则将监控的所有信息数据交由上游的 VNFM 节点进行判断，若是达到伸缩条件，则由 VNF 进行弹性伸缩处理，并发送相关请求给 VNFM，当系统整体负荷较低的情况下，如果达到了缩容条件，则发送第二请求消息给 VNFM，并触发弹性伸缩请求，VNFM 发送 VM 资源变更请求给 NFVO，通过对 VM 进行一系列状态迁移操作，释放 VM 资源，然后对 VNF 进行 SLB 处理，降低系统 TCO；当系统整体负荷较高的情况下，如果达到了扩容条件，则发送第二请求消息给 VNFM，VNFM 然后发送 VM 资源变更请求给 NFVO，并根据 VNF 绑定的 VM 配置，通过对 VM 进行一系列状态迁移操作，启动相应的 VM，然后对 VNF 进行 SLB 处理，以此扩展、扩大系统处理能力。

由上述可知，NFVO 配合 VNFM 完成 VM 的弹性伸缩。该过程由 VM 发起，VNFM 收到 VNF 的创建请求后，计算所需修改的资源列表，向 NFVO 申请分配资源，完成资源的预占和分配，VNFM 启动新增 VM 并向 VNF 通知完成。图 9 是根据本发明实施例的 VNF 弹性收缩过程图，如图 9 所示，该过程包括如下步骤：

步骤 S902：云数据中心平台部署完成，VNF 资源创建好弹性伸缩规则，虚拟网络部署完成，VNF 信息采集完成；

步骤 S904：VNF 自身弹性判断，若 VNF 自身无法进行判断是否达到了弹性伸缩规则的条件，则将监控的信息数据交由上游的 VNFM 节点进行判断；

步骤 S906：若达到伸缩条件，则由 VNF 发送请求进行 VM 资源弹性伸缩处理的弹性伸缩处理消息给 VNFM；

步骤 S908：VNFM 解析 VNF 发起的创建 VM 请求，请求中携带 VM 名称和操作、以及创建 VM 的个数；

如果 VM 已经存在，当 VM 处于异常状态时，直接上报错误；

如果 VM 已经存在，当 VM 处于运行状态时，直接上报成功；

如果 VM 已经存在，VM 处于其他正常状态，则 VNFM 启动 VM 到正常运行状态，然后上报成功；

5 如果 VM 不存在，则进入下一步；

VNFM 向 NFVO 发送第二请求消息，请求全局的资源访问接口进行预申请资源，资源申请成功上报消息并进入下一步；若 NFVO 无应答超时，则流程终止并上报错误；如果预申请资源失败，则上报失败，流程终止；

VNFM 发送 VM 资源变更请求给 NFVO，资源分配成功上报消息并进入下一步。

10 若 NFVO 无应答超时，则流程终止并上报错误；如果资源分配失败，则上报失败，流程终止；

步骤 S910：NFVO 成功创建 VM，最终使 VM 达到初始状态，VNFM 生命周期业务向 VNF 发送启动创建的 VM；若启动 VM 失败，则上报错误；VNFM 上报消息返回 VNF 创建结果，携带操作动作和 VM 名称，以及每个 VM 的操作结果；VNF 对新增 VM 进行 SLB 处理。

优选实施例三 当上述第一节点为 VNFM 时的 VM 资源弹性伸缩处理：

VNFM 节点除了有可能需要判断下游的 VNF 指标信息外，还需要同时判断并监控自身的指标信息，这里的监控指标主要涉及内存、CPU、硬盘、带宽等系统信息以及网元处理能力，若是达到伸缩条件，则进行弹性伸缩处理，当系统整体负荷较低的情况下，如果达到了缩容条件，则触发弹性伸缩请求，发送 VM 资源请求变更消息给

20 NFVO，通过对 VM 进行一系列状态迁移操作，释放 VM 资源，然后对 VNFM 进行 SLB 处理，降低系统 TCO；当系统整体负荷较高的情况下，如果达到了扩容条件，则触发弹性伸缩请求，发送资源请求变更消息给 NFVO，并根据 VNFM 绑定的 VM 配置，通过对 VM 进行一系列状态迁移操作，启动相应的 VM，然后对 VNFM 进行 SLB 处理，以此扩展、扩大系统处理能力。

综上可知，NFVO 配合完成 VNFM 的弹性伸缩。该过程由 VM 发起，VNFM 收到 VNF 的创建请求后，计算所需修改的资源列表，向 NFVO 申请分配资源，完成资源的预占和分配，VNFM 启动新增 VM 并向 VNF 通知完成。图 10 是根据本发明实施例的 VNFM 弹性伸缩过程图，如图 10 所示，该过程包括如下步骤：

步骤 S1002：云数据中心平台部署完成，VNFM 资源创建好弹性伸缩规则，虚拟网络部署完成，VNFM 信息采集完成；

步骤 S1004：VNFM 自身弹性判断，若 VNFM 自身无法进行判断是否达到了弹性伸缩规则的条件，则将监控的信息数据交由上游的 NFVO 节点进行判断；若达到伸缩

5 条件，则由 VNFM 发送用于请求进行 VM 资源弹性伸缩处理的弹性伸缩处理消息给 NFVO；

步骤 S1006：VNFM 解析自身发起的创建 VM 请求，请求中携带 VM 名称和操作、以及创建 VM 的个数：

如果 VM 已经存在，当 VM 处于异常状态时，直接上报错误；

10 如果 VM 已经存在，当 VM 处于运行状态时，直接上报成功；

如果 VM 已经存在，VM 处于其他正常状态，则 VNFM 启动 VM 到正常运行状态，然后上报成功；

如果 VM 不存在，则进入下一步；

VNFM 向 NFVO 发送第三请求消息，请求全局的资源访问接口进行预申请资源，
15 资源申请成功上报消息并进入下一步；若 NFVO 无应答超时，则流程终止并上报错误；如果预申请资源失败，则上报失败，流程终止；VNFM 发送 VM 资源变更请求给 NFVO，资源分配成功上报消息并进入下一步；若 NFVO 无应答超时，则流程终止并上报错误；如果资源分配失败，则上报失败，流程终止；

步骤 S1008：NFVO 成功创建 VM，上报消息返回 VNFM 创建结果，携带操作动作和 VM 名称，以及每个 VM 的操作结果；VNFM 对新增 VM 进行 SLB 处理。
20

在上述各实施例中，各个层级的 VM 通过自身及上游节点的实时监控及弹性伸缩判断，尽量减少各层级的频繁实时信息交互，更加精确的实时检查并监控 VM 状态指标，检查如果达到弹性伸缩的条件，触发弹性伸缩动作，对 VM 自动进行状态迁移，
25 调整 VM 的状态，通过云数据中心操作内存、CPU、硬盘、带宽等 VM 资源，达到扩展网元能力或者降低 TCO。

相对于相关技术，本发明实施例中提出的电信设备虚拟化技术下的弹性伸缩状态迁移方法，不仅能减轻对上游的冲击，还能有效消除单点故障，减少消息交互，同时让采集的指标信息更加精确，在提升弹性伸缩服务安全性的同时，而且能降低设备资

源的使用成本，而且可以得到各种定制化的服务，包括虚拟服务动态迁移、VM 按需分配。进而实现了更加安全、可靠、灵活、高效的信息运作的目的。

显然，本领域的技术人员应该明白，上述的本发明的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，并且在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本发明不限制于任何特定的硬件和软件结合。

以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

工业实用性

如上所述，通过上述实施例及优选实施方式，解决了相关技术中存在的对 NFVO 信息交互过多，冲击大的问题，进而达到了减少与 NFVO 的信息交互，降低对该 NFVO 的冲击的效果。

权 利 要 求 书

1. 一种虚拟机 VM 资源弹性伸缩处理方法，包括：

由第一节点或第一节点的相邻上游节点根据所述第一节点中的 VM 的指标信息确定所述第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；

利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理。

2. 根据权利要求 1 所述的方法，其中，所述 VM 的指标信息包括以下信息至少之一：

内存信息、中央处理器 CPU 信息、硬盘信息、宽带信息、所述 VM 所承载的业务指标信息、所述 VM 的处理能力。

3. 根据权利要求 1 所述的方法，其中，当所述第一节点为 VM 时，利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理包括：

所述 VM 向虚拟化网络功能 VNF 发送用于请求进行所述 VM 资源弹性伸缩处理的第一请求消息；

所述 VNF 根据所述第一请求消息向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；

所述 VNFM 根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；

所述 NFVO 根据所述 VM 资源变更请求对所述 VM 进行所述 VM 资源弹性伸缩处理。

4. 根据权利要求 1 所述的方法，其中，当所述第一节点为虚拟化网络功能 VNF 时，利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理包括：

所述 VNF 向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；

所述 VNFM 根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；

所述 NFVO 根据所述 VM 资源变更请求对所述 VNF 进行所述 VM 资源弹性伸缩处理。

5. 根据权利要求 1 所述的方法，其中，当所述第一节点为虚拟化网络功能管理 VNFM 时，利用所述第一节点的上游节点对所述第一节点进行所述 VM 资源弹性伸缩处理包括：

所述 VNFM 向网络功能虚拟化编排 NFVO 发送用于请求全局的资源访问接口进行预申请资源的第三请求消息；

在进行所述预申请资源申请成功的情况下，所述 VNFM 向所述 NFVO 发送 VM 资源变更请求；

所述 NFVO 根据所述 VM 资源变更请求对所述 VNFM 进行所述 VM 资源弹性伸缩处理。

6. 一种虚拟机 VM 资源弹性伸缩处理装置，包括：

确定模块，位于第一节点或第一节点的相邻上游节点中，设置为根据所述第一节点中的 VM 的指标信息确定所述第一节点达到了进行 VM 资源弹性伸缩处理的预定阈值；

处理模块，位于所述第一节点的各个上游节点中，设置为对所述第一节点进行所述 VM 资源弹性伸缩处理。

7. 根据权利要求 6 所述的装置，其中，所述处理模块包括：

第一发送单元，位于 VM 中，设置为当所述第一节点为所述 VM 时，向虚拟化网络功能 VNF 发送用于请求进行所述 VM 资源弹性伸缩处理的第一请求消息；

第二发送单元，位于所述 VNF 中，设置为根据所述第一请求消息向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；

第三发送单元，位于所述 VNFM 中，设置为根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；

第一处理单元，位于所述 NFVO 中，设置为根据所述 VM 资源变更请求对所述 VM 进行所述 VM 资源弹性伸缩处理。

8. 根据权利要求 6 所述的装置，其中，所述处理模块包括：

第四发送单元，位于虚拟化网络功能 VNF 中，设置为当所述第一节点为所述 VNF 时，利用所述 VNF 向虚拟化网络功能管理 VNFM 发送用于请求全局的资源访问接口进行预申请资源的第二请求消息；

第五发送单元，位于所述 VNFM 中，设置为根据所述第二请求消息向网络功能虚拟化编排 NFVO 发送 VM 资源变更请求；

第二处理单元，位于所述 NFVO 中，设置为根据所述 VM 资源变更请求对所述 VNF 进行所述 VM 资源弹性伸缩处理。

9. 根据权利要求 6 所述的装置，其中，所述处理模块包括：

第六发送单元，位于虚拟化网络功能管理 VNFM 中，设置为当所述第一节点为所述 VNFM 时，利用所述 VNFM 向网络功能虚拟化编排 NFVO 发送用于请求全局的资源访问接口进行预申请资源的第三请求消息；

第七发送单元，位于所述 VNFM 中，设置为在进行所述预申请资源申请成功的情况下，向所述 NFVO 发送 VM 资源变更请求；

第三处理单元，位于所述 NFVO 中，设置为根据所述 VM 资源变更请求对所述 VNFM 进行所述 VM 资源弹性伸缩处理。

10. 一种虚拟机 VM 资源弹性伸缩处理系统，包括权利要求 6 至 9 中任一项所述的装置。

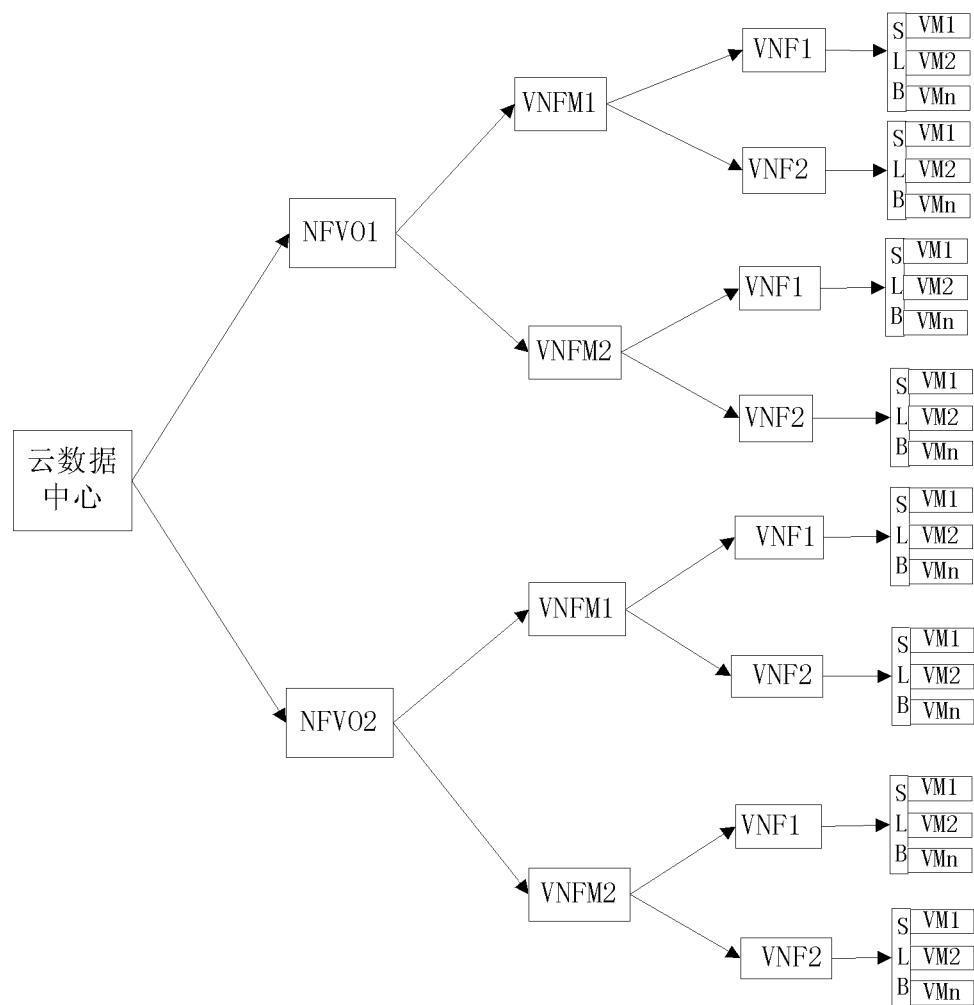


图 1

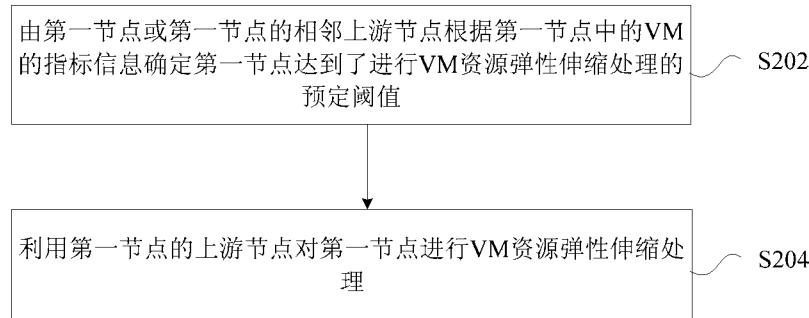


图 2



图 3



图 4

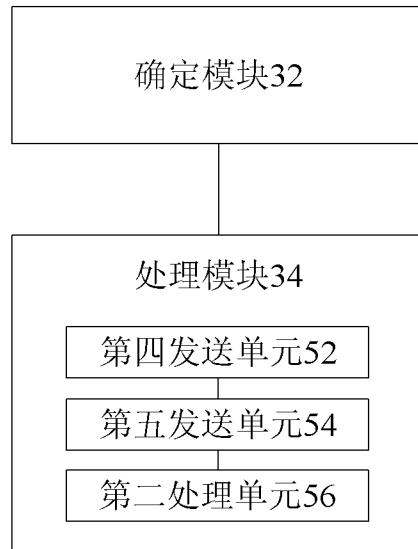


图 5

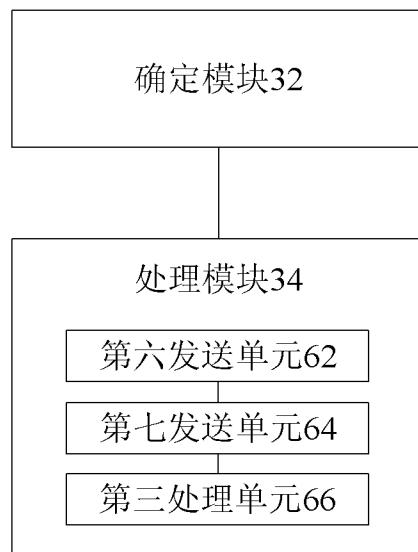


图 6

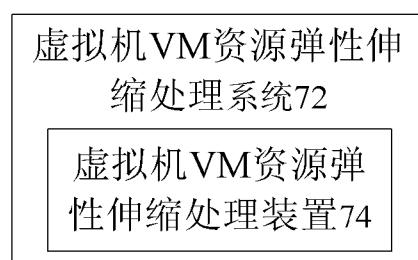


图 7

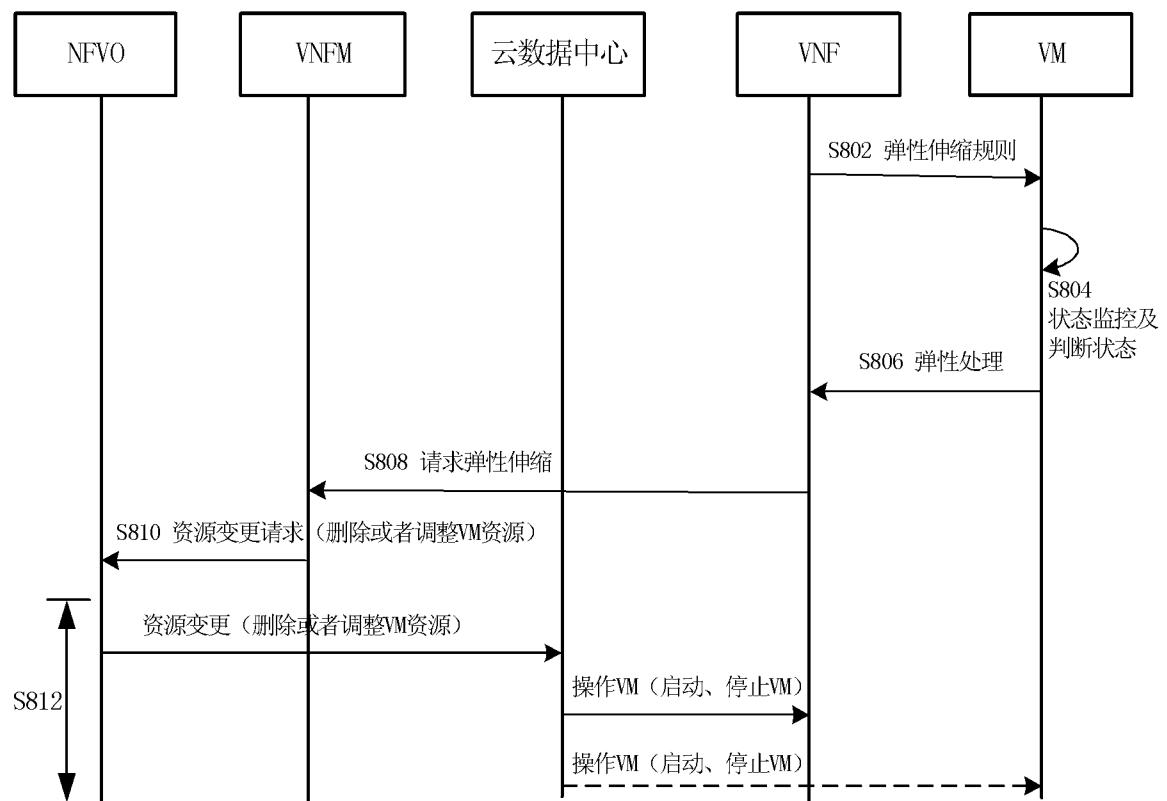


图 8

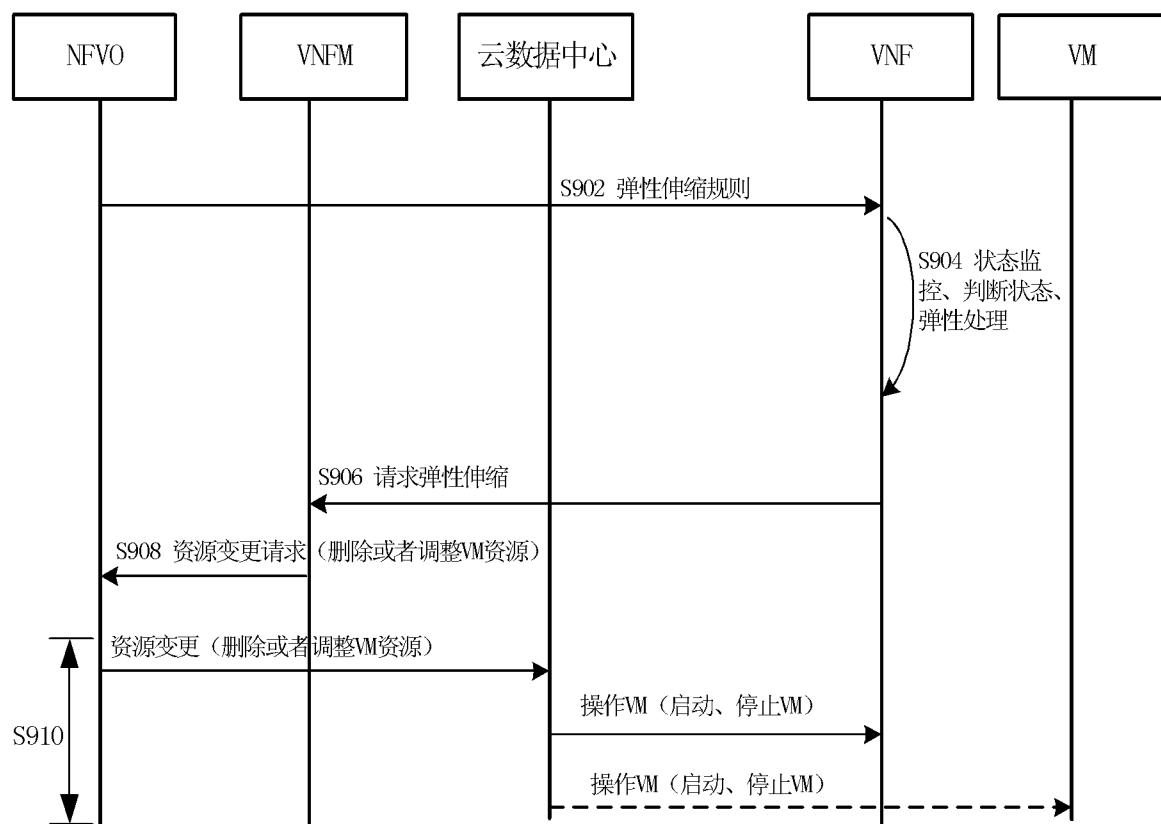


图 9

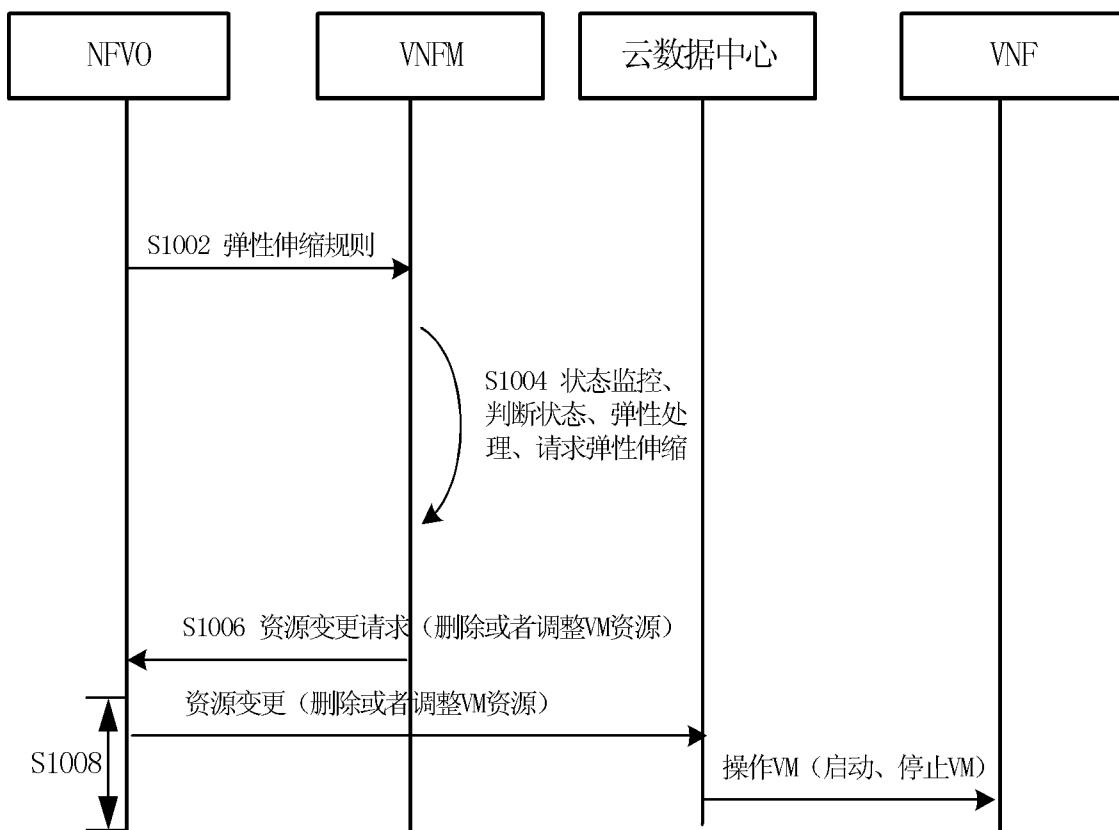


图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/073964

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/24 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI: VNP, VNFM, NFVO, upstream, virtual machine, resource, elasticity, flexible, dilatation, scale, add, increase, reduce, decrease, threshold, limit, virtual, network, function, management, orchestrator

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101488098 A (UNIV HUAZHONG SCIENCE TECH) 22 July 2009 (22.07.2009) description, page 4, line 7 from the bottom to page 6, line 8, claims 2 to 4	1, 2, 6, 10
Y	CN 101488098 A (UNIV HUAZHONG SCIENCE TECH) 22 July 2009 (22.07.2009) description, page 4, line 7 from the bottom to page 6, line 8, claims 2 to 4	3-5, 7-10
X	US 2013047158 A1 (ESDS SOFTWARE SOLUTION PVT. LTD.) 21 February 2013 (21.02.2013) description, paragraphs [0015] to [0017]	1, 2, 6, 10
Y	US 2013047158 A1 (ESDS SOFTWARE SOLUTION PVT. LTD.) 21 February 2013 (21.02.2013) description, paragraphs [0015] to [0017]	3-5, 7-10
Y	CN 104050045 A (HUAWEI TECH CO., LTD.) 19 September 2014 (19.09.2014) description, paragraphs [0104], [0130] and [0132], and figure 1	3-5, 7-10
A	CN 102646062 A (GUANGDONG ELECTRONICS INDUSTRY INST CO., LTD.) 22 August 2012 (22.08.2012) the whole document	1-10

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&”document member of the same patent family

Date of the actual completion of the international search
30 June 2015

Date of mailing of the international search report
22 July 2015

Name and mailing address of the ISA
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No. (86-10) 62019451

Authorized officer
YAN, Yan
Telephone No. (86-10) 62413415

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/073964

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103678251 A (CHENGDU TIANZHI DAXING INFORMATION TECHNOLOGY CO., LTD.) 26 March 2014 (26.03.2014) the whole document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2015/073964

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101488098 A	22 July 2009	None	
US 2013047158 A1	21 February 2013	IN MUM201102308 A	22 February 2013
CN 104050045 A	17 September 2014	None	
CN 102646062 A	22 August 2012	None	
CN 103678151 A	26 March 2014	None	

国际检索报告

国际申请号

PCT/CN2015/073964

A. 主题的分类

H04L 12/24(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, EPODOC, WPI:虚拟机, VM, 资源, 弹性, 伸缩, 扩容, 缩容, 阈值, 门限, VNF, VNFM, NFVO, 上游, virtual machine, resource, scale, add, increase, reduce, decrease, threshold, limit, virtual, network, function, management, orchestrator

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 101488098 A (华中科技大学) 2009年 7月 22日 (2009 - 07 - 22) 说明书第4页倒数第7行-第6页第8行, 权利要求2-4	1-2, 6, 10
Y	CN 101488098 A (华中科技大学) 2009年 7月 22日 (2009 - 07 - 22) 说明书第4页倒数第7行-第6页第8行, 权利要求2-4	3-5, 7-10
X	US 2013047158 A1 (ESDS SOFTWARE SOLUTION PVT. LTD.) 2013年 2月 21日 (2013 - 02 - 21) 说明书第[0015]-[0017]段	1-2, 6, 10
Y	US 2013047158 A1 (ESDS SOFTWARE SOLUTION PVT. LTD.) 2013年 2月 21日 (2013 - 02 - 21) 说明书第[0015]-[0017]段	3-5, 7-10
Y	CN 104050045 A (华为技术有限公司) 2014年 9月 17日 (2014 - 09 - 17) 说明书第[0104]、[0130]、[0132]段, 图1	3-5, 7-10
A	CN 102646062 A (广东电子工业研究院有限公司) 2012年 8月 22日 (2012 - 08 - 22) 全文	1-10

 其余文件在C栏的续页中列出。 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“0” 涉及口头公开、使用、展览或其他方式公开的文件

“&” 同族专利的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

国际检索实际完成的日期

2015年 6月 30日

国际检索报告邮寄日期

2015年 7月 22日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)

北京市海淀区蓟门桥西土城路6号

100088 中国

传真号 (86-10)62019451

受权官员

阎岩

电话号码 (86-10)62413415

国际检索报告

国际申请号

PCT/CN2015/073964

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 103678151 A (汉柏科技有限公司) 2014年 3月 26日 (2014 - 03 - 26) 全文	1-10

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/073964

检索报告引用的专利文件		公布日 (年/月/日)		同族专利		公布日 (年/月/日)	
CN	101488098	A	2009年 7月 22日		无		
US	2013047158	A1	2013年 2月 21日	IN	MUM201102308	A	2013年 2月 22日
CN	104050045	A	2014年 9月 17日		无		
CN	102646062	A	2012年 8月 22日		无		
CN	103678151	A	2014年 3月 26日		无		

表 PCT/ISA/210 (同族专利附件) (2009年7月)

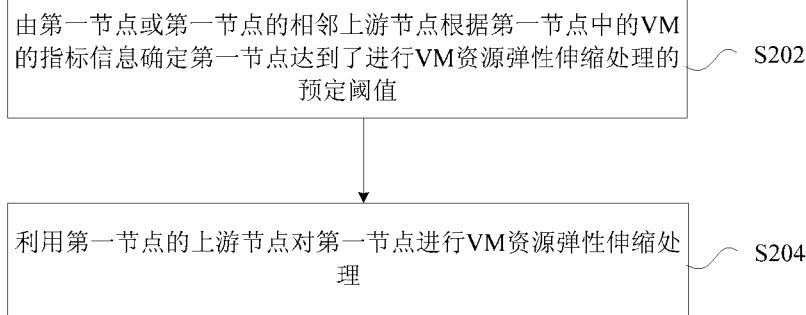


图 2 / FIG. 2 图 1 / FIG. 1 权利要求
替换页(细则第26条)

图 2 / Fig. 2