第 28 卷 第 11 期
2005 年 11 月
计 算 机 学 报
CHINESE JOURNAL OF COMPUTERS
Vol. 28 No. 11
Nov. 2005

移动 Agent 系统中的排队机制研究

杨公平 曾广周 卢朝霞
(山东大学计算机科学与技术学院 济南 250061)

摘 要 针对现有的移动 Agent 系统缺乏排队机制的课题,定义了 Agent 排队系统的概念,然后分别给出了单工作节点和复合工作节点的排队模型,讨论了其中的排队规则、服务规则和 Agent 派遣机制等关键技术. 实验和分析表明,引入排队机制可以明显改善 Agent 和工作节点的运行质量.

关键词 移动 Agent 系统;排队机制;排队模型;动态优先队列;模型实验
中图法分类号 TP301

Study on the Queuing Mechanism of Mobile Agent System

YANG Gong-Ping ZENG Guang-Zhou LU Zhao-Xia
(*School of Computer Science & Technology, Shandong University, Jinan* 250061)

Abstract The current mobile Agent systems are lack of queuing mechanism. If a large number of Agents migrate to one work-node within a short period of time, the average executive time for each Agent would be much longer due to the competition for the limited resources. In this paper, the author introduces the queuing conception into mobile Agent systems to solve this problem. Two queuing models are created for corresponding two different types of work-nodes went by the name of sole work-node and compound work-node. For each model, a set of detailed queuing rules, service rules and the Agent dispatching mechanism are defined. Authors choose $P/N/m/\infty$ queuing model to implement queuing mechanism in sole work-node, and use dynamic priority queue to manage waiting Agents. Compound work-node is made up of a lot of sole work-nodes, when a Agent migrated to one compound work-node, it would be dispatched to one sole work-node according to the task type which will be executed. The effectiveness of the queuing mechanism is evident in that it can alleviate the load of work-nodes, which are illustrated with the experiment results at the end of this paper.

Keywords mobile Agent system; queuing mechanism; queuing model; dynamic priority queue; model experiment

1 引 言

排队是日常生活和工作中常见的现象,排队问题由两方面构成:一方有需求,要求得到服务,要求服务的人或设备称为顾客;另一方要设法满足对方的要求给予服务,服务人员或服务机构称为服务员或服务台;顾客与服务台构成一个排队系统. 顾客渊源而来,受许多因素的影响,顾客到达服务台的时刻是随机的,服务台也受许多因素的影响,服务完一个

收稿日期:2003-12-26;修改稿收到日期:2005-06-07. 本课题得到国家自然科学基金(60473123)、山东省科学技术发展计划项目基金(031110123)资助. 杨公平,男,1970 年生,博士研究生,副教授,主要研究方向为智能计算. E-mail: gpyang@sdu.edu.cn. 曾广周,男,1947 年生,教授,博士生导师,主要研究领域包括 CSCW、智能计算、移动计算及其应用技术. 卢朝霞,女,1974 年生,博士研究生,主要研究方向为工作流管理.

顾客的时间长短也是随机的. 因此,在任一时刻要求服务的顾客数超过服务台的服务容量时,顾客就必须排队等待,直到服务台出现空闲时才得到服务,服务完后离开.

在移动 Agent 系统中,Agent 为完成创建者赋予的任务要在其它工作节点之间进行迁移. 从角色上划分,移动 Agent 系统主要由移动 Agent 本体和 Agent 服务设施即 Agent 服务器组成,每个工作节点的 Agent 服务器对迁移来的 Agent 进行安全认证,为 Agent 分配和提供所需的资源及服务,控制 Agent 的运行. 由于工作节点所能够提供的资源和服务总是有限的,因此,当多个 Agent 在同一时间段内迁入工作节点时,Agent 服务器应该引入排队机制,以便对 Agent 的资源竞争做出限制,防止过多 Agent 的并发运行影响工作节点的服务质量.

FIPA(Foundation for Intelligent Physical Agents)从不同方面详细规定或建议了 Agent 在体系结构、通信、移动、知识表达、管理和安全等内容,对 Agent 在运行方面的调度策略未做规定;现有的移动 Agent系统如 aglet[1]、ajanta[2]、concordia[3] 均采用这样的调度策略:对迁移过来的 Agent 做必要的安全认证后,即让其投入运行,资源和服务方面的调度和管理工作交由操作系统承担. 缺乏排队机制的移动 Agent 系统其缺点是明显的:如果在同一时间段内有过多的 Agent 迁移到同一个工作节点,则拥挤和资源竞争必定会引起该工作节点运行负载过重,致使在该工作节点上运行的各 Agent 任务完成时间延长,甚或影响到系统的稳定性和可靠性.

针对上述问题,本文把排队机制引入移动Agent 系统,在下面的第 2 节给出了移动 Agent 排队系统的概念;第 3 节把工作节点划分为单工作节点和复合工作节点,论述了复合工作节点的结构和功能;第 4 节和第 5 节分别给出了基于两类工作节点的排队模型,讨论了其中的排队规则、服务规则、队列管理、复合工作节点结构、Agent 派遣机制等关键技术;第 6 节给出了模型实验与分析;结合正在进行的基于移动计算范型的协同产品商务平台研究,本文最后指出了移动 Agent 排队机制中进一步的工作.

2 移动 Agent 排队系统的概念模型

结合排队论思想,一个移动 Agent 排队系统可抽象为 Agent 到达 Agent 服务器后,Agent 服务器根据当前资源状况和服务需求情况决策该 Agent 能否满足运行条件. 若满足,便立即让其运行,否则让其等待,服务完后 Agent 离开工作节点. 因此可以用图 1 来表示一个移动 Agent 排队系统.

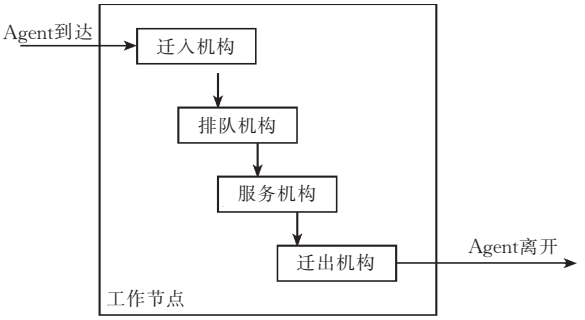

图 1 Agent 排队系统模型图

定义 1. 移动 Agent 排队系统 MAQS 是一个四元组(IM,QM,SM,OM),其中,

IM 是迁入机构. 它主要负责移动 Agent 的迁入管理,对到达的 Agent 进行安全认证,认证完毕提交给排队机构处理;

QM 是排队机构. 它对到达的移动 Agent 按排队规则进行排队;

SM 是服务机构. 主要是指服务台、服务规则和服务方式等;

OM 是迁出机构. 它主要负责移动 Agent 的迁出管理.

对于 Agent 排队系统来讲,移动 Agent 到达工作节点的过程称之为输入过程,可以用如下定义来描述输入过程.

定义 2. 输入过程 IP 是一个三元组(AN,AAS,ATC),其中,

AN 为到达排队系统的移动 Agent 数量;

AAS 为移动Agent到达排队系统的方式,$AAS=1$ 表示成批到达,$AAS=0$ 表示单个到达;

ATC 为相继到达排队系统的移动 Agent 之间的间隔时间分布,即输入分布 $A(t)$.

定义 3. SM 是一个六元组(SR,ST,SS,SAT,STC,SAI),其中,

SR 为服务规则. 用来描述 Agent 服务器 如何从排队队列中选取 Agent 投入运行;服务规则主要有先到先服务(FCFS)、短进程/线程优先(SPF/STF)服务、有优先权的服务(PS)等;

ST 为服务台集合. $ST=\{s_1,s_2,s_3,\cdots,s_m\}$,其中 $s_i(1\leq i\leq m)$代表第 i 个服务台,m 是服务机构中的服务台个数,即工作节点允许并发运行的最大移动 Agent 个数;

SS 为服务台集合中各个服务台之间的服务方式，$SS=0$ 为并行服务，表示各服务台可以独立地同时分别向多个 Agent 提供服务，$SS=1$ 为串行服务，表示各服务台只能按一定顺序为同一个 Agent 提供服务；

SAT 为每个服务台一次接受的移动 Agent 个数；

STC 为每个服务台为 Agent 提供服务的服务时间分布，即服务分布 $B(t)$；

SAI 为对同一个服务台，Agent 之间接受服务的时间是否独立，$SAI=0$ 独立，$SAI=1$ 不独立.

定义 4.　输入分布 $A(t)$：对于输入过程，设 C_n $(n=1,2,\cdots)$ 表示到达排队系统的第 n 个移动 Agent，其到达的时刻为 t_n，则 $\tau_n=t_n-t_{n-1}$ 表示 C_n 与 C_{n-1} 之间的时间间隔. 假定 $t_0=0$，则 $\tau_1=t_1$. 如果 Agent 到达是相互独立的，则 $\{\tau_n\}$ 是独立的随机变量序列. 假定它们有相同的概率分布函数，记作 $A(t)$，称为输入分布.

定义 5.　服务分布 $B(t)$. 设 $C_n(n=1,2,\cdots)$ 表示到达排队系统的第 n 个移动 Agent，C_n 的服务时间为 V_n. 如果同一服务台为各 Agent 服务的时间是独立的，则 $\{V_n\}$ 是独立的随机变量. 假定它们有相同的概率分布函数 $B(t)$，称为服务分布.

在图 1 所示的排队系统模型中，迁入机构和迁出机构是 Agent 服务器提供的基本功能模块，本文不再赘述.

3　移动 Agent 系统中工作节点的划分

移动 Agent 为完成任务需要在主机之间迁移，这些主机称为工作节点. 如果工作节点在物理上表现为一台计算机，即该节点不仅负责移动 Agent 的安全认证、资源分配和任务调度，而且也是移动 Agent 的实际运行场所，则称这种工作节点为单工作节点(sole work-node).

在基于移动计算范型的迁移工作流系统中[4]，Agent 迁移的目标是多个不同的企业站点，Agent 在某一企业站点的任务执行可能在多台物理计算机上进行. 换句话说，Agent 的工作节点是一个网络，而不是单台计算机，本文称这种工作节点为复合工作节点(compound work-node)，复合工作节点可以用图 2 表示.

复合工作节点由停靠站服务器和工作机网络构成，停靠站服务器不仅负责迁入 Agent 的安全认证和管理工作，而且根据 Agent 所需的资源表和服务表来选择工作机，把 Agent 派遣到合适的工作机上执行特定的任务. 执行完任务后 Agent 返回到停靠站服务器，进而迁出该节点到达下一个工作节点.

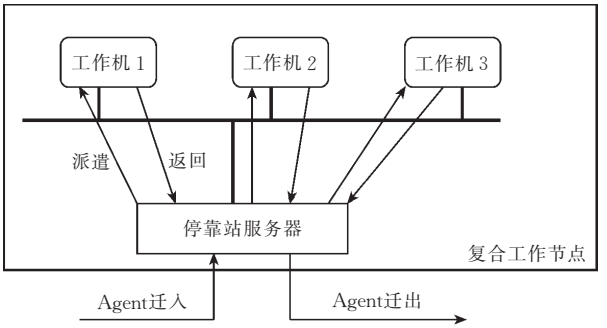

图 2　复合工作节点结构图

在图 2 所示的结构中，所有工作机对外界都是不透明的，工作机所能提供的资源和服务由停靠站服务器向外界公布，至于某一资源和服务由哪个具体的工作机提供则对外界保密. 复合工作节点利用停靠站服务器实现了内部工作机的安全，弱化了工作机的 Agent 认证职能，使各工作机能够更专注于协同完成 Agent 携带的工作任务.

4　单工作节点的 Agent 排队模型

单工作节点的排队模型如图 3 所示，下面说明模型的定义.

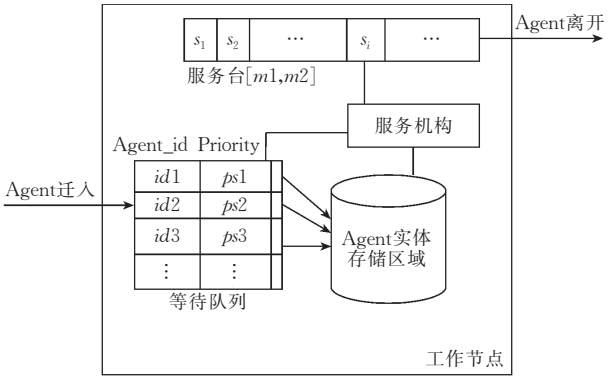

图 3　移动 Agent 的 $P/N/m/\infty$ 排队模型

4.1　$P/N/m/\infty$ 排队模型

参照本文第 2 节给出的定义，描述 $P/N/m/\infty$ 排队模型如下：

(1) 输入过程 IP.

AN 取无穷大，Agent 的来源是无限的，工作节点可以源源不断地为 Agent 提供服务，模型中记作 ∞；

$AAS=0$，Agent 是单个到达工作节点的，即不同的 Agent 独立地到达同一工作节点；

ATC:输入分布函数 $A(t)$ 为泊松分布[5],模型中记作 P;

(2)排队规则. 每个Agent按照优先权进行排队;

(3)服务机构 SM.

SR 实行有优先权的服务,服务台空闲时,从等待队列中选取优先权最高的 Agent 投入运行;

ST 取服务台的数目 m 在范围$[m1,m2]$内动态可调,$m1\geq 1$;

SS 为 0,即服务台之间互相独立,各服务台可以并行地为 Agent 提供服务;

SAT 为 1,即每个服务台一次只能接受一个 Agent;

STC:服务分布函数 $B(t)$ 为负指数分布[5],模型中记作 N;

SAI 为 0,即对同一服务台而言,Agent 之间接受服务的时间相互独立.

本文用动态优先队列管理等待的Agent(见图 3). 每个 Agent 都有一个优先数(priority),优先数可以根据情况更改. 动态优先队列的项只存储 Agent 标识($Agent_id$)及其优先数,Agent 实体由 Agent 服务器专门开辟另外区域存储. 服务机构从动态优先队列中选择出即将投入运行的 Agent 后,根据队列中 $Agent_id$ 标识从 Agent 实体存储区域中检索该 Agent 并调度其运行.

4.2 动态优先队列

动态优先队列用抽象数据类型 DPQ 描述如下:

抽象数据类型 $DPQ\{$

$Agent$ 对象$D=\{q_i\in ElemSet, i=1,2,\cdots,n, n\geq 0\}$;

$ElemSet:=\langle Agent_id, Priority\rangle$;

优先关系 $R=\{\langle q_{i-1},q_i\rangle \mid q_{i-1},q_i\in D, q_{i-1}.Priority>q_i.Priority, i=2,3,\cdots,n\}$;

基本操作

$Create()$:创建一个空的优先队列;

$IsEmpty()$:如果队列为空,则返回 true;否则返回 false;

$Add(q)$:向队列中添加元素 q;

$Take_out(q)$:从队列中取出具有最大优先权的元素;

$Adjust(id,ps)$:调整标识为 id 的 Agent 的优先数为 ps;

$\}$

Agent 服务器启动后,首先调用 $Create()$ 创建一个空的优先队列;当 Agent 通过迁入机构的安全认证后,排队机构计算该 Agent 的优先数,然后调用 $Add(q)$ 把该 Agent 插入到优先队列;若有 Agent 迁出造成某个服务台空闲,服务机构的调度模块则调用 $Take_out(q)$ 选择具有最高优先权的 Agent 投入运行.

优先级的确定可参考以下因素:

(1)Agent 任务的紧迫程度(初始优先数). 由 Agent 的创建者在创建时为其赋予一个紧迫值(紧迫值在整个系统中具有全局可比性). Agent 迁移到达某个工作节点后,Agent 的创建者还可以通过与该工作节点的协商更改紧迫值.

(2)Agent 的资源和服务占用程度. Agent 迁移到达某个工作节点后,该节点的排队机构可以根据自己的资源情况和优先策略为其设置一个优先数,例如所需资源少、任务简单的 Agent 优先运行等.

(3)Agent 的等待时间. 为保证调度的公平性,随着 Agent 等待时间的延长,其优先级应该能动态增长,以便保证每个处于等待状态的 Agent 总有被选择运行的机会.

动态优先队列的特性不仅体现在 $Add(q)$ 操作时要通过计算优先数而确定 Agent 的插入位置,而且体现在 $Adjust(id,ps)$ 操作上,排队机构可以根据上述因素进行动态的优先权调整.

描述动态优先队列最简单的方法是采用无序线性表,$Add(q)$ 操作的时间复杂度为 $O(1)$,$Take_out(q)$ 和 $Adjust(id,ps)$ 操作的时间复杂度均为 $O(n)$. 另一种比较高效的描述方法是采用最大堆(max heap),由于堆对应一棵完全二叉树,拥有 n 个元素的堆其高度为 $\log_2(n+1)$[6],因此 $Add(q)$,$Take_out(q)$ 和 $Adjust(id,ps)$ 操作的时间复杂度为 $O(\log_2 n)$.

5 复合工作节点的 Agent 排队模型

对于复合工作节点(图 2),移动 Agent 既可以在停靠站服务器上排队,也可以在各工作机上分别排队. 如果在停靠站服务器上排队,那么服务器排队机制仍然可以使用 $P/N/m/\infty$ 模型,每台工作机相当于一个服务台,此时需要解决的主要问题是如何探测工作机空闲状态的出现并且依据工作机的空闲状态适时进行 Agent 派遣. 如果以上一个 Agent 的返回为契机派遣下一个 Agent,那么两次迁移过程不仅会引起服务延迟,也不利于充分发挥服务台的潜力. 因此,本文约定:Agent 排队在各工作机上按照单工作节点模型 $P/N/m/\infty$ 进行. 这样,复合工作节点 Agent 排队模型的核心问题就变成:停靠站服务器如何根据 Agent 的需求派遣它到合适的工

作机上排队.

当复合工作节点中同时有多个工作机可以满足某个 Agent 的任务需求时，停靠站服务器需要根据自己的派遣策略为 Agent 优选工作机. 本研究采用的策略是：如果工作机 A 上等待队列中所有 Agent 需要的资源和服务量最少，则派遣 Agent 到 A，其中 Agent 对资源和服务的需求量在 Agent 设计时与任务一起指定，由停靠站服务器统计.

当 Agent 的任务可以划分成子任务并且需要由多个工作机协同完成时，停靠站服务器要考虑派遣机制. 本研究采用的 Agent 派遣机制有 3 种类型：

(1)停靠站服务器执行多次派遣. 比如某个 Agent 有 3 个串行的子任务：$SubTask_1$，$SubTask_2$，$SubTask_3$，需要分别在 3 个工作机 WS_1，WS_2，WS_3 上执行. 停靠站服务器先把 Agent 派遣到 WS_1 上执行 $SubTask_1$，返回后再派遣到 WS_2 上执行 $SubTask_2$，最后派遣到 WS_3 上执行 $SubTask_3$，所有任务执行完毕后 Agent 返回到停靠站服务器并迁出该工作节点.

(2)停靠站服务器为 Agent 创建一个临时旅行计划，由 Agent 携带. 以上面 3 个工作机执行 3 个串行子任务为例，停靠站服务器把 Agent 派遣到 WS_1 执行 $SubTask_1$，执行完毕后 Agent 根据旅行计划自主迁移到 WS_2 上执行 $SubTask_2$，执行完毕后再自主迁移到 WS_3 上执行 $SubTask_3$，最后返回到停靠站服务器并迁出该工作节点.

(3)克隆多个 Agent，同时派遣. 如果各子任务之间是可以并行执行的，为提高执行效率，停靠站服务器可以根据 Agent 创建者的许可克隆多个 Agent 并同时派遣到不同的工作机上执行. 假定 $SubTask_1$，$SubTask_2$，$SubTask_3$ 是 3 个可并行执行的子任务，可以分别在 3 个工作机 WS_1，WS_2，WS_3 上执行. 停靠站服务器首先克隆 3 个 Agent：Agent_c_1，Agent_c_2，Agent_c_3，然后分别把 Agent_c_1 派遣到 WS_1 执行 $SubTask_1$，Agent_c_2 派遣到 WS_2 执行 $SubTask_2$，Agent_c_3 派遣到 WS_3 执行 $SubTask_3$. Agent_c_1，Agent_c_2，Agent_c_3 执行完毕后各自返回到停靠站服务器，向原 Agent 提交执行结果. 最后 Agent 整合结果并迁出工作节点.

6　模型实验和分析

单工作节点排队模型为 $P/N/m/\infty$，最大堆作为动态优先队列，复合工作节点中 Agent 排队在各工作机上进行. 下面的实验内容为每个移动 Agent 都执行以下排序任务：排序元素为[0，2000]之间的随机整数，元素个数 $n=1000$. 每个 Agent 有 3 个子线程，分别执行冒泡排序、选择排序和快速排序[6].

实验 1.　无排队机制. WS_1 连续向 WS_2 派遣 6 个 Agent，这 6 个 Agent 到达 WS_2 后立即执行. 表 1 记录了各 Agent 的任务开始时刻、任务结束时刻和任务执行时间.

表 1　无排队机制时各 Agent 的任务执行情况表

Agent 名称	任务开始时刻 (h，min，s)	任务结束时刻 (h，min，s)	任务执行时间 (min，s)
Agent1	10，33，21	10，37，14	3，53
Agent2	10，33，23	10，37，16	3，53
Agent3	10，33，24	10，37，19	3，55
Agent4	10，33，26	10，37，18	3，52
Agent5	10，33，27	10，37，19	3，52
Agent6	10，33，29	10，37，19	3，50

实验 2.　$P/N/m/\infty$ 排队机制. WS_2 上服务台个数 $m=3$. WS_1 连续向 WS_2 派遣 6 个 Agent，这 6 个 Agent 到达 WS_2 后排队执行. 表 2 记录了各Agent 的任务开始时刻、任务结束时刻和任务执行时间.

表 2　有排队机制时各 Agent 的任务执行情况表

服务台	Agent 名称	任务开始时刻 (h，min，s)	任务结束时刻 (h，min，s)	任务执行时间 (min，s)
1	Agent1	10，42，12	10，43，51	1，39
	Agent4	10，43，51	10，45，24	1，33
2	Agent2	10，42，14	10，43，56	1，42
	Agent5	10，43，56	10，45，25	1，29
3	Agent3	10，42，15	10，43，52	1，37
	Agent6	10，43，52	10，45，22	1，30

比较表 1 和表 2 可以看出：在设置排队机制的移动 Agent 系统中，Agent 执行性质相同的排序任务所需的时间比在没有排队机制的移动 Agent 系统中的明显减少. 由于移动 Agent 在任务执行完毕后随即离开工作节点，因此排队可以整体上改善工作节点的负载性能.

对复合工作节点实验，也可以得到类似的结果.

7　结束语

在移动 Agent 系统中，工作节点是 Agent 的运行场所. 如果工作节点缺乏 Agent 排队机制，那么过多 Agent 的并发运行会导致工作节点负载过重，使在该工作节点上的 Agent 任务执行时间延长，影响工作节点的服务质量. 因此，移动 Agent 系统引入排队机制，限制过多 Agent 在同一工作节点上的并发运行是必要的. 本文从排队论的基本思想出发，

首先定义了 Agent 排队系统的概念，然后把工作节点划分成单工作节点和复合工作节点，并分别给出了基于这两类工作节点的排队模型，讨论了排队模型中排队规则、服务规则、队列管理、复合工作节点结构及 Agent 派遣机制等关键技术，最后给出了模型实验与分析. 实验结果表明，排队机制可以明显提高工作节点的服务水平和质量.

排队机制是移动 Agent 系统中一个新的研究内容，文中提出的模型在基于移动计算范式的迁移工作流平台上得到了初步实现. 结合正在进行的基于移动计算范型的协同产品商务平台研究，本文进一步的工作包括以下几个主要方面：(1) $P/N/m/\infty$ 排队模型中服务台数目的动态调整规则及实现机制；(2) $P/N/m/\infty$ 排队模型中优先数的动态调整算法；(3) 复合工作节点排队模型中停靠站服务器对工作机服务能力的动态评估和优选算法；(4) 复合工作节点排队模型中停靠站排队方式与工作机排队方式的对比分析等.

参 考 文 献

1 Aridor Y., Lange D. B.. Agent design patterns: Elements of Agent application design. In: Proceedings of the Autonomous Agents'98, Minneapolis, 1998, 108～115

2 Karnik Neeran, Tripathi Anand. Agent server architecture for the Ajanta mobile-Agent system. In: Proceedings of the 1998 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'98), Lasvegas, 1998, 66～73

3 Walsh T., Paciorek N., Wong D.. Security and reliability in Concordia. In: Proceedings of the Thirty-First Annual Hawaii International Conference on System Sciences, Kona, Hawaii, 1998, 7: 44～53

4 Zeng Guang-Zhou, Dang Yan. The study of migrating workflow based on the mobile computing paradigm. Chinese Journal of Computers, 2003, 26(10): 1343～1349(in Chinese)
(曾广周，党 研. 基于移动计算范型的迁移工作流研究. 计算机学报, 2003, 26(10): 1343～1349)

5 Meng Yu-Ke. Queueing System Basic and Its Application. Shanghai: Tongji University Press, 1989(in Chinese)
(孟玉珂. 排队论基础及应用. 上海：同济大学出版社，1989)

6 Wang Shi-Li, Sun Xiao-Dong. Data Structures, Algorithms, and Applications. Beijing: China Machine Press, 2002(in Chinese)
(汪诗林，孙晓东译. 数据结构、算法与应用. 北京：机械工业出版社, 2002)

YANG Gong-Ping, born in 1970, Ph. D. candidate, associate professor. His research interest focuses on intelligent computing.

ZENG Guang-Zhou, born in 1947, professor, Ph. D. supervisor. His research interests include CSCW, intelligent computing theory and technology, mobile computing and its application technology.

LU Zhao-Xia, born in 1974, Ph. D. candidate. Her research interest is mainly work-flow management.

Background

This work is mainly supported by the National Natural Science Foundation of China under grant No. 60473123. Title of the project is "Study on the Model Theory and Systemic Method of Migrating Workflow". This project aims at providing a bottom-up method to organize business process via multi-Agent negotiation in a mobile Agent system.

Agent servers can become a bottleneck in a large-scale Agent system. When a large number of Agents migrate to one Agent server within a short period of time, the average executive time for each Agent would severely suffer due to the competition for the limited resources. This paper solves the problem by introducing the queuing conception into the mobile Agent system. Experiments show that the queuing mechanism is effective in improving both the running quality of Agents and the performance of Agent servers.

The research direction of this group includes mobile Agent system, workflow management, artificial intelligence and CSCW. The research group has published near 40 papers on the research direction.