1 Introduction

With the heightening demand for network services, the needed computation power has increased. In traditional systems, an increase in computing power corresponds to an increase in number of machines used. Each extra machine added to a system causes the system to draw more electricity to run. This increase in energy demand is most often undesirable for two main reasons. First, an increased consumption rate of energy causes an increased energy bill. Secondly, the negative environmental impact of additional energy consumption is a growing concern with the recent talk of global warming. For these reasons, among others, both the research community and industry have been exploring the ways that more can be done with less. That is, finding methods of getting more computing power while using less electrical power.

Since the early days of computing systems, there has been the push to build larger and larger systems with more and more computing power. This push has led to the increased use and improvement of technologies that allow clusters of computing machinery to be created. These clusters can give performance that approaches the sum of each of the machines individually. Additionally, these clusters have been developed using both homogeneous and heterogeneous computing hardware, adding greater flexibility. These flexible parts can be arranged into clusters of many different shapes and sizes with an extraordinary amount of configuration options. Typically, without special configuration, cluster machines add energy costs proportional to their performance.

With the rapid growth of computing clusters, companies and researchers quickly realized that their static configurations were inflexible and had high energy usage. Both of these factors drove up the cost associated with running the cluster. In order to solve these problems, researchers begun to come up with and investigate ways of dynamically reconfiguring computing clusters. These methods employed both software and hardware techniques that gave administrators numerous ways that they could optimize their system. One such technique, called Preboot Execution Environment allows a machine to receive its boot image dynamically at the time the machine is turned on. This and other techniques used in dynamic configuration of clusters are discussed in the Background section of this paper.
Using dynamic configuration techniques, administrators get the ability to optimize their cluster system over many different metrics. These metrics include performance, power consumption, and resource utilization. It is well known that optimization comes with a price. That is, in order to get the benefit of the optimized metric, you often have to give up some amount of another metric. For the purposes of this paper, I focus on the research that has resulted in dynamically configuring clusters in order to save power. This research will be used to find ways to optimize for power consumption without giving up much performance.

The drive for reduced energy consumption has led to the development of low power processors. While these chips use significantly less electricity than traditional CPUs, their performance does suffer. However, market competition between vendors of low power chips is pushing the industry to increase the performance of the chips. Specifically, optimizations and features such as on-chip AES encryption, out of order instruction execution, and on-chip random number generation are expected to appear in the next generation of low power processors.

One of the market leading chips is the Intel Atom processor. This chip has gained notoriety as the processor powering many of the netbook computers currently available on the consumer market. These chips have also garnered the attention of the research community. A team of computer scientists at the Parallel Data Laboratory at Carnegie Mellon University used Atom and other low power chips to develop their low-power computing cluster which they named FAWN (Fast Array of Whimpy Nodes). This initial work at using low power processors in a cluster gave some inspiration for this thesis. The focus of this work was increasing the number of queries that could be answered per unit power by using a statically configured cluster.

Bridging all of these technological advances and research achievements together, a natural question develops. Namely, how can we take the lessons and techniques from traditional cluster systems, including dynamic reconfiguration, and apply it to clusters of low power processors? To address the issue, this thesis goes through an in-depth look at how traditional dynamic cluster reconfiguration techniques can be used and improved for clusters of low power nodes. Additionally, this thesis proposes and evaluates techniques for creating and dynamically reconfiguring clusters of intermixed low power and traditional nodes. This approach allows a system to maintain low power usage during most of the server’s operation while getting the best performance during (usually relatively brief) periods of high demand.

**TODO:** Discuss briefly what sort of things came out of the work and reinforce their importance as research contributions.

The rest of this paper is organized as follows. In chapter 2, I elaborate on this introduction by providing the background information needed to understand this thesis in the context of previous work. In the third chapter, I present my thesis idea. Next, I discuss the experiments that I performed on my ideas in chapter 4. Following that, in chapter 5, I give the results of the experiments that I performed as well as analyze the meaning and causes of the results. In
chapter 6, I summarize the contributions of this paper with concluding remarks. Finally, in chapter 7, I propose areas in which this research could be extended.