SMS 云网关负载感知弹性伸缩控制算法①

谈龙兵[1], 高 洋[2], 张成业[2], 韩曾帆[2]

[1](中国移动通信集团河南有限公司, 郑州 450003)
[2](东软集团股份有限公司, 沈阳 110179)

摘 要: SMS 短信网关(Short Message Service, SMS)具有并发量大、峰值高、难预测以及消耗资源量大的特点, 如何保障系统高效、稳定运行一直是电信运营商致力解决的关键问题之一. 本文提出了一种能够通过云计算平台动态资源控制能力, 实现负载感知的弹性 SMS 云网关设计方案及弹性控制算法, 提升了 SMS 短信网关服务质量目标保障效率, 以及资源利用率. 仿真环境下测试结果显示相比现有算法, 弹性控制算法执行效率更高, 集群规模随资源利用率抖动性更小.

关键词: SMS 短信网关; 云应用; 云计算; 弹性伸缩

Workload Aware Elastic-Scaling Algorithm for Short Message Service Cloud Gateway

TAN Long-Bing[1], GAO Yang[2], ZHANG Cheng-Ye[2], HAN Zeng-Fan[2]

[1](China Mobile Corporation(Henan Branch), Zhengzhou 450003, China)
[2](Neusoft Corporation, Shenyang 110179, China)

Abstract: SMS (Short Message Service) gateway has characteristics such as high concurrent capacity, high peaks, difficult to predict and high resources consumption. How to improve the system efficiency and stability has always been one of the key issues for the telecom operators to solve. This paper presents elastic SMS cloud gateway design and an elastic control algorithm, which can realize load sensing by dynamic resource control through the cloud computing platform, and improves SMS gateway service quality and the efficiency of resource utilization. Results of test on simulation environment show that, comparing with existing algorithms, the proposed algorithm is more efficient on scaling control, and has less resource utilization jitter on the condition of cluster size change.

Key words: short message service; cloud application; cloud computing; elastic scaling

SMS 短信网关(Short Message Service, SMS)[1]是电信网络中负责收发短信的核心系统. 在运行期短信网关负载具有并发量大、峰值高、难预测以及消耗资源量大的特点. 短信网关、彩信网关系统是移动通信网的典型业务平台, 主要定位于移动互联网 SP/CP/企业用户的接入移动短信、彩信业务的通信通道, 并对接入的业务进行有效的控制、管理. SMS 短信网关系统在移动通信网中的位置如图 1 所示.

通常, 短信、彩信网关业务平台软件可以抽象为如下主要功能:

① 服务通信接口: 为 EC/SP/企业用户其下游网元设备提供短信、彩信的通信接口, 各省网关外部网元数量通常在 2000 到 6000 不等, 外部网元众多并发量大.

② 消息路由及转发: 对收到的短彩信按照一定的路由原则寻找转发的目的网元, 并进行转发处理.

③ 消息缓存与匹配: 对于转发完毕的短彩信, 根据最终发送状态结果查找到原始的消息记, 更新其状态, 并将最终结果按照原路由返回给发送端.

④ 原始数据存储: 对于经过短彩信网关处理的消息, 存储其消息原始处理记录, 用于用户计费核查.

短信、彩信网关软件需要 7*24 小时运行在服务器

① 基金项目: 国家重大科技专项项目(2013ZX03002006)
收稿时间:2016-05-10;收到修改稿时间:2016-06-12 [doi:10.15888/j.cnki.csa.005572]

设备上, 系统整体处理访问成功率达到 99.999%, 系统故障恢复时间 <30 分钟, 满足电信级服务要求. 如何实现短信网关高效、稳定运行, 使得系统性能、稳定性保障工作极具挑战一直是电信运营商致力解决的关键问题之一. 近年来, 云计算技术的成熟与完善为解决此问题提供了新思路.

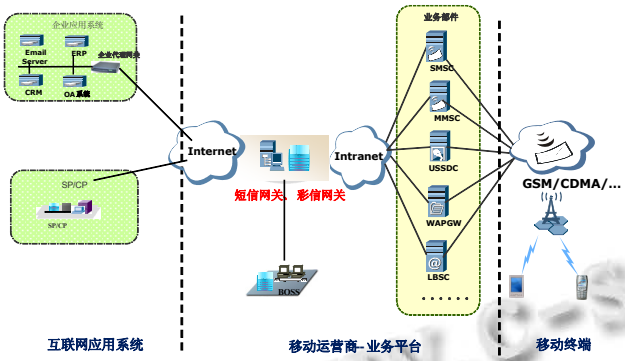

图 1 SMS 短信网关系统网络拓扑结构

美国国家标准化与技术协会(NIST)将云计算定义为一种可以通过网络以无处不在、随需即取的方式申请访问 IT 资源(如计算、存储、网络、应用、服务等)的模式[2]. 云环境下的资源利用率更高, 交付与释放的速度更快, 管理成本更低. 在此基础上, 结合我国移动通信网络建设的实际情况, 提出并建立基于于云计算虚拟化技术的移动通信网业务平台系统架构, 明确业务平台系统云化移所应具备的能力与特性, 对主流虚拟化产品进行评估, 研究资源监控及资源动态管理技术, 并针对支持动态伸缩的分布式集群、数据存储与匹配等关键技术进行分析. 中国移动对于扩充现网系统容量、建设完备容灾系统的云方案需求日渐强烈, 由此, 结合现有的网关分布式系统实现的云方案将有很大的市场前景.

1 弹性计算SMS云网关概述

云计算平台除了以 IaaS 的服务形式提供虚拟硬件资源的统一分配与管理、动态调度以及负载均衡之外, 还提供了一些诸如并行计算、分布式云存储 等通用的平台服务能力, 旨在简化平台业务开发者的开发门槛, 让开发者将精力更多地集中在问题域和业务域上.

弹性计算云平台提供的分布式计算和存储等平台服务能力具有平滑扩展、负载均衡、自动容灾等特性, 对使用者屏蔽了虚拟硬件、OS 底层以及相关实现的细节, 提供了定义良好、灵活方便的 API 供使用者直接使用. 然而, 由于网关系统运行期服务质量保障存在的特殊性问题, 解决负载感知的随需弹性伸缩问题是实现 SMS 云网关需要攻克的关键问题之一[3,4].

对资源调度问题的研究已经有数十年的历史, 出现了很多有效的调度方法. 如 Maui[5] 中实现的 FCFS(First Come First Serve)调度机制, ORCA 实现的以“租约”为核心的资源调度模式[6]. 文献[7]和文献[8]在租约交付模式的基础上提出了一种以租约为核心, 面向私有云的资源调度策略, 给出了开源实现 Haizea[9]. 文献[10]针对私有云环境中资源交付与调度的高效实现问题, 提出一种基于分裂聚类的云应用的资源交付与配置方法, 优化了虚拟机与虚拟设备资源配置. 然而, 传统的移动通信业务平台多为集中计算架构, 系统的功能、扩展性和容灾备份还主要依赖于硬件平台本身以及系统自身的架构实现方案, 并且其中的一些功能和方案在业务系统迁移到云平台上后将受到限制, 甚至是不可用. 因此研究和验证哪些业务平台功能可以通过云计算平台所提供的平台服务实现是本文探讨的一个核心问题.

在短信网关中, 每天 0 点到 7 点的这个时间段内, 短消息量比其他时间段少很多, 波谷值业务量一般只有波峰值的 3%左右, 如图 2 所示. 而每年的节假日业务高峰期又会出现不同程度的业务激增, 甚至可以达到的日均量的 10 倍, 峰值最多的省份可以达到 12000 多条/秒短信, 如图 3 所示. 按峰值业务量部署服务器节点的方式造成了服务器资源的极大浪费, 业务闲时多余的服务器开销也消耗了大量的电能.

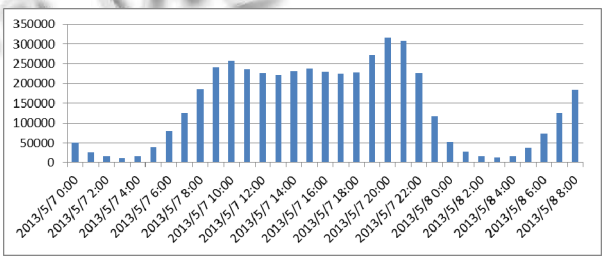

图 2 某省一天中每小时的短信业务量

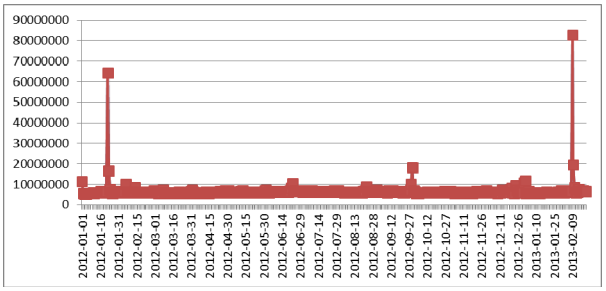

图 3 某省一年中每天的短信业务量

基于虚拟化的短信网关、彩信网关系统的部署容量可快速伸缩, 以自动适应业务负载的动态变化. 保障用户使用的资源同业务的需求相一致, 避免了因为服务器性能过载或冗余而导致的资源浪费. 网关业务量增加时, 系统自动增加核心业务模块, 以满足性能需求; 当业务量下降时, 业务模块自动卸载、模块卸载之前能够将其未处理完毕消息自动迁移到其他机器上.

2 SMS云网关系统设计

2.1 云网关解决方案

根据实际运行环境下的需求和约束, 云环境下网关系统架构需要满足如下要素:

(1) 管理节点

功能类似 Hadoop Zookeeper, 但又有不同, 功能主要功能有:

① 管理集群节点: 负责集群内网元的上线和下线的控制, 例如集群内网元数无法满足系统性能需求时, 通过集群管理节点的协调处理, 在不影响系统应用的情况下, 动态增加一个内部网元.

② 内部网元协调调度: 使各内部网元业务处理尽量均衡, 避免集群内业务处理失衡导致业务故障, 也是云计算模块弹性伸缩的基础, 如果业务量无法均衡处理, 动态增加网元也无法起到业务分担的作用.

③ 负责集群弹性伸缩管理: 管理节点实时获得各内部网元的流量信息, 了解当前系统负载情况, 这些信息是弹性伸缩算法的输入, 管理节点实时进行弹性伸缩算法判断, 下达动态增删网元指令.

(2) 动态负载分担通用代理

动态负载分担通用代理(Common Agent, CAgent)是分布式系统的基础, 兼容四层交换设备(F5、RADWARE 或开源 LVS(Linux Virtual Server)[11]), 网关架构中的 CAgent 隔离了外部网元和内部网元的链接, 与四层设备相比, 不进行连接级的负载均衡, 而是具备了业务级别的负载均衡, 以保障能更快接收到外部网元信息, 以更合理的算法向内部网元进行推送, 根据各内部网元的处理能力, 进行负载分担, 保障了内部网元的业务量均衡, 为应用系统云计算架构弹性伸缩管理提供了基础.

(3) 存储与业务的分离

基本架构层面的设计, 最简单的架构是单机是完整业务形态, 通过部署多个业务模块, 配合负载均衡代理, 完成分布式系统的研发, 配合管理节点, 完成分布式向云架构演进. 这是最理想的也是比较理想的架构形态, 但由业务形态的不同, 例如如果数据是有状态的, 即如果多个不同的数据需要放在一起进行处理, 负载均衡代理将消息无序发送到各个系统, 那么就需要分离出一些节点, 将各个数据分类汇总.

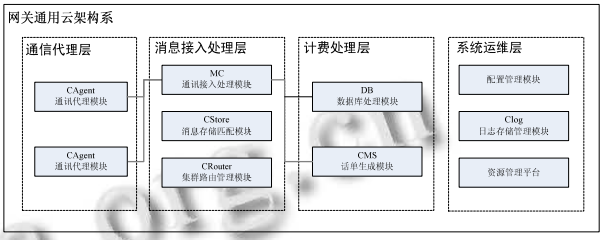

图 4　网关通用云架构示意图

为解决以上问题, 本项目提出网关通用云架构系统, 它将云概念引入网关系统架构中, 形成一套通用并具有云特性的网关新架构系统. 以适应移动互联产业的迅猛发展. 系统架构设计如图 4 所示.

2.2 云网关优势

网关云系统架构采用四层结构, 各业务处理层间具有比较独立的功能划分. 网关各功能模块采用集群部署模式. 网关架构层次化与集群化部署方式, 使网关具备水平扩展能力, 当业务数据量增大时, 可以通过增加相应模块部署数量的方式对系统的整体处理能力增容.

针对目前网关产品特性, 形成了新的网关系统架构. 新架构同时适用于三套网关系统. 网关系统架构以行业、梦网、互通网关为蓝本, 研发统一的通用架构系统. 此架构系统具有可复用性, 可将其应用于同类网关产品中. 使各网关系统拥有统一的系统架构, 方便新技术的研发、降低网关后续的维护工作量、适应于云计算技术要求.

提出了具备云计算特性的网关系统架构. 新架构支持带来超大规模、虚拟化、高可靠性等云特性功能. 本项目将云计算特性的网关架构理念融入网关系统架构的研发中, 对网关架构作了重新策划与调整. 架构更好的支持了后续“云”的演进, 即消息存储转发类系统向云特性架构演进.

3 弹性伸缩算法设计

3.1 算法原理

目前, 在计算机集群系统应用过程中, 常常为了满足峰值业务量而部署或配置相应的资源共享池(资源包括网络, 服务器, 存储, 应用软件, 服务等). 而应用系统运行过程的大部分时间里, 实际承载的业务量远远没有达到资源利用率的警戒阀值(通常定义为70%), 或者经常处于低于资源利用率的空闲阀值(通常定义为 10%)的状态.

现有的技术中, 通常采用服务器的 CPU、内存、磁盘 IO 和网络带宽等硬件和网络资源利用率的方法判断业务负载, 单纯采用其中的某些指标, 在一些特定的情况下, 不足以确定当前系统是否能够满足业务需求. 例如, CPU 利用率高可能是因为当前操作系统正在进行系统资源整理, 而不是业务量升高导致的, CPU 高无法体现出业务量高, 只要系统能够满足业务性能要求, CPU 高也可认为是一种常态; 在短信网关集群系统应用中, 通常采用预先申请固定大小和个数的消息块内存空间的方法, 以提高接收消息后系统响应速度, 不必每次都申请内存, 这样就无法通过服务器内存利用率的方式判断业务负载; 同样, 在短信网关等实时的应用系统中, 一般采用内存存储消息块的方式, 对磁盘 IO 利用率较低, 所以不能通过磁盘 IO 的方法来判断业务量; 目前, 集群内部的网络流量由于包含内部通讯过程, 不能统计出业务节点对外提供的真实网络流量, 而集群外部的路由器等网络设备常常是与其他系统共用或者根本没有权限去采集路由器信息, 也就无法通过网络带宽利用率的方式统计业务负载.

从而可以得出, 现有技术中通过服务器硬件资源利用率的方法无法真实反映业务负载情况或方法不适用于短信网关集群系统, 而通过网络带宽利用率的方法则受限于集群内部, 只能通过集群外部统计, 集群内部无法实现. 鉴于上述判断系统业务量的方法中存在的问题, 提出了一种基于服务器 CPU 利用率和实时业务量的判断算法.

3.2 评估指标定义

云网关弹性伸缩的基本输入为节点弹性伸展或缩减的指标, 表给出了算法判断所需的必要指标定义.

表 1 弹性伸缩算法评估指标

符 号	代表意义
Y	当前业务量下需要的业务节点数
(n+1)	当前用于生产的业务节点数 n 及至少 1 个冗余节点
X	动态增删业务节点数量算法判定结果, 分别为 1、-1 或 0
M	数据样本连续采集次数, 该参数控制着弹性伸缩节点的频率, 范围可设置为 10 到 100 次, 推荐取 m 为 20 次, 采样间隔为 30 秒, 即 10 分钟进行一次弹性伸缩节点的判断
CPU	节点主机的 cpu 利用率
s	节点消息存储量利用率, 本例中 12G 内存可预先申请最大消息块数为 1000 万, 通过已用消息块数除以最大消息块数即为消息存储量利用率
f	节点当前每秒消息流速
F	节点最大每秒消息流速
e	节点预测每秒消息流速, 根据历年节假日期间流速数据乘以每年业务量增长系数后估算今年流速
a	业务闲时系数, 一般采用 0.01 到 0.2 之间, 推荐取 a 为 0.1
b	业务忙时系数, 一般采用 0.6 到 0.9 之间, 本例中取 b 为 0.7

各个指标及弹性伸缩指标定义如下:

弹性伸展指标 1: 在 m 次采集结束后, 当节点 CPU 利用率大于警戒阀值, 并且消息流速也大于最大每秒消息流速的警戒阀值的次数超过 $m*b$ 次时, 则进行弹性扩展操作.

$$\sum_{i=1}^{m}\left(\mathrm{CPU}_i > b \,\mathrm{I}\ f_i > F*b\right) > (m*b) \tag{1}$$

弹性伸展指标2: 在m次采集结束后, 当节点消息存储量利用率大于业务忙时系数的次数超过 $m*b$ 次时, 则进行弹性扩展操作.

$$\sum_{i=1}^{m}(s_i > b) > (m*b) \tag{2}$$

弹性伸展指标3: 在m次采集结束后, 当节点预测消息流速大于最大每秒消息流速的警戒阀值的次数超过 $m*b$ 次时, 则进行弹性扩展操作.

$$\sum_{i=1}^{m}(e_i > F*b) > (m*b) \tag{3}$$

弹性缩减指标 4: 在 m 次采集结束后, 当节点 CPU 利用率小于空闲阀值, 并且消息流速也小于最大每秒消息流速的空闲阀值的次数超过 $m*a$ 次时, 则进行弹性缩减操作.

$$\sum_{i=1}^{m}\left(\mathrm{CPU}_i < a \,\mathrm{I}\ f_i < F*a\right) > (m*a) \tag{4}$$

弹性缩减指标5: 在m次采集结束后, 当节点消息存储量利用率小于业务闲时系数的次数超过 $m*b$ 次时, 则进行弹性扩展操作.

$$\sum_{i=1}^{m}(s_i < a) > (m*a) \tag{5}$$

3.3 弹性控制策略

当前业务量下需要的业务节点数由已部署的节点数和弹性伸缩算法决定, 节点个数需要大于等于最小

节点数配置2，并且小于等于最大节点数配置M，弹性伸缩算法运行的结果分别为1、-1和0，当满足指标1、指标2或指标3中任意一个指标时判断结果为1，即弹性扩展；当满足指标4或指标5时判断结果为-1，即弹性缩减；否则为0，即节点个数保持不变.

$$Y = n+1+X,\ M \geq Y \geq 2, M > n \geq 1,\ (Y, M, n \in N+)$$

$$X = \begin{cases} 1, & \text{指标1+指标2+指标3+任意节点异常宕机} \\ -1, & \text{指标4+指标5} \\ 0, & \text{其他} \end{cases} \quad (6)$$

3.4 算法流程设计

主控节点根据负载变化特征，定时重复运行弹性伸缩算法，根据算法运行结果，判断当前业务节点数是否满足业务量需求，如果满足需求，则不作任何处理，继续下一次判断；如果运行结果为增加一个业务节点，则向集群内任意一个未启用的业务节点发送启用指令，业务节点收到启用指令后会主动加入集群，处理业务，如图5所示；如果运行结果为减少一个业务节点，则向集群内业务量最小的节点发送停用指令，业务节点收到停用指令后，会停止处理业务，将已有业务推送给其他业务节点，主动从集群中退出，如图6所示.

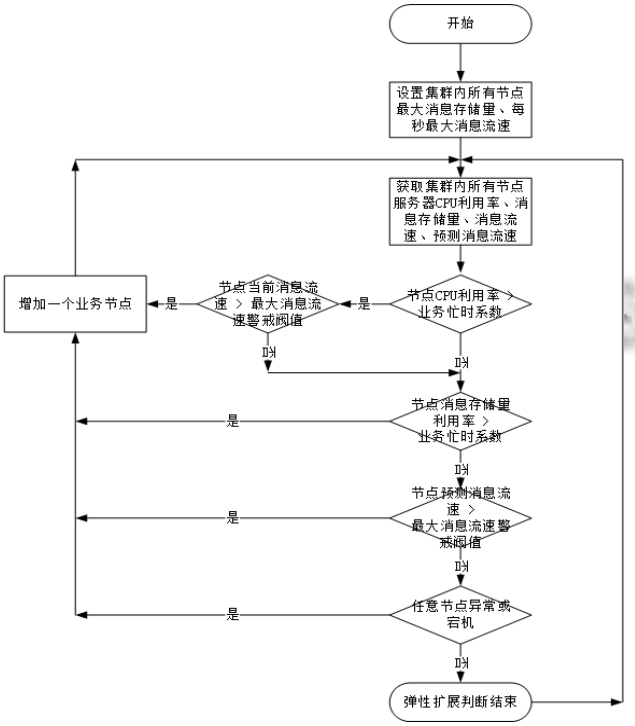

图5　弹性扩展判断流程

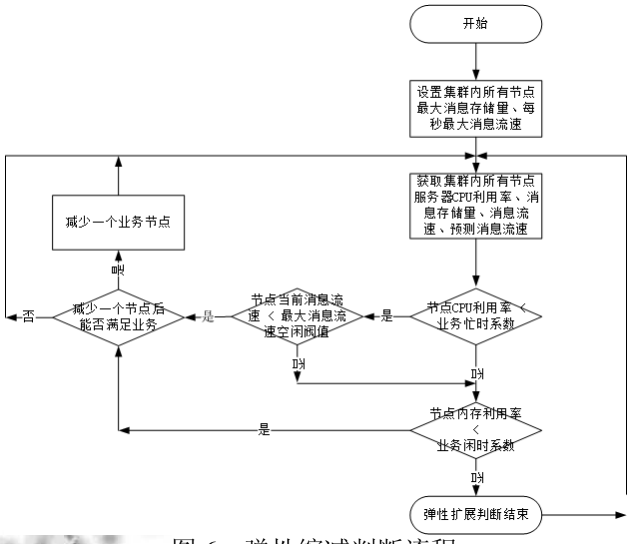

图6　弹性缩减判断流程

4　实验与结果分析

4.1 试验环境

弹性云网关的测试环境组网情况如图7所示，包括业务处理资源池、管理资源池和数据资源池，其中，业务资源池负责各个业务系统处理业务，数据资源池负责海量日志存储和查询，管理资源池负责所有虚拟化资源的分配、管理和调度. 测试环境软硬件列表如表2所示.

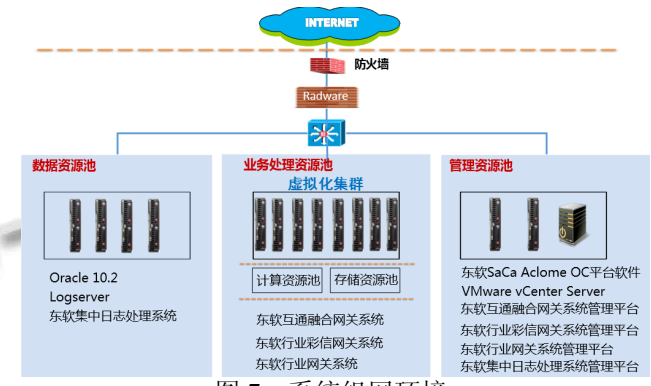

图7　系统组网环境

表2　测试环境软硬件列表

系统	硬件名称及型号	部署软件名称	数量
网关系统	HP BL460C G7	VMware vSphere Hypervisor (ESXi) 5.0 彩信网关系统(MPIAG-MMS V2.0) 短信网关系统(MPIAG-MMS V2.0)	8
	HP BL680c G5	Oracle 10.2	1
	虚拟机	SaCa Aclome OC 平台 SaCa Aclome OC V5.0	1
	虚拟机	VMware vCenter Server Appliance 5.1.0.5100 Build 799730	1

4.2 结果分析

在中国移动某省的实际生产环境里, 在 8 台主机集群环境下, 系统资源使用变化情况如图 8、图 9 所示. 其中黑色线为系统总体业务量变化曲线, 业务量从 10->5000->8000->12000 条, 考虑到容灾需求, 该产品最小规模主机数为 3, 开始时主机一、二、三启动满足业务需求, 流量增长到 8000 时, 主机四自动启动, 业务量增长到 12000 条时, 主机五、六启动, 资源利用率主机平稳.

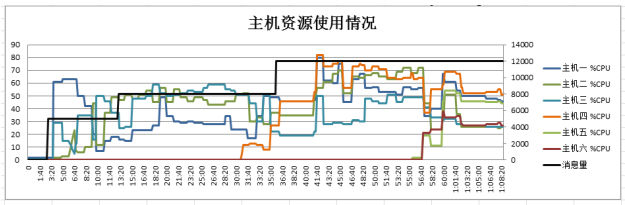

图 8 主机资源使用情况

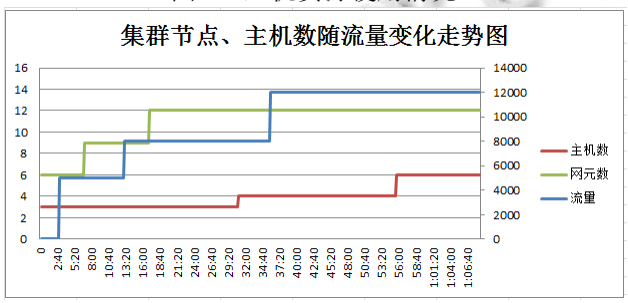

图 9 集群节点、主机数随流量变化走势图

如图 8、图 9 所示, 相比虚拟机自带的弹性伸缩算法, 该算法效率更高, 当业务量发生变化时, 本算法能及时进行模块的弹性伸缩, 满足业务需求. 该算法以业务量、积压量等作为算法判断依据, 当主机物理资源占用率变化时, 主机规模不随主机资源利用率变化而抖动.

4.3 应用效果和意义

网关通用云架构系统中各模块支持在线弹性扩展功能. 网关系统服务的规模可快速伸缩, 以自动适应业务负载的动态变化. 保障用户使用的资源同业务的需求相一致, 避免了因为服务器性能过载或冗余而导致的服务质量下降或资源浪费.

中国移动某省的短信网关利用本算法以及应用本算法的系统, 在满足业务量需求的前提下, 提高了集群节点服务器资源利用率; 节约电能, 降低服务器集群成本, 在短信网关中, 通过算法判断可以减少非节假日期间和平时每天 0 点到 7 点时间段内的服务器数量; 采用本算法后, 系统可以自适应业务变化, 无需人工干预.

5 结束语

本文重点研究了基于弹性云平台实现负载感知的 SMS 云网关核心问题. 通过构建多业务云集成验证平台, 集中部署云化的短信网关、彩信网关系统, 并加入业务群发应用系统, 进行负载均衡、过负荷控制及容灾等方案的研究与验证, 形成整体的业务平台云技术解决方案. 最后, 基于上述业务平台云化关键技术的研究与实践, 结合移动通信核心网的实际特点, 形成移动通信核心网应用云计算的负载感知弹性控制核心算法及关键技术解决方案. 通过实际环境下测试验证了算法及方案效果. 当前使用负载感知策略采用分析历史数据中负载随时间变化特征触发弹性控制, 并未与核心网关系统并发量监控实时监控数据实现联动, 因此弹性控制灵活性和有效性不高, 存在提升空间, 有待后续研究解决.

参考文献

1 Short message service. https://en.wikipedia.org/wiki/Short_Message_Service.

2 Mell P, Grance T. The NIST definition of cloud computing. Gaithersburg: National Institute of Standards and Technology, 2011-1-25.

3 Vaquero LM, Rodero-Merino L, Buyya R. Dynamically scaling applications in the cloud. ACM SIGCOMM Computer Communication Review, 2011, 41(1): 45–52.

4 Dastjerdi AV, Garg SK, Rana OF, et al. CloudPick: A framework for QoS-aware and ontology-based service deployment across clouds. Software: Practice and Experience, 2015, 45(2): 197–231.

5 Maui administrator's guide. http://www.adaptivecomputing.com/re-sources/docs/maui/mauiadmin.php, 2010-5-16.

6 Irwin D, Chase J, Grit L. Sharing networked resources with brokered lease. USENIX Technical Conference. 2006.

7 Sotomayor B, Montero RS, Foster I. Resource leasing and the art of suspending virtual machines. Seoul, Korea: The 11th IEEE International Conference on High Performance Computing and Communications. 2009.

8 Sotomayor B, Montero RS, Foster I. Virtual infrastructure management in private and hybrid clouds. IEEE Internet Computing, 2009, 13(5): 14–22.

9 Sotomayor B. The haizea manual. http://haizea.cs.uchicago.edu/, 2010.

10 许力,周进刚,张霞,谭国真.云应用资源交付与分裂聚类调度方法.计算机工程,2011,37(11):52–55.

11 Linux virtual server. https://en.wikipedia.org/wiki/Linux_Virtual_Server.