SMSC负荷状态检测方法研究*

王 田*1,2

(1. 重庆大学通信工程学院 重庆 400044; 2. 重庆工商大学计算机科学与信息工程学院 重庆 400067)

【摘要】对短消息服务中心的负荷状态检测方法作了探讨。在研究系统负荷状态与内部资源关系的基础上，给出了负荷状态检测参数、检测模型及基于多参数数据融合的检测方法，并通过实验测试结果对其有效性进行了分析。研究表明，系统的负荷状态可以通过对多种内部资源消耗量的检测进行评价，多参数冗余信息融合后可以提高结果的可靠性和置信度。

关 键 词 短消息服务中心; 负荷检测; 数据融合; 负荷控制

中图分类号 TN919.2 文献标识码 A

Research on Method for Load State Detection in SMSC

Wang Tian[1,2]

(1. School of Communication Engineering, Chongqing University Chongqing 400044;
2. Chongqing University of Industry and Commerce Chongqing 400067)

Abstract Some thoroughly research on load state detection in SMSC entity is introduced in this paper. Load detection parameters and model are presented after analyzing the relationship between inter resource and load state in the system. Moreover, a novel strategy to detect system load state based on fusion of multi detection data is presented. By some test, the theory and method are proved most feasible. Our research show that load state of SMSC can be recognized by detecting utility of internal resource utilizing, and multi redundant detecting data fusion is effective approach to improve the reliability and trust degree of the result.

Key words short message service center; load detection; data fusion; load control

短消息服务中心(short message service center，SMSC)作为移动通信网中的一个逻辑和功能实体，在实现时采用了智能网思想。将分散的业务进行集中处理带来了业务实现的便利性，但也带来了业务处理点上的高负荷性[1, 2]。目前，SMSC的业务量正在迅速增长，特别是在某些特定时段内受到大业务量的冲击，当系统外部业务请求大于系统处理能力时，SMSC处于过负荷状态，若得不到控制和缓解将使系统性能迅速恶化，则导致响应时延急剧增加 、数据丢失甚至系统崩溃，本文对SMSC的负荷状态检测方法进行了深入研究。

1 负荷状态的形式化定义

负荷是系统瞬间能力状态的反映，与处理的业务量、业务特性、任务调度策略、内部资源、资源分配和管理策略有密切关系。在对SMSC进行负荷控制时，主要考虑与处理能力有关的资源，包括处理机资源、存储资源、消息总线资源、I/O资源等。为了分析的方便，下面先给出负荷的形式化定义和描述[3~5]。

2002年6月5日收稿
* 重庆市攻关项目，编号：99-03D
** 男 32岁 博士后 主要从事移动智能网、多媒体通信、流量控制等方面的研究

定义1 若系统在给定的资源集R下处理对象$X(t)$的能力容限为$X_{\max}$，对于任意时刻t系统的负荷状态定义为：$L(t)=X(t)/X_{\max}$，其中资源集R是资源元素r_i的集合，即$R=[r_1,r_2,\cdots,r_n]$。

若系统处理对象$X(t)$消耗的资源为$R(t)=[r_1(t),r_2(t),\cdots,r_m(t)]$，$m\in n$，则系统的负荷状态可以表示为

$$L_t = f(\cdot)\left(\frac{r_1(t)}{r_1},\frac{r_2(t)}{r_2},\cdots,\frac{r_m(t)}{r_m}\right) \tag{1}$$

式中 $f(\cdot)$表示资源元素的消耗量与负荷状态之间的映射关系；r_i表示资源元素i的最大配置限度。系统设计时，根据处理能力的设计指标配置资源元素，处理能力的设计指标为理论容限。

定义2 给定系统的理论容限对应的负荷阀值L_0，对于任意时刻t，若$L(t)>L_0$，则称当前状态为过负荷状态。在不同负荷控制阀值$L_i(L_i>L_0)$下，系统过负荷状态可以分为不同的级别。导致系统过负荷的因素相互影响，系统的负荷状态与资源元素的消耗量并不呈线性关系，具有典型的非线性特征。

2 负荷状态的检测方法

3.1 负荷源

SMSC的基本结构如图1所示，主要包括SMPP Agent(short message peer to peer agent)、IW/GMSC (interworking/gateway mobile switch center)、SP(service processor)、DB、SR(service receiver)、OM&M (operater maintenance and management)等子系统。SMSC的业务处理基于请求/响应模式，处理负荷主要来源于下面三个方面：

1) 外部实体的业务请求 GSM网络侧MS起呼的短消息、Alert通知等业务请求经过IW/GMSC转发给SP进行处理；来自ESME(external short message Entity)侧的操作请求(如短消息提交、查询、替换、删除等)经SMPP Agent转发给SP进行处理；

2) 外部实体的响应消息 外部实体对消息投递返回的结果直接影响系统下一步处理流程和系统内部的资源管理，系统高负荷期间也要保证响应消息的接收和处理；

3) 系统内部消息 如定时消息、内部用户广播等。

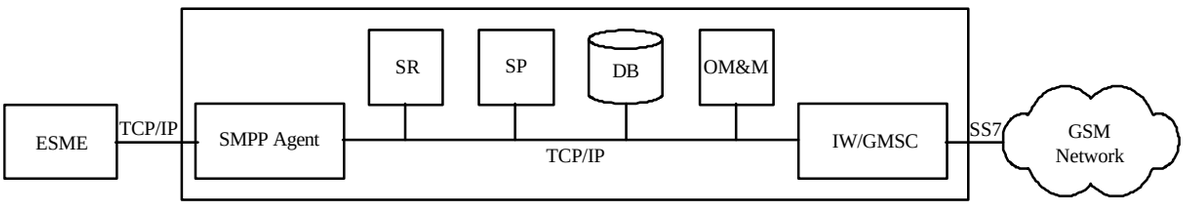

图1 SMSC的基本结构

2.2 负荷状态评价参数

从定义1可知，系统的负荷与资源消耗及系统处理能力容限有关，资源的消耗量或占用率可以作为负荷状态评价参数。作为负荷评价参数，参数量变化对负荷的变化必须很敏感。为了描述系统负荷状态和评价参数之间的关系，可以引入参数灵敏度的概念。

定义3 对于任意时刻t系统负荷状态$L(t)$和评价参数$V(t)$，参数灵敏度$D(t)$为

$$D(t)=\frac{\mathrm{d}V(t)}{\mathrm{d}L(t)} \tag{2}$$

显然，$D(t)$的均值越大，参数对负荷状态变化的反应越灵敏。作为负荷状态评价参数应该选择灵敏度较高、时移较小、容易检测的参数。

SP、IW/GMSC和SMPP Agent的负荷状态可以由各子系统内部的处理器利用率、缓存占用率、平均业务队列长度、内部总线利用率、I/O通道利用率、单位时间内平均业务处理量等参数进行评价，这些参数对处理机作业响应时间的影响较大，是度量处理机负载的重要指标[3, 4]。

2.3 负荷状态检测模型和方法

对负荷状态的准确评价和判断至关重要，误判断将导致严重的负效应。在SMSC中对负荷状态的判断涉及多个参数，每个参数都是一个独立的信息源，根据单参数检测结果进行负荷状态评价有一定的局限性，具有较高的误报率和信息不完整性。可根据多个参数进行信息融合综合判断系统的负荷状态，获得更完整的信息，降低误报率。用数据融合的方法把来自多个检测参数的信息进行联合 、相关 、组合以获取精确的负荷状态估计和严重程度的完整评价，在可靠性和置信度上具有很大的优势[6]。

本文考虑一个包含n个检测参数、n个局域决策点和１个融合中心的分布式负荷状态检测模型。若各检测参数仅向融合中心传送独立的判决，该判决信息不能充分反映参数的信息量，如过负荷严重程度、可信度等，信息损失较大，融合系统的性能有所下降，因而采用冗余信息融合方法，即单个参数同时向融合中心传送其是否处于过负荷状态的判决、负荷状态级别及判决的可信度，融合中心根据各个参数的判决及其可信度形成最终的判决[4]。

若用H_1表示系统出现过负荷的概率，H_0表示系统正常的概率，C_i表示第i个参数的检测，u_i表示其判决，u_i=1表示系统处于过负荷状态，u_i=0表示系统处于正常状态，K_i表示负荷阀值。单参数的判决是将其检验统计量$L(C_i)$与负荷阀值K_i进行比较的过程：当$L(C_i)>K_i$时，判决系统处于过负荷状态；当$L(C_i)<K_i$时，判决系统处于正常状态。

$L(C_i)$超过或低于负荷阀值K_i的程度反映了负荷状态的级别和此参数判决的可信度。单参数判决的可信度可通过对$L(C_i)$相对于K_i的偏差进行有限级别的量化来获得，量化级别由系统预定义的负荷级别决定。

当$L(C_i)\geq K_i$时，为了反映$L(C_i)$超过K_i的程度，需要对区间$[K_i,\infty]$进行有限量化。若存在某一数值A_i，$A_i \in [K_i,\infty]$，使得$P(L(C_i)>A_i \mid H_0)\approx 0$，则充分反映判决$u_i$=1的可信度仅需对区间$[K_i,A_i]$进行量化。

当$L(C_i)<K_i$时，若存在某一数值B_i，$B_i \in [-\infty,K_i]$，使得$P(L(C_i)<B_i \mid H_1)\approx 0$，则充分反映判决$u_i$=0的可信度仅需对区间$[B_i,K_i]$进行量化。对于一个任意分布的检验统计量$L$，根据契比雪夫不等式可以求出$A_i$和$B_i$。

若单参数的判决可信度T_i用N个级别进行描述，当$L(C_i)\geq K_i$时，令

$$K_{i,j}^{1}=K_i+j\frac{A_i-K_i}{N} \qquad j=0,1,2,\cdots,N-1 \tag{3}$$

记$K_{i,j}^{1}=\infty$，则判决u_i=1的可信度为

$$T_i=j \qquad K_{i,j}^{1}\leq L(C_i)\leq K_{i,j+1}^{1} \tag{4}$$

当$L(C_i)<K_i$时，令

$$K_{i,j}^{0}=K_i-j\frac{K_i-B_i}{N} \qquad j=0,1,2,\cdots,N-1 \tag{5}$$

记$K_{i,j}^{0}=-\infty$，则判决u_i=0的可信度为

$$T_i=j \qquad K_{i,j+1}^{0}\leq L(C_i)\leq K_{i,j}^{0} \tag{6}$$

多参数负荷状态检测系统由N个参数构成，各参数的判决信息为

$$\boldsymbol{U}_i=\begin{bmatrix}u_i\\T_i\\O_i\end{bmatrix} \tag{7}$$

式中　u_i、T_i、O_i为负荷状态的判决、可信度和负荷级别，融合中心根据向量$\boldsymbol{U}_T=(U_1,U_2,\cdots,U_n)$进行假设检验形成最终的判决。

3 实验与分析

3.1 实验环境

针对SMSC的不同子系统，在实验测试中采用如图2、3所示的两组实验环境测试IW/GMSC、SP和SMPP Agent的负荷性能。

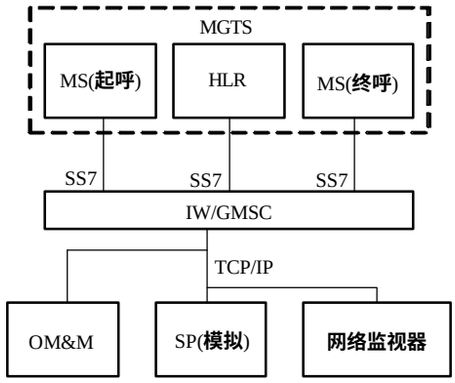

图2 实验环境1

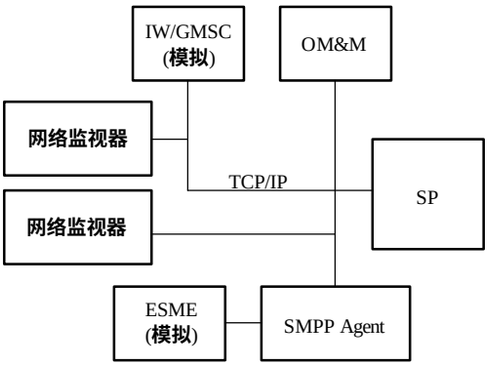

图3 实验环境2

大话务量模拟测试信令仪(message generator traffic simulator，MGTS)能够同时模拟HLR、起呼MSC和终呼MSC功能。测试时，加载50 000个用户数据，模拟的IW/GMSC用来模拟MS起呼短消息(MS-MS)，在用户中随机选取源用户和目的用户(起呼速率可根据要求调节)；模拟的ESME用来模拟扩展短消息实体提交短消息(ESME-MS)，在用户中随机选取目的用户(提交速率可根据要求调节)；消息长度平均为40个字节，对于每种速率系统连续测试时间大于15 min。

3.2 测试结果

测试结果如图4、5所示，从图中可以看出系统的负荷状态可以用资源占用率h进行度量和评价。系统资源消耗与业务量存在直接的关系，业务量增加时，资源占用率同步增长，反映了系统负荷状态的实时变化。CPU占用率(处理器利用率)、缓存占用率(缓冲区利用率)、总线占用率(内部总线利用率)等参数具有很高的灵敏度，可以作为实时负荷状态检测参数，并能通过与系统拥塞间的映射关系确定过负荷的程度或级别。由于系统内存管理策略的差异，内存利用率灵敏度较小，不宜作为检测参数。SMSC中各模块间内部消息通信消耗的网络资源很少(占用率不到高速以太网带宽的5%)，模块间通信不是系统性能的瓶颈，内部消息流量的变化范围极大，不宜直接作为负荷状态检测参数。

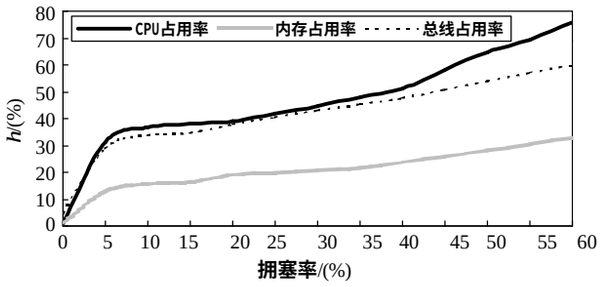

图4 IW/GMSC中拥塞率与资源占用率的关系

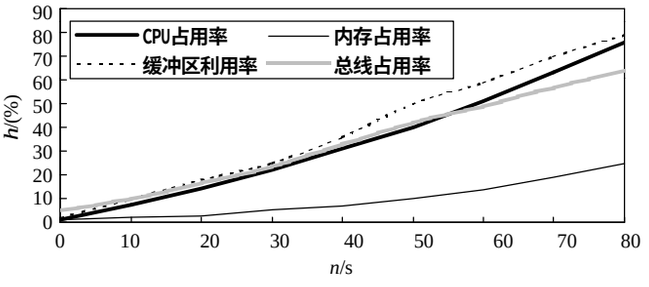

图5 SP中负荷量与资源占用率的关系

参 考 文 献

[1] Rumsewicz M. Load control and load sharing for heterogeneous distributed systems[M]. Proceedings of the International Teletraffic Congress'99(ITC-16) Edinburgh, 1999

[2] Maria K. Overload control strategies for distributed communication networks:[Ph.D. dissertation][M]. Sweden:Lund Institute of Technology. 1999

[3] 王 田, 曹长修, 汪纪锋. 短消息服务中心过负荷控制机制研究[J]. 电信科学. 2001. 17(9): 12-16

[4] 王 田. 移动智能网的过负荷控制机制研究:[博士学位论文][D]. 重庆: 重庆大学, 2002

[5] Ranganathan M. Acharya A, Saltz J. Distributed resource monitors for mobile objects[C]. Proceedings of the Fifth International Workshop on Object Orientation in Operating Systems. Seattle. WA. October 1996:19-23

[6] Tenney R R, Sandell N R. Detection with distributed sensors[J]. IEEE Transactionson AES. 1981, 17(4):501-510

编 辑 徐培红