Smart, Adaptive Energy-efficient Data Centre Computing

Kaiqiang Zhang (zhangkq0413@gmail.com)
Prospective supervisor: Dr. Zheng Wang

Abstract

The landscape of computing is changing. Much of today's computation cycles are provided by data centres which suffer embarrassing energy inefficiencies: only a small fraction of a typical data centre's power is spent on useful work. The inefficiencies lead to worldwide energy waste measured in billions of pounds and tens of millions of metric tons of CO2. This has to change.

This proposal **aims to improve the energy efficiency of data centres through innovative compiler and runtime optimisation**. The underlying issue of data centres is that the computing systems are conservatively provisioned for rare utilisation peaks, leading to energy waste in underutilised systems and over-provisioning of hardware. What is needed is way to allow system software to precisely schedule hardware resources to reduce energy waste and dynamically tailor application tasks to achieve the best possible performance accordingly. This project offers precisely that. As a departure from prior work, this research investigates the use of predictive modelling techniques to predict live server traffic, enabling us to build new scheduling policies for precise resource scheduling. It will also create a novel, energy-aware compilation system to dynamically tailor applications on the fly through e.g. recompilation, and replacing algorithms or data structures as the program runs. One of the key innovations I wish to explore is **putting portability and adaptation at the core of this approach** by using machine learning to automate design space exploration and allowing self-adaptation of system software over time.

As knowledge of the underlying platform, workloads and applications grows, the system software will enhance the performance and eliminate the energy waste of data centres. The operating cost will be cut down, programs will run faster, and data centres will draw less power, having less environmental impact. These are ambitious goals which will tackle an enormous number of difficulties that must be addressed. There is enormous scope for innovation and novel research in this area, which will open up new possibilities and challenges as never before.

1 Introduction

1.1 Background

Data centres provide the IT infrastructure that powers many of today's computing environments from the Internet to the mobile cloud and are currently a $50 billion market with an exponential growth rate [**?**]. Energy consumption is the fundamental issue for data centres [**?**, **?**]. Electricity costs typically make up 70% of operating expense of high end servers []. In the UK data centres account for 3% of total domestic electricity use, or equivalent to one nuclear power station [**?**]. The total carbon footprint of the world's data centres is approximately the same as the carbon emissions of the entire Czech Republic and will triple by 2020 [**?**]. Given the increasing demand of computing power of data centres and the rising cost of energy, there is a critical need to optimise data centre applications for both raw performance and energy consumption.

Hardware manufacturers have responded to this issue by making enormous strides in managing power demands of processors []. Sophisticated power management features, such as dynamic voltage and frequency scaling and heterogeneous processing units, are now placed across system components, hoping that software techniques will be found to make use of them. Software developers, however, are struggling to cope with this dramatic increase in complexity, the existing tools and languages are simply inadequate to the task. As a result, todays data centres suffer embarrassing energy inefficiencies: it is not unusual for less than 20% of the power is used to perform useful work for data centres []. This situation has to change.

1.2 Research Challenges

However, improving energy efficiency for data centres is challenging. Firstly, the hardware resources are often conservatively provisioned (e.g. machines operate on the peak power mode as unnecessarily) for rare utilisation peaks because of being unable to accurately predict the load peaks [**?**, **?**]. This leads to energy waste in underutilised systems. Secondly, server applications are currently hard-coded and optimised for a fully utilised system, but energy efficiency generally degrades outside that range [**?**, **?**]. As such, techniques that can dynamically tailor programs according to workload and resource changes are essential for energy efficiency [**?**]. Unfortunately, servers are a truly multi-tasking environment (where thousands of tasks running at the same time, competing for shared resources), often with the support of virtualisation technology [**?**], creating enormous difficulties for software power optimisation. This situation becomes more difficult as more and more heterogeneous hardware are used now days [**?**]. What is needed is a technique that **evolves and adapts to the changing workload and resource and delivers scalable energy-efficient performance**.

1.3 Proposed Research

This research will use predictive modelling techniques to model and predict data centres workloads. It will design a new Linux scheduler to make use of the accurate workload prediction to quickly react to the change of workload demands and precisely allocate server resources. Furthermore, to harness the potential of dynamic resource scheduling, applications must be tailored to fit the drastically changing computing environment, as a poorly optimised program can overshadow any energy benefits that would have been obtained. To this end, this research will investigate program tuning techniques that enable us to tailor applications on-the-fly (e.g. dynamically replacing the algorithm or data structure as the program runs). On top of these, we would like the system software to catch up with the evolving hardware, applications and workloads and to deliver "future scalable" performance. To achieve this, I will explore an innovative **machine-learning-based** approach that self-adapts to future hardware and applications, and improves its performance over time.

1.4 Expected Outcome and Impact

This project will create a novel predictive modelling based task and resource scheduler in the Linux operating system. It will also develop energy-aware compilation heuristics and optimisation algorithms in the Jike JVM [], targeting heterogeneous many-core systems. When this project is successful, the tool-chain it creates will show how to harness the potential of heterogeneous many-cores (for example, a CPU-GPU based system). The techniques will be demonstrated on real world applications in the domains of data-intensive applications such as Hadoop based database system. We will show that our automatic optimisations are able to increase performance and energy efficiency for those applications by **50%** compared to unoptimised code, and are not more than 5% away from hand optimisation. We will prove the **portability** of our system by showing that our applications remain optimised for different architectures, **without the developer needing to alter a single line of code**.

1.5 Novelty

Machine learning has recently been proved to be effective at learning how to optimise programs in a static environment [**?**, **?**]. The way is now open, for the first time, for those techniques t**o be tried in a large-scale, ever-changing server environment; continuously improve over time**. This project cuts across different layers of the software stack, bringing together workload modelling, runtime program optimisation and resource scheduling to address a critical problem in a way that has never been attempted. There are great challenges ahead and enormous scope for interesting, high impact research to be carried out in this new area.

2 Research Programme and Methodology

The main goal of this PhD research is to develop the software, which can reduce the energy consumption while making programs faster, using machine learning algorithms. Considering such motivations, there are two main tasks to solve this problem: a) develop a precise workload prediction approach; b) intelligent tasks scheduler for optimising the computational resources usage. Due to the complexity of reality and high-speed hardware development, the key point here is ensuring those two functions to be adaptive to changeable hardware and complex situations. Moreover, the investigations around different modelling methods and scheduling approaches, which will help relative multi-core based system study, will be conducted as an important part. The detailed objectives are listed as following:

- Modelling the data centres systems. Efficient models help to save time for analysis and future computation. As introduced data centres are very complicated systems, it is valuable to develop an efficient model to express the workload, resource usage, and tasks runtime. Such a modelling method has to be available for different systems. In order to achieve an improving performance, the scheduling logs and power measurements need to be considered as key features in the model.

- Workload Prediction approach development. In order to achieve high computing recourse utilisation, an accurate estimation of workload has to be developed as the foundation of optimal tasks scheduling. The predicting accuracy needs to increase overtime with more system information collected. This is expected to achieve through machine learning algorithms, which can provide reliable probabilistic predictions with a certain error tolerance.

- Adaptive optimal tasks scheduler for data centres. The proposed tasks scheduling software is required to manage the optimal physical recourse setting and provide program scheduling, including power modes of hardware, server time allocation, tasks distribution and migration. The employed machine learning method should adapt to new circumstances and continuously improve itself in the end-user's environment. The proposed adaptive optimal task scheduler can optimise the recourse use and improve the performance overtime, which will open up new possibilities and challenges.

- Investigations around different possible scheduling methods. With an appropriate model, task schedule schemes may influence the management efficiency in terms of scheduling time consumption and tasks managing performance. The investigations here are trying to find a scheduling method, which can provide considerable optimisation result for energy-efficiency data centres with high speed. The evaluation will be conducted around different machine learning algorithms (including supervised learning and reinforcement learning) and operation management methods (such as queuing theory, or game theory e.g.[**?**]). The evaluation result will help the future research to choose appropriate management method on efficient tasks scheduling on not only data centres but also on other multi-core based systems.

In order to achieve above tasks, there are several significant developments on scheduler design and brilliant methodologies recently. For instance, the state-of-the-art work, which models parallel tasks into a series of sequential or parallel segments, shows the possibilities of real-time scheduling in multi-core processor systems[**?**][**?**]. Applying such scheduling method, huge programs are split into small tasks. In this way, the optimal runtime management is achievable by optimising each short computing time segment. The development on performance feature-awaring system realised energy-efficiency in specific multi-core processors, e.g. memory aware system[**?**] and energy aware system[**?**]. As a consequence, the observability of system features show the possibilities to apply feedback methodology for optimal runtime scheduling. The energy consumption or memory can be the criteria data to train the machine learning algorithms. Such system features can be the cost function, which the optimisation programs learn to employ the minimising cost optimisation scheme using machine learning algorithm from. Similar methodology has been applied on smart electric grid management[**?**] and distributed sensors optimisation problems[**?**]. Thus, the feedback methodology is expected to implement optimal runtime scheduling with machine learning algorithms, like reinforcement learning.

3 Related Work

Energy-efficiency for Data Centres FIXME!Describe the energy issues of data centres and existing hw/software techniques

Performance-Energy Tradeoff: Saving energy without compromising performance is increasingly difficult. Even in single task, single core systems, energy and performance do not always correlate strongly []. One often cannot optimise code to maximise performance *and* to minimise energy use [**?**]. Ideally, program optimisation would respond to changes in the environment, so that a server application, for example, would use different program versions depending on the program input, availabilities of hardware resources and other constraints like power cap and responsive time.

Heterogeneous Many-cores: Heterogeneous may-cores, such as mixed CPU/GPU systems, provide potential for reducing energy consumption. Some devices can run some tasks more efficiently [?]. Currently, the task to core mapping is hard-corded when applications are built or is statically determined when the program launches [?]. Dynamically adjusting the task to core mapping can improve energy efficiency [?]. However, we must over come significant challenges to realise this benefit. Tasks must be optimised for and scheduled onto different cores and core frequencies must be set. Cores in a heterogeneous system have their own energy and performance characteristics that will change on a per task basis. Schedules need to be reconstructed as tasks enter, exit, or change phase. These challenges require the OS to understand how the tasks will respond to different cores and the compilation system to tailor the program dynamically in order to fit the change of task to core mapping well. We will explore compiler and run-time techniques that dynamically adjust the task to core mapping and resource allocation and processor frequency in FIXME!Theme 2.

Multi-tasking: Multi-tasking brings a host of problems that we must tackle. Tasks will have different resource requirements and will cause interference with other tasks [?]. Prior work on compilers and task mapping mainly targets single applications on unloaded machines without external workloads [?, ?] or dedicate servers running one single type of applications [?]. Energy and performance tradeoffs on one task must not affect schedule fairness, throughput, responsiveness, etc. for other tasks [?]. Moreover, tasks will have different priorities so they will have different goals for the energy/performance tradeoff. In Theme 2, we will build an energy efficient multi-task scheduler. The scheduler will use the precise workload prediction information from FIXME!Theme 1 to predict the effect of different scheduling decisions and to derive an optimised scheduling plan.

Evolving Computing Environment: New architectures, application domains and workloads constantly emerge, making it hard to keep optimisation strategies current [?]. Any optimisation system must support fast, easy re-tuning. Predictive modelling has emerged as a viable means to automate the construction of optimisation strategies for homogeneous multi-cores and single applications, outperforming human experts, and across architectures [?, ?]. The predictive modelling techniques of FIXME!Themes 1 and 2 will support models that can be retrained whenever the environment changes.

4 Brief Research Plan

The mile stones and matching time consumption is planed as:

- Modelling the multiple tasks problem for data centres considering energy consumption and physics recourse management. (6 months)
- To develop an adaptive accurate perdition of work load in data centres based on machine learning algorithm. The work needs to be conducted with sufficient data and the prediction method is expected to be tested in simulation and real circumstances (The cooperation between companies or data centres is needed). (6 months)
- Study around the machine learning algorithms [?], including reinforcement learning and supervised learning, to realise the way to adaptive optimal tasks scheduling approach. (6 months)

- Develop the algorithms and exam the performance in simulation based on real time data. (6 months or less)
- Investigate around different scheduling approaches, including game theory or queuing theory. Evaluation is needed for tasks scheduling in data centres real applications.(6 months or more)
- Apply the proposed scheduling approach on real systems. There is always a gap between theoretical design and A cooperation between data centres or product companies is highly expected. (6 months)
- Conduct all finished work, push forward for possible improvements. Prepare and write up the thesis. (6 months)

References