Quality of Service-Aware, Scalable Cache Tuning Algorithm in Consumer-based Embedded Devices
M. Hammam Alsafrajani and Ann Gordon-Ross
Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

INTRODUCTION

Background & Motivation
- Consumer-based Embedded Devices (CEDs) are ubiquitous
- High consumer-defined quality of service (QoS) expectations (consumer goals) and stringent energy constraints, often contending

Goal: Innovate a CEDs design approach that enables CEDs to adhere to consumer goals and energy requirements, independently of applicationsdeployment

Challenges
- Consumer goals must be known during design time
- Impractical with unknown third party applications
- Impractical, given rapid growth of unknown third party applications (~1.5 million apps for Android)

Methodology

Using configurable hardware and associated tuning algorithms
- Configurable hardware contains modifiable tunable parameters
- Hardware tuning may be used to change application's hardware requirements
- Parameter values dictated by a tuning algorithm
- Tuning algorithm monitors application execution, evaluates energy consumption and quality of service
- Adjustable to suit the needs of the applications

Contribution

Application-scalable hardware and runtime tuning algorithm
- Scalable hardware
  - Compressed tuning information into auxiliary tables
  - Employed LRU policy to enable scalability of application tuning
  - Only 4% area (hardware) overhead
- Dynamic, general purpose CED cache-tuning algorithm
  - Requires no a priori knowledge of applications
  - For disparate CEDs

Hardware Specifications
- Configurable, private level 1 cache memory
  - Possible configurations: 2KB 1-way; 4KB 1- or 2-way; and 8KB 1-, 2-, or 4-way

Tuning Algorithm
- Adjusts parameter values such that CEDs can be completely avoided
- Only consumes the lowest energy

Evaluation Methodology

System Platform
- Base system:
  - Processor: Quad-core system
  - Clock, voltage, memory, etc.
  - Prior work degraded 7X more, compared to our approach
  - Can be completely avoided only with a priori knowledge of applications

Future expansion

Interdisciplinary collaboration
- Psychology behind consumer expectations and its impact on design constraints
- Quantification of consumer feelings, moods, and thoughts
- Other factors such as time of use (day, night, etc), place of operation (office, construction site, on-route, etc.), and more

Architecture innovations
- Performance & energy expectations vs. privacy & security
- Adaptability to Internet of Things (IoT) devices
- Tuning through IoT

References

Conclusion

Designing approach for CED's
- Saving energy does not require a priori knowledge of applications
  - Enables scalability to an arbitrary number of unknown (future), third-party applications
  - Meets consumer goals
- Tuning modes provide flexible adherence to QoS expectations
- Can incorporate tuning modes for disparate conditions/situations, such as

Comparison to prior work

Prior work [1][2] saved more energy, however
- Incurred as much as 109.2% and 132.5% more tuning overhead while tuning
- Requires long applications execution to characterize the tuning overhead

Energy savings and QoS degradation
- Energy savings: Measured application's energy consumption of best configuration determined by the tuning algorithm, calculated energy percent decrease with respect to the base system, and averaged the energy savings for all applications (34 total)
- Quality of Service: Calculated the number of QoS degradation occurrences while tuning each application, compared the result to a perfect (no QoS degradation) system, and averaged the QoS degradation for all applications (34 total)

Comparison to prior work: Incorporated a tuning mode which represents prior work approach; an aggressive mode which determined the lowest-energy configuration without considering QoS degradation.