

# 虚拟多端口的短信分发调度系统模型<sup>①</sup>

黄声勇, 范冰冰, 黄兴平

(华南师范大学 计算机学院, 广州 510631)

**摘要:** 针对短信虚拟运营商短信多端口下发问题, 提出一种基于模型的动态负载均衡下发调度系统模型。首先根据所设计的系统总体架构引出了系统的应用模型, 接着通过排队论知识进行模型的分析与论证, 并且引入动态负载均衡和多线程调度策略。通过理论和实践证明, 该系统模型是可以确保短信应用的高并发性、实时性和负载均衡的。

**关键词:** 排队论; 负载均衡; 实时; 多端口

## SMS Dispatch System Model Based on Virtual Multi-Port

HUANG Sheng-Yong, FAN Bing-Bing, HUANG Xing-Ping

(College of Computer Science, South China Normal University, Guangzhou 510631, China)

**Abstract:** Dedicated to the dispatch of SMS based on the virtual multi-port, this paper puts forward on the SMS Dispatch System Model with dynamic load balancing strategy based on the queuing model. It extracts the system application model on the basis of the overall system architecture, then analyzes and demonstrates the model with the knowledge of Queuing Theory. It brings in dynamic load balancing and multi-thread scheduling strategy. Through the theory and practice, we conclude that the system model can ensure the high concurrency, real-time and load balance of SMS application.

**Key words:** queuing theory; load balance; real-time; multi-port

### 1 概述

移动通信业务中的短信息业务 (Short Message Service, SMS) 日益普及, 其用户和业务量也在飞速增长, 所以利用 SMS 进行信息处理的企业必须在最短的时间内来推送短信, 并确保其高并发量和实时性的应用。

由于专线接入网关的短信下发方式其实现细节复杂, 而 Model 接入方式收发速率低, 即使是多 Model 集群方式也由于单个设备的受限而整体下发性能不高<sup>[1]</sup>, 所以基于虚拟运营商的短信端口接入方式成为当前短信下发的主流应用, 特别是短信多端口的接入方式更具有实际的研究意义。

本文的应用背景就是基于虚拟运营商短信多端口接入方式, 针对以往短信应用系统中缺乏数学模型论

证和动态负载均衡策略, 不能够充分保证 SMS 应用的实时性、并发性及系统利用率的情况, 引入了基于排队模型的动态负载均衡和多线程调度相结合的短信分发系统模型 (SMS Dispatching System Model, SMSDSM), 利用其动态负载均衡策略和并发性把大量的短信息在最短的时间内进行分发, 确保 SMS 应用的实时性和提高系统的利用率。

### 2 系统的总体架构

本论文研究的 SMSDSM 是在基于虚拟运营商短信多端口的基础上, 把每条分发短信的号码及相对应的内容封装成为指定的短信实体, 然后利用短信端口所提供的 Web Service 接口进行异步提交, 并由短信端口推送到短信网关。SMSDSM 的系统总体方案如下图

① 基金项目: 广东省科技支撑计划(200922-D261)

收稿时间: 2010-12-17; 收到修改稿时间: 2011-01-21

1 所示。

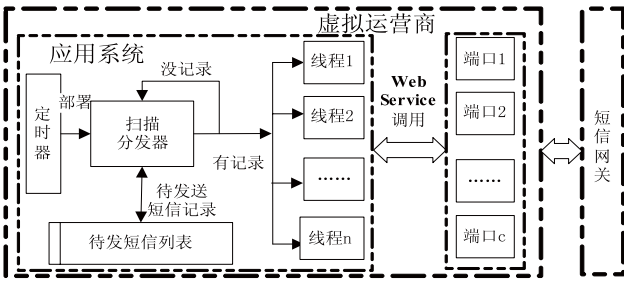


图 1 SMSDSM 总体设计架构

在上述的架构图中，短信的到达都是随机的和要求被实时处理的，则需要引入多线程处理<sup>[2]</sup>。SMSDSM 的处理流程是：当系统部署运行后，全局定时器根据系统设定好的间隔时限不断地触发前端扫描分发器，执行扫描待发短信列表，当扫到有可以发送的短信记录，把每条记录封装成为一个对象消息放进前端扫描分发器的队列中，线程 1, 2.....n 根据各个短信端口的动态消息队列负载均衡策略进行任务派遣，端口 1, 2.....c 代表负载均衡为 c 个短信组件接口并行处理。

### 3 系统的应用模型

#### 3.1 应用模型描述

在排队系统模型中，采用多服务员、单队列的系统方案，其各项运行指标都优于多队列的排队系统<sup>[3]</sup>。根据上述 SMSDSM 系统总体架构的设计，系统的应用模型可采用多服务员、单队列的排队系统模型，其应用模型如下图 2 所示：

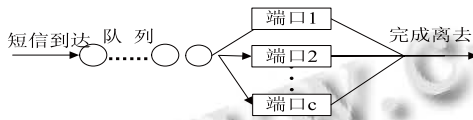


图 2 SMSDSM 的排队系统应用模型

在上述图 2 中，根据排队论的三要素角度<sup>[4]</sup>，可做如下假设：

#### (1) 输入过程

待发短信看作输入的顾客，定时器启动后，在不相交的时间区间内到达的待发短信数是相互独立的，而且在某段时间内到达的待发短信数只与时间长度有关，与时间段的起点无关，并以参数的泊松分布到达。

#### (2) 排队规则

1) 顾客：等待制；2) 服务：FCFS（先来先服

务)；3) 排队空间：即前端扫描分发器的队列长度；4) 队列数目：1 个，即前端扫描分发器，根据每个短信端口的负载情况进行动态负载均衡派遣。

#### (3) 服务机构

1) 服务台有 c 个，即负载均衡后虚拟运营商短信端口的数目；2) 每个服务台在处理短信发送时所进行的工作都一样，从平均服务率的角度来看，由于配置相同，顾客的到达概率均等，每个服务台所表现的出的服务能力基本相同，而且到达与服务是彼此独立的，即服务时间  $T$  服从平均服务率  $\mu$  的负指数分布。

根据上述的假设，可以把 SMSDSM 看作一个具有负载均衡的  $M/M/c$  排队系统。

### 3.2 应用模型的数学论证及分析

#### 3.2.1 输入流的泊松分布验证

设  $N(t)$  表示在时间  $(0, t](t > 0)$  内到达的短信数，令  $P_n(t_1, t_2)$  表示在时间区间  $(t_1, t_2](t_2 > t_1)$  内有  $n(n \geq 0)$  条短信到达的概率，即  $P_n(t_1, t_2) = P\{N(t_2) - N(t_1) = n\}$ 。

在 SMSDSM 的排队模型有如下三个条件成立：

(1)  $N(0) = 0$ ，即表示在定时器未开始工作时，前端扫描分发器有 0 条待发短信。

(2) 在不重叠的时间区间内短信到达数相互独立；即  $\{N(t), t \geq 0\}$  有独立增量，对任取的  $m$  个时刻： $0 < t_1 < t_2 < \dots < t_m$ ，随机变量， $N(t_2) - N(t_1)$ ， $\dots$ ， $N(t_m) - N(t_{m-1})$  是相互独立的。

(3)  $P_n(t)$  即在  $t$  时间内达到  $n$  条短信的概率，则  $P_n(t) = (\lambda t)^n e^{-\lambda} / n!, t > 0, n = 0, 1, 2, \dots$ <sup>[4]</sup>。

综上所述的三个条件，根据泊松分布的定义<sup>[4]</sup>，可知 SMSDSM 的待发短信输入流服从泊松分布。

#### 3.2.2 服务时间的负指数分布验证

在 SMSDSM 中，设随机变量  $T$  为服务时间，则  $T$  具有如下的性质：

(1)  $T$  具有独立性（无记忆性），即每个服务时间是相互独立的。(2)  $T$  具有连续性，而且是非负取值，因为在每个服务台的配置相同的情况下，服务时间  $T$  可以看作是连续的。

根据(1)、(2)的性质，任取  $t > 0$ ，令  $g(t) = P\{T > t\}$ ，则  $g(t)$  是右连续单调的。则有： $g(t) = e^{-\mu t}, t > 0$ <sup>[4]</sup>

上式说明服从相同参数的负指数分布。

综上所述，SMSDSM 是一个具有负载均衡的  $M/M/c$  排队系统模型，队列的输入过程是扫描分发器扫到待发短信，服务机构是短信端口。

### 3.2.3 M/M/c 排队系统的性能指标

M/M/c 排队系统的性能指标反映了 SMSDSM 的性能,其主要指标有<sup>[4]</sup>:

(1) 队长,指系统中的待发短信数,期望值记作  $\bar{N}$ ,包括排队等待服务的短信数(期望值记为  $\bar{N}_1$ )和正在接受服务的短信数(期望值记为  $\bar{N}_2$ ),则有:  
 $\bar{N} = \bar{N}_1 + \bar{N}_2$ 。一般情况下,  $\bar{N}$  或者  $\bar{N}_1$  越大,服务率越低,说明系统的并发处理能力低。

(2) 等待时间与逗留时间,等待时间的期望值记为  $\bar{w}_q$ ;逗留时间指待发短信在系统中的等待时间和服务时间之和,期望值记为  $\bar{w}$ 。等待时间和逗留时间是系统最关心的指标,一般情形下(统计平衡下),逗留时间越短,短信的实时性越好,系统的并发性越好。

(3) 系统的忙期与闲期,在统计平衡下,闲期越短,系统的并发性越好,系统的负载均衡达效果越好。

上述的性能指标直接就是 SMSDSM 在短信实时性、系统并发性及负载均衡性方面的效果指标。

## 4 动态负载均衡的策略

### 4.1 问题描述

在 M/M/c 排队系统中,我们是假设了系统的容量为无穷大,即排队队列可以容纳无限数的待发短信;另外,我们也假设了当有待发短信到达排队队列时,有空闲的端口服务台即刻进行服务。而在实际的应用中,排队队列可以容纳的数目是有限的,当有大量待发短信到达时,会导致部分输入消息阻塞在前端,以至产生丢包的问题;而且调度任务给空闲的服务台需要有实际的判断信息,要实时的知道哪个服务台是有空闲的,尽可能让每个服务台所承受的负载近似相同,提高每个服务台资源的利用率。

已有研究表明<sup>[5]</sup>,无论使用静态负载均衡调度还是动态负载均衡调度都可以提高集群系统的性能。另外,在实际应用的负载均衡环境中要考虑延迟和计算负载耗时<sup>[6]</sup>,没有哪种算法是最优的,只有选择符合系统自身和能达到负载效果的算法。

### 4.2 动态负载均衡的策略

针对上述的问题,在基于 M/M/c 排队模型的 SMSDSM 中引入基于服务台队长实时计算的动态负载均衡策略来解决问题。其数学模型如下图 3 所示,在每个服务台中引入一个缓冲队列 Q 和 Q 队列长度的计数器, Q 的作用是解决 SMSDSM 中输入队列可能溢

出的问题,及时接收来自输入队列的待发短信;另外,与每个 Q 相对应的队列长度计数器值就是每个服务台的负载情况,计数器值越大表明该服务台的负载越大,那么输入队列调度任务的时候就派遣给计数器值小的服务台。

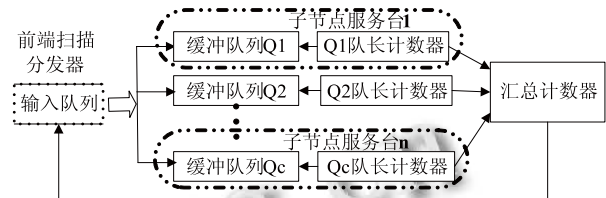


图3 基于服务台队长实时计算动态负载均衡的数学模型

基于服务台队长实时计算动态负载均衡策略的工作原理是:

(1) 在任务启动时,初始化每个服务台 Q 的计数器为 0;

(2) 一开始输入队列按照服务台的顺序把待发短信推送给计数器为 0 的服务台;

(3) 经过一段时间  $t(t > 0)$  后,当 Q 接收到来自输入队列的待发短信,或者服务台消化了 Q 中的待发短信,则立刻把该缓冲队列的存放数目(计数器值)传递给汇总计数器,汇总计数器把汇总信息反馈给前端扫描分发器,接着输入队列根据 c 个缓冲队列的汇总信息,把待发短信推向计数器值最低的服务台,让每个服务台所承受的负载近似相同。这样,输入队列在进行待发短信分发到短信端口时,只需检查汇总计数器而不需要进行复杂的计算,降低分发时间,实现了负载均衡,提高服务台的处理效率。

(4) 当然 Q 的队列长度 l (任务开始后的计数器值)与上述的  $\bar{w}$  和  $\mu$  密切相关,有如下关系:

$$\bar{w} \uparrow \Rightarrow l \uparrow, \mu \uparrow \Rightarrow l \downarrow; \bar{w} \downarrow \Rightarrow l \downarrow, \mu \downarrow \Rightarrow l \uparrow$$

上述关系表明了 l 与  $\bar{w}$  成正比关系,与  $\mu$  成反比关系,所以进行基于服务台队长实时计算的动态负载均衡的本质就是按照 SMSDSM 排队模型中的  $\bar{w}$  和  $\mu$  来进行的。

## 5 并发线程的调度策略

在 SMSDSM 中同一时刻存在着多个线程在并发运行,所以必须得有线程同步机制和调度策略来确保整个系统的运行效率。

线程同步采用新增变量和共享资源加锁的机制进行控制，即采用了“空间和时间”相结合的方式，而非单一的消耗空间或时间，避免了大量浪费空间和线程高阻塞情况的出现。

SMSDSM 需要在很短时间内迅速处理大批任务，如果频繁地创建、销毁线程则会耗费大量的系统资源。所以在 SMSDSM 中预先建立了线程池，池中元素是处于睡眠状态的线程，当有任务到达的时候从池中取出线程，任务执行完毕后放回池中而不作销毁，这样可以提高系统并发处理的性能。

### 6 系统的性能分析

依据上述基于多端口的 SMSDSM 及其使用的策略算法，可知输入过程的待发短信服从泊松分布，服务时间服从负指数分布，SMSDSM 是一个具有负载均衡和多线程并发处理的单队列多服务台  $M/M/c$  的排队系统。

将上述的 SMSDSM 系统方案应用到某大型的社交平台短信系统中，该系统每月有近短信  $9 \times 10^8$  发送量，则单位时间内到达的短信数  $\lambda = 30.47$  条/秒（即短信平均到达率）；在该系统中，每个短信端口单位时间内的处理能力（平均服务率） $\mu = 26$  条/秒（中国移动开放给该系统的短信端口处理速率上限为 52 条/秒），则每个端口的服务强度  $\rho = \lambda / \mu = 1.1719$ ；系统采用了 4（ $c=4$ ）个端口并行处理，其服务强度  $\rho_c = \lambda / 4\mu = 0.2930$ ，因  $\rho_c < 1$ ，则处于统计平衡条件下的，所以代入 SMSDSM 排队模型，可得如下数据：

$P_0$ ，即一段时间内没短信到达的概率，也是短信端口空闲概率，反映系统的利用程度：

$$P_0 = \left[ \sum_{j=0}^{c-1} \frac{\rho^j}{j!} + \frac{c\rho^c}{c!(c-\rho)} \right]^{-1} = 30.12\%$$

系统中排队等待服务的平均短信数  $\bar{N}_1$ ：

$$\bar{N}_1 = \frac{\rho_c}{(1-\rho_c)^2} \cdot \frac{\rho^c P_0}{c!} = 0.0139 \text{ 条}$$

系统的平均队长  $\bar{N}$ ，即包括等待服务的短信数和正在被服务的短信数：

$$\bar{N} = \bar{N}_1 + \bar{N}_2 = \bar{N}_1 + \rho = 1.1858 \text{ 条}$$

短信在系统中的平均等待时间  $\bar{W}_q$ ：

$$\bar{W}_q = \frac{\rho_c}{\lambda(1-\rho_c)^2} \cdot \frac{\rho^c P_0}{c!} = 0.0005 \text{ 秒}$$

短信在系统中的平均逗留时间  $\bar{W}$ ：

$$\bar{W} = \bar{W}_q + \frac{1}{\mu} = 0.0630 \text{ 秒}$$

短信到达后必须等待的概率  $P_n$

$$P_n(n > c) = \frac{\rho^n P_0}{c^{n-c} \cdot c!} = 6.9\%, \quad n = 5$$

另外，在系统运行时，记录了 4（ $c = 4$ ）个端口服务台 CPU 利用率的情况如图 4 所示：

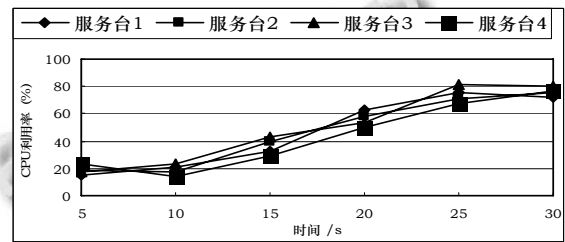


图 4 服务台的负载均衡情况

在 4 个服务台的情况下，从上述的实验结果可以得出：

(1) 短信端口的空闲概率不大，闲期短而忙期长，系统处于繁忙而又不是高负载的工作状态下，提高了系统的利用率和负载均衡的效果。

(2) 根据上述的系统平均队长，每个待发短信几乎不用排队可以接受服务，说明系统的并发处理能力高。

(3) 待发短信在系统中的平均逗留时间小，即一条待发短信产生到发送成功的时间短，一个小时将近有 57143（ $3600/0.0630$ ）条短信发送成功，说明 SMS 应用的实时性得到保证。

(4) 在图 4 的记录中，可知道 4 个服务台集群的负载比较均衡，即短信发送的任务负载能够被均衡的调度到各个服务台处理，提高了系统的整体利用率和并发处理能力。

### 7 总结

本文提出的基于短信多端口的 SMSDSM 把排队模型和动态负载均衡在实际应用中相结合，并进行了相关的数学论证和分析，这是本文的难点和创新点所在。理论和实践表明，基于排队模型和动态负载均衡的 SMSDSM 在大数据量和高并发量的情况下有明显的性能优势，确保了系统的整体利用率，在尽可能短的时间内保证 SMS 应用的实时性；特别是在排队理论

(下转第 11 页)

互作用。

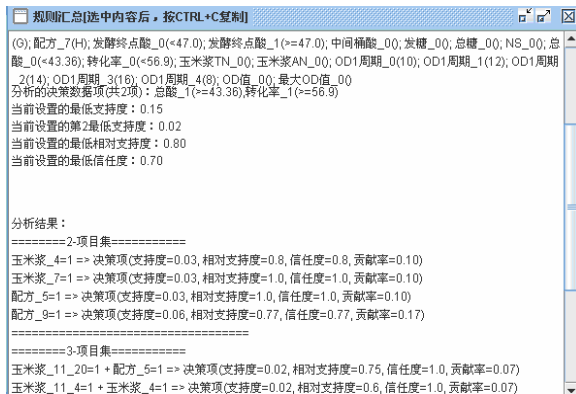


图 5 数据项关联规则挖掘结果

### 3 结语

对关联规则知识发现的研究一直是数据挖掘研究的重点领域,其市场分析、客户心理分析及环境质量评价中具有较广泛的应用<sup>[5-7]</sup>,而用关联规则挖掘来分析生物化工生产数据、并从中提取规律的研究现有资料较少<sup>[4-8]</sup>。本文以生物化工生产中的共性特征出发,研究并完成一个生化企业生产数据关联规则挖掘的应用系统,并对企业生产数据的现有数据集进行了初步试验和测试,取得较满意的效果。

本文设计和开发的生化企业关联规则挖掘分析系统,可以完成分析项目建立、数据预处理和数据挖掘

算法实现等过程,特别是在现有数据中存在稀有数据项时,提出了基于相对支持度的解决策略,较好地满足了企业的实际数据情况,所设计开发的软件系统,也可以应用于类似的生产企业数据分析中。

### 参考文献

- 1 曹月芹,林枫,陈国浪.基于 Apriori 分类事务库关联规则算法.计算机系统应用,2010,19(8):62-65.
- 2 陈申燕,曹旻.多层关联规则挖掘算法的研究及应用.计算机工程与设计,2010,4:885-888.
- 3 殷剑锋,徐建城,李伟强.改进 Apriori 挖掘算法的网格实现.计算机仿真,2010,2:145-148,268.
- 4 周永生,熊结青,沙宗尧.基于关联规则挖掘的生化企业数据分析及其应用研究.计算机与应用化学,2010,27(9):1252-1256.
- 5 管乐,王纯.多维关联规则挖掘在彩铃推荐中的应用.计算机系统应用,2009,18(4):155-157.
- 6 章杰鑫,张烈平.基于时序关联规则的商品需求预测.计算机工程,2009,35(22):65-67.
- 7 Chen CY, Shyue SW, Chang CJ. Association rule mining for evaluation of regional environments: case study of dapeng bay, taiwan. International Journal of Innovative Computing Information and Control, 2010,6(8):3425-3436.
- 8 张泉灵,金晓明,荣冈,苏宏业.化工生产过程数据挖掘系统的研究与应用.计算机与应用化学,2008,25(7):769-772.

(上接第 42 页)

的指导下,增加服务台数目(大于 5)时,系统处理能力表现更加优秀。

本文的 SMSDSM 设计方案对大中型企业的短信系统具有一定的参考和借鉴价值;另外,本文实际应用中的 SMSDSM 使用 java 实现,消息队列采用 EJB 组件规范中的 MDB 来支持,可以方便地被集成到不同操作系统的平台中使用。

### 参考文献

- 1 沈斌,李兴国,钟金宏,沈丽娜.基于多队列和多线程的短信实时并发控制算法.计算机工程,2008,34(8):62-65.
- 2 Wang ZT, Guo ZY, Wang ZQ. Analysis of Multi-task Scheduling Based on SMS. 2008 International Conference on

Computer Science and Software Engineering. IEEE. 2008. 1087-1089.

- 3 归敏丹,蒋毅飞,张志敏,吴锡生.多服务员时两种等待队列性能的比较.计算机工程与应用,2008,44(13):44-46.
- 4 唐应辉,唐小我.排队论基础与分析技术.北京:科学出版社,2006.50-57.
- 5 曾东海,刘海,金士尧.集群负载调度算法性能评价.计算机工程,2006,32(11):78-79.
- 6 Meng QY, Qiao JZ. A Dynamic Load Balancing Method Based on Stability Analysis. 2008 International Symposium on Computer Science and Computational Technology. IEEE. 2008.404-408.