

基于服务策略的 Parlay 网关过载控制仿真研究

刘博¹ 党大伟¹ 杜新辉² 宋俊德¹

(1. 北京邮电大学 PCN &CAD 部级重点实验室, 北京 100876; 2. 新疆军区 69026 部队 69 分队, 乌鲁木齐 830002)

摘要 针对目前电信业务网络的业务差异性大、业务突发性高等造成系统过载现象,研究了未来 NGN 网络的业务网关-Parlay,及目前的系统过载控制技术。结合下一代网络中 Parlay 网关的特征,根据服务质量 QoS 对业务进行分类和建模,给出了基于服务策略的业务过载控制模型,并详细描述了过载控制算法。仿真试验结果表明,该算法在保证业务服务策略的基础上,有效控制了业务应用服务器的过载现象,并具有一定的公平性。

关键词 Parlay; 业务模型; 过载控制; 服务策略; 下一代网络

中图分类号 TP 393

文章编号 1007-4333(2007)02-0076-04

文献标识码 A

Simulation of Parlay overload control based on service policy

Liu Bo¹, Dang Dawei¹, Du Xinhui², Song Junde¹

(1. Department Key Lab. of PCN &CAD, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. 69 Unit of Army 69026, Military Area of Xinjiang, Urumqi 830002, China)

Abstract Based on the phenomena of current Telco service network system overload caused by service differences and strong burst, a technology of service gateway of the next generation network-Parlay and its overload control mechanisms are presented in this paper. A service overload control model based on service policy was established with consideration of operating features of a Parlay Gateway. Telco services were classified and modeled by QoS, and the overload control algorithm was analyzed in detail. Based on the simulation results, we concluded that the algorithm could be used for controlling system overload based on service policy effectively and fairly.

Key words Parlay; service model; overload control; service policy; NGN

下一代网络(NGN)是以 IP 协议为基础、数据为中心、业务为驱动力,综合传统的固定网、移动网、Internet 网以及广电网的统一开放网络,因此快速高效地开发和部署各种综合业务服务于终端用户十分重要。传统的电信业务开发对网络依赖严重,要求 IT 开发人员不但要熟悉底层异构网络和复杂的信令交互,而且要考虑业务的平台跨越问题;因此,如何打破封闭的传统电信业务生成模式,加速个性化 3G(第三代移动通信)业务开发和部署的过程,已经成为各电信运营商以及标准化组织的研究焦点。由 Parlay、3 GPP、ETSI 组织联合成立的 Joint Work Group 共同制定了 Parlay/OSA/ETSI 技术标准^[1-2],该技术的产生摒弃了传统封闭的电信业务

生成模式,使 IT 开发人员不需要熟悉底层异构网络以及复杂的信令交互就可以利用电信网络资源进行增值业务的开发,从而成为业界广泛认可的下一代业务网络的核心技术。

Parlay 的业务逻辑结构主要包括业务应用服务器(application server,简称 AS),业务能力服务器(service capability server,简称 SCS)以及 parlay 框架/framework)。Parlay 的业务能力服务器的功能可以视为一个通向核心网络的代理网关,负责抽象网络资源,屏蔽底层电信设备的差异,并向业务服务器开放各种网络资源,包括:呼叫控制、用户交互、移动位置服务、计费等^[3],这些网络资源被定义为标准的 Parlay API 接口。业务应用服务器是实现业务逻辑

收稿日期:2006-09-18

基金项目:国家自然科学基金资助项目(60372098)

作者简介:刘博,工程师,博士研究生,主要从事 Parlay/OSA、IMS 技术和 NGN 业务平台方面的研究,E-mail:hongmingliu@vip.sina.com

辑的主体,通过业务能力服务器提供的标准 Parlay API 接口使用下层的网络资源,向终端用户提供最终的业务服务。Parlay 框架为业务应用服务器提供接入电信核心网络所需的安全、管理和控制等功能,并支持业务能力服务器到 Parlay 的注册。作为下一代网络中电信级的设备,大量的业务请求和处理很容易形成瓶颈,所以 Parlay 结构需要高效的过载控制功能,而未来业务种类繁多、差异大、突发性高、服务策略不同,使其业务过载控制远远复杂于传统的智能网^[4-9];因此如何在新一代业务网络环境下提供有效的业务过载控制能力已成为当前的研究焦点。为此,本文中针对下一代网络中 Parlay 网关过载控制的需求,结合当前运营商的业务服务策略,设计了下一代业务网关过载控制模型,并给出详细的算法描述。

1 过载控制技术及管理模型

传统电信行业的过载控制技术起源于智能网建设初期,即公用交换电话网(PSTN)和窄带综合业务数字网(N-ISDN)实现智能业务时,SCP(业务控制点)的业务量过载^[4,8-9]问题已引起人们的重视。在 IN CS21 标准中,针对现有 PSTN 和 ISDN,使用 CG(Call Gap,呼叫间隙)算法来实现 SCP 的业务量过载控制,一般认为 CG 与 Percent 算法、令牌桶算法等属于接纳控制算法。另外一种算法叫做过载检测算法,主要包括:基于队列长度的检测、基于呼叫数目的检测、基于平均响应时间的检测等。应用实际表明,一个好的过载控制算法必须满足以下要求:

- 1) 有效性。也就是说,在负载非常重的情况下,Parlay 网关的吞吐量不会下降。
- 2) 公平性。在多业务环境中,当某类业务应用服务器过载时,Parlay 网关不能对没有产生过载的业务进行限制。
- 3) 对运营策略的支持。业务运营者可根据自己的需要,如每种业务的处理时间和从每种业务得到的收益,规定在过载情况下各种不同业务的处理比率。
- 4) 对不同的过载模块、不同的业务流、不同的用户行为(用户重试)和 Parlay 结构都适用。
- 5) 算法必须易于实现。

在传统的单业务环境中检测系统是否过载时,可根据消息的排队长度、平均到达率、业务应用服务

器 CPU 利用率和消息队列是否有消息排队超时来进行,类似于传统智能网中的 SCP^[10-11]以及互联网的 Web Server。Parlay 网关在业务提供方面起着承上启下的作用^[12],其过载控制的重要地位是显而易见的,而且 Parlay 网关需要运行的业务种类更多,除基本的两方呼叫外,还包括多方呼叫、消息类业务、位置业务等其他业务形式。这些业务具有不同的 QoS(服务质量)要求,呼叫类业务有较高的实时性要求,而消息类业务时延要求较低。业务请求的发起者也增多,除接受下一代业务网络 IMS 发起的业务请求外,来自应用服务器侧的互联网业务请求也占相当的比例,另外,基于国家和政府应急安全通信请求也成为下一代网络业务请求的重要组成部分。业务服务策略与运营需求绑定得越来越紧密,针对不同时间(如节假日、重大事件发生时刻)、地点等业务服务的比例因子需要灵活调整,因此过载控制问题更为复杂。在这种多业务环境中,由于每种业务的处理时间存在差异,基于消息排队长度和消息到达率的方法不再适合,相比较而言,基于业务应用服务器 CPU 利用率和消息队列是否有消息排队超时 2 种方法比较适合。考虑到实现的复杂度,本文中通过业务应用服务器 CPU 利用率的大小来判断系统是否过载。针对 Parlay 网关的本身特性,给出过载控制模型(图 1)。

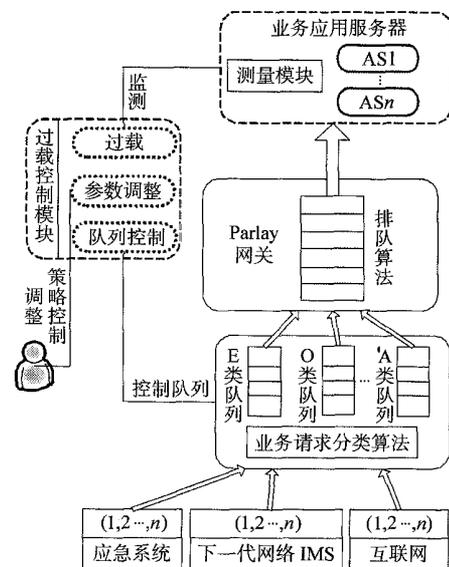


图 1 Parlay 网关过载控制模型

Fig. 1 Parlay gateway overload control model

业务请求分类模块将来自国家应急安全通信系统、下一代网络 IMS 以及互联网的业务请求分为紧

急事件业务请求(E类业务)、具有QoS的业务请求(Q类业务)和尽力而为的业务请求(B类业务),不同的业务请求放到不同的队列中排队,过载控制模块通过漏桶算法的参数 R 对业务呼叫队列进行控制。不同业务请求根据Parlay网关的队列调度算法在网关进行排队,然后通过Parlay API请求业务应用服务器。系统测量模块负责根据算法需要对CPU利用率、不同优先级业务占用的CPU资源,以及业务到达率等参数进行测量。过载检测按照一定的检测算法检测系统的负载状态,它不仅负责判断过载状态的开始,还负责判断过载状态的结束,并与过载控制和队列调度交互以完成过载控制功能。参数调整模块主要根据运营商的业务服务策略及时调整业务服务比例因子。过载控制的主要任务是,当检测到过载时对业务呼叫队列进行控制。

2 过载控制算法及仿真

业务应用服务器针对不同的业务请求进行相应的服务,不同的业务对应不同的系统进程,其CPU占用率设为 i 。处理的业务请求量较小时, i 值比较低,随着请求量的增加,它会逐渐增大,业务完成率也相应增加;但达到一定程度后,随着请求量的进一步增大而出现过载时,业务完成率反而下降。其原因是一方面CPU需要丢弃排队时间过长的消息,另一方面消息的响应时间也会增加,导致很多消息因为超时而失效。由此可见,处理进程存在一个最佳值 i^* ,当 i 超过 i^* 时,业务应用服务器已经达到处理极限,系统开始过载; i 小于或接近 i^* 时,系统不过载。系统过载如果是单种类型的业务造成的,只需对该种业务请求进行微调控制;如果是多种类型业务造成的,则需要统一对各种业务请求进行控制。过载控制算法的变量声明如下:

μ_i 为第 i 类业务的服务速率; r_i 为第 i 类业务的服务比例因子, $i=1, \dots, n$; $E_i(k)$ 、 $Q_i(k)$ 、 $B_i(k)$ 分别为第 k 个周期 T 紧急优先级类、具有QoS类、尽力而为类业务 i 的到达速率; $E(k)$ 、 $Q(k)$ 、 $B(k)$ 分别为第 k 个周期 T 紧急优先级类、具有QoS类、尽力而为类业务占用的应用服务器CPU资源的利用率; $\rho(k)$ 为业务应用服务器CPU资源的总体利用率。

1) 过载控制主算法。

步骤1:算法初始化,环计数器 k 置为1。

步骤2:在周期 T 时间内,利用应用服务器的测量模块对各类业务占用的CPU资源的利用率进行测量计算,获取的利用率分别为

$$\begin{aligned} E(k) &= \sum_i E_i(k) / \mu_{Ei} \\ Q(k) &= \sum_i Q_i(k) / \mu_{Qi} \\ B(k) &= \sum_i B_i(k) / \mu_{Bi} \end{aligned}$$

计算业务应用服务器CPU资源的总体利用率 $\rho(k) = E(k) + Q(k) + B(k)$ 。

步骤3:根据CPU资源利用率的门限值、当前各类业务服务策略的比例因子进行业务过载的检测和控制,具体流程如下。

IF ($B(k) < B_r^*$) && ($Q(k) < Q_r^*$) && ($\rho(k) < \rho_r^*$)

系统没有过载, $k = k + 1$,转到步骤1;

ELSE IF ($B(k) > B_r^*$) && ($Q(k) < Q_r^*$) && ($\rho(k) < \rho_r^*$)

尽力而为类业务过载,调用尽力而为类业务速率微调算法;

ELSE IF ($B(k) < B_r^*$) && ($Q(k) > Q_r^*$) && ($\rho(k) < \rho_r^*$)

QoS类业务过载,调用QoS类业务速率微调算法;

ELSE IF ($\rho(k) > \rho_r^*$)

业务应用服务器系统过载,调用业务速率控制算法。

2) 业务速率微调算法。

根据运营商的业务规则,依照当前各类业务的比例因子对该类业务的呼叫接入速度进行微调。由于尽力而为类业务与QoS类业务的微调算法逻辑上是一致的,故只描述尽力而为类业务微调算法。

$$B(k) = \rho_r^*(k) - E(k) - Q(k)$$

IF [$B(k) / \rho_r^*(k) < B$] 新的参考输入速率 $B = B \times \rho_r^* \times \mu_B$;

ELSE IF [$B(k) / \rho_r^*(k) > B$] 新的参考输入速率 $B = B \times B(k) \times \mu_B$ 。

3) 业务速率控制算法。

根据对紧急类业务的定义,可以认为只此一类业务并不能使系统处于过载状态,从而 $\rho_r^* - E(k) > 0$,算法描述如下。

步骤1:在保证处理紧急类业务能力的前提下,获取可用的系统资源

$$r^*(k) = r^* - e(k)$$

步骤 2: 依照各类业务的比例因子计算该类业务新的参考输入速率。尽力而为类业务的参考输入速率

$$B = \frac{B}{B+Q} \times r^*(k) \times \mu_B$$

QoS 类业务的参考输入速率

$$Q = \frac{Q}{B+Q} \times r^*(k) \times \mu_Q$$

采用上述 Parlay 网关过载控制模型进行仿真试验, 假定业务请求包括 E 类、Q 类、B 类 3 种业务, CPU 参考过载利用率 $r^* = 95\%$, $B_r^* = 40\%$, $Q_r^* = 60\%$, 对 E 类业务不作限制, 系统忙时对 B、Q 2 类业务的服务比例为 2:3; 过载控制模块的检测周期 $T = 10\text{ s}$ 。从图 2 可以看出, 该过载控制模型很好地控制了应用服务器 CPU 的利用率, 并在全面支持 E 类业务的同时, 保证了业务运营的比例因子。

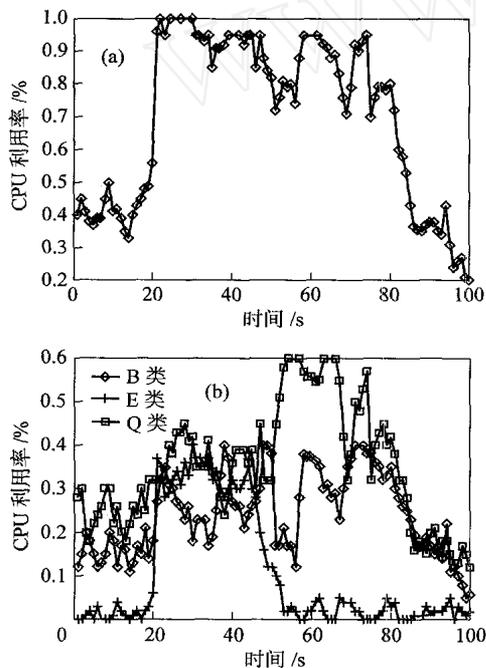


图 2 Parlay 网关过载控制模型仿真试验结果:(a) 总 CPU 利用率;(b) 各类业务 CPU 占有率

Fig. 2 Simulation result of Parlay gateway model: (a) utilize ratio of CPU; (b) each service utilize ratio of CPU

3 结束语

与传统智能网的业务过载控制模型相比, 本研究充分考虑了下一代网络业务多样性的特点, 根据

业务服务质量的不同建模, 相应的业务过载控制措施也不同; 此外, 根据运营商的现行业务提供策略设置了业务服务比例因子, 从而增强了算法对现有网络的适应性。试验结果表明, 该算法具有良好的有效性和公平性。当然, 未来业务网络对用户的服务可能需要软交换、IMS、媒体网关、媒体服务器、数据服务器等设备协同工作, 任何设备的过载都将会造成网络的拥塞和业务服务能力的下降, 这需要进一步的研究。

参 考 文 献

- [1] 董斌, 苏森. 融合电信网络和企业应用的统一业务体系结构研究[J]. 计算机工程与设计, 2004, 11(11): 1869-1875
- [2] The Parlay Group Web Services Working Group. Parlay Web Services architecture comparison: version 1.0[EB/OL]. [2002-10-31]. <http://www.parlay.org/specs/ParlayWebServicesArchitectureComparison1-0.pdf/>
- [3] ETSI ES 203 915-1. Open Service Access (OSA) [J/OL]. Application Programming Interface: Part 1: Overview, 2005-04
- [4] Kawahara R, Asaka T. Overload control for intelligent networks based on an estimation of maximum number of calls in a node [J]. Intelligent Network Workshop IEEE, 1996, 1(3): 364-369
- [5] Stathopoulos V M, Polykalas S E, Venieris I S. An algorithm for the efficient utilization of service control capabilities in intelligent networks [C]. Proceedings of 10th Mediterranean Electrotechnical Conference, Piscataway (NJ, USA): IEEE, 2000: 102-106
- [6] Venieris I S, Hussmann H. 智能宽带网[M]. 廖建新, 龙元香, 王晶, 译. 北京: 人民邮电出版社, 2001
- [7] Smith D E. Ensuring robust call throughput and fairness for scp overload controls [J]. IEEE/ACM Transaction on Networking, 1995, 3(10): 538-548
- [8] 李彤红. 移动智能网设计及过载控制问题研究[D]. 北京: 北京邮电大学, 1998
- [9] Northcote B S, Smith D E. Service control point overload rules to protect intelligent network services [J]. IEEE/ACM Transaction on Networking, 1998, 6(2): 71-81
- [10] 王玉龙, 廖建新. 移动智能网 SCP 多业务环境下的过载控制研究[J]. 电子学报, 2005, 10: 1849-1852
- [11] 刘彦明, 马玉祥, 易克初. 多 SCP 智能网的负载控制模型与算法[J]. 通信学报, 2003, 24(2): 125-131
- [12] Ard-Jan Moerdijk, Lucas Klostermann. Opening the networks with Parlay/OSA: standards and aspects behind the APIS[J]. IEEE Network, 2003, 17(3): 58-64