

多处理器片上系统任务调度研究进展评述

李仁发 刘彦 徐成
(湖南大学计算机与通信学院 长沙 410082)
(lirenfa@vip.sina.com)

A Survey of Task Scheduling Research Progress on Multiprocessor System-on-Chip

Li Renfa, Liu Yan, and Xu Cheng
(School of Computer and Communication, Hunan University, Changsha 410082)

Abstract Multiprocessor is very common in embedded computing systems because it can meet the performance, cost and energy/power consumption goals. Multiprocessor system-on-chip is often heterogeneous multiprocessors and integrates multiple instruction-set processors on a single chip that implements most of the functionality of a complex electronic system. Current trends indicate that multiprocessor system-on-chip is being increasingly used in application such as image processing, network multimedia, embedded system, and so on. Scheduling and mapping of tasks are important key problems in multiprocessor system-on-chip design, and are substantially more difficult than scheduling a uniprocessor. The basic architecture and design challenge of multiprocessor system-on-chip task scheduling algorithm are introduced. In particular, the current research progresses are summarized according to scheduling algorithm analysis and implementation framework. The scheduling algorithm analysis is classified into three categories, and scheduler implementation framework is classified into two categories by using task modeling. Many open research problems are pointed out. Because of the large variety of timeliness requirements in real-time applications, an important goal is to find canonical representations of task considering timing constraints. It is an important target to implement high-effect's scheduler based on multiprocessor system-on-chip platform. By comparing and analyzing these different projects and algorithms, researchers of related topic can gain useful information about task scheduling problem.

Key words multiprocessor system-on-chip; real-time task scheduling; embedded system; architecture exploration; scheduling model

摘要 多处理器片上系统在单芯片上集成了多种指令集处理器,可完成复杂完整的功能,在图像处理、网络多媒体和嵌入式系统等应用领域前景广阔.任务映射与调度是多处理器片上系统设计的关键问题之一.介绍了多处理器片上系统的基本结构和面临的挑战,从调度算法分析和实现框架两个方面着重探讨了近年来多处理器片上系统任务调度的国内外研究进展情况,分析了当前亟待解决的问题与下一步主要的研究方向,可为多处理器片上系统相关研究提供参考.

关键词 多处理器片上系统;实时任务调度;嵌入式系统;体系结构探索;调度模型

中图法分类号 TP316.4

市场对计算平台性能需求的不断提升,使得多处理器片上系统(MPSOC)成为新一代 SOC 的主流设计趋势^[1-2]. MPSOC 一般由多个处理器单元、专用功能模块甚至混和信号电路组成,构建一个复杂的集成计算系统,从而满足市场对于系统在计算性能、功耗、实时性与成本等方面的需求. MPSOC 不是简单的片上多处理器(chip multiprocessor),后者强调将更多的处理器放在单片上提高单位面积晶体管密度,并不考虑平衡应用的需求. MPSOC 则通过定制体系结构来满足不同应用在成本和功耗等方面的需求,已广泛地用于通信、消费类电子产品和网络多媒体等诸多领域^[3-4].

近年来, MPSOC 相关研究引起了广泛的关注,一些企业和科研院所围绕编程模型、任务调度、系统开发、调试环境与系统安全等方面开展研究,在 IEEE 和 ACM 相关会议上发表了一些重要的研究成果. 2006 年 8 月 14 日在美国举行了 MPSOC 2006 会议,进行了多处理器系统芯片设计的理论和战略研讨. AMD 公司在 2006 年提出了基于“融合”(fusion)技术的多核处理器产品路线图,其实质是由当前的同构片上多处理器转向更加具有“硬件加速器”概念的异构片上多处理器系统. 英特尔、Sun 和 IBM 等大公司也已投身到多核或者多线程技术的研发之中,并投入巨资建立了各自庞大的研究计划. 美国斯坦福大学^[5]、普林斯顿大学^[6]、威斯康辛大学^[7]等著名学府也一直关注 MPSOC 相关的研究工作,启动了一系列的研究计划;法国的 TIMA、新加坡、意大利等大学和研究室也发表了相关研究成果;国内中国科学院、中国科技大学等也持续跟踪相关研究.

把一个复杂的任务分解,将其映射到异构多处理器上,满足实时性、功耗与成本等条件的约束是 MPSOC 系统设计的关键问题之一. 很多任务调度的研究是基于抽象模型给出任务的形式化定义,然后进行调度策略与算法的设计. 由于多处理器任务调度问题是 NP 难问题^[8],改进算法的效率并构建任务调度实现机制成为研究重点. 同时,体系结构上的异构性使得 MPSOC 任务调度特别需要关注任务在不同体系结构的处理单元上执行时的特点,并在调度器实现上提出了很多新的挑战. 本文首先介绍了 MPSOC 体系结构与关键的设计挑战,然后综述了 MPSOC 任务调度领域目前的研究状况,主要针对实时任务调度的算法分析、实现机制、面临的挑战与亟待解决的问题等方面开展深入讨论.

1 MPSOC 体系结构与设计挑战

1.1 MPSOC 体系结构

MPSOC 体系结构常用于嵌入式计算系统中,并在性能、成本和功耗等方面有明确要求,其体系结构具有明显的异构性,如图 1 所示. 典型的 MPSOC 由 3 类组件组成:处理器单元,包括处理器、DSP 或 ASIC 用于执行计算任务;存储模块用于保存数据;互连网络用于组件间通信:

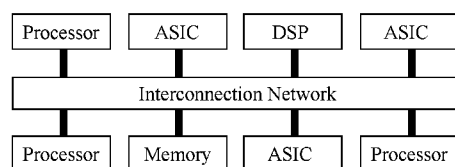


Fig. 1 MPSOC architecture.

图 1 MPSOC 体系结构

异构性是嵌入式多处理器显著的特点,即将计算能力赋予需要它的地方来满足性能需求,移除不必要的组件来降低功耗和成本. 系统中可以有不同类型的可编程处理器(如 CPU 和 DSP),也可以有专用处理器单元(如 ASIC)来完成单个功能. 片上存储结构和通信互连结构也是异构和定制的,从而满足嵌入式应用在性能和实时性上的需求. MPSOC 在嵌入式应用中已经逐渐展示其强大的生命力.

1.2 MPSOC 设计挑战

虽然可从传统的并行计算机系统中吸取经验,但 MPSOC 在应用中的实际需求使其面临独特的设计挑战,这仍将困扰 MPSOC 设计者相当长的时间. 实际上,设计复杂的 MPSOC 对工程师和学者带来新的挑战^[2-4,9-10].

1) 设计方法学: MPSOC 设计者需要一种结构化的设计方法支持系统的设计,从而减少系统设计时间,并对整个系统的开发周期和资源进行更好的估计. 目前 MPSOC 设计方法学的研究不断进展,并在未来随着新技术和工具的出现更加的重要.

2) 体系结构:包括处理单元、存储结构和片上互连方法需要在可编程性与灵活性之间取得平衡. 异构存储结构、可定制处理器和互连方式逐渐在体系结构设计中采用,但相关理论研究依然是体系结构的重要课题.

3) 编程模型:与传统的顺序编程方式相比, MPSOC 更能符合算法的并行处理方式. 但分解

一个顺序描述的应用成为可预测的互相协作的一组并行任务仍然是非常困难的事情。

4) 同步与控制: MPSOC 中将任务映射到处理器或者功能单元, 任务之间的同步与控制仍具有挑战。任务的控制是集中式还是分布式, 任务间是否有大量的处理器间通信以及优先级翻转、资源冲突等问题同样存在。

5) 系统的调试与工具: MPSOC 上运行的任务使用不同的编程模型和语言, 运行在不同结构的处理器上, 任务间通信方式也多种多样, 从而使得设计者对整个系统保持一个统一的视角, 进行调试, 这是一个非常困难而又急需解决的问题。

6) 操作系统与中间件: MPSOC 平台的操作系统和中间件应该在代码容量受到限制的情况下, 充分利用异构多处理器的特性进行优化, 提供调度、内存管理、IO 和通信管理的同时满足高性能、实时和低功耗的要求。

7) 嵌入式系统安全: MPSOC 与网络互连带来严重的安全问题。此外, 对于嵌入式系统应用领域, 通过攻击降低服务质量, 从而使得系统失去实时性成为有别于传统安全领域问题的一个新课题。软硬件结构设计者需要合作研究来提供安全的 MPSOC 结构。

2 MPSOC 任务调度研究进展

任务调度是经典问题, 很多相关综述^[11-12]从各个角度对其进行了详细的介绍。本文主要关注多处理器片上系统的任务调度问题, 将从任务调度算法分析和任务调度实现框架两个新的角度总结近年来嵌入式多处理器计算平台上任务调度相关研究进展。

2.1 任务调度算法分析

任务调度算法分析主要针对抽象任务模型进行可调度性分析和调度算法设计, 提高算法效率和系统的可调度性是重要问题^[13]。本节将对近年来基于不同任务模型的多处理器片上系统调度算法相关研究进行讨论。

2.1.1 周期模型类相关算法

任务调度的最早相关研究工作之一是 Liu 和 Layland^[14]提出的, 其周期任务模型是很多实时任务模型的基础, 并扩展至多处理器环境下用于任务调度可行性分析与算法设计, 其主要优势在于忽略某些实现细节, 抽象层次较高。

周期任务模型中任务由有限个执行时间为 e 的作业(jobs)组成, 其后续作业在至少 p 个单位时间后产生。在该任务模型中任务之间不共享资源。Baruah 描述了一种多处理器周期任务调度模型, 其主要内容如下^[15]:

- 1) 处理器集合用 Π 表示, 任务集合用 Γ 表示;
- 2) 有一个 $(|\Gamma| \times |\Pi|)$ 大小的实数矩阵 $[u_{i,j}]$, 其中 $u_{i,j}$ 表示 j 处理器用于执行 i 个任务所需要的计算能力比例; 如果 i 任务不能在 j 处理器上执行, 那么 $u_{i,j}$ 为 ∞ ;
- 3) 任务在处理器上执行的速度用速率矩阵 $[r_{i,j}]$ 表示。 $r_{i,j}$ 表示 j 处理器执行任务 i 的作业的速度; 如果 i 任务被分配到 j 处理器上执行, 则其每个作业将执行 $(e_i/r_{i,j})$ 时间单位;
- 4) 计算平台的异构性通过任务在不同处理器上执行时间的不同来描述多处理器。

多处理器任务划分问题^[15-16]即为有一组任务和一组可以执行任务的处理器, 根据处理器执行任务的速度、调度的目标看能否在符合约束条件的情况下将任务分配到各个处理器上执行。

上述的任务形式化模型有一些重要的假设, 即任务的独立性、任务划分执行和无通信开销。文献[16]首次从理论上分析异构多处理器平台上的周期任务的调度问题, 证明多处理器任务划分问题是 NP 难问题。作者用整数线性规划(NP 难问题)建模, 并进一步用线性规划解调度问题, 从而提出一种多项式时间调度算法。文献[17]中研究对象是多处理器(uniform multiprocessors)平台上的周期任务模型。实际上将调度问题划分执行, 从而对于已经调度好的任务在单处理器上执行时用 EDF(优化)策略。

任务的数据依赖、任务迁移和通信等方面的限制使得周期任务模型未能很好地描述实际系统, 研究者不断放松限制条件研究任务调度问题^[15, 18-21]。当允许任务抢占和迁移时一般具有更好的可调度性。Baruah 重点研究了允许任务可抢占和迁移的情况下异构多处理器任务调度问题^[15], 并提出了异构多处理器平台上周期任务的多项式时间全局调度算法。文献[20]考虑在多处理器计算平台上的全局调度问题, 允许任务抢占和任务迁移。文中针对多媒体应用等软实时系统, 研究具有一定延迟的任务模型中基于 EDF 调度原则进行全局任务调度, 并进一步提高多处理器计算平台整体的利用率。实际上, 任务调度是实现应用系统根据所需要的性能和服务等级

动态、自适应地从单处理器实现扩展到多处理器实现的关键问题^[21],特别在嵌入式应用领域,由于多处理器结构相对简单,因此使用允许任务迁移的全局调度有一定优势。

近年来针对松散任务和非周期任务模型上的调度问题研究广泛^[19,22-24]。文献[19]对具有期限约束的松散周期任务在多处理器上的调度问题进行研究,设计了多项式时间算法,并从理论上进行了分析和证明。文献[22]提出了一种实时异构系统的非周期任务动态分配调度算法。该算法采用的是在当前局部调度时,按一定规则在待调度的任务集中选取一批任务并构造目标函数,将问题转化为非平衡分配问题,以增大任务调度的可行性。文献[23]提出了一种非周期任务动态调度算法——分组适度算法。该算法包括分组策略和适当选取策略,可以提高平台资源的利用率和处理器利用率,性能优于近似算法和节约算法。

嵌入式计算系统始终关注可靠性与功耗。利用任务冗余提高系统的可靠性和调度成功率是近年来多处理器任务调度研究的一个方向。Gopalakrishnan 等人在文献[25]中通过复制任务在不同的处理器上运行,从而在多处理器平台上通过任务级的冗余来提高可靠性。该文的主要贡献是提出一个任务划分与复制的全多项式时间近似算法。国内研究者^[26-27]也提出多处理器容错调度算法,其中阳春华等人^[26]将 RM 和 EDF 相结合,同时调度周期任务和非周期任务,在处理器周期任务的预留处理器时间内调度非周期任务,使得系统到达的非周期任务并不需要对周期任务重新调度,在保证系统实时性的情况下降低了在线调度算法的复杂度。台湾国立大学的 Hsu Heng-Ruey 等人^[28]在给定能耗的约束下研究实时周期任务的调度问题,针对多处理器任务分配这个 NP 问题,提出一个近似优化算法,并可将研究成果应用动态电压调度技术从而进一步降低功耗。

2.1.2 DAG 模型类相关算法

实时系统中可用一个有向无环图 $DAG(V, E)$ 对任务进行建模,其中 V 是节点的集合,用于表示用于分配的任务,而 E 是一组有向边,用于表示节点间的优先约束或通信开销。基于 DAG 图任务调度问题可以描述为^[29]将任务分配处理器并协调执行,使得在满足实时性约束条件下任务的整体执行时间、功耗或其他指标最优,它同样是 NP 难问题。

有向无环图任务模型与周期任务模型相比更加实用,能直接表达任务间的优先约束及通信开销等,

在并行多处理器任务调度研究中广为使用。近年来,基于 DAG 任务模型的实时任务调度相关研究,着重针对嵌入式计算环境下多处理器系统的各种应用(如网络多媒体应用)对任务调度提出的新需求开发高效算法,并引入新的计算方法和工具解决传统调度问题。

Kuang 等人^[30]基于 DAG 任务模型,考虑了任务间的优先约束和通信开销,针对多媒体应用进行实时任务的划分和流水化调度,目标是在保证应用性能的前提下降低整个系统的成本。文献[31]对嵌入式多处理器分簇调度算法进行了评估和研究,将其与传统调度算法进行了比较。文中实验结果表明,对实时任务进行分簇预处理以降低任务调度时的通信开销,应用于多处理器任务调度效果很好。但分簇算法的选择和性能将很大程度影响后续任务调度的结果。Becchi 和 Crowley 在文献[32]中认为任务管理是提高多处理器平台计算性能的关键,并开发了一个运行时的监控程序捕捉进程动态行为,允许进程在多个处理器之间迁移。作者将进程动态分配与静态分配原则在同构多处理器和异构多处理器平台上进行了大量的对比实验,结果表明,在异构多处理器平台上使用动态进程分配方法可以显著地提高整体性能。

研究者十分关注嵌入式系统功耗和成本的约束,文献[33]则首次将温度因素作多处理器片上系统设计时的约束,提出温度敏感多处理器任务分配与调度算法,并开发了相关的启发式算法和工具。实验表明,对于各种结构的多处理器片上系统,温度敏感调度算法在降低系统峰值温度和平均温度上都有良好的表现。

近年来有一些研究者引入新的计算方法来解决多处理器调度问题。文献[34]中提出了一种用于多处理器任务调度的遗传算法,该算法具有灵活和自适应的特定,与传统的多处理器任务调度算法相比调度性能更好,但是运行时间开销较大。Krzysztof Rzdca 等人^[29]用 DAG 图来形式化表示任务,并利用差异演化(differential evolution)来优化任务的优先级。文献[35]提出了基于细胞自动机的多处理器任务调度算法,任务也是用 DAG 图表示。该调度算法分为学习、一般操作和重用操作 3 个阶段,并引入遗传算法和人体免疫系统模型来学习任务调度的规则。新的计算方法的引入使得问题求解的精度有所提高,但算法效率有待进一步改进。

2.1.3 其他扩展模型算法

针对不同应用领域,研究人员陆续提出新的任务模型,增强表达实际应用的能力,不断提高系统的可调度性和调度成功率。

Multi-frame^[36]模型用于对执行时间有变化的任务进行建模.比如在流媒体应用中由于每一帧的大小可能不同,若仍用周期任务模型则过于保守.文献[37]中对 Multi-frame 模型进行了进一步的总结.

循环实时任务模型(recurring real-time task model)^[38]将 DAG 图任务模型扩展为具备条件分支表达能力.文献[39]分析了基于上述模型在单处理器上任务调度的充分必要条件,并推导出多项式时间算法,并可再进一步推广到多处理器计算平台上.文献[40]中提出一种用于加速调度分析与评估速度的交互式模型方法,并基于循环实时任务模型进行了实验,结果表明应用交互式调度分析方法可以有 20 倍的性能提升,该方法也可以扩展至其他任务模型.

基于流的任务模型^[41]是一种面向流数据的任务模型.该模型由两部分组成,其一是用于准确描述流媒体应用软件的任务结构,其二是用于表达系统中数据到达的事件模型,Chakraborty 进行了调度分析研究,给出了可调度性判定条件,表明该模型相比于传统的周期任务模型表达能力更强.文献[42]则针对控制领域的应用中任务调度可能会对数据获取及过程处理带来延迟,提出了理想瞬间任务模型,可解决一些传统任务模型无法很好处理的控制问题,提高控制器的性能.文献[43]提出一个新的任务模型,除了任务执行速度外还增加了“粒度”参数.

2.2 任务调度实现框架

任务调度实现框架更多从系统和实际应用角度看待任务调度问题,着重研究多处理器平台上任务划分、调度机制和调度器的实现.本文将从体系结构探索与调度器实现两个方面讨论近年来在 MPSOC 任务调度实现框架领域的研究进展.

2.2.1 体系结构探索与任务映射

文在进行系统设计时如何将应用划分和分配到 MPSOC 平台,并考虑成本、性能、功耗和安全^[44]等约束是体系结构探索的重要内容.

Kempf 等人在文献[45]中提出了一种基于 SystmeC 的模拟框架,可以对将应用映射到平台的过程进行定量评估.该方法的关键是用一个可配置的事件驱动 VPU 来捕捉多处理器/多线程 MPSOC 平台的时间行为.任务由功能和时间模型来表示,系统由事件模型描述.文中实验结果显示该框架的模

拟速度、接口交互性均有显著优势,并能较好地在“应用到平台”的任务分解和映射中进行异构体系结构的设计空间探索和进行性能分析.但该文在任务映射、调度方法及成本功耗等方面尚未考虑进一步优化. Sesame^[46]是一个用于异构多处理器系统进行高效体系结构探索的软件框架,利用多目标非线性混和整数规划来评估处理时间、功耗和成本等设计约束.文献[47]提出一种通用和可扩展的体系结构探索框架,为多处理器平台上的各个处理单元选择合适的调度策略,对系统整体的功耗和成本进行评估.在实际多媒体终端进行的实验表明,设计者使用文中所提框架可以更好地降低系统的整体功耗.

由于 MPSOC 上的任务分配和调度必须考虑性能和功耗两个方面的因素,文献[48]结合基于 NoC 通信模型上的性能评估和功耗评估方法,提出一个基于任务图的 MPSOC 仿真器,底层基于 Mesh 结构的 NoC,并提供动态电压调度机制,可以对多处理器片上系统设计时在计算能力与整体功耗之间进行权衡.文献[49]中实时调度框架由两部分组成,一部分是设计时离线的任务分配,计算量较大;另一部分是运行时任务调度.这种二阶段的任务调度框架增加了系统设计的灵活性,减少了多处理器片上系统的设计时间,降低整个系统的功耗.文献[50]为了减少形式化调度模型与实际多处理器调度技术和通信机制之间的不匹配,基于输出事件模型提出一种系统级调度方法,可以很好地利用已有的调度分析技术.

2.2.2 调度机制与调度器实现

将调度算法理论分析的成果进一步应用到实际平台上,需要高效、可执行的调度机制与调度器,研究涉及体系结构、操作系统等多个方面.

Chen 等人在文献[51]中开发了一个基于异构多核片上系统的嵌入式微内核,主要是为多个核上运行的内核提供通信机制.内核中的进程调度是基于优先级的. Martino Ruggiero 等人在文献[52]中实现了一个基于问题分解的优化调度器.首先作者根据通信的开销来安排分配和调度的方法;其后对总线行为进行高层建模,并通过精确的模拟确定可以确保系统确定性行为的条件.映射和调度算法确保所有任务满足这些条件.作者对于映射和调度采用不同的方法进行处理,并使得两者的过程可以交互,从而得到更好的调度结果,并证明了其算法的收敛性. Cho 等人在文献[53]中实现了一个 MPSOC 的静态调度器.从实现的角度来说,作者考虑了同步

相关开销的细节.文章的主要贡献是在 MPSOC 系统的实现中,对于分布式和集中式调度器使用的权衡和评估.同时,在平台上进行了 H. 263 编码的分布式和集中式调度器的实现和对比分析.

Xue 等人在文献[54]中考虑的是多个应用在多处理器片上系统上运行时进行优化与资源管理的问题.作者主要分两层来实现上述想法:第 1 层是离线地(offline)对应用进行概要描述并注释,第 2 层则是根据上一步的结果进行实时(online)地资源分配.本文与其他相关研究比较显著的不同是以“应用”代替“任务/线程”作为管理和调度的对象,而用于分配的资源也将计算能力和存储统一考虑.作者通过对实际程序的分析发现并不是一直对每个应用赋予更多的资源就一定会带来性能提升.因此,应该给资源划分找到一个合适的结合点.文中使用的“离线+在线”调度方法是比较合理的.可以先将计算量大的优化运算在离线部分完成,得到一组优化的解;然后在线调度时则可以保证实时性的情况下获得较好的调度效果.

Ma 等人在文献[55]中探讨开发了子任务级的并行,即为了更好地利用处理器,多个任务的子任务可以调度交错运行. Ma 提出一个设计时算法进行交错运行调度.该算法可以作为一个层次化调度器的一部分.实际上该项研究的主要思路是将任务分解为子任务.这样可以通过一个任务由于数据相关性带来的处理器空等待由另外一个任务的子任务来“插空”,从而进一步提高处理器的利用率.文中阐述了作者提出的子任务分解技术和调度算法,并使用 VTC, IDCT 和 FIR 测试代码进行了实验.

任务迁移目前已成为 MPSOC 领域的研究热点,重点在降低迁移开销和针对不同目标的多处理器任务迁移与对系统性能的提升. Nollet 等人在文献[56]中介绍了一种使用处理器的调试寄存器来降低任务运行时开销和迁移反映时间的技术. MPSOC 中实时任务的迁移在系统实现时是一个关键问题.文中描述了一种便于实现的工程技术,充分利用空余的寄存器来降低任务迁移的开销. Stefano Bertozzi 等人在文献[57]中也讨论了 MPSOC 平台上的任务迁移问题,提出了一种适合 MPSOC 操作系统的任务管理机制.与前面提到的技术不同的是作者是基于代码断点进行任务切换的.在基于 SystemC 的 MPARM 仿真平台上进行的实验证明了作者方法的有效性.任务/数据迁移^[58]是一种适合异构处理器平台的任务调度方案.针对功耗、通信、成本等

不同目标,通过“离线分析、在线调度”对不同应用领域定制设计调度机制,可提供较好的整体性能提升.

3 MPSOC 任务调度亟待解决的问题

多处理器片上系统由于其系统的异构性和高集成度,使得任务调度与映射问题与传统并行系统有很大的不同.虽然在 MPSOC 任务模型、调度实现机制等方面已有部分成果发表,但其真正走向应用成熟还需要解决一系列重要问题.

3.1 任务分配、映射与体系结构探索问题

系统体系结构探索时将逐渐重视任务的划分与映射对系统结构的影响. MPSOC 体系结构的异构性使得任务分配与映射问题复杂性大大增加,处理器结构和功能的不同、片上互连方式的不同、存储结构的异构性都直接影响任务划分与调度的效果.将任务划分算法、任务调度机制与结构设计相结合,从而降低成本、提高性能、增加系统可靠性等是体系结构探索的目标,迫切需要研究高效快速的探索机制.通过高层次系统抽象模型,综合考虑片上通信、任务执行与切换、存储空间分配等因素,构建快速 MPSOC 体系结构设计支撑平台将是下一步相关研究的重要方向.

3.2 任务模型的改进

目前的任务形式化模型还不能完全适应 MPSOC 应用领域的任务描述的需要,需要进一步研究表达能力更为完整的形式化模型.新的任务模型应该与实际应用结合更紧密,并且便于在实际的实现平台上进行快速验证.根据嵌入式系统面向应用开发的特点,在特定领域研究专用任务模型,并提出一套与之配套的调度分析算法、形式化验证工具等将是未来研究的重点.例如,针对网络多媒体应用,综合考虑软、硬实时任务以及服务质量,提出新的任务模型将是很有意义的研究工作.

3.3 任务迁移等问题

任务调度时是否允许任务迁移将在很大程度上影响可调度性分析和调度算法的复杂性. MPSOC 体系结构的异构性又使得任务迁移的实现机制相比 SMP 要复杂很多.不同的体系结构使得任务的迁移所需开销差别很大,特别是动态可重构计算平台的出现,使得软硬件任务协同工作成为可能.如何建立一套较为有效的任务迁移评价体系与方法,对面向特定应用的嵌入式系统中任务迁移成本进行评估,还是需要进一步加以研究和探讨.同时,任务复制可

提高调度成功率和可靠性,并可更多地满足任务实时性需求.但在MPSOC平台上任务复制研究刚刚起步.

3.4 调度器的实现

MPSOC应用驱动的特点十分明显,因此任务调度的研究成果最终要应用到实际平台上并提高系统性能.近年来对MPSOC任务调度算法的相关研究仍然停留在调度分析和算法效率改进上,无论是离线调度还是在线实时调度都与实际应用还有不小的距离.目前各种多核片上系统平台及开发工具较多,在多核操作系统任务管理、调度器实现方面还没有重大突破,还需要结合新的任务模型进一步改进算法效率与实现机制,并在具体的应用平台上实现.多处理器任务调度器的实现研究将是多核操作系统相关研究中非常重要的内容.

4 总 结

多处理器片上系统(MPSOC)已经成为嵌入式应用领域的主流体系结构,但相关软件支撑系统、编程模型和多核操作系统的研究仍显不足,任务调度相关课题的研究是其关键问题之一.本文介绍了MPSOC的关键设计挑战,并从调度算法分析和实现框架两个方面综述了近年来在MPSOC平台上的任务调度研究进展,并展望了进一步的研究方向,期望本文的介绍能够为同行学者提供一定参考.

参 考 文 献

- [1] Tilak Agerwala, Siddhartha Chatterjee. Computer architecture: Challenges and opportunities for the next decade[J]. IEEE Micro, 2005, 25(3): 58-69
- [2] Wayne Wolf. The future of multiprocessor systems-on-chips [C] //Proc of the 41st Annual Conf on Design Automation. New York: ACM Press, 2004: 681-684
- [3] Ahmed Jerraya, Hannu Tenbunen, Wayne Wolf. Multiprocessor systems-on-chips [J]. IEEE Computer, 2005, 38(7): 36-40
- [4] Tim Kogel, Heinrich Meyr. Heterogeneous MP-SoC-The solution to energy-efficient signal [C] //Proc of the 41st Annual Conf on Design Automation. New York: ACM Press, 2004: 686-691
- [5] Kunle Olukotun. Stanford hydra single-chip multiprocessor [EB/OL]. [2006-10-30]. <http://www-hydra.stanford.edu/>
- [6] Ruby B Lee. Princeton architecture laboratory for multimedia and security [EB/OL]. [2006-10-30]. <http://palms.ee.princeton.edu/index.html/>
- [7] Hill Mark, Wood David. Wisconsin multifacet project [EB/OL]. [2006-10-30]. <http://www.cs.wisc.edu/gems/>
- [8] Garey M R, Johnson D S. Computers and Intractability: A Guide to the Theory of NP-Completeness [M]. New York: Freeman W H, 1979
- [9] Grant Martin. Overviews of the MPSoC design challenge [C] //Proc of the 43rd Annual Conf on Design Automation. New York: ACM Press, 2006: 274-279
- [10] Hwang David D, Schaumont Patrick, Tiri Kris, et al. Securing embedded systems [J]. IEEE Security & Privacy, 2006, 4(2): 40-49
- [11] Liu Sha, Tarek Abdelzaher, Karl-Erik Arzen, et al. Real time scheduling theory: A historical perspective [J]. Real-Time Systems, 2004, 28(2): 101-155
- [12] Wang Qiang, Wang Hong'an, Jin Hong, et al. Overview of scheduling algorithms of aperiodic tasks in real-time systems [J]. Journal of Computer Research and Development, 2004, 41(3): 385-392 (in Chinese)
(王强, 王宏安, 金宏, 等. 实时系统中的非定期任务调度算法综述[J]. 计算机研究与发展, 2004, 41(3): 385-392)
- [13] Baruah S, Nathan F. The feasibility analysis of multiprocessor real-time systems [C] //Proc of the EuroMicro Conf on Real-Time Systems. Los Alamitos, CA: IEEE Computer Society Press, 2006: 85-94
- [14] Liu C, Layland J. Scheduling algorithms for multiprogramming in a hard real-time environment [J]. Journal of the ACM, 1973, 20(1): 4-61
- [15] Baruah S. Feasibility analysis of preemptive real-time systems upon heterogeneous multiprocessor platforms [C] //Proc of the 25th IEEE Int Real-Time Systems Symposium. Los Alamitos, CA: IEEE Computer Society Press, 2004: 37-46
- [16] Baruah S. Task partitioning upon heterogeneous multiprocessor platform [C] //Proc of the 10th IEEE Real-Time and Embedded Technology and Applications Symposium. Los Alamitos, CA: IEEE Computer Society Press, 2004: 536-543
- [17] Funk Shelby, Baruah S. Task assignment on uniform heterogeneous multiprocessors [C] //Proc of the Euromicro Conf on Real-Time Systems. Los Alamitos, CA: IEEE Computer Society Press, 2005: 219-226
- [18] Shen Zhuowei, Wang Yun. A schedulability analysis algorithm for EDF-based end-to-end real-time systems [J]. Journal of Computer Research and Development, 2006, 43(5): 813-820 (in Chinese)
(沈卓伟, 汪芸. 基于 EDF 调度策略的端到端实时系统可调度性分析算法[J]. 计算机研究与发展, 2006, 43(5): 813-820)
- [19] Baruah S, Fisher Nathan. The partitioned multiprocessor scheduling of deadline-constrained sporadic task systems [J]. IEEE Trans on Computers, 2006, 55(7): 918-923

- [20] Paolo Valente, Giuseppe Lipari. An upper bound to the lateness of soft real-time tasks scheduled by EDF on multiprocessors [C] //Proc of the 26th IEEE Int Real-Time Systems Symposium. Los Alamitos, CA: IEEE Computer Society Press, 2005: 311-320
- [21] Marko Bertogna, Michele Cirinei, Giuseppe Lipari. Improved schedulability analysis of EDF on multiprocessor platforms [C] //Proc of the 17th Euromicro Conf on Real-Time Systems. Los Alamitos, CA: IEEE Computer Society Press, 2005: 209-218
- [22] Li Jianguo, Chen Songqiao, Lu Zhihui. A dynamic scheduling algorithm based on group optimization in real-time heterogeneous systems [J]. Chinese Journal of Computers, 2006, 29(6): 976-984 (in Chinese)
(李建国, 陈松乔, 鲁志辉. 实时异构系统的动态分批优化调度算法[J]. 计算机学报, 2006, 29(6): 976-984)
- [23] Bin Xuelian, Yang Yuhai, Jin Shiyao. A new dynamic scheduling algorithm for real-time multiprocessor systems based on grouping and properly choosing policies [J]. Chinese Journal of Computers, 2006, 29(1): 81-91 (in Chinese)
(宾雪莲, 杨玉海, 金士尧. 一种基于分组与适当选择策略的实时多处理器系统的动态调度算法[J]. 计算机学报, 2006, 29(1): 81-91)
- [24] Rodolfo Pellizzoni, Giuseppe Lipari. Improved schedulability analysis of real-time transactions with earliest deadline [C] //Proc of the 11th IEEE Real Time and Embedded Technology and Applications Symposium. Los Alamitos, CA: IEEE Computer Society Press, 2005: 66-75
- [25] Gopalakrishnan S, Caccamo M. Task partitioning with replication upon heterogeneous multiprocessor systems [C] //Proc of the 12th IEEE Real-Time and Embedded Technology and Applications Symposium. Los Alamitos, CA: IEEE Computer Society Press, 2006: 197-207
- [26] Yang Chunhua, Gui Weihua, Ji Li. A fault-tolerant scheduling algorithm of hybrid real-time tasks based on multiprocessors [J]. Chinese Journal of Computers, 2003, 26(11): 1480-1486 (in Chinese)
(阳春华, 桂卫华, 计莉. 基于多处理机的混和实时任务容错调度[J]. 计算机学报, 2003, 26(11): 1480-1486)
- [27] Zhou Shuang'e, Yuan Youguang, Xiong Bingzhou, *et al.* An algorithm of processor pre-allocation based on task duplication [J]. Chinese Journal of Computers, 2004, 27(2): 216-223 (in Chinese)
(周双娥, 袁由光, 熊兵周, 等. 基于任务复制的处理器预分配算法[J]. 计算机学报, 2004, 27(2): 216-223)
- [28] Hsu Heng-Ruey, Chen Jian-Jia, Kuo Tei-Wei. Multiprocessor synthesis for periodic hard real-time tasks under a given energy constraint [C] //Proc of the Conf on Design, Automation and Test in Europe. Munich, Germany: European Design and Automation Association Press, 2006: 1061-1066
- [29] Krzysztof Rzdca, Franciszek Seredynski. Heterogeneous multiprocessor scheduling with differential evolution [C] //Proc of the 2005 IEEE Congress on Evolutionary Computation. Piscataway, NJ: IEEE Press, 2005, 3: 2840-2847
- [30] Kuang S -R, Chen C -Y, Liao Ren-Zheng. Partitioning and pipelined scheduling of embedded system using integer linear programming [C] //Proc of the 11th Int Conf on Parallel and Distributed Systems. Los Alamitos, CA: IEEE Computer Society Press, 2005: 37-41
- [31] Kianzad Vida, Bhattacharyya Shuvra S. Efficient techniques for clustering and scheduling onto embedded multiprocessors [J]. IEEE Trans on Parallel and Distributed Systems, 2006, 17(7): 667-680
- [32] Becchi M, Crowley P. Dynamic thread assignment on heterogeneous multiprocessor architectures [C] //Proc of the 3rd Conf on Computing Frontiers. New York: ACM Press, 2006: 29-40
- [33] Xie Yuan, Hung Wei-lun. Temperature-aware task allocation and scheduling for embedded multiprocessor systems-on-chip (MPSoC) design [J]. Journal of VLSI Signal Processing, 2006, 45(3): 177-189
- [34] Annie S Wu, Han Yu, Shiyuan Jin, *et al.* An incremental genetic algorithm approach to multiprocessor scheduling [J]. IEEE Trans on Parallel and Distributed Systems, 2004, 15(9): 824-834
- [35] Anna Swiecicka, Franciszek Seredynski, Albert Y Zomaya. Multiprocessor scheduling and rescheduling with use of cellular automata and artificial immune system support [J]. IEEE Trans on Parallel and Distributed Systems, 2006, 17(3): 253-262
- [36] Mok A K, Chen D. A multiframe model for real-time tasks [J]. IEEE Trans on Software Engineering, 1997, 23(10): 635-645
- [37] Baruah S, Chen D, Gorinsky S, *et al.* Generalized multifraem tasks [J]. Real-Time Systems, 1999, 17(1): 5-22
- [38] Baruah S. Dynamic- and static-priority scheduling of recurring real-time tasks [J]. Real-Time System, 2003, 24(1): 99-128
- [39] Baruah S, Chakraborty S. Schedulability analysis of non-preemptive recurring real-time tasks [C] //Proc of the 14th Int Workshop on Parallel and Distributed Real-Time Systems. Los Alamitos: IEEE Computer Society Press, 2006
- [40] Bordoloi U D, Chakraborty S. Interactive schedulability analysis [C] //Proc of the 12th IEEE Real-Time and Embedded Technology and Applications Symposium. Los Alamitos: IEEE Computer Society Press, 2006: 147-156
- [41] Chakraborty S, Thiele L. A new task model for streaming applications and its schedulability analysis [C] //Proc of the Conf on Design, Automation and Test in Europe. Los Alamitos: IEEE Computer Society Press, 2005: 486-491

- [42] Fábio Rodrigues de la Rocha, Rômulo Silva de Oliveira. A real-time task model based on ideal instant [C] //Proc of the IEEE 10th Conf on Emerging Technologies and Factory Automation. Piscataway, NJ: IEEE Press, 2005: 1085-1088
- [43] Michael A Palis. The granularity metric for fine-grain real-time scheduling [J]. IEEE Trans on Computers, 2005, 54(12): 1572-1583
- [44] Tao Xie, Xiao Qin. Scheduling security-critical real-time applications on clusters [J]. IEEE Trans on Computers, 2006, 55(7): 864-879
- [45] Kempf T, Doerper M, Kogel T, *et al.* A modular simulation framework for spatial and temporal task mapping onto multiprocessor SoC platform [C] //Proc of the Conf on Design, Automation and Test in Europe. Los Alamitos: IEEE Computer Society Press, 2005: 876-881
- [46] Erbas C, Cerav-Erbas S, Pimentel A D. Multiobjective optimization and evolutionary algorithms for the application mapping problem in multiprocessor system-on-chip design [J]. IEEE Trans on Evolutionary Computation, 2006, 10(3): 358-374
- [47] Minyoung Kim, Sudarshan Banerjee, Nikil Dutt, *et al.* Design space exploration of real-time multi-media MPSoCs with heterogeneous scheduling policies [C] //Proc of the 4th Int Conf on Hardware/Software Codesign and System Synthesis. New York: ACM Press, 2006: 16-21
- [48] Fabio Wronski, Eduardo Wenzel Briao, Flavio Rech Wagner. Evaluating energy-aware task allocation strategies for MPSOCs [C] //Proc of Int Federation for Information Processing. Berlin: Springer, 2006: 215-224
- [49] Peng Yang, Chung Wong, Marchal P, *et al.* Energy-aware runtime scheduling for embedded-multiprocessor SOC [J]. IEEE Design & Test of Computers, 2001, 18(5): 46-58
- [50] Richter K, Racu R, Ernst R. Scheduling analysis integration for heterogeneous multiprocessor SoC [C] //Proc of the 24th IEEE Real-Time Systems Symposium. Los Alamitos, CA: IEEE Computer Society Press, 2003: 236-245
- [51] Chen Jing, Liu Jian-Hong. Developing embedded kernel for system-on-a-chip platform of heterogeneous multiprocessor architecture [C] //Proc of the 12th IEEE Int Conf on Embedded and Real-Time Computing Systems and Applications. Los Alamitos, CA: IEEE Computer Society Press, 2006: 246-250
- [52] Ruggiero Martino, Guerri Alessio, Poletti Francesco, *et al.* Communication-aware allocation and scheduling framework for stream-oriented multi-processor systems-on-chip [C] //Proc of the Conf on Design, Automation and Test in Europe. Munich, Germany: European Design and Automation Association Press, 2006: 3-8
- [53] Cho Y, Yoo S, Choi Y, *et al.* Scheduler implementation in MP SoC design [C] //Proc of the 2005 Conf on Asia South Pacific Design Automation. New York: ACM Press, 2005: 151-156
- [54] Xue Liping, Ozturk Ozcan, Li Feihui, *et al.* Dynamic partitioning of processing and memory resources in embedded MPSoC architecture [C] //Proc of the Conf on Design, Automation and Test in Europe. Munich, Germany: European Design and Automation Association Press, 2006: 690-695
- [55] Ma Zha, Cathoor Francky, Vounckx Johan. Hierarchical task scheduler for interleaving subtasks on heterogeneous multiprocessor platform [C] //Proc of the 2005 Conf on Asia South Pacific Design Automation. New York: ACM Press, 2005: 952-955
- [56] Nollet V, Avasare P, Mignolet J -Y, *et al.* Low cost task migration initiation in a heterogeneous MPSoC [C] //Proc of the Conf on Design, Automation and Test in Europe. Los Alamitos, CA: IEEE Computer Society Press, 2005: 252-253
- [57] Bertozzi S, Acquaviva A, Bertozzi D, *et al.* Supporting task migration in multi-processor system-on-chip: A feasibility study [C] //Proc of the Conf on Design, Automation and Test in Europe. Munich, Germany: European Design and Automation Association Press, 2006: 15-20
- [58] Ozturk O, Kandemir M, Son S W, *et al.* Selective code/data migration for reducing communication energy in embedded MpSoC architectures [C] //Proc of the 16th ACM Great Lakes Symposium. New York: ACM Press, 2006: 386-391



Li Renfa, born in 1956. Ph. D., Professor and Ph. D. supervisor of the School of Computer and Communication of Hunan University. His main research interests include embedded computing, wireless network, network & digital medium.

李仁发, 1956年生, 博士, 教授, 博士生导师, 主要研究方向为嵌入式计算、无线网络、网络与数字媒体。



Liu Yan, born in 1979. Ph. D. candidate of computer application technology since 2005. His main research interests include computer architecture and embedded system.

刘彦, 1979年生, 博士研究生, 主要研究方向为体系结构、嵌入式系统(qmttz@163.com)。



Xu Cheng, born in 1962. Ph. D., Professor and Ph. D. supervisor of the School of Computer and Communication of Hunan University. His main research interests include embedded system and automated mechanical design and manufacturing.

徐成, 1962年生, 博士, 教授, 博士生导师, 主要研究方向为嵌入式系统、机械制造及其自动化(cheng_xu@yeah.net)。

Research Background

Multiprocessor system-on-chip is often a heterogeneous multiprocessors and integrates multiple instruction-set processors on a single chip that implements most of the functionality of a complex electronic system. Scheduling and mapping of tasks are important key problems in multiprocessor system-on-chip design, and are substantially more difficult than scheduling a uniprocessor. This paper introduces the basic architecture and design challenge of a multiprocessor system-on-chip and the multiprocessor system-on-chip task scheduling algorithm. The scheduling algorithm analysis is analytically classified into three categories and the scheduler implementation framework is classified into two categories by task modeling. Many open research problems are pointed out. By comparing and analyzing these different projects and algorithms, researchers of related topics can gain useful information about task scheduling problem.

(上接第 1553 页)

4) 管理信息系统,决策支持系统,地理信息系统,信息安全技术,数据库技术,搜索引擎技术,图形图像处理,商务智能技术,遥测遥控遥感,环境科学信息技术,管理科学信息技术,计算机应用技术。

5) 关于信息技术与应用的教学思想、教学理论、教学方法、教学研究、教学改革、教学创新、教学探索、教学经验等。

论文出版

1) 录用的优质英文论文,将发表在有 IEEE Computer Society 出版的《IFITA 2009 论文集》,并被 ISTP、EI 检索;会后,特优论文还将被推荐到国际 SCI 源刊“Journal of Algorithms and Computational Technology”、EI 源刊“Journal of Computer”正刊优先发表。

2) 录用的其余优质中/英文论文,将发表在中国国家级核心期刊《计算机科学》(专刊);会后,特优论文还将被推荐到《计算机科学》正刊优先发表。

3) 录用并注册的论文发表后,将挂号赠寄该文所在的《IFITA 2009 论文集》或《计算机科学》(专刊)1 册。

论文投稿

1) 所有来稿须为 Word 电子文档,均应主题鲜明、资料可靠、数据准确、书写规范、图表清晰。投稿前,作者顺序须确定;投稿时,须提交投稿联系人信息表。投稿方式,可为电子邮箱(ifita2009@163.com 或 xxjsyy@swufe.edu.cn)并注明“IFITA2009 投稿”,或论坛网站(<http://www.ifita.cn>)。

2) 英文来稿:必须严格按英文论文模板要求进行编排(版面一般不超过其模板规定板式页面 5 页),且必须有:单栏排版的题目、作者姓名、作者单位、作者地区、作者邮编、作者国家或地区;双栏排版的摘要、关键词(以上可另附中文版),正文、参考文献(至少 8 项)。

3) 中文来稿:必须严格按中文论文模板要求进行编排(版面一般不超过其模板 5 号字 A4 页面 5 页),且必须有:单栏排版的题目、作者姓名、作者单位、作者地区、作者邮编、作者国家或地区,摘要、关键词(以上均需有中、英文版);双栏排版的正文、参考文献(至少 4 项)。

4) 获重要课题或项目资助的来稿:英文来稿应于文末致谢段落,注明其所受资助的课题名或项目名;中文来稿应于文题页左下方,注明其所受资助的课题名或项目名。

5) 文稿严禁抄袭,文责自负。所有来稿,必电子邮件回复:

① 来稿收到后,均在收稿之日起 3 个工作日内,回复作者“来稿收妥”;

② 来稿无论被录用与否,都在收稿之日起 20 个工作日内,通知作者“审稿结果”。

论文注册

1) 《IFITA 2009 论文集》(IEEE)注册费:每篇英文论文 4 页内为人民币 2400 元(或 450 美元);超过 4 页者,每超一页另收人民币 500 元(或 80 美元)(注:IEEE 会员、学生,凭有效证件,9 折优惠)。

2) 《计算机科学》(专刊)发表费:每篇中文/英文论文 4 页内为人民币 800 元(或 150 美元);超过 4 页者,每超一页另收人民币 200 元(或 40 美元)(注:出席 IFITA 2009 会议者,需另交会务费人民币 700 元或 140 美元)。

3) “Journal of Algorithms and Computational Technology”、“Journal of Computer”、《计算机科学》正刊发表费:按照各正刊论文发表费标准执行。

关键日期

投稿截止日期:2009 年 1 月 10 日;

录用截止日期:2009 年 1 月 15 日;

注册截止日期:2009 年 1 月 20 日。

论坛咨询

1) 咨询邮箱: ifita2009@163.com 或 xxjsyy@swufe.edu.cn

2) 电话咨询:中国内: 028-87353517, 028-88088880, 0-13060035282, 0-13880890751

中国外: 86-28-87353517, 86-28-88088880, 86-28-0-13060035282, 86-28-0-13880890751