WORKLOAD-AWARE ADAPTIVE SCHEDULING ALGORITHMS FOR HETEROGENEOUS MULTI-CORE SYSTEMS

Upasakage Saneth Maduranga

229351T

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2024

WORKLOAD-AWARE ADAPTIVE SCHEDULING ALGORITHMS FOR HETEROGENEOUS MULTI-CORE SYSTEMS

Upasakage Saneth Maduranga

229351T

Thesis submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2024

DECLARATION OF THE CANDIDATE & SUPERVISOR

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: U.S.Maduranga Date: 30/06/2024

The above candidate has carried out research for the Masters thesis under my supervision.

Name of the supervisor: Dr.Chathuranga Hettiarachchi

Signature of the supervisor: Date : 5th July 2024

ABSTRACT

WORKLOAD-AWARE ADAPTIVE SCHEDULING ALGORITHMS FOR HETEROGENEOUS MULTI-CORE SYSTEMS

Microprocessors have undergone a revolutionary transformation since their inception, with transistor count exploding from 2,300 in the Intel 4004 to a 90 billion in the Apple M3 Ultra. This surge in processing power coexisted with challenges like performance limitations and power constraints. Multi-core processors emerged as a solution, but the "dark silicon era" highlighted the inability to fully utilize all available cores due to thermal limitations. Heterogeneous Multi-Core Processors (AMPs) addressed this by incorporating cores with diverse capabilities, maximizing power efficiency. AMPs are now all-over in the latest smartphones, tablets, and laptops. Still, the software and scheduler support remain questionable on the AMPs. Existing schedulers, designed for homogeneous systems, struggle to manage the varied capabilities of AMP cores. This research tackles this challenge by proposing a Workload-Aware Adaptive Scheduling (WAAS) algorithm. This novel approach leverages machine learning to create a workload-aware scheduler specifically designed for heterogeneous multi-core systems, aiming to unlock the full potential of AMPs and bridge the gap between processor capabilities and scheduling efficiency.

Keywords: AMP, Heterogeneous Multi Cores, Workload Aware Scheduler

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Description
AMP	Asymmetric Multi-core Processors
GHz	GigaHertz
KHz	KiloHertz
ILP	Instruction Level Parallelism
IPC	Instructions Per Cycle
TDP	Thermal Design Power
IPC	Instructions Per Clock
DRAM	Dynamic Random Access Memory
TLP	Thread Level Parallelism
L2	Level 2
SSE	Streaming SIMD Extensions
ITRS	International Technology Roadmap for Semiconductors
DVFS	Dynamic Voltage-Frequency Scaling
CMOS	Complementary Metal-Oxide-Semiconductor
I/O	Input/Output
WPU	Wearable Processing Unit
GPU	Graphics Processing Unit
FPGA	Field-Programmable Gate Array
API	Application Programming Interface
PMC	Performance Monitoring Counter
TD	Thread Director
WAAS	Workload-Aware Adaptive Scheduling
CFS	Completely Fair Scheduler
RTMS	Random Task Migration Scheduler
CPU	Central Processing Unit
MIPS	Million Instructions Per Second
EAS	Energy-Aware Scheduling
CAS	Capacity-Aware Scheduling
SMT	Simultaneous Multi Threading

GTS	Global Task Scheduler
PELT	Per-Entity Load Tracking
TLB	Translation Lookaside Buffer
ML	Machine Learning
GB	Giga Bytes
JWT	Json Web Token
ID token	Identifier Token
DB	DataBase
L1	Level 1
LLC	Least Level Cache
L1D	Level 1 Data
L1I	Level 1 Instructions
MLC	Multi-Level Cell

LIST OF APPENDICES