

文章编号:1000-582X(2004)03-0032-04

# MIN 负荷状态检测方法\*

王 田<sup>1,2</sup>

(1. 重庆大学 自动化学院, 重庆 400030; 2. 重庆工商大学 计算机科学与信息工程学院, 重庆 400067)

**摘 要:**论文对移动智能网(MIN)的负荷状态检测方法进行了深入研究。在基于对系统负荷状态与内部资源关系的深入分析基础上,对 MIN 中 SCP 和 SSP 的负荷源进行了分析,给出了负荷状态检测参数、检测模型以及基于多参数数据融合的检测方法,并通过测试结果对其有效性进行了分析。研究表明,移动智能网单节点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,采用多参数冗余信息融合可以提高检测结果的可靠性和置信度。

**关键词:**移动智能网; 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

智能网技术将原来分散的业务逻辑控制进行集中处理带来了业务实现的便利性,另一方面也将分散的业务处理量高度集中,带来了业务控制点上的高负荷性<sup>[1-2]</sup>。第三代移动通信网络的一个重要的特征就是智能化,在 IMT-2000(UMTS)中将实现智能网与移动网的完全融合。SCP 是智能网的核心,在第三代移动通信系统中 SCP 在整个网络中的地位将更加突出,原有移动网的功能如切换、位置更新、鉴权等都主要由 SCP 来负责实现,SCP 将承受巨大的业务量。目前,移动智能网(MIN, Mobile Intelligent Network)的业务量正在迅速增长,在某些特定时段内极易受到突发性的巨大业务量冲击,使系统和网络处于过负荷(overload)状态,若过负荷状态得不到及时控制和缓解将使业务 QOS 和系统性能迅速恶化,导致响应时延急剧增加、数据丢失甚至系统崩溃。寻求有效的负荷状态检测方法是进行过负荷控制的关键课题之一<sup>[3-5]</sup>。

## 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。

MIN 中的资源主要包括两大类:信道资源和网络节点资源。在第三代移动通信系统中,信道资源由用户无

线接入信道和固定传输信道两个部分组成。固定传输信道能力基本上能保证用户和业务的需要,不是主要的瓶颈。用户无线接入信道所用频谱资源也有了很大改善。ITU 在 WRC92 上为 IMT-2000 规划了 1 920~1 980 MHz 和 2 110~2 170 MHz 的对称频谱资源共 1 20MHz 供 FDD 使用,1 900~1 920 MHz 和 2 010~2 025 MHz 的非对称频谱资源共 35 MHz 供 TDD 使用,在 WRC2000 会议上又增加了 800 MHz 频段(806~960 MHz)、1.7 GHz 频段(1 710~1 885 MHz)、2.5 GHz 频段(2 500~2 690 MHz)供 IMT-2000 业务使用。以上频谱资源为未来的业务应用预留了较大的资源空间。网络节点资源是本文主要的研究对象,主要包括:处理机资源、存储资源、消息总线资源、I/O 资源等几个方面。在 SSP 和 SCP 这两类节点中处理机资源和存储资源是影响处理能力的最主要因素。

为了分析的方便,下面先给出负荷的形式化定义和描述<sup>[5-6]</sup>。

**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: $R = [r_1, r_2, \dots, r_n]$ 。

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

\* 收稿日期:2003-09-17

基金项目:重庆市科委应用基础项目资助(8084)

作者简介:王田(1971-),男,湖北荆州人,博士后,重庆工商大学副教授,研究方向:移动智能网,多媒体通信。

以表示为:

$$L_i = f(\cdot) \left( \frac{r_1(t)}{r_1}, \frac{r_2(t)}{r_2}, \dots, \frac{r_m(t)}{r_m} \right) \quad (1)$$

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阈值  $L_0$ , 对于任意时刻  $t$ , 若  $L(t) > L_0$ , 则称当前状态为过负荷状态。在不同负荷控制阈值  $L_i (L_i > L_0)$  下, 系统过负荷状态可以分为不同的级别。

系统过负荷与资源元素消耗量有直接关系, 并可能出现不同现象和后果。

1) 当业务量超过处理机的处理能力后, 待处理的业务队列长度增加, 业务处理时延增加, 业务 QoS 下降, 处理机可能出现性能不稳定;

2) 等待处理业务数据超过缓存容量, 则数据可能溢出导致业务请求数据丢失;

3) 内部消息量过大导致总线阻塞, 则处理时延增加;

4) 频繁读写存放在外部 SDP 的业务数据和逻辑控制信息导致 I/O 阻塞, 则处理时延增加。

移动智能网中各类业务的特性存在着差异, 导致系统过负荷的因素各不相同, 各种因素之间是相互影响的, 如总线阻塞或 I/O 阻塞必然导致处理时延增加, 从而引起等待处理的业务数据积压而出现缓存溢出。系统的负荷状态与资源元素的消耗量并不是线性关系, 具有典型的非线性特征。

## 2 MIN 的负荷源<sup>[5]</sup>

### 2.1 MIN 中的信令消息

MIN 中节点处理的主要对象是物理实体之间的交互消息, 物理实体之间通过 SS7 的底层协议 TCAP 传递 IN 业务所需的应用层 INAP 消息。根据 X. 208 的规定, INAP 由 ASN. 1 标记描述, INAP 的编码是按照 X. 209 中定义的基本编码规则和 Q. 773 中规定的限制进行的。MIN 物理实体之间的消息交互方式(协议栈)如图 1 所示。

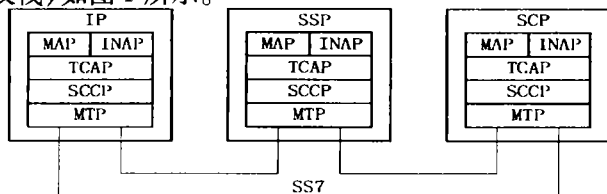


图 1 SSP、IP 和 SCP 间消息交互方式(协议栈)

### 2.2 SSP 的主要负荷源

根据 MIN 中 SSF/CCF 定义的功能, SSP 节点的负荷主要来源于下面几个方面:

1) 用户的业务请求消息。CDMA 网络侧的 MS 起呼业务经过 BS 转给 SSP(MSC) 进行处理。MS 终呼业务经过其他 MSC 转给用户端的 SSP(MSC) 进行处理。呼叫业务请求要求在较短的时间内完成处理并返回响应。

2) 与 SCP 间的交互消息。业务处理基于请求/响应模式, 在用户业务请求处理中, 检测到设置的 DP 点并满足触发条件, 当与 SCP 相应关系是控制关系时, SSP 挂起呼叫向 SCP 发出请求下一步操作指令的请求消息流, 并等待 SCP 的响应; 当与 SCP 相应关系是监视关系时, SSP 不挂起呼叫, 但向 SCP 发出通知消息, 不等待 SCP 的响应(在 SSP 中的 DP 处理规则要求通知类的 DP 优先于请求类 DP 进行处理)。接收到 SCP 响应后, SSP 继续处理。超过一定时间, 没有收到 SCP 的响应, SSP 将重发请求消息。SCP 的响应消息丢失时, SSP 中会话占用的资源不能及时释放。

3) 与其他网络实体的交互消息。SSP 在呼叫处理过程中要和其他网络实体(包括 AC、VLR、HLR 和 SMSC) 进行消息交互, 以获取业务处理需要的信息。SSP 检测到设置的 DP 点满足触发条件, 会根据预先的定义, 向确定的网络实体发送消息流, 请求业务处理需要的信息, 如: 客户描述文件及鉴权信息。与其他其他网络实体的交互消息也可能是转发的业务信息, 如: 转发短消息等。

4) 系统内部消息。如系统内部的定时处理、定时重试消息、内部广播消息等。

### 2.3 SCP 的主要负荷源

SCP 主要同 SSP 进行交互, 控制业务所需要的业务逻辑, 同时和其他智能网实体进行消息交互, 获业务处理需要的必要数据和辅助信息。

SCP 的处理机负荷主要来源于下面 3 个方面:

1) SSP 的业务请求消息。SCP 对 SSP 发送的业务逻辑指示请求进行响应。SCP 收到的消息有两类: 一类是 SSP 发送的业务请求指示消息, 另一类是 SSP 发送的事务通知。若收到的是请求消息, 经过处理后向 SSP 发回响应消息, 指示下一步的操作, 求消息需要系统在较短的时间内完成处理并返回响应; 收到的是事务通知时, 不用向 SSP 发回响应。

2) 其他智能网实体(包括 SDP、IP、SMP 及其他 SCP) 的交互消息。SCP 同 SDP 进行交互, 获取业务逻辑处理需要的数据和操作信息; 同 IP 进行交互, 面向

业务流程控制 IP 中的资源服务;同 SMP 进行交互,提供系统管理需要的数据和信息;需要时,可以跟其它 SCP 进行,实现分布式业务控制和自愿的业务通知。

3)系统内部消息和数据处理。如定时处理、定时重试消息、内部广播任务等。同时,为保证系统的正常运行和正确评价系统的服务质量,SCP 需要对 SMP 和自身管理需要的数据进行采集、监测和统计,并实时进行发送。

### 3 负荷状态评价参数

#### 3.1 参数的灵敏度

从定义 1 中可以知道,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评价的参数。网络节点资源中曾经被研究及用作负荷指标的主要包括 CPU 利用率、内存占用率、业务队列长度、某段时间内平均队列长度、内部总线占用率、I/O 通道利用率等。

作为负荷评价的参数(或指标),参数量的变化对负荷的变化必须很敏感。为了描述系统负荷和评价参数之间的关系,下面引入灵敏度的概念。

定义 3 对于任意时刻  $t$  系统的负荷  $L(t)$  和负荷评价参数  $V(t)$ ,则参数的灵敏度  $D(t)$  为:

$$D(t) = \frac{dV(t)}{dL(t)} \quad (2)$$

显然  $D(t)$  越大,参数对负荷的变化越灵敏。作为负荷评价的参数应该选择  $D(t)$  值较大,受时间影响较小的参数。

#### 3.2 MIN 的负荷评价参数<sup>[5,7]</sup>

在 MIN 中,业务集中在 SCP 上进行业务逻辑处理和信令控制,SCP 的负荷量最大。SCP 节点处理机的负荷状态一般可以由几个参数进行综合评价:系统处理资源占用率(CPU 利用率)、用户数据区(缓存)的使用率、业务请求平均等待时间(平均业务队列长度)、内部总线利用率、I/O 通道利用率、单位时间平均业务处理量。

MIN 的 SSP 在 BCSM 中 DP 上,将呼叫控制转向 SCP,主要功能以基本呼叫管理、WIN 交换管理、特征交互管理为主,其节点处理机的负荷状态也可以由几个参数进行判断:系统处理资源占用率、用户数据区(缓存)的使用率、用户业务请求平均等待时间和平均业务队列长度、单位时间平均业务量

这些参数都是对系统的负荷状态的反映,每个参数与负荷状态存在着一定的映射关系。这种映射关系并不具有精确性,而是一个区间映射。这种映射关系根据处理资源调度、缓冲区管理策略、队列排队规则以

及软硬件结构的差异而不同。

### 4 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 MIN 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。可以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。用数据融合的方法可以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。

考虑一个包含若干  $n$  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<sup>[5-6]</sup>。

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。 $K_i$  表示负荷阈值。单参数的判决就是将其检验统计量  $L(C_i)$  与负荷阈值  $K_i$  进行比较过程:

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。

$L(C_i)$  超过或低于负荷阈值  $K_i$  的程度实际上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  $P(L(C_i) > A_i | H_0) \approx 0$ ,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  $P(L(C_i) < B_i | H_1) \approx 0$ ,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

$$K_{i,j}^1 = K_i + j \frac{A_i - K_i}{N}, \quad j = 0, 1, 2, \dots, N - 1 \quad (3)$$

记  $K_{i,j}^1 = \infty$ , 则判决  $u_i = 1$  的可信度为

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

当  $L(C_i) < K_i$  时,令

$$K_{i,j}^0 = K_i - j \frac{K_i - B_i}{N}, \quad j = 0, 1, 2, \dots, N - 1 \quad (5)$$

记  $K_{i,j}^0 = -\infty$ , 则判决  $u_i = 0$  的可信度为

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

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  $U_i$

$$U_i = \begin{bmatrix} u_i \\ T_i \\ O_i \end{bmatrix} \quad (7)$$

式中  $u_i, T_i, O_i$  分别表示负荷状态的判决、可信度和负荷级别,融合中心根据向量  $U_T = (U_1, U_2, \dots, U_n)$  进行假设检验形成最终的判决。

### 5 实验与分析

#### 5.1 实验环境

笔者分别对 SCP 和 SSP 的负荷性能进行了测试和实验。其中针对 SCP 的实验主要采用模拟 SSP 处理机、网络监视器、信令接口单元(SIU)、本地维护终端和 SCP 模块进行组网,如图 2 所示:

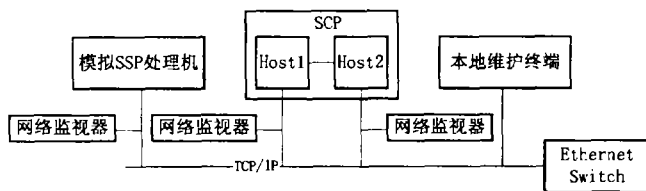


图 2 SCP 负荷性能测试和实验组网结构

SCP 双处理机均采用 TCP/IP 协议通过高速以太网交换机分别独立与模拟 SSP 处理机交换数据。在本地维护终端上动态监视主用机和备用机的 CPU、内存、总线消息占用率等负荷信息,实时统计智能业务的处理次数以及高负荷期间备用机参与处理业务的次数。在主用机(Host1)、备用机(Host2)和模拟 SSP 处理机输出端设置网络监视器监视通信流量的变化。

模拟 SSP 处理机随机地从预置的移动用户中选择源用户和目的用户,向 SCP 处理机发起预付费业务呼叫,呼叫速度可以进行调节。SCP 处理机根据模拟 SSP 处理机的业务请求进行实际业务逻辑控制处理并响应。对于模拟呼叫的每种速度,系统连续测试时间不少于 15 min,备用机在测试中不参与业务处理。

#### 5.2 结果分析

实验结果如图 3 所示,从图中可以看出系统的负荷状态可以用资源的占用率进行度量和评价。系统资源消耗与业务量存在着直接的关系,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程度之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。MIN 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 8%),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。

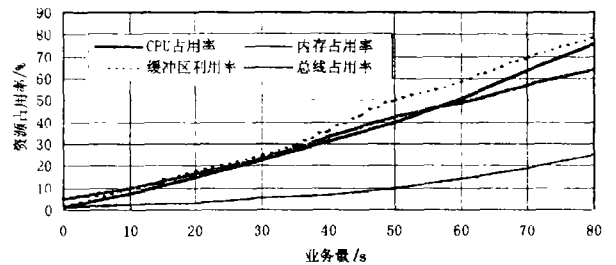


图 3 SCP 中负荷量与资源占用率的关系

#### 参考文献:

- [1] RUMSEWICZ M. Load Control and Load Sharing for Heterogeneous Distributed Systems[A], Proc. of the International Teletraffic Congress99(ITC-16)[C], Edinburgh,1999.
- [2] KIHLMARIA. Overload Control Strategies for Distributed Communication Networks[D], Sweden: Lund Institute of Technology Ph. D. dissertation, 1999.
- [3] WELSH M, CULLER D. Overload Management as a Fundamental Service Design Primitive[A], Proc. of the Tenth ACM SIGOPS European Workshop[C], Saint - Emilion, France, 2002.
- [4] 廖建新,李彤红,陈俊亮. 移动智能网和宽带智能网的研究现状及其展望[J],电子学报,1999,27(4):106-111.
- [5] 王田. 移动智能网的过负荷控制机制研究[D],重庆:重庆大学博士学位论文,2002.
- [6] WANG TIAN. Research on Method for Load State Detection in SMSC[J], Journal of UEST of China, 2003, 2: 125-128.
- [7] RANGANATHAN M, ACHARYA A, SALTZ J. Distributed Resource Monitors for Mobile Objects[A], Proc. of the Fifth International Workshop on Object Orientation in Operating Systems[C], Seattle, WA, 1996:19-23.

## The Implementation Study on The Modeling of Directed-graph With Weightings Based on The BOM of Product

SHI Wei-ren, PENG Shi-qiang, KANG Jing, XIAO Qian

(College of Automation, Chongqing University, Chongqing 400030, China)

**Abstract:** The nodes of BOM may appear repeatedly in the different layers. While the structure of traditional tree model can not express the relation between these nodes expediently. By analyzing the localization of tree model BOM, this paper studies the essential structure of BOM and puts forward the directed-graph with weightings of BOM. Then it provides corresponding realization of relation database and studies the method of maintenance. It also discusses the expansion of attribute based on enterprise's requirement. Finally it provides the thinking of searching, relevant algorithm and the method of programming language realization. This kind of BOM offers foundation for the integration of information and data in PDM/CAPP/CIMS/ERP.

**Key words:** BOM; directed-graph with weightings; modeling

(编辑 张小强)

~~~~~  
(上接第36页)

## Approach for Load State Detection in MIN

WANG Tian<sup>1,2</sup>

- (1. College of Automation, Chongqing University, Chongqing 400044, China;
2. Chongqing Technology and Business University, Chongqing 400067, China)

**Abstract:** Some thoroughly research on the basic method for load state detection in MIN is introduced. Load sources of SCP and SSP in MIN are analyzed in detail after investigating the relationship between inter resource and load state in the system. And load detection parameters and model are presented. 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 MIN can be recognized by detecting utility of internal resource utilizing of single node in MIN, and multi redundant detecting data fusion is effective approach to improve the reliability and trust degree of the result.

**Key words:** mobile Intelligent network; load detection; data fusion; load control

(编辑 吕赛英)