

NGN 中基于策略控制的过载控制体系结构

王开西, 杨放春

(北京邮电大学网络与交换国家重点实验室, 北京 100876)

摘要: 给出特定条件下过载控制的含义, 分析了下一代网络(NGN)过载控制的需求, 在 ETSI TISPAN 组织提出的 NGN 过载控制体系结构的基础上增加了策略组件, 方便业务的开发、部署和管理, 实现了在 NGN 中快速、经济、有效地开发业务, 给出 NGN 过载控制组件的 4 种部署方案, 分析这些组件协作完成过载控制的过程和相关信息流。

关键词: 下一代网络; 策略控制; 过载控制; 业务开发; 服务等级协定

Overload Control Architecture Based on Policy Control in NGN

WANG Kai-xi, YANG Fang-chun

(State Key Laboratory of Networking & Switching Technology, Beijing University of Posts & Telecommunications, Beijing 100876)

【Abstract】 With defining the overload control under the given situation, this paper analyzes the requirements for NGN overload control, and enhances the overload control architecture proposed in an ETSI TISPAN draft by adding SLA-based policy component. It facilitates the development, deployment and management of services and makes services development quick, economical and effective in NGN. Four deployments about NGN overload control component are discussed, and the control procedure is given, as well as the information flow.

【Key words】 Next Generation Network(NGN); policy control; overload control; service development; Service Level Agreement(SLA)

下一代网络(Next Generation Network, NGN)是业务驱动的网络, 如何快速、经济地向用户提供业务是各运营商首先要解决的问题, 因此, 如何缩短业务开发周期是关键所在。目前已有很多改善业务开发环境、缩短业务开发周期的方法和技术, 如图形开发技术、API 或脚本开发技术。此外, 如何将业务的核心业务特征和其所需的非功能特征(如网络传输需求)分开也是促进业务开发的有效手段, 因为这样有利于业务开发者专注于开发业务功能特征, 而无须考虑业务所需的 QoS 要求、服务等级协定(Service Level Agreement, SLA)等网络传输和运营要求。NGN 向用户提供具有质量保证的服务, IP 多媒体系统(IP Multimedia Subsystem, IMS)是其提供多媒体业务的子系统。IMS QoS 管理体系结构基于区分服务、集成服务及 MPLS 流量工程等技术, 采用策略方式为用户提供具有质量保证的传输服务, 但不能保证对 IMS 各功能实体的请求不超过其处理能力。针对 3GPP QoS 管理体系结构和核心网络带宽相对过剩的现状, 本文的过载控制有 2 层含义: (1)各功能实体的过载控制, 即请求总数超过 IMS 服务实体的处理能力; (2)业务合约的过载, 即某个业务的请求总数超过业务提供商和网络运营商之间的 SLA 合约。后者在 3GPP QoS 管理体系框架内是无法解决的。过载控制作为独立于业务的非功能特征, 是所有业务运行时必需的。

1 NGN 过载控制的相关研究

过载控制被认为是网络建设中必须考虑的问题。人们在研究了传统电信网、基于 IP 的因特网和提供智能业务的智能网中的过载控制^[1-2]后, 提出了基于请求队列长度、响应时间、处理时间差等不同的过载控制检测方法, 通过自动呼叫限制、比例限制和漏桶算法等方法限制呼叫到达率。NGN 分布式、相对扁平的体系结构和多协议、提供多种业务的特点使过载控制成为当前 NGN 的研究热点。

目前, 对 NGN 过载控制的研究主要集中在点过载方面, 如文献[3]提出针对软交换的控制算法: 基于预测的自适应接入控制算法和多目标接入控制算法。文献[4-5]对 Parlay 网关的过载控制进行了研究, 前者以在 AS、业务能力服务器(SCS)和网络的总呼叫数或仅以 SCS 中的呼叫数为负载水平, 根据负载水平静态或动态地拒绝呼叫; 后者提出基于运营收入最大化的 Parlay 网关过载控制模型和基于 Agent 线性规划理论的过载控制系统结构, 并通过实验验证模型具有较好的可伸缩性, 提高了控制灵敏度。文献[6]基于策略模式设计了 NGN 业务逻辑执行环境的过载控制框架结构, 提出了基于施密特触发器原理和 ATM 的信元同步 2 种不同的过载检测方法, 以 CPU 资源为稀缺资源, 改造令牌桶算法来保证不同优先级的业务具有提供一定质量保证的服务。

NGN 使用多种协议提供业务, 各协议对过载控制的支持程度不同^[7]。IMS 子系统基于 SIP 协议, 目前对该协议过载控制的研究主要有草案^[8], 但只停留在需求层次上; 有些 SIP 扩展由于自身的特殊性而单独考虑了过载控制问题, 如 SIP 消息扩展; 对媒体网关和媒体网关控制器的过载控制主要在 ETSI TISPAN 组织内进行。除了研究具体协议的过载控制外, 标准化组织对 NGN 通用过载控制机制的研究主要有 ETSI TISPAN 组织和多业务论坛(MSForum)的 2 个草案^[7,9]。它们分析了 NGN 网络过载控制的必要性和需求, 考虑了 NGN 网络体系结

基金项目: 国家“973”计划基金资助项目(2003CB314806); 国家杰出青年科学基金资助项目(60125101); 高等学校博士学科点专项科研基金资助项目(20020013004); 教育部长江学者和创新团队发展计划基金资助项目“通信网的网络理论和技术”

作者简介: 王开西(1971-), 男, 博士研究生, 主研方向: 通信软件, 智能网, 下一代网络; 杨放春, 教授、博士生导师

收稿日期: 2007-02-10 **E-mail:** kxwang@qdu.edu.cn

构和通信协议对NGN过载控制的影响,并提出设计原则。ETSI TISPAN草案提出不同组件构成的过载控制体系结构^[7]后,比较这些组件不同部署方案的优劣,并提出适合NGN网络的通用过载控制应用协议(GOCAP)。本文在ETSI TISPAN提出过载控制体系结构的基础上,为使NGN能快速、经济、有效地提供业务,提出了基于策略控制方式实现过载控制的方案。

2 NGN 对过载控制的需求

上述 2 个草案^[7,9]详细分析了过载控制对于NGN的必要性,综合考虑多种因素后认为,过载控制需要一种自动、快速、有效、经济、可配置、可扩展的方案。基于NGN的分层、分布式网络体系结构和多协议、多服务的特点,NGN过载控制的需求有:(1)有效性:最大化系统的有效吞吐量;(2)自动性:自动检测系统过载并启动过载控制(如会话释放请求导致的过载);(3)优先级:区分对待不同优先级消息(如释放会话消息、会话中消息、创建会话消息),区分可拒绝消息与不可拒绝消息;(4)可配置:对响应时间、交叉控制到达稳态、过载控制组件等配置;(5)协作性:与按需呼叫路由、负载均衡和负载分发的协作;(6)公平性:竞争的业务请求流;(7)约束性:根据SLA过滤业务请求;(8)适应性:适用于业务提供者内和不同业务提供者间,也可用于NGN的一个子系统内和在NGN不同子系统间;(9)可观察性:过载状态和过载控制结果的可观察性。

3 基于策略的过载控制体系结构

为满足上述需求和 NGN 过载控制的设计原则,文献[7]提出的过载控制体系结构由下列组件构成:过载检测和自适应限制组件(M),过载信息分发组件(D),过载控制执行组件(R),内部接纳和负载控制组件(A)等。该体系结构基于反馈机制具有自适应性的特点。过载控制除了具备快速、自动、有效等特点,还应便于管理人员根据不同情况部署过载控制策略,并根据 SLA 控制负载,因此,本文基于文献[7]给出的体系结构,增加了策略机制,在服务 and 用户之间提供了一种表达合约的机制,即根据业务提供者和业务平台之间的服务合约以及系统环境对业务请求进行接纳控制。这里的服务可能是为最终用户提供的服务,也可能是系统某层为上层提供的服务;用户可能是最终用户,也可能是使用系统某层服务的上层实体。如图 1 所示。

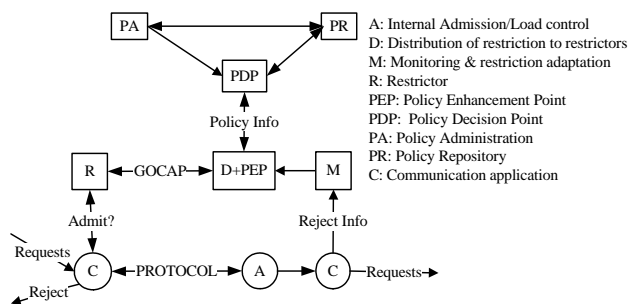


图 1 基于策略的 NGN 过载控制体系结构

该体系结构中,管理员可使用策略管理接口(PA)配置策略库(PR)或策略决策点(PDP),然后根据这些策略和系统状态计算过载限制水平。PDP 在遇到 D 发来的请求时,根据策略配置和负载状况决定负载分发信息。业务开发者仅须关注业务功能特征的开发,而无须考虑容错处理、过载控制等业务共有的需求,把业务的这些公共需求抽象出来,由业务平台

实现,并提供给业务使用,从而简化了业务的开发,缩短了业务的开发和部署周期。从商务角度看,网络运营商和业务提供商之间基于合约的过载机制既有利于网络运营商合理规划、部署网络,也有利于按照二者间的 SLA 协议保证业务提供商所需的网络资源,从而为最终用户提供具有 QoS 保证的业务。同时,业务运营商和最终用户的合约机制必须向用户提供透明的业务,使其能够感知或度量业务质量,这也有助于提高用户满意度。应用基于 SLA 的策略控制部署上述合约,便于业务的开发、部署及业务合约管理,实现过载控制。

4 过载控制组件的部署

执行过载控制时,策略相关组件只涉及 PEP 和 PDP。因此,过载控制组件的部署不考虑策略管理的部署,只考虑 PEP 和 PDP。PEP 就是将 PDP 的策略决定分发给各过载控制点,所以它和 D 组件是一起的,下文中的 D 组件含有 PEP 的功能;PDP 策略组件往往部署在其他管理实体上。A 是内部过载控制组件,部署在过载点上的。若 R 也部署在过载点上,则其功能和 A 重复,并且降低了系统有效吞吐量,根据系统要有内部和外部过载控制的设计原则,R 不适合放在过载点上,而应放在过载源处。基于上述分析,本文主要考虑 M, D 两个组件的部署,其部署方式主要有 4 种:(1)在过载点部署 M, D, 如图 2 所示;(2)将 M, D 部署在请求源上,如图 3 所示;(3)将 M 部署在过载点、而将 D 部署在单独管理实体上,如图 4 所示;(4)将组件 M 和 D 组成单独的集中控制中心,如图 5 所示。

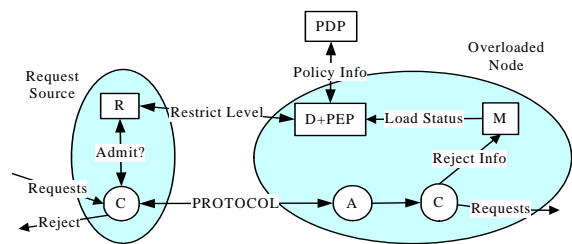


图 2 过载控制组件部署方案 1

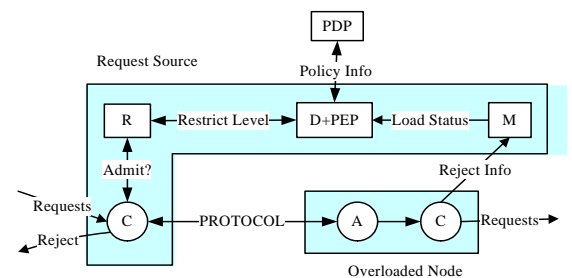


图 3 过载控制组件部署方案 2

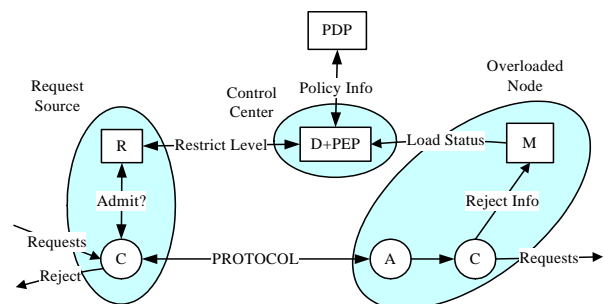


图 4 过载控制组件部署方案 3

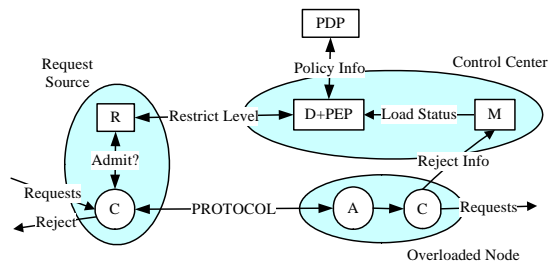


图5 过载控制组件部署方案4

第1种方案中,请求源和过载点之间传递的信息是过载限制阈值;第2种方案中传递的是负载或拒绝状况;第3种方案中,过载点和控制中心之间传递系统负载状况,再由控制中心根据策略信息决定各请求源的过载限制阈值;第4种方案中,过载点向控制中心提供负载状况,控制中心检测这些负载信息,然后决定并分发各请求源的过载限制阈值。通过对比前2种方案得出,第1种部署方案相对于第2种方案具有全局性、公平性等优点^[7],但由于增加了策略控制功能,策略决定可能会消耗更多的资源,从而降低过载点的有效吞吐量。因此,推荐将决策功能独立出来的后2种方案。虽然这2种集中方案存在单点失效的弊端,但其对硬件配置要求不高,因此,可采用其他备份方案解决。另外,第4种方案相比第3种方案需要集中检测多个服务实体的负载状况,它要求每个服务实体提供相关负载查询/报告接口,而不是在每个服务实体上部署过载控制检测组件。这种方式将过载检测功能转移到控制中心后,减轻了过载服务实体的负载,并可实时检测,无须考虑何时启动过载检测的问题。

5 基于策略的过载控制信息流

尽管各过载控制组件的部署有多种方案可选择,其协调工作的基本过程是一致的。图6是基于外包策略模型实现的过载控制各组件间的交互过程。

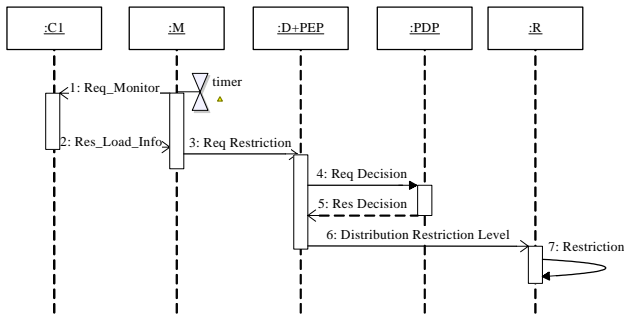


图6 过载控制过程

由基于策略的过载控制体系结构可以看出,各组件间传递的信息主要有:请求/报告服务实体的负载信息,报告负载水平、过载源与过载业务,过载策略请求/决策信息,过载限制阈值发布。各组件间遵守的协议在相关NGN通信协议栈上运行,与协议具体的过载控制配合使用。为方便操作人员表达策略,本文采用人机可读的XML来表示这些交换的信息。限于篇幅,其具体定义略。各组件间的信息流的方向、含义及其遵照的协议如下:

(1)监视请求信息流,方向为从M到应用/主机;协议为自定义的XML表示。监视组件M向应用C或主机发出监视请求,具体请求信息由检测方法和检测要求决定,主要根据CPU利用率因素或业务拒绝率等比较成熟的检测方法进行检测^[2]。若需要对单个业务的负载状况进行检测,则信息流中要含有

识别业务标志。时间间隔由负载变化决定。

(2)监视响应信息流,方向为从应用/主机到M;协议为自定义的XML表示。应用或主机向M响应,将根据检测条件得到的检测结果返回M。

(3)请求过载控制流,方向为从M到PEP;协议为自定义的XML表示。监视组件M向PEP(D)发出过载控制请求。

(4)请求过载控制策略流,方向为从PEP到PDP;协议为自定义的XML策略。PEP将检测的结果发给PDP,向PDP发出策略决定请求。

(5)过载控制策略决定流,方向为从PDP到PEP(D);协议为自定义的XML策略。PDP根据SLA约束以及系统负载状况等,作出决定,并通知给PEP(D)。

(6)分发过载控制限制阈值流,方向为D(PEP)到R;协议为GOCAP。D向各请求源分发过载限制阈值,请求过载控制。

(7)各请求源根据限制阈值限制业务请求,方向:C内部。各请求源有多种方法实现过滤业务请求^[2],本文采用健壮性和效率较高的漏桶算法。

6 结束语

从自然灾害、恐怖袭击等对人类造成巨大损害的事件来看,NGN网络过载控制是必须解决的问题。ETSI TISPAN和MSForum等都在加紧分析NGN网路过载控制的需求,讨论相关机制,制定相关规范。本文在分析了NGN过载控制的需求及其各种场景、NGN网络体系结构及多协议对过载控制带来的负面影响后,基于文献[7]的过载控制体系结构,增加了策略控制组件,较好地满足了业务提供商的可配置和SLA约束需求,有利于简化业务的开发、部署及管理。同时采用策略标准和GOCAP协议,有效地解决了多协议带来的一致性问题。在比较了4种集中和分布式NGN过载控制的分布式过载控制部署方案后,给出了过载控制的过程和信息流。

参考文献

- [1] Manfield D R, Millsted G, Zukerman M. Congestion Controls in SS7 Signaling Network[J]. IEEE Communications Magazine, 1993, 31(6).
- [2] Ackerley R. Telecommunications Performance Engineering[M]. London: Institution of Electrical Engineers, 2004.
- [3] 李亚波. 下一代网络中软交换关键技术的研究——体系结构、话务量建模和过载控制[D]. 北京: 北京邮电大学, 2003.
- [4] Andersson J K. Overload Control and Performance Evaluation in a Parlay/OSA Environment[D]. Sweden: Lund Institute of Technology, 2004.
- [5] 张云勇, 张智江, 李国杰, 等. 基于运营收入最大化的 Parlay 网关过载控制[J]. 通信学报, 2005, 26(3): 92-98.
- [6] 武威. 下一代网络业务执行环境及过载控制的研究[D]. 北京: 北京邮电大学, 2003.
- [7] ETSI DTR. Next Generation Networks: Architecture for Control of Processing Overload[Z]. [2006-10-23]. http://webapp.etsi.org/exchan-gefolder/tr_182015v010101p.pdf.
- [8] Rosenberg J. Requirements for Management of Overload in the Session Initiation Protocol[Z]. [2006-11-11]. <http://jdrosen.net/papers/draft-ietf-sipping-overload-reqs-00.txt>.
- [9] British Telecommunications PLC. NGN Control Plane Overload and Its Management[Z]. [2006-02-10]. <http://www.msforum.org/techinfo/reports/MSF-TR-ARCH-007-FINAL.pdf>.

