软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(7):2116−2132 [doi: 10.13328/j.cnki.jos.005555]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

分布式云的研究进展综述*

张晓丽[1,2], 杨家海[2,3], 孙晓晴[1,2], 吴建平[1,2,3]

[1](清华大学 计算机科学与技术系,北京 100084)

[2](清华大学 网络科学与网络空间研究院,北京 100084)

[3](清华信息科学与技术国家实验室(清华大学),北京 100084)

通讯作者: 杨家海, E-mail: yang@cernet.edu.cn

摘 要: 云计算作为全新的计算模式,将数据中心的资源包括计算、存储等基础设施资源通过虚拟化技术以服务的形式交付给用户,使得用户可以通过互联网按需访问云内计算资源来运行应用.为面向用户提供更好的服务,分布式云跨区域联合多个云站点,创建巨大的资源池,同时利用地理分布优势改善服务质量.近年来,分布式云的研究逐渐成为学术界和工业界的热点.围绕分布式云系统中研究的基本问题,介绍了国际、国内的研究现状,包括分布式云系统的架构设计、资源调度与性能优化策略和云安全方案等,并展望分布式云的发展趋势.

关键词: 云计算;分布式云;云架构;资源调度;云安全

中图法分类号: TP393

中文引用格式: 张晓丽,杨家海,孙晓晴,吴建平.分布式云的研究进展综述.软件学报,2018,29(7):2116−2132. http://www.jos.org.cn/1000-9825/5555.htm

英文引用格式: Zhang XL, Yang JH, Sun XQ, Wu JP. Survey of geo-distributed cloud research progress. Ruan Jian Xue Bao/ Journal of Software, 2018,29(7):2116−2132 (in Chinese). http://www.jos.org.cn/1000-9825/5555.htm

Survey of Geo-Distributed Cloud Research Progress

ZHANG Xiao-Li[1,2], YANG Jia-Hai[2,3], SUN Xiao-Qing[1,2], WU Jian-Ping[1,2,3]

[1](Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

[2](Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100084, China)

[3](Tsinghua National Laboratory for Information Science and Technology (TNList) (Tsinghua University), Beijing 100084, China)

Abstract: Cloud computing as a new computing model, provides computing and storage services to users by the virtualization technology. Users then can request and access elastic cloud resources on demand to run their applications via the Internet. Recently geo-distributed cloud has been a hot research topic in both academia and industry. It federates multiple cloud sites to maintain huge and distributed virtual resources, so as to provide better services for users. Based on fundamental research problems in geo-distributed cloud systems, this acticle describes the international and domestic research progress of this area, including architecture design, resource scheduling schemes, and security mechanisms. It also points out some of the research trends in the field.

Key words: cloud computing; geo-distributed cloud; cloud architecture; resource scheduling; cloud security

云计算(cloud computing)是指数据中心(data center)通过互联网将应用(application)、系统软件和硬件作为服务提供给用户的新型的计算模式.近年来,云计算得到了迅猛的发展,越来越多的公司(应用服务提供商)倾向于将其应用部署在云系统中.从硬件配置和价格的角度考虑,云计算的优势主要有 3 点:(1) 提供无限大的可用

* 基金项目: 国家自然科学基金(61432009); 国家重点基础研究发展计划(973)(2012CB315806)

Foundation item: National Natural Science Foundation of China (61432009); National Key Basic Research and Development Program of China (973) (2012CB315806)

收稿时间: 2016-06-03; 修改时间: 2017-04-21; 采用时间: 2018-01-13; jos 在线出版时间: 2018-02-08

CNKI 网络优先出版: 2018-02-08 11:56:00, http://kns.cnki.net/kcms/detail/11.2560.TP.20180208.1155.010.html

资源池,且具备负载激增的适应能力,使得云用户不需要对资源使用量进行预先计算;(2) 按需供给模型提高了业务的扩展能力;(3) 基于用户的资源使用情况进行细粒度的计费(计算资源以小时为单位进行计费,存储资源以天为单位进行计费),降低成本支出和操作开销.

虽然云计算得到了广泛的推广,但其依旧存在严峻的挑战.

(1) 服务的高可用性需求.应用服务商将应用部署在云环境中,其最关心的莫过于服务的高可用性,以保证终端用户的体验.虽然云服务提供商通过与云用户之间签订服务等级协议(service level agreement,简称 SLA)来保障服务的可用性和可靠性,然而这并不能从根本上解决问题.同时,由单一云站点提供的云服务始终受到单点失败(a single point of failure)的威胁,许多不可抗因素,如能源供应中断、自然灾害、人为攻击等都会导致服务不可用.即使对于拥有多个数据中心的公司来讲,虽然基于不同的网络提供商的服务,其实现了基础设施的构建,但是一般情况下,其基础设施和计费系统都是相同的,使得其风险规避能力较弱.因此,学术界越来越关注联合多个云提供商共同提供服务来保证服务的高可用性[1].

(2) 应用请求的分布变化.随着全球化的发展,应用的受众不再局限于某一区域,其往往呈现出离散的分布状况,文献[2]分析了商业云服务的用户追踪信息,发现对于特定的云应用,例如邮件系统,其不需要大规模的计算服务和通信服务,但是需要利用地理上的多样性来改善其在扩展性、可靠性和性能等方面的需求.因此,将应用部署在分布式的云系统中,有利于缩短应用到终端用户的平均距离[3],能够有效降低访问时延和带宽开销[4].

分布式云系统(geo-distributed cloud,federated-cloud,multi-cloud,intercloud,简称 GDC),能够有效地解决以上问题,其联合多个云站点,提供更好性能的服务,保障了服务的质量和可靠性,同时降低了资源消耗,节约了开销,达到云服务提供商和云用户的双赢局面.

每个云都有其独一无二的基础架构,其中的存储接口处于专有的状态,用户不能方便地将数据从一个站点导入到另一个云站点中.同时,对于数据密集型应用来说,其部署在多个云中会使得数据放置问题和传输问题变得十分复杂.因此,在云环境中,基于多个云服务提供商的基础平台,不仅需要提供云间的互操作标准,建立统一的接口,设计合理的分布式云架构,促使应用提供商能够跨云部署服务和放置数据,还需要优秀的资源调度策略来优化数据的放置问题,以降低数据传输的开销和时延.同时,云站点间的资源互访问往往引入广域网带宽消耗,因此,其中的网络架构和带宽分配策略都值得考虑.在多云环境中,虚拟机迁移是其内部常见的跨云操作,这相较于在单一云站点的内部迁移,具有更大的挑战性.另外,云服务提供商间在安全策略方面也存在着一定的差异,异构的云环境往往也会引入一些安全问题.对于分布式云环境下数据存储一致性的问题,目前工业界已经有了比较成熟的技术,其中包括谷歌公司内部应用状态的分布式文件系统 GFS[5]、Amazon 公司的简单存储服务 S3(simple storage service)[6]和分布式数据存储系统 Dynamo[7]以及微软公司的 WAS(windows azure storage)系统等.虽然这些模型全都应用于同一云提供商的多数据中心中,但是对于不同云服务提供商共同搭建的云环境,从保证数据一致性的技术层面来分析,并没有什么不同,其均为设计最终一致性模型并为模型添加部分强一致性属性来实现.因此,本文对于数据一致性方面不进行过多的介绍.

综上,本文试图通过介绍分布式云系统架构、资源调度策略以及安全策略的相关研究现状,分析并对比其中典型策略的特点和适用范围,并指出已有方案的局限性,进而讨论分布式云在各个方面的发展趋势.

1 分布式云架构

随着互联网基础设施的发展,具备大型计算、存储服务能力的云数据中心得到广泛部署.在此基础上提出的“分布式云系统”旨在联合各数据中心,通过策略性信息交互和资源调配,在实时性、鲁棒性、节能性、易用性和安全性等方面提供更优质的云服务.分布式云系统需要解决的首要问题——如何实现异构数据中心间的通信与互操作,成为学术界与工业界的重点研究课题.本节将针对分布式云的系统设计架构和站间网络部署方案进行详细介绍.

1.1 系统设计架构

截至目前已提出的众多分布式云架构模型,可依照耦合程度、数据中心类型和应用场景分为混合架构、对等架构和聚合架构 3 类.其中,耦合程度(松耦合、部分耦合、紧耦合)体现为各数据中心相互在资源管理、监控、迁移等方面不同程度的操作权限[8];数据中心类型有依照云服务的部署方式划分的公有云、私有云和混合云;应用场景则表现为分布式云系统中各数据中心间的业务关系,如从属、租用、代理、合作等.本节将依据上述 3 方面对混合、代理、聚合和多层 4 类分布式云架构模型进行详细介绍.

1.1.1 混合架构模型

混合架构常见于私有云与公有云之间,如图 1 所示,私有云进行动态规模扩展,通过在需求过载情况下向一个或多个公有云请求资源的方式,达到保障云服务稳定性的同时,降低日常资源储备开销的目的.此类场景下,私有云与公用云之间仅有租用业务,可对虚拟资源进行简单操作,无高级控制权限,属于松耦合架构.

Fig.1 Hybrid architecture

图 1 混合架构模型

Sitaram 等人[9]基于云计算系统分层模型提出 SCF(simple cloud federation)框架,针对资源访问路径和请求路由时间制定交互策略,并通过联合 OpenStack 和 Amazon 上的云资源验证了 SCF 架构的可行性.工业界中,许多开源云计算管理平台,如 OpenNebula、Eucalyptus、OpenStack 等也提供了 API 支持混合架构分布式云平台构建[10].

1.1.2 对等架构模型

对等架构常见于多个业务关系紧密但耦合程度较低的云数据中心之间,各云服务提供商通过签署协议,聚合各方云资源,解决高峰期用户需求过载导致的服务性能下降问题和低谷期空闲资源浪费问题,实现双赢.根据是否有第三方介入,可将对等架构模型分为自主合作(如图 2(a)所示)和代理合作(如图 2(b)所示)两种形式,后者通过一个代理服务器联合各公有云资源.与传统意义上负责协调云提供商与用户间交互的云代理不同,对等分布式云中的代理服务器往往需要一个软件适配层,屏蔽不同云提供商的接口差异,同时需要基于性能、开销、能耗等优化指标,寻找最佳数据中心,完成应用部署.

针对自主合作方式,Rochwerger[11]等人根据云的伸缩性与服务水平协议(SLA),提出 Reservior 模型.该模型通过服务管理器、虚拟执行环境管理器(VEEM)和虚拟执行环境宿主机(VEEH)这 3 部分组件交互合作,形成了 3 层抽象,将面向用户的服务管理与底层的基础设施调度解耦,实现了多个云数据中心的动态联合.文献[12]则提出 InterCloud 模型,该模型由云间交换器、云间根实例以及云站点构成.其中,云间交换器负责促进异构云环境之间的谈判与合作,云间根实例负责保存 Intercloud 架构中所有实体的当前信息,各部分通过云间网关完成通信.针对代理合作方式,文献[13]设计了包含资源调度器和虚拟设施管理器的云代理服务器,其中,资源调度器基于用户的基础设施需求和各云数据中心可用资源信息,利用优化算法生成服务部署方案.虚拟设施管理器则基于 OpenNebula 的虚拟资源管理器,完成上层抽象,屏蔽接口差异,实现对虚拟资源的监控和管理.文献[14]则通过分别在集中式和分布式代理云架构下进行实验,阐明了采用代理服务器实现云数据中心联合的可行性与高效性.

1.1.3 聚合架构模型

聚合架构常见于隶属于同一机构的云数据中心之间,如图 3 所示,聚合分布式云通过上层的核心云操作系

统全面控制各数据中心资源,并提供统一的对外访问接口,属于紧耦合架构.此类架构更易于实现负载均衡、容错冗余等高级管理功能,适合企业充分利用旗下分布于不同地理位置的数据中心构建分布式云系统.目前,多层架构模型已在亚马逊、谷歌、阿里巴巴等大型云服务公司中得到应用,如亚马逊公司的 CDN 服务 CloudFront[5].此外,OpenCirrus[15]也联合了 6 个隶属于不同区域的数据中心,为研究人员提供丰富的云服务资源.

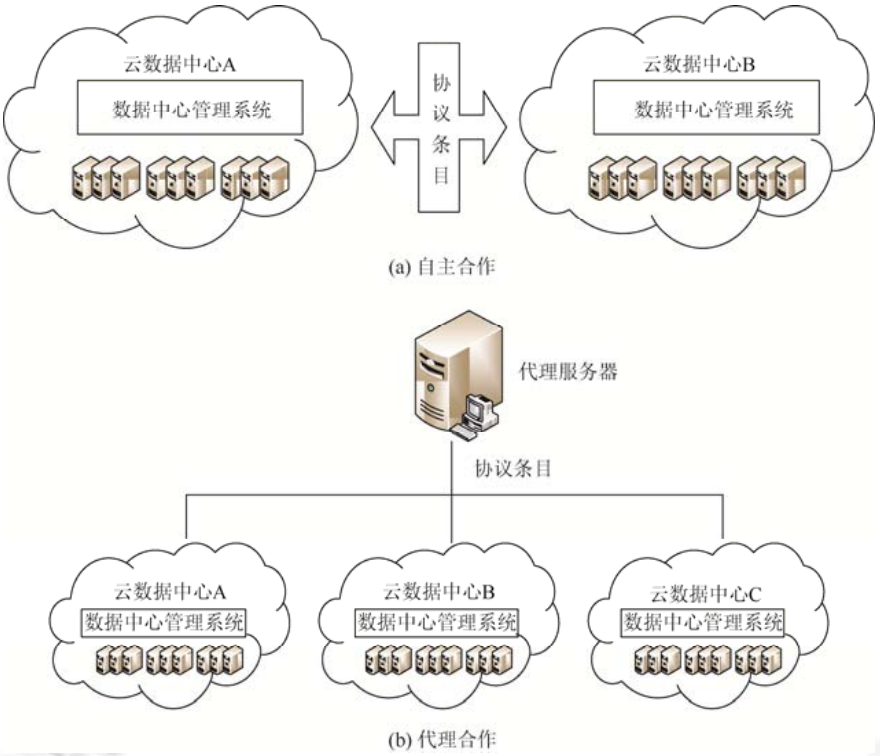


Fig.2 Peer-to-Peer architecture

图 2 对等架构模型

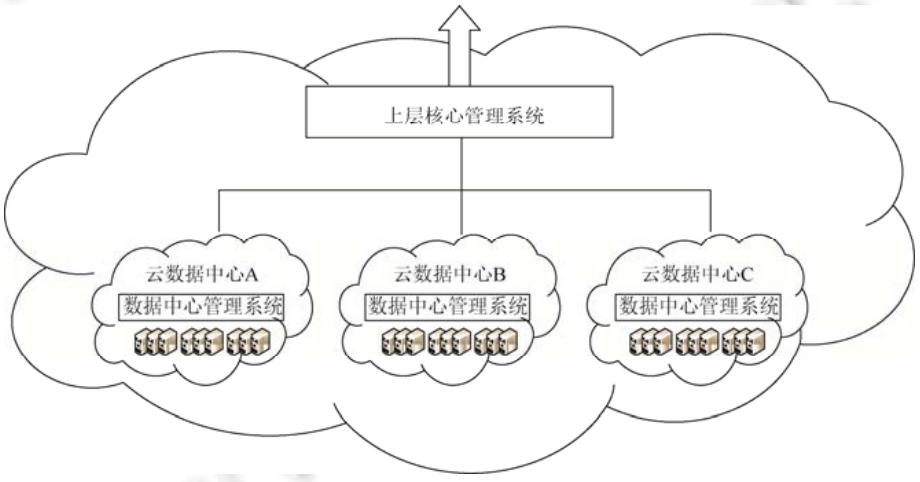


Fig.3 Aggregated architecture

图 3 聚合架构模型

综上,分布式云架构模型的对比情况见表 1.目前,混合架构与聚合架构已有广泛应用,而对等架构由于涉及多方合作等原因实际部署较少.而边缘计算、雾计算等概念的提出与发展,对数据中心分布广泛性提出要求,各方合作需求日益增长,对等架构分布式云系统也将得到更多应用.

Table 1 Comparison of cloud architecture model

表 1 云架构模型对比

模型	云数据中心类型	应用场景	耦合等级
混合架构	私有云与多个公有云资源	租用业务,需求过载期间规模扩展	松耦合
对等架构	不同 CSP 的私有云/公有云	合作业务,适应峰值需求、利用空闲资源,实现双赢	部分耦合
聚合架构	拥有多个数据中心的大型云	从属关系,提高整体可扩展性、隔离性及多站点支持	紧耦合

1.2 分布式云网络方案

分布式云系统的网络分为数据中心间网络(inter-datacenter,简称 Inter-DCN)和数据中心内部网络(intra-datacenter,简称 Intra-DCN),二者在通信规模、架构拓扑以及流量特性等方面存在较大差异.一般而言,Inter-DCN 以固定的拓扑结构搭建在互联网的骨干网中,若无明显人为干预,难以对其进行物理更新.Intra-DCN 则在网络拓扑、传输协议以及无线通信技术等方面都已有比较细致的研究[16,17],本节更多关注对 Inter-DCN 网络架构方案的探讨.

随着波分复用技术(wavelength division multiplexing,简称 WDM)在骨干网络中的广泛应用,文献[18]提出通过光路灵活构建虚拟 Inter-DCN 拓扑.此外,软件定义网络(software defined networking,简称 SDN)的兴起,也为互联网骨干服务提供了实时修正控制的可能[18−20].Peng 等人[18]利用一组全光通路基于不规则的 WDM,建立了逻辑立方体拓扑结构,如图 4 所示.在交换技术的选择方面,Inter-DCN 的流量大多是稳定的大流,同时有较松的时延要求,采用固定网格 WDM 交换技术.其控制平面采用通用多协议标签交换(generalized multiplexing protocol label switching,简称 GMPLS)技术实现 Inter-DCN 的包转发.同时其根据网络中链路的上行和下行流量特性,设计了不对称的网络以节省流量开销.

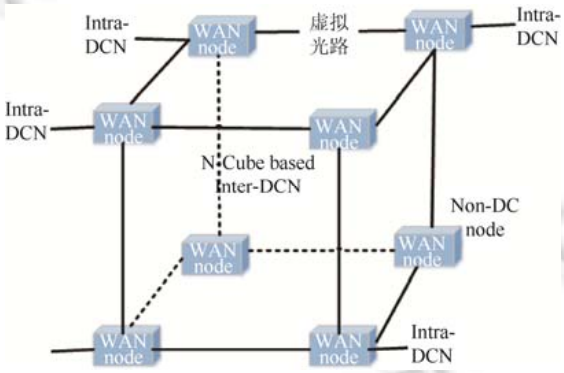


Fig.4 An integrated Intra-DCN and Inter-DCN architecture with multiple layers[18]

图 4 多层数据中心域内及域间网络架构[18]

不同于 Peng 等人设计的通用分布式云的网络架构,商业分布式云系统主要有以下特点[19]:(1) 所有应用服务和网络连接由各自数据中心统一管理.(2) 大多数带宽敏感的应用需要数据中心间的大规模数据拷贝.(3) 数据中心数目约在几十到一百数量级,给全局式的带宽管理提供了可行性.谷歌和微软公司针对其分布式数据中心的业务特点,分别设计了应用于广域网的网络方案 B4[19]和 SWAN[20],并采用 SDN 作为 Inter-DCN 网络控制平面实现 WAN 互联控制.

B4[19]通过部署全新的路由、调度、监控和管理功能及协议,简单、高效地管理数据中心间网络.在 B4 网络中,SDN 架构被分为 3 个层次,如图 5 所示.其中,交换机硬件层(switch hardware layer)仅提供最简单的路由和转

发操作.站点控制层(site controller layer)由网络控制服务器(NCS)管理的 OpenFlow 控制器及网络控制应用共同组成.全局层包含中央逻辑应用,如 SDN 网关以及 TE 服务器.这些应用控制着不同数据中心间的网络流量,通过全局控制来管理云数据中心间的带宽和数据传输.SWAN[20]不仅引入了 SDN,还将域间流量划分为背景流量和用户流量.针对其流量特点,SWAN 提出一个流量整形的策略,保证整体链路利用率恒定的情况下,通过忙时背景流量保证用户请求.此外,SWAN 还以单个数据中心作为节点,设计了整体的域间路由策略以及阶段性的动态调整策略以实现实时的链路切换,最终达到云数据中心域间带宽的有效利用.

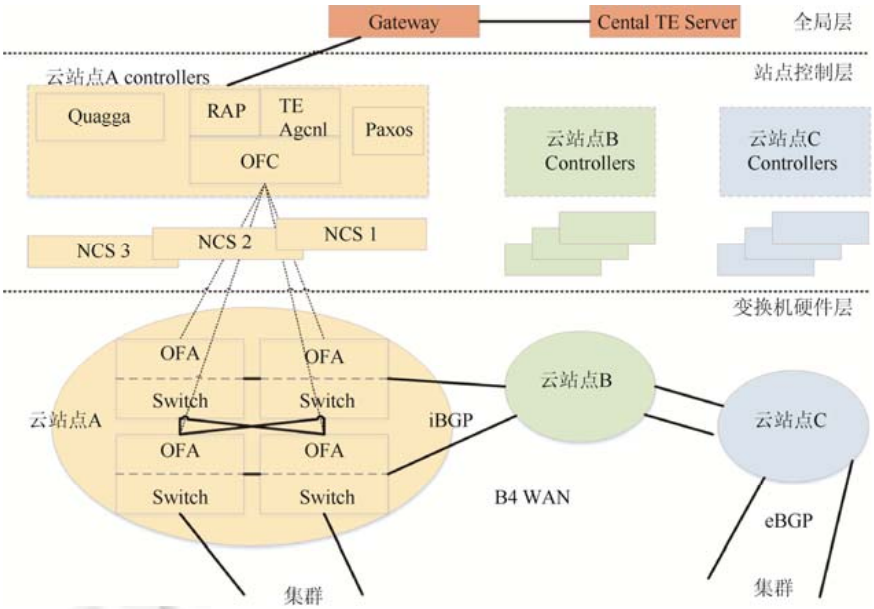


Fig.5 The architecture of B4[19]

图 5 B4 架构[19]

本节对现有的分布式云网络方案进行介绍.这些方案都将 SDN 集中控制技术应用于 Inter-DC 网络方案的控制层面设计,实现广域网的流量调度.其中,B4 和 SWAN 分别在谷歌和微软公司进行了部署.因此,虽然目前广域网的网络方案较少,且 B4 和 SWAN 分别针对的是不同公司的流量特性而设计,但由于这些系统均进行了可靠的部署和实验,可以估计的是,SDN 会在广域网的网络方案中得到更加广泛的应用.

2 分布式云的资源调度策略

根据第 1 节的介绍,分布式云架构联合多个分布在不同地区的云站点提供虚拟资源,这样的云架构不仅能够降低开销[21],也可以避免服务提供商的商业垄断[22].通过在不同的云站点上部署应用,不仅能够帮助应用提供商更好地适应其用户数据管理方面越来越严格的规定,也能够增加应用的潜在用户量,推进应用的全球化发展,为全球范围的用户提供良好的用户体验.文献[23]综合考虑了服务的响应时间、维护数据一致性的传输时延、服务的可用性以及二氧化碳的排放量等因素,设计多种算法以解决分布式云中云站点的选址问题.本节主要说明在完成分布式云的部署后,其中的资源调度策略.

分布式云系统由多个云服务提供商联合多个云站点的资源提供服务,使其不仅具有单一云节点的服务,如可扩展的服务、按需分配的资源以及灵活的计费系统(pay-as-you-go),同时还能够利用各个数据中心地理上的差异性来优化数据放置策略,提供优质服务.文献[1]指明了云环境中的用户和提供商之间的角色关系,GDC 系统一般由终端用户(end user)、云用户(cloud user)、云服务提供商(cloud provider,简称 CSP)这 3 个实体构成[1,24],其之间的关联如图 6 所示.云用户(也称为应用提供商)通过利用云服务提供商的虚拟资源来部署应用,为终端用

户提供服务.终端用户作为应用的消费者,通过接口访问应用的各个功能,其行为特征会影响应用部署方案.云提供商作为基础设施的拥有者,管理物理和虚拟资源来为云用户提供服务部署环境.其管理的应用种类一般根据云用户的需求呈现出多样性的特点.另外,特殊情况下,云用户也可作为终端用户.

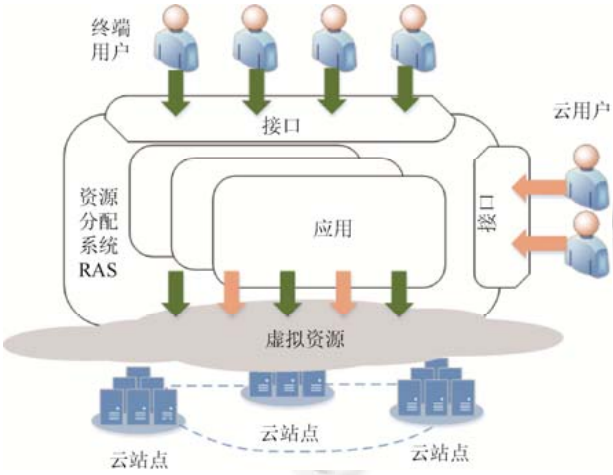


Fig.6 Entities in the cloud computing ecosystem[24]

图 6 云环境中的实体[24]

资源分配系统(resource allocation system,简称 RAS)通过整合 CSP 的物理资源和虚拟资源,将应用有效地部署到 GDC 基础设施中,为终端用户提供服务.根据应用类型的不同,RAS 的资源分配对象不仅要考虑应用服务类型,也要考虑应用的用户数据特征[25].与此同时,GDC 因其特有的架构特性,即整合不同的云站点资源提供服务的能力[24],使得 RAS 策略成为学术界比较关注且具有挑战性的课题.首先,不同的云站点拥有不同的架构、软件和硬件环境.基于应用的业务需求,RAS 需要设计合理的资源模型,屏蔽资源的异构性,混合并匹配不同云站点的资源来提供服务(即,资源建模).其次,面向用户,云服务提供商声明其提供的服务种类和资源类型,即资源提供和处理方式,能否匹配应用需求,直接决定了 RAS 满足云用户的 QoS 需求的能力.另外,资源发现和监控是 RAS 进行资源分配和优化的先决条件.由于资源需求变化或者站点故障等,高效的资源动态迁移也是 RAS 的重要组成部分.由于资源建模、资源提供和处理、资源发现和监控在文献[24]中已有充分讨论,本文则重点关注资源选择和优化、资源迁移和分配两个模块在学术界的研究进展状况.

2.1 资源选择和优化

普遍情况下,云中资源分配的优化目标包含 4 点.

(1) 保证云应用的 QoS 需求[26,27].云用户(应用提供商)作为应用服务的所有者,根据需求定义专有的服务约束和类型,如数据密集型、时间敏感型、错误敏感型等等.CSP 根据其 QoS 需求来实现虚拟资源分配.在 GDC 环境下,保证 QoS 的基本手段有,在数据中心资源紧张的情况下,获取其他云的资源满足应用的峰值需求;将应用副本部署在多个云站点中提高失败恢复的能力或访问时延.

(2) 最大化 CSP 的经济利益[25,26].从云服务提供商自身角度考虑,提高应用部署请求的通过率,增大基础设施中资源的利用率,以提高成本效益,也是其基本优化目标.

(3) 减少应用提供商的资本支出[25].例如,当其在多个云站点中放置用户数据的多个副本时,需要权衡存储开销和服务的可用性需求.

(4) 能源消耗和二氧化碳排放量.

Qureshi 等人[28]首先提出了利用不同地区的电费差异来减少分布式云系统整体的电能消耗.Rao 等人[29]设计了分散用户请求的策略来降低数据中心的用电消耗.Xu 等人[30]综合考虑不同云站点(数据中心)的电费和 ISP

带宽差异,以优化支出.对于一些特殊的应用,如典型的计算密集型应用——高性能计算应用(high performance computing,简称 HPC),在云环境中进行部署时,不考虑数据传输等问题,其优化目标往往是,在满足 QoS 的基础上,减少应用运行过程中的能源消耗和二氧化碳排放,同时尽可能地提高 CSP 的收入[31].社交网络应用的数据放置问题也需要考虑绿色节能方面的优化[25,32].

RAS 的资源选择和优化模块,也称为虚拟资源分配模块(virtual resource allocation,简称 VRA),基于以上优化目标,从资源候选集合中选择最优的资源配置,最终将应用的计算和存储等请求匹配到最适合的云站点、机架和物理主机中.由于资源的动态性以及需求的多样性,往往导致算法复杂性很高,VRA 也变成一个 NP 难问题[25,26,31].对应于具体应用场景和算法特性,本文将目前学术界提出的 VRA 算法划分为 4 个类别,即,计算节点的扩展性、请求响应的实时性、资源分配的动态性以及资源部署的鲁棒性.

(1) 计算节点的扩展性

根据计算节点的扩展性,VRA 算法可以分为集中式算法和分布式算法.这两类算法都有其独特的优缺点.集中式的解决方案一般采用一个专有的服务器来计算优化结果.其优势在于,计算实体在每一步执行过程中,都能够根据全局的资源信息来计算,其最终结果也会更逼近理想值.其劣势在于,集中式的计算实体存在单点失败的威胁,这会降低整个分配过程的可靠性.另外,该方案也存在可扩展性的问题.当资源池资源或者应用请求过多时,单一的计算节点可能会不堪重负.Volley 系统[2]为集中式算法的典型代表.其主要解决时延敏感性应用的动态迁移问题.各个云站点向专用服务器 Cosmos 提交其数据中心某应用的请求日志,以及资源和时延等约束.基于计算模式 SCOPE 以加权球形平均(weighted spherical means)方法,迭代计算最优的数据迁移结果,并将迁移建议返回给云服务提供商.其处理过程如图 7 所示.

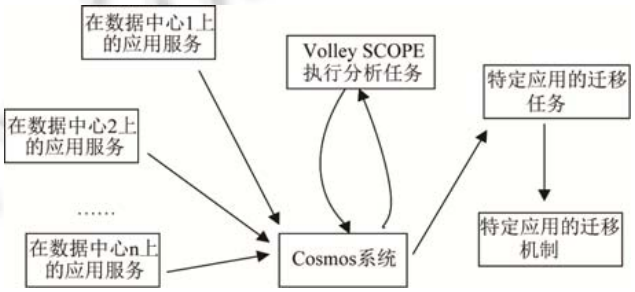


Fig.7 Dataflow for an application using Volley[2]

图 7 Volley 优化的数据流[2]

相较于集中式算法,分布式算法通过联合多个实体来计算嵌入结果.该方案明显具有更好的可扩展性,降低了每个实体的计算压力.然而,每个节点需要掌握足够的全局信息来进行优化计算,增加了同步开销.因此,这个方案一般都需要在通信开销和优化质量之间加以权衡.

在分布式云架构中,一般是两种策略的结合,即多个云服务提供商之间采用分布式结构,各自计算部分资源嵌入请求,而每个云服务商在其云站点内部,采用集中式的方法.文献[33]指出,在 Reservior 模型(属于对等模型)中,每个 IaaS 云都是一个独立自治域,当解决 VRA 问题时,每个云站点能够基于其本地的能力和远程可访问的资源状况,结合自身的管理政策来最大化 VRA 的效用.这是典型的分布式计算模式与集中式计算模式相结合的案例.

(2) 请求响应的实时性

根据能否对请求及时响应,VRA 算法分为在线算法和离线算法.在线算法是指 RAS 能够在请求达到时刻,运行 VRA 算法在较短的时间内得到放置的最优解.离线算法则是指 RAS 在假设已知所有请求的情况下,为某个请求分配最优的资源.显然,离线算法的分配结果一定程度上更优于在线算法.然而实际环境中,请求的到达往往是无法预知的,VRA 算法必须在其到达时及时反应并分配相应资源,因此,VRA 算法一般都是在线算

法.Hao 等人基于多层的云架构模型,提出了资源(VM)在线分配算法[25].在不对未来请求作任何假设的情况下,针对每个到达的请求分别构造原始方案(primal solution)和对偶方案(dual solution),联合多个资源(CPU、内存、硬盘等),应用α近似算法(α-approximation algorithm)找到 VM 的最优放置位置.为了改善在线算法的优化结果,学术界也有考虑对未来结果进行预测以校准当前的放置策略[3].而对于某些特定的应用,如 HPC 应用,用户提交一个或多个应用请求时,告知 RAS 系统其所需的执行时间和处理器数目,VRA 算法根据当前和未来请求的时间槽来匹配最优放置,因此,HPC 应用的分配算法是离线算法[31].

(3) 资源分配的动态性

根据是否能够根据请求实时变化动态分配资源,VRA 算法可分为静态算法和动态算法.其中,静态算法是指根据应用的峰值需求提供恒定资源,分配结束后,不会对虚拟资源重新分配来改善基础设施中资源的利用率.很明显,在非峰值阶段,静态算法会使得 GDC 环境中的资源利用率较低,对云应用提供商来讲是不划算的[34].同时,GDC 环境中也有很多因素会导致部分或全部虚拟资源重新分配.

① 基础设施及其资源变化.服务提供商扩展其云站点规模,或者有新的云服务提供商加入分布式云环境,这都会导致基础设施的物理资源进行更新.当前的 VRA 有可能在新的物理资源中找到更优的分配方案.另外,云中资源随着请求到期而动态释放,使得资源池中的资源出现分段,即碎片化[35].这会降低虚拟资源嵌入请求的通过率,从而降低 CSP 的收入.

② 应用需求变动.在请求的生命周期到期之前,运行应用的虚拟集群会随着应用提供商的需求发生变动,而在拓扑结构、资源规模等方面发生变化.

动态算法则是根据应用的实际需求而为其提供弹性的资源,在分配结束后,能够重新组织资源对应用的虚拟集群进行重新分配和调整[36].这提高了基础设施的资源利用率和云提供商的成本效益.同时需要注意,尽量减少资源的移动以降低迁移开销[35].对于时延敏感型应用或用户变化动态性很强的应用(如社交网络应用),往往会利用动态算法调整其数据放置位置以降低终端用户的访问时延[27,37].Wu 等人[3]分析社交流媒体应用的特性,提出了对应用的用户数据进行实时动态迁移的方案,以降低用户的访问时延.动态算法往往需要时刻监控应用变化,为了降低相应的监控测量和分析开销,Abusaki 等人[37]提出了分阶段的广域试探伸缩放置(wide area tentative scaling)策略,即将部分应用和相关数据重复性地备份到其他的数据中心,试探性地动态改变应用部署的云站点集合,根据终端代理返回的应用响应时间得到平均响应时间,不断调整,从而找到最好的集合.这样试探伸缩的方法有效地排除了精确建模中性能分析的困难.该系统的缺点是,每次试探性地进行数据迁移都会造成比较大的计算和存储开销.

(4) 资源部署的鲁棒性

根据资源部署的鲁棒性,即是否对资源进行多副本存储,VRA 算法可分为单一副本策略和多副本策略.单一副本的策略是,只为应用部署分配保证其运行的必要资源.相较于多副本的策略,其为应用分配的资源不存在冗余,一定程度上提高了资源申请的通过率.但其不能保证从失败中自动进行恢复.对于失败敏感型应用,在部署时往往创建冗余的资源来应对主节点失败的情况.然而,高可靠性的应用,需消耗更多的资源,同时也降低了云中资源请求通过率.因此从副本角度考虑,需要权衡应用的可靠性和资源嵌入开销.另外,在多副本方案中,当虚拟资源嵌入后,RAS 需要实时监控基础架构.一旦实体出现失败的情况,需要启动一个援助机制,将主节点的实例切换到其备份实例中,以保证应用正常工作.

对于社交网络类型的应用,用户之间存在大量的读写交互,其采用多副本的数据放置策略,不仅能够提高可靠性,更多的是利用地域的差异性来降低数据访问的时延[3,25].文献[25]依托于一个符合企业实际备份的 master-slave 范式的场景设计数据放置策略.根据不同用户数据的读和写的关系,建立了用户之间的关联关系.进而利用对偶思想设计固定点迭代算法分别基于 s-t 最小图割算法和贪心算法计算 master 副本和 slave 副本的放置位置,找到优化模型的最优解.

表 2 给出资源分配策略的对比情况.

Table 2 Comparison of virtual resource allocation strategy
表 2 虚拟资源分配策略对比

文献	年份	适用场景	云架构模型	方法论	在线/离线	静态/动态	集中式/分布式	单一副本/多副本
[26]	2014	一般应用	聚合架构	α近似算法	在线	静态	–	–
[31]	2013	HPC 应用	混合/对等架构	基于 Pareto 的多目标遗传算法	离线	静态	集中式	单一副本
[25]	2014	社交媒体应用	–	结合贪心和图割算法的固定点迭代方法	在线	静态	集中式	多副本
[3]	2012	社交流媒体应用	所有架构均适用	混合整数规划算法和次梯度算法	在线	动态	集中式	多副本
[27]	2012	社交网络应用	–	COSPLAY(贪心算法)	在线	动态	集中式	多副本
[38]	2010	有截止时间且不迁移的工作负载	混合架构	二进制整数线性规划	离线	静态	集中式	单一副本
[39]	2012	高性能应用	混合架构	Knowledge-Free 方法	离线	静态	分布式	单一副本
[4]	2012	独立的虚拟资源或松耦合的多组件服务(组件之间基本没有通信需求)	对等(代理)架构	二进制整数规划	离线	静态	集中式	单一副本
[36]	2013	多层服务(HPC 应用和 Web 服务集群)	对等(代理)架构	二进制整数规划	在线	动态	集中式	单一副本
[33]	2011	一般应用	对等架构(Reservior 模型)	整数线性规划,2-近似算法	离线	静态	分布式	单一副本
[40]	2012	科学应用	对等架构(Aneka)	Deadline 驱动的资源分配算法	离线	静态	集中式	单一副本
[41]	2012	支持多种类型的应用	对等架构	两阶段基于约束的算法	在线	静态	集中式	单一副本
[37]	2014	时延敏感型应用	所有架构均适用	分阶段的广域试探伸缩放置(wide area tentative scaling)策略	在线	动态	集中式	多副本
[35]	2012	时延敏感型应用	对等架构	2-近似算法	在线	动态	分布式	单一副本
[2]	2010	时延敏感型应用(Live Mesh, Live Messenger, Facebook)	–	加权球形平均(weighted spherical means)方法	离线	动态	集中式	单一副本

2.2 资源迁移策略

RAS 根据系统和应用需求实施完成资源选择和优化算法后,同时还能保证应用的可靠运行.应用运行的资源消耗往往呈现动态变化的特性,因此,当运行应用的虚拟机过载或宕机时,需要对相应的资源进行迁移来保证应用运行的质量.而云计算环境下的资源主要分为 3 个部分:计算资源、存储资源和网络资源.其中,计算资源迁移涉及到广域网环境下的虚拟机迁移问题,同时,存储资源和计算资源的迁移主要涉及到广域网网络资源的消耗问题,因此,本节首先讨论计算资源的迁移策略,其次介绍分布式云场景下网络资源的分配策略.

2.2.1 计算资源迁移

云环境中计算资源的迁移,即虚拟机迁移,分为“冷”迁移和“热”迁移两种[42],其中,“热”迁移又称为实时迁移,迁移过程中虚拟机保持原有状态,整个过程对用户透明.相反,“冷”迁移中则会明显察觉到服务中断情况.由于“冷”迁移用户体验感较差,故在研究数据中心间虚拟机迁移问题时,学术界大多关注跨云站点的实时迁移过程的实现与优化.

为保证用户的透明性,迁移前后的虚拟机在 IP 地址、MAC 地址等网络配置方面需要保持一致,这就造成传统条件下虚拟机迁移只能在同一子网中进行,严重限制了其应用范围.一些隧道技术可以连通两个虚拟层间的二层网络,但这些技术没有优化网络路由,致使通信效率低下.文献[43–45]针对解决全网通信问题提出系统模型,其中,Travostino 等人[43]提出的模型中包括“迁移支持服务器”和“代理服务器”,前者负责计算迁移最佳路径,后者负责 IP 隧道资源配置问题.该模型无需拓展宿主机或虚拟机功能即可实现全网迁移,但是由于每次迁移都需要重新搭建 IP 隧道,会消耗大量资源,为宿主机带来负担,在实用性与高效性方面存在一定缺陷.文献[44]中提出的 HyperMIP 模型为虚拟层引入 IPV4 移动性支持,可直接由虚拟层进行全网实时迁移.但该模型中虚拟层需为每个虚拟机进行包转发,且为所有迁移的虚拟机分配 IPV4 地址,同样存在宿主机负载过大的缺陷.文献[45]提出的 MIP6 模型则直接对虚拟机引入 IPV6 移动性支持,无需宿主机额外运行进程,且由于模型中每个虚拟机都独自进行网络管理,一台宿主机上可以实现多虚拟机同时迁移.相较于上面两个模型,MIP6 模型更适合进行虚

拟机全网实时迁移.文献[46,47]进一步优化了 MIP6 模型,舍弃本地代理组件,提出了 MAT 架构,保证节点间通过最佳路由通信.如图 8 所示,每个虚拟机都会有永久地址(HoA)和移动地址(MoA)两个 IP 地址,其中,永久地址专有且固定,是与外界通信的标识.MoA 则随虚拟机的位置变动而改变,两地址间的映射由 IP 地址映射服务器(IMS)组件管理.这样就保证了虚拟机面向用户网络配置的一致性,同时,由于 HoA 固定不变,通信节点间可以通过最优路由进行访问,保证了通信的高效性.

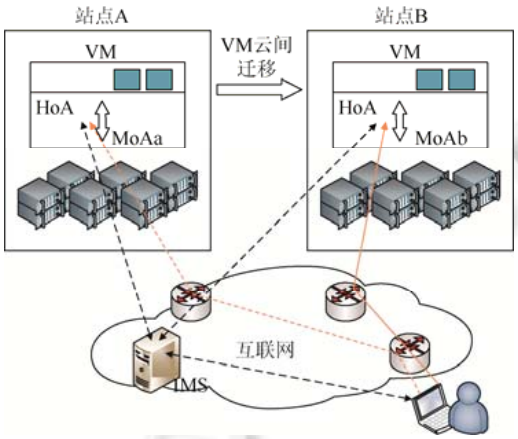


Fig.8 The architecture of MAT[47]

图 8 MAT 架构[47]

2.2.2 网络资源优化

分布式云数据中心之间的数据交互,大多数情况下都是通过广域网来实现的.相比于数据中心内部的可用带宽,数据中心间的网络带宽是非常匮乏的,存在很高的超额订购现象.并且,由于广域网链路距离非常长,网络环境多样性非常大,整个广域网的丢包和出错概率都相对较高,一般链路利用效率只有 30%~40%左右[19].为了降低虚拟机迁移过程中的网络资源消耗,Celesti 等人[48]提出一种组合镜像克隆(CIC)方式,将虚拟机从一个整体拆分开,视为由若干“可组合模块”和“用户数据模块”组成.目的站点只需要从源站点复制其缺少的“可组合模块”,结合内存及状态信息即可完成虚拟机迁移,这样大大减少了网络传输数据量.理想情况下,站点间只需传输“用户数据模块”即可.

另外,有研究表明,流量花费往往占到云服务提供商总的操作开销的 15%[49].因此,在去除了数据中心间不必要的资源传输后,如何对 Inter-DCN 网络中带宽进行合理分配也是学术界的一个研究点,其主要解决的问题是, 在 Inter-DCN 网络中如何协调具有不同特性的流量进行高效传输.对于一个云站点,其传输带宽要么用来响应用户的实时请求,要么用来实现与远程数据中心的 server-to-server 的数据传输.因此,根据时延敏感性,Inter-DCN 流量可以划分为 3 类:(1) 交互式流量,其对时延十分敏感;(2) 长传输流量,需要传输在限制的时间内完成;(3) 背景流量,没有严格的时间限制[19].Chen 等人[50]发现,在 Yahoo!公司的 Inter-DCN 流量中背景流量占主导地位,同时谷歌公司也部署了大规模的 Inter-DCN 备份服务[51],这类服务的特点是实时性要求较低.因此,利用闲置的带宽资源来传输背景流量(运行备份应用),就能够有效提高云服务提供商已购买带宽的利用率.

在分布式云系统中,云站点中应用的需求变化呈现出很强的昼夜模式,但是不同的云站点位于不同的地区,其往往具有一定的时差,这就意味着云站点 A 的峰谷时间可能与云站点 B 的峰谷时间不吻合,因此,如何高效地利用时间区间不重合的闲置带宽来传输数据,是一个亟需解决的问题.Laoutaris 等人[52]提出了 NetStitcher 系统,将批数据分块,应用存储转发(store-and-forward)算法,设置转发存储节点来对数据进行临时存储和重新路由,由此实现了对不同地区和时间的闲置带宽的有效结合.基于 NetStitcher 的存储转发思路,Jetway 系统[53]设计了应用于 Inter-DCN 网络中视频流量的路由策略.同时,相较于 NetStitcher 考虑的单一文件的传输,Jetway 在问题建模时,考虑了多视频流的共同存在现象,这更加符合现实要求.文献[54]则基于 NetStitcher,在进行调度时,考虑了

传输流的优先级,并赋予数据流严格的截止时间.Zhang 等人[55]则设计了基于截止时间的网络抽象,为租户提供确定其传输截止时间的接口,并设计了 Amoeba 系统来保证服务提供商实现截止时间内的有效调度.更加全面地,BwE[56]通过设计一个全局的层次化的系统来实现对不同优先级的流量控制转发策略.最终,该系统应用于谷歌公司的 B4 系统中.表 3 给出各个带宽分配策略的对比情况.

Table 3 Comparison of bandwidth allocation strategy

表 3 带宽分配策略对比

文献	系统	应用/场景	流量特性	方法论
[52]	NetStitcher	备份服务	背景流量	存储转发(store-and-forward)算法
[53]	Jetway	视频流量	背景流量	对视频流进行切分并路由
[54]	–	具有截止时间的数据流	长传输流量	结合带宽预留、动态调整、未来需求友好的动态调度算法
[55]	Amoeba	具有截止时间的数据流	长传输流量	时空分配算法
[56]	BwE	不同优先级的流量调度	所有流量	扩展现有的最大-最小方法,实现多路径路由和分层公平性

3 分布式云安全

一套完整的云安全系统必须是针对从硬件资源到应用软件的全方位综合方案,其中,包括硬件安全、虚拟机管理安全、客户操作系统安全、应用安全以及网络安全.随着跨云服务和多云合作相关概念和构架的出现,探究多云环境下的云安全问题也逐步成为研究热点之一.部分学者认为[57],多云环境一定情况下提高了数据的安全性,通过从应用的系统层面、逻辑层面以及用户数据层面进行分块,将其存储在分布式云系统中,能够比较有效地避免恶意的数据篡改、泄露等事件的发生.Coady 等人[21]指出,分布式云系统通过将数据存储在本地或者近邻的地方,从理论上讲,能够提高云计算环境下数据的隐私性.然而,多云合作提供的数据冗余服务在提升了服务可用性的同时,用户被迫扩大了自己的信任域,这就使得对访问权限控制和数据存储过程的安全性提出了更高要求.其中,分布式数据存储的安全加密算法在文献[58]中已给出充分研究,本文主要对云间访问控制策略展开介绍.一般访问控制主要由认证和授权构成,分布式云系统不仅考虑用户访问多云系统的安全模型,还要考虑云站点和云站点之间的信任模型及认证机制.因此,本节主要从用户信任模型、云间信任模型两方面展开介绍.

3.1 用户信任模型

在云计算环境中,数据托管在云服务器中,导致数据拥有者不能对数据和对其的访问进行直接控制,因此对不同安全等级的数据分别存储是一个比较直观的解决方案.Cascella 等人[59]提出了用户信任关系模型,并设计了一种基于隐私的以用户为中心的 personal cloud 模型.在这个模型中,用户能够直接控制其数据的存储地方,而不需要将其敏感数据的管理权力派发给公有云服务提供商.本质上,该方案通过强化本地性的方式确保了隐私性.基于同样的思路,Oliver 等人在欧盟的(scalable and secure infrastructures for cloud operations,简称 SSICLOPS)项目中允许用户指定哪些数据在本地进行存储和处理,以降低数据泄漏的风险.

扩大用户的信任域,另一种思路则是通过强化认证、密钥或者访问控制的方式,强化数据连接与存储的安全来保障隐私.在文献[60]中,作者假定系统由数据拥有者、数据消费者以及多个云服务提供商构成.将每个数据文件关联一组属性,结合基于属性的加密技术(KP-ABE)、代理重加密(PRE)和惰性重加密技术,实现了基于数据属性定义的访问政策.同时,允许数据拥有者通过更细粒度的数据访问控制,在不暴露底层数据内容的情况下,将计算任务派发给不受信任的云服务提供商,由此实现了安全的、可扩展的、细粒度的数据访问控制机制.文献[60]仅提供了对存储数据本身进行机密性保护的机制,并未涉及对数据访问信息或数据访问模式信息的保护.在文献[61]中,作者采用 B+-树的结构,并应用 cover、cache 和 shuffle 技术,设计了新的数据结构——洗牌索引(shuffle index),其假定数据被存储在叶子之间没有任何连接的 B+-树中,应用节点级别的加密技术来保证实际数据不被云存储服务器获取到.客户端(数据拥有者)不仅能够将实际请求隐藏在 cover 请求和缓存节点中,也能够将存储在服务器中的数据块进行洗牌打乱.由此包括服务器在内的任何观察者都不能重建数据块和访问数据之间的关联,确保了数据的机密性以及数据查询操作的安全性,同时能够有效保护单一或连续的访问操作.

另外,作者还验证了洗牌索引机制能够在可容忍的通信和计算开销下实现在广域网即分布式云系统中的数据访问.

3.2 云间信任模型

由第 1 节可知,分布式云根据云间的互操作性和访问权限,可以分为松耦合、部分耦合以及紧耦合 3 类.对于紧耦合的多云系统,其由一个 CSP 控制,云站点之间基本是同构的,因此云站点间一般不需要严格的信任模型.而对于松耦合和部分耦合的系统,不同的云站点往往隶属于不同的云服务提供商,异构系统之间的互操作会引入一些安全问题.因此,学术界在云间的信任模型方面进行了一定的研究.目前的解决方案按照是否引入第三方可信机构[62],分为:(1) SSO(single-sign on)认证,云站点 A 一旦获取到对云站点 B 的访问权限后,再次访问也不需要进行身份认证.(2) 数字认证和第三方,当云站点 A 要访问云站点 B 时,需要利用第三方提供的数据证书.

文献[12]基于 Intercloud 模型提出了基于公钥基础设施 PKI(public key infrastructure)的信任模型.PKI 信任模型依赖于几个 leader node 来保证整个系统的安全,同时,leader node 的认证有效性由认证中心 CA(certificate authority)设定.其中,证书不仅能够认证云站点,也可以认证云所提供的资源,因此,这需要 CA 根据云中动态的资源和负载生成相应的证书.Intercloud Root Systems 作为信任权威(trust authority)提供静态的 PKI CA,Intercloud Exchange 作为 CA 中介,提供实时的具有有限生命周期的信任证书.在该模型中,作者利用具有时间限制的 Trust Index 来表示 CSP 的信任级别(40%、50%,等等),并将不同的云站点划分成几个信任域,在同一信任域的云站点间有较高的信任等级.其认证和访问管理流程如图 9 所示.在这样一个典型的联邦认证模型中,云服务商之间建立安全通信,首先需要向对应的信任提供商请求一个信任令牌.信任提供商返回给其信任服务的加密校验令牌 P1 以及加密的请求令牌 T1,以完成后续的操作.该方案是数字认证和第三方解决方案的应用,其利用统一的规则为每个云站点分配令牌.文献[62]提出的信任与认证模型则属于 SSO 认证模型,该模型也划分了信任域,并且每个信任域由 IdP(identity provider)管理.其认证管理流程与上一个模型的差异在于,当云站点获取到某个 IdP 的信任之后,即可访问该 IdP 管理域的所有云站点,同时,该模型中也允许每个云站点使用其独立的认证机制.两个方案对比来看,文献[12]的模型中每个云站点虽然不能利用其特有的认证机制,但其考虑了分布式云系统中资源的动态性,且强调其 CA 的生命周期,具有更强的保密机制,更适用于实际场景.

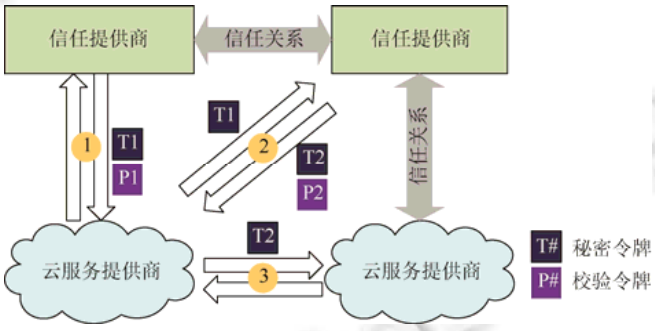


Fig.9 Intercloud identity federation model[12]

图 9 Intercloud 认证联邦模型[12]

4 发展趋势和展望

相对于目前拥有大量计算资源的集中式单一数据中心,分布式云计算的提出,主要是为了利用软件将分散在不同地理位置的数据中心的计算/存储/网络资源整合起来,通过实现数据本地化服务,为用户提供较近的计算和存储资源,有效地降低通信开销,提高用户体验和系统稳定性.本文分别从分布式云架构、资源调度与性能优化策略、分布式云安全这 3 个方面进行介绍.通过分析可以看出,现有分布式云系统中部分研究还处于初级阶段,主要原因在于各个云服务提供商出于利益等方面的考虑并未实现云间联合,更多的是云服务提供商自己

多数据中心基础设施的搭建与实现以及混合架构模型的实现.根据本文的分析,未来可以更多地关注以下几点.

1.分布式云架构模型中可进一步关注的研究点有:

(1) 对等架构各云服务提供商通过签署协议,聚合各方云资源,形成庞大的云资源.该架构是最契合分布式云提出思路的模型,但目前成果较少.云服务设计过程中未充分考虑互操作性问题,而学术界多数研究尚处于初级阶段,未真正落地.在这一方面,通过定义云服务标准接口或设置代理实现云间互操作性的问题亟待解决.

(2) 云站点间的联合对站间网络方案提出了很大挑战,网络架构设计、网络管理、带宽资源分配、一致性更新等是在数据中心网络中备受关注的研究方向,其研究成果是否可以得到沿用,而又将有怎样的改进,值得进一步关注.

2.分布式云中资源分配策略一直备受关注,在资源选择和优化算法方面有较多的研究成果.其中存在的问题主要有:

(1) 托管在云中的应用千差万别,其中有 Web 应用、视频应用、存储应用等,同时,这些应用的服务要求也存在较大的差异.因此,基于 GDC 的底层基础设施能力,面向不同应用的软件架构及需求而实现的通用资源提供平台有待设计,其中尤其需要考虑资源粒度和应用需求的匹配,以提高资源利用率.

(2) 为了保证应用严格的服务等级协议,对应用进行部署时往往需要在实验环境中测量响应时延和响应结果.很多科研人员在优化算法的研究思路上普遍集中于精确建模,如文献[3],然而公有云的性能随着具体运行的应用而变化,同时,任意两个数据中心的通信情况也有很大的差别,因此,在联合云中很难精确得到不同云站点间性能和通信延迟等参数,这会造成其中建模分析以及仿真的结果不够准确.因此在资源优化过程中,保证应用和系统性能监控分析的准确性,需要一个明确的解决方案.

(3) 多云环境中的带宽分配策略,其调度目标比较明确,根据不同流量的时延敏感需求,设计相应的算法,实现交互式实时流量、长传输流量以及背景流量的协调.然而,在目前的研究中,能够综合所有流量进行统一调度的流量工程工作比较单一,如谷歌公司的 B4 系统的 BwE 机制,还可以作进一步的探索.

3.分布式云中访问控制相对于普通的单一云场景下的访问控制,提出了用户信任模型和云间信任模型.整体来说,云安全方面的研究还比较少,有待进一步深入,目前的局限性有:

(1) 分布式云相对于单一云环境,云中的租户关系发生转变,衍生出跨云租户关系.而目前对于跨云租户间的授权访问模型研究得较少,其可能的原因是不同云服务提供商的云间联合架构没有得到实际应用和推广.未来可能的研究方向之一是,将单一云环境下的 RBAC(role-based access control)模型灵活迁移到跨云的租户访问场景下,以实现租户之间的跨云安全访问.

(2) 多云环境下的攻击模式尚不明确,这也导致了该环境下安全策略的发展缓慢.另外,云间的访问控制模型往往需要跨站点获取权限,这在一定程度上降低了云间的访问效率.因此,在保证云间访问安全性的前提下,对如何提高云间访问效率可以进行深入的探索和研究.同时,现有的用户信任模型比较简单,没有考虑用户信任域和 CSP 信任域不匹配的情况.

综上,目前分布式云研究还处于初级阶段,在分布式云领域的研究与工业界结合紧密,可以预计在未来数年内分布式云的研究会慢慢成为焦点.学术界关于分布式云的研究将成为云计算发展的有力助推器,并推动计算机网络体系架构的创新.

5 结束语

目前,越来越多的应用都倾向于部署在云平台之上,这对单一云服务提供商的服务提出了挑战.分布式云通过联合多个云站点,构建接近无限大的资源池,这一巨大优势得到了学术界的关注.本文从分布式云架构、资源调度策略以及分布式云安全这 3 个方面展开综述,以期能够为分布式云的未来研究做出一些有价值的探索.

致谢 在此,衷心感谢为本文提出宝贵意见的老师和同学.

References:

[1] Armbrust M, Fox A, Griffith R, *et al*. A view of cloud computing. Communications of the ACM, 2010,53(4):50–58.

[2] Agarwal S, Dunagan J, Jain N, *et al*. Volley: Automated data placement for geo-distributed cloud services. In: Proc. of the NSDI. 2010. 17–32.

[3] Wu Y, Wu C, Li B, *et al*. Scaling social media applications into geo-distributed clouds. In: Proc. of the INFOCOM. IEEE, 2012. 684–692.

[4] Tordsson J, Montero RS, Moreno-Vozmediano R, *et al*. Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers. Future Generation Computer Systems, 2012,28(2):358–367.

[5] Amazon. Amazon CloudFront. 2012. http://aws.amazon.com/cloudfront/

[6] Amazon EC. Amazon Web services. 2015. http://aws. amazon. com/es/ec2/(November 2012)

[7] DeCandia G, Hastorun D, Jampani M, *et al*. Dynamo: Amazon's highly available key-value store. ACM SIGOPS Operating Systems Review, 2007,41(6):205–220.

[8] Moreno-Vozmediano R, Montero RS, Llorente IM. IaaS cloud architecture: From virtualized datacenters to federated cloud infrastructures. Computer, 2012,45(12):65–72.

[9] Sitaram D, Phalachandra HL, Harwalkar S, *et al*. Simple cloud federation. In: Proc. of the 8th Asia Modelling Symp. (AMS). IEEE, 2014. 83–89.

[10] Moreno-Vozmediano R, Montero RS, Llorente IM. Multicloud deployment of computing clusters for loosely coupled MTC applications. IEEE Trans. on Parallel and Distributed Systems, 2011,22(6):924–930.

[11] Rochweger B, Breitgand D, Epstein A. The reservior model and architecture for open federated cloud computing. Technical Report, 0018-8646/09, IBM, 2009.

[12] Bernstein D, Vij D. Intercloud security considerations. In: Proc. of the 2nd IEEE Int'l Conf. on Cloud Computing Technology and Science (CloudCom). IEEE, 2010. 537–544.

[13] Tordsson J, Montero RS, Moreno-Vozmediano R, *et al*. Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers. Future Generation Computer Systems, 2012,28(2):358–367.

[14] Khanna P, Jain S, Babu BV. Cloud broker: Working in federated structures. In: Proc. of the Int'l Conf. on Advances in Computing. IEEE, 2014.

[15] Campbell R, Gupta I, Heath M, *et al*. Open cirrusTM cloud computing testbed: Federated data centers for open source systems and services research. In: Proc. of the 2009 Conf. on Hot Topics in Cloud Computing. USENIX Association, 2009. 1.

[16] Fares M, Loukissas A, Vahdat A, Scalable A. Commodity data center network architecture. In: Proc. of the SIGCOMM. Washington, 2008.

[17] Li D, Chen GH, Ren FY, Jiang CL, Xu MW. Data center network research progress and trends. Chinese Journal of Computers, 2014,37(2):259–274 (in Chinese with English abstract).

[18] Peng L. On the future integrated datacenter networks: Designs, operations, and solutions. Optical Switching and Networking, 2016,19:58–65.

[19] Jain S, Kumar A, Mandal S, *et al*. B4: Experience with a globally-deployed software defined WAN. ACM SIGCOMM Computer Communication Review, 2013,43(4):3–14.

[20] Hong CY, Kandula S, Mahajan R, *et al*. Achieving high utilization with software-driven WAN. ACM SIGCOMM Computer Communication Review, 2013,43(4):15–26.

[21] Coady Y, Hohlfeld O, Kempf J, *et al*. Distributed cloud computing: Applications, status quo, and challenges. ACM SIGCOMM Computer Communication Review, 2015,45(2):38–43.

[22] Zhygmanovskyi A, Yoshida N. Distributed cloud bursting model based on peer-to-peer overlay. In: Proc. of the 3rd Int'l Conf. on Future Internet of Things and Cloud (FiCloud). IEEE, 2015. 823–828.

[23] Le K, Guitart J, Torres J, *et al*. Intelligent placement of datacenters for internet services. In: Proc. of the 31st Int'l Conf. on Distributed Computing Systems (ICDCS). IEEE, 2011. 131–142.

[24] Endo PT, de Almeida Palhares AV, Pereira NN, *et al*. Resource allocation for distributed cloud: Concepts and research challenges. Network, 2011,25(4):42–46.

[25] Jiao L, Lit J, Du W, *et al*. Multi-Objective data placement for multi-cloud socially aware services. In: Proc. of the INFOCOM. IEEE, 2014. 28–36.

[26] Hao F, Kodialam M, Lakshman TV, *et al*. Online allocation of virtual machines in a distributed cloud. In: Proc. of the INFOCOM. IEEE, 2014. 10–18.

[27] Jiao L, Li J, Xu T, *et al*. Cost optimization for online social networks on geo-distributed clouds. In: Proc. of the 20th IEEE Int'l Conf. on Network Protocols (ICNP). IEEE, 2012. 1–10.

[28] Qureshi A, Weber R, Balakrishnan H, *et al*. Cutting the electric bill for Internet-scale systems. ACM SIGCOMM Computer Communication Review, 2009,39(4):123–134.

[29] Rao L, Liu X, Xie L, *et al*. Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment. In: Proc. of the INFOCOM. IEEE, 2010. 1–9.

[30] Xu H, Li B. Joint request mapping and response routing for geo-distributed cloud services. In: Proc. of the INFOCOM. IEEE, 2013. 854–862.

[31] Kessaci Y, Melab N, Talbi EG. A Pareto-based metaheuristic for scheduling HPC applications on a geographically distributed cloud federation. Cluster Computing, 2013,16(3):451–468.

[32] Gao PX, Curtis AR, Wong B, *et al*. It's not easy being green. ACM SIGCOMM Computer Communication Review, 2012,42(4): 211–222.

[33] Breitgand D, Marashini A, Tordsson J. Policy-Driven service placement optimization in federated clouds. IBM Research Division, 2011,9:11–15.

[34] Hu Y, Wong J, Iszlai G, *et al*. Resource provisioning for cloud computing. In: Proc. of the 2009 Conf. of the Center for Advanced Studies on Collaborative Research. IBM Corp., 2009. 101–111.

[35] Alicherry M, Lakshman TV. Network aware resource allocation in distributed clouds. In: Proc. of the INFOCOM. IEEE, 2012. 963–971.

[36] Lucas-Simarro JL, Moreno-Vozmediano R, Montero RS, *et al*. Scheduling strategies for optimal service deployment across multiple clouds. Future Generation Computer Systems, 2013,29(6):1431–1441.

[37] Yabusaki H, Nakagoe H, Murayama K, *et al*. Wide area tentative scaling (WATS) for quick response in distributed cloud computing. In: Proc. of the 2014 IEEE Conf. on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2014. 31–36.

[38] Van den Bossche R, Vanmechelen K, Broeckhove J. Cost-Optimal scheduling in hybrid iaas clouds for deadline constrained workloads. In: Proc. of the 3rd IEEE Int'l Conf. on Cloud Computing (CLOUD). IEEE, 2010. 228–235.

[39] Javadi B, Abawajy J, Buyya R. Failure-Aware resource provisioning for hybrid cloud infrastructure. Journal of Parallel and Distributed Computing, 2012,72(10):1318–1331.

[40] Vecchiola C, Calheiros RN, Karunamoorthy D, *et al*. Deadline-Driven provisioning of resources for scientific applications in hybrid clouds with Aneka. Future Generation Computer Systems, 2012,28(1):58–65.

[41] Wright P, Sun YL, Harmer T, *et al*. A constraints-based resource discovery model for multi-provider cloud environments. Journal of Cloud Computing, 2012,1(1):1–14.

[42] Cerroni W. Multiple virtual machine live migration in federated cloud systems. In: Proc. of the 2014 IEEE Conf. on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2014. 25–30.

[43] Travostino F, Daspit P, Gommans L. Seamless live migration of virtual machines over MAN/WAN. Future Generation, Computer System, 2006, 901–907.

[44] Li Q, Huai J, Li J, Wo T, Wen M. HyperMIP: Hypervisor controlled mobile IP for virtual machine live migration across network. In: Proc. of the IEEE High Assurance Systems Engineering Symp. 2008. 80–88.

[45] Harney E, Goasguen S, Martion J. The efficacy of live virtual machine migrations over the Internet. In: Proc. of the 2nd Int'l Workshop on Virtualization Technology in Distributed Computiing. Reno, 2007.

[46] Watanabe H, *et al*. A performance improvement method for the global live migration of virtual machine with IP mobility. In: Proc. of the 5th Int'l Conf. on Mobile Computing and Ubiquitous Networking (ICMU 2010). 2010. 194–199.

[47] Kondo T, Aibara R, Suga K. A mobility management system for the global live migration of virtual machine across multiple sites. In: Proc. of the 38th IEEE Annual Int'l Computers, Software and Applications Conf. Workshops. 2014.

[48] Celesti A, Tusa F, Villari M. Improving virtual machine migration in federated cloud environments. In: Proc. of the 2nd Int'l Conf. on Evolving Internet. 2010.

[49] Greenberg A, Hamilton J, Maltz DA, *et al*. The cost of a cloud: Research problems in data center networks. ACM SIGCOMM Computer Communication Review, 2008,39(1):68–73.

[50] Chen Y, Jain S, Adhikari VK, Zhang ZL, Xu K. A first look at inter-data center traffic characteristics via Yahoo! datasets. In: Proc. of the IEEE INFOCOM 2011. 2011.

[51] Ford D, Labelle F, Popovici FI, *et al*. Availability in globally distributed storage systems. In: Proc. of the OSDI. 2010. 1–7.

[52] Laoutaris N, Sirivianos M, Yang X, *et al*. Inter-Datacenter bulk transfers with netstitcher. ACM SIGCOMM Computer Communication Review, 2011,41(4):74–85.

[53] Feng Y, Li B, Li B. Jetway: Minimizing costs on inter-datacenter video traffic. In: Proc. of the 20th ACM Int'l Conf. on Multimedia. ACM, 2012. 259–268.

[54] Wu Y, Zhang Z, Wu C, *et al*. Orchestrating bulk data transfers across geo-distributed datacenters. IEEE Trans. on Cloud Computing, 2017,5(1):112–125.

[55] Zhang H, Chen K, Bai W, *et al*. Guaranteeing deadlines for inter-datacenter transfers. In: Proc. of the 10th European Conf. on Computer Systems. ACM, 2015. 20.

[56] Kumar A, Jain S, Naik U, *et al*. BwE: Flexible, hierarchical bandwidth allocation for WAN distributed computing. In: Proc. of the 2015 ACM Conf. on Special Interest Group on Data Communication. ACM, 2015. 1–14.

[57] Bohli JM, Gruschka N, Jensen M, *et al*. Security and privacy-enhancing multicloud architectures. IEEE Trans. on Dependable and Secure Computing, 2013,10(4):212–224.

[58] AlZain MA, Pardede E, Soh B, *et al*. Cloud computing security: From single to multi-clouds. In: Proc. of the 45th Hawaii Int'l Conf. on System Science (HICSS). IEEE, 2012. 5490–5499.

[59] Cascella RG, Morin C, Banâtre JP, *et al*. Private-by-Design: Towards personal local clouds [Ph.D. Thesis]. Inria Rennes, 2014.

[60] Yu S, Wang C, Ren K, *et al*. Achieving secure, scalable, and fine-grained data access control in cloud computing. In: Proc. of the Infocom. IEEE, 2010. 1–9.

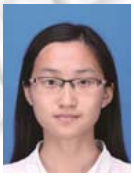
[61] Vimercati SDCD, Foresti S, Paraboschi S, *et al*. Shuffle index: Efficient and private access to outsourced data. ACM Trans. on Storage (TOS), 2015,11(4):19.

[62] Celesti A, Tusa F, Villari M, *et al*. Security and cloud computing: intercloud identity management infrastructure. In: Proc. of the 19th IEEE Int'l Workshop on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE). IEEE, 2010. 263–265.

附中文参考文献:

[17] 李丹,陈贵海,任丰原,蒋长林,徐明伟.数据中心网络的研究进展与趋势.计算机学报,2014,37(2):259–274.

张晓丽(1993－),女,山西运城人,博士,主要研究领域为云计算,网络功能虚拟化,网络安全.



孙晓晴(1995－),女,博士,主要研究领域为计算机网络,网络管理与测量,网络空间安全.

杨家海(1966－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机网络,网络管理与测量,网络空间安全,云计算.

吴建平(1953－),男,博士,教授,博士生导师,中国工程院院士,CCF 会士,主要研究领域为计算机网络体系结构,下一代互联网,网络协议工程学.