第23卷 第5期 自动化学报 Vol. 23, No. 5
1997年9月 ACTA AUTOMATICA SINICA Sep., 1997

研究简报

一般客户/服务器系统的负载分析[1)]

金尊和 吴 澄 王 然
(清华大学自动化系国家 CIMS-ERC 北京 100084)

关键词 客户/服务器,系统建模,马尔科夫过程,排队网络.

1 系统描述

目前,有关客户/服务器系统分析和建模方面的工作不多[1,2]. 本文首次用马尔科夫排队网络分析、计算具有多类服务器的一般网络系统稳态性能的解析解,给出了服务器负载均衡的条件.

考虑由 K 个客户机、M 个服务器群组成的一般网络系统. 假设:1)客户机独立、随机地提出任务请求,请求的时间间隔服从参数为 λ 的负指数分布;2)客户请求和服务结果均经网络传送,传送时间服从负指数分布,参数为 μ_{net};3)服务器群相互独立,第 i 个服务器群由 $m_i(m_i\geqslant 1)$个独立的相同服务器组成,服务时间服从负指数分布,参数为 μ_i. 第 i 个服务器群接收请求的比例为 $p_{s_i}\left(\sum\limits_{i=1}^{M}p_{s_i}=1\right)$.

2 系统性能求解

2.1 一般客户/服务器系统的数学模型及其求解

按上述假设,这时系统实际上是一个有限源的马尔科夫排队网络(图1). 图中每个圆圈代表一个服务器群节点. 按文献[3]的定理1,得系统稳态时的联合分布概率为

$$p(k_{\text{net}},k_{s_1},\cdots,k_{s_M})=\left[\sum_{K\in A}\frac{\left(\dfrac{0.5\mu_{\text{net}}}{\lambda}\right)^{(N-k_{\text{net}}-\sum_{i=1}^{M}k_{s_i})}}{(N-k_{\text{net}}-\sum\limits_{i=1}^{M}k_{s_i})!}\prod_{i=1}^{M}\frac{\left(\dfrac{0.5p_{s_i}\mu_{\text{net}}}{\mu_{s_i}}\right)^{k_{s_i}}}{\beta_{s_i}(k_{s_i})}\right]^{-1}$$

$$\frac{\left(\dfrac{0.5\mu_{\text{net}}}{\lambda}\right)^{(N-k_{\text{net}}-\sum_{i=1}^{M}k_{s_i})}}{(N-k_{\text{net}}-\sum\limits_{i=1}^{M}k_{s_i})!}\prod_{i=1}^{M}\frac{\left(\dfrac{0.5p_{s_i}\mu_{\text{net}}}{\mu_{s_i}}\right)^{k_{s_i}}}{\beta_{s_i}(k_{s_i})},\tag{1}$$

---

1) 国家自然科学基金资助项目.
收稿日期 1995-05-02

其中 $\beta_{s_i}(k_{s_i})=\begin{cases} k_{s_i}!, & k_{s_i}<m_i, \\ m_i!m_i^{k_{s_i}-m_i}, & k_{s_i}\geqslant m_i. \end{cases}$

k_{s_i}为节点 s_i的客户机数.

2.2　系统的负载指标

利用稳态时系统的联合分布概率，计算系统的性能指标.

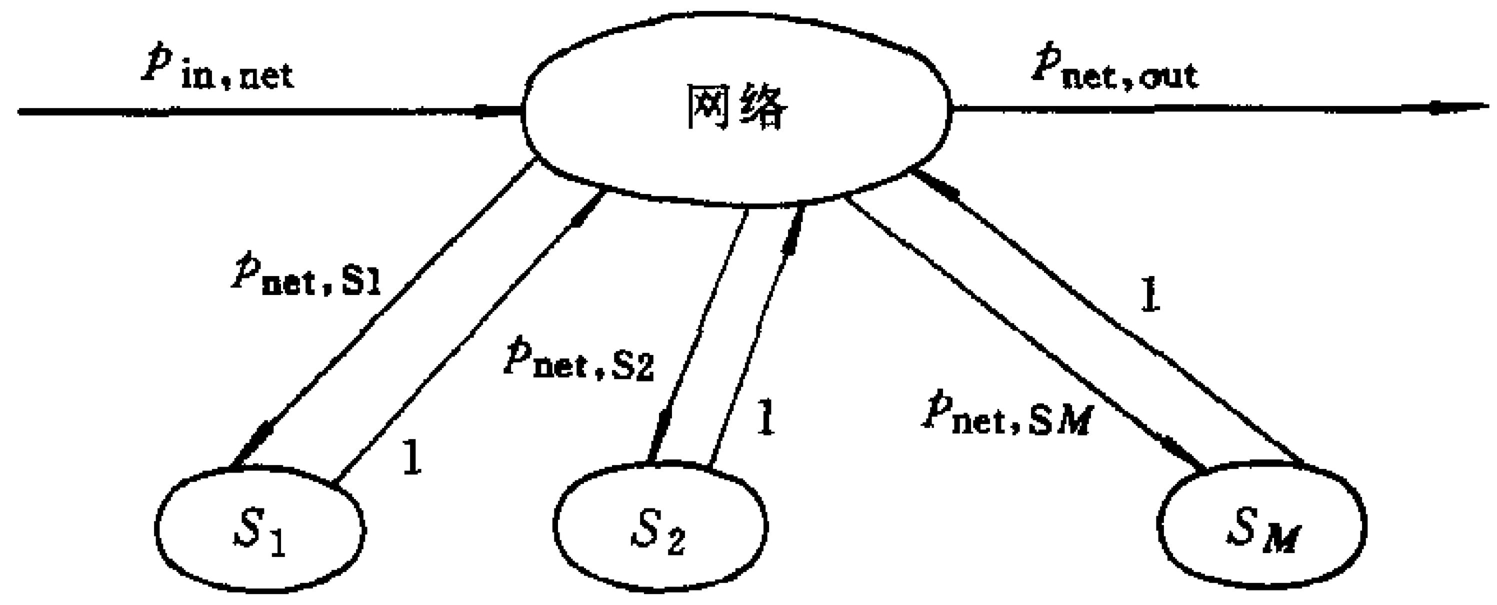

图1　系统的马尔科夫排队网络模型

网络的队长为

$$L_{Q_{\text{net}}}=\sum_{i=1}^{N}(i-1)p(i)=\sum_{i=1}^{N}(i-1)\sum_{K\in A_{\text{net}}}p(i,k_{s_i},\cdots,k_{s_M}). \tag{2}$$

网络的平均利用率为

$$q_{\text{net}}=1-p(0)=1-\sum_{K\in A_{\text{net}}}p(0,k_{s_1},\cdots,k_{s_M}), \tag{3}$$

其中 $A_{\text{net}}=\{(i,k_{s_1},\cdots,k_{s_M})\mid\sum_{i=1}^{M}k_{s_i}=N-i,i\in[0,N]\}$是网络分量固定为 i 的状态子空间.

第 i 个服务器的排队队长为

$$L_{s_i}=\sum_{i=1}^{N}(i-1)p_{s_i}(i)=\sum_{i=1}^{N}(i-1)\sum_{K\in A_i}p(k_{\text{net}},k_{s_1},\cdots,k_{s_i},\cdots,k_{s_M}). \tag{4}$$

平均利用率为

$$q_{s_i}=\sum_{j=0}^{m_i-1}\frac{j}{m_i}p_{s_i}(j)+\sum_{j=m_i}^{N}p_{s_i}(j)=1-\sum_{j=0}^{m_i-1}\frac{m_i-j}{m_i}\sum_{K\in A_i}p(k_{\text{net}},k_{s_1},\cdots,j,\cdots,k_{s_M}), \tag{5}$$

其中 $A_i=\{(k_{\text{net}},k_{s_1},\cdots,l,\cdots,k_{s_M})\mid k_{\text{net}}+\sum_{j\neq i}k_{s_j}=N-l,l\in[0,N]\}$是第 i 个服务器分量固定为 l 的状态子空间.

3　各类服务器负载均衡的条件

系统运行时，希望系统资源充分利用，各类服务器群的负载均衡，即 $q_{s_1}=q_{s_2}=\cdots=q_{s_M}$. 为此，有如下的结论：

定理1. 所有服务器的负载均衡的充分条件是

$$\mu_{s_1}:\mu_{s_2}:\cdots:\mu_{s_M}=p_{s_1}:p_{s_2}:\cdots:p_{s_M}. \tag{6}$$

定理2. 当各类服务器数均为1时，负载均衡的充要条件是

$$\mu_{S_1}:\mu_{S_2}:\cdots:\mu_{S_M}=p_{S_1}:p_{S_2}:\cdots:p_{S_M}, \tag{7}$$

也就是说，服务器负载均衡的条件是各个服务器的服务速度与接到请求的比例成正比，这与我们的直观理解是一致的.

4　计算实例

例1. 假设有一个基于客户/服务器结构的数据库系统，其参数如下：

$N=15,M=2,\lambda=0.3,\mu_{\text{net}}=10,m_2=1,m_1=1,\mu_{s_1}=2,\mu_{s_2}=2,p_1=1/3,p_2=2/3.$

由式(2)—(5)计算可得

$$L_{net}=0.420, q_{net}=0.491, L_{s_1}=2.130, L_{s_2}=2.130, q_{s_1}=0.850, q_{s_2}=0.850.$$

由结果可知,此时两个服务器的负载是均衡的,与定理1是一致的.

例2. 改变第二个服务器的参数,令 $\mu_{s_2}=3$,得到

$$L_{net}=0.523, q_{net}=0.533, L_{s_1}=2.862, L_{s_2}=0.711, q_{s_1}=0.891, q_{s_2}=0.663.$$

本文建立了一般客户/服务器系统的马尔科夫排队网络模型,得出了系统稳态性能的解析结果,并分析、论证了系统服务器负载均衡的条件.这不仅有助于认识一般客户/服务器系统的稳态特性,而且对一般客户/服务器系统的配置设计和优化具有重要作用.

参 考 文 献

[1] Moriguchi S. Performance and Performability Evaluation of Client/Server:System by Petri Nets. *Trans. A of IFIP*. 1992:622—628.

[2] Petriu D C,Woodside C M. Approximate MVA from Markov Model of Software Client/Server System. In:Proc. of 3rd IEEE Symposium on Parallel and Distributed Processing,322—329.

[3] Jin Zunhe,Wu C, Wang R. Analytical Results of Server Loads and Network Load in a General Client/Server System. *Science and Technology of Tsinghua*,1996,**1**(2):193—196.

SYSTEM LOAD ANALYSIS FOR A GENERAL CLIENT/SERVER SYSTEM

JIN ZUNHE WU CHENG WANG RAN

(CIMS-ERC,Tshinghua University,Beijing 100084,*PRC)*

Key words Client/server,System modeling,Markov process,queuing network.