Online Learning of Timeout Policies (OLTP) for Dynamic Power Management

Dr. Umair Ali Khan

Outlook

- About me
- Research Objectives
- Contribution & Achievements
- MobiTrick (Mobile Traffic Checker)
- Online Learning of Timeout Policies (OLTP)
- Online Adaptation of Power/Performance (OAPP)
- Implementation Platform
- Evaluation
- Conclusion
- Future Directives

About Me

- PhD (Energy efficiency in portable embedded devices) - 2013
  - Pervasive computing group, Alpen-Adria Universität Klagenfurt, Austria
- MS (Intelligent Transportation Systems) - 2010
  - Institute of Smart Systems Technologies, Alpen-Adria Universität Klagenfurt, Austria
- B.E (Computer Systems Engineering) - 2004
  - QUEST, NawabShah
- Research Interests
  - Energy-efficient embedded platforms
  - Machine learning
  - Image processing

Research Objectives

- A portable traffic surveillance platform
- Generic DPM[1] framework
- Model independency
- Addressing non-stationarity
- Power/performance trading
- Constraints adaptation
- Implementation on MobiTrick[2]

[1]Dynamic Power Management

[2]Mobile Traffic Checker

Contribution & Achievements

- A portable, heterogeneous embedded traffic surveillance platform
- Online, RL[3] based DPM
- Learning, optimization & control
- Pareto-optimal tradeoff
- Constraints adaptation
- Scalability
- Implementation and evaluation

[3] Reinforcement Learning

MobiTrick (1/4)

- Salient Features
  - Portable, easy deployment
  - Multiple, heterogeneous image sensors
  - High-level stereo image processing
  - 3D reconstruction
  - Vehicel detection & classification
  - Over-height estimation
  - License plate detection
  - Toll-collection
- Project Partners
  - NES, KLU, Austria (http://www.uni-klu.ac.at/tewi/ict/nes/)
  - ICG, TU-Graz, Austria (http://www.icg.tu-graz.ac.at/)
  - EFKON, Austria (http://www.efkon.com/)
- Funded by Austrian Research Foundation (http://www.ffg.at)

MobiTrick's sensing platform

MobiTrick (2/4)

- Heterogeneous architecture
  - Exploiting the redundancy
  - Getting multiple views of a scene
  - Covering wider range of lighting conditions
  - Eliminating the need of larger sensors

MobiTrick's sensing platform

MobiTrick (3/4)

- Work Packages
  - Development of the sensor head
  - Envisaging power-efficient algorithms

- Requirements
  - Portability, compactness
  - Low-power design
  - Local processing
  - Autonomous operation
  - Auto-calibration & adaptation
  - CUDA based image processing
  - Online power management

MobiTrick's sensing platform

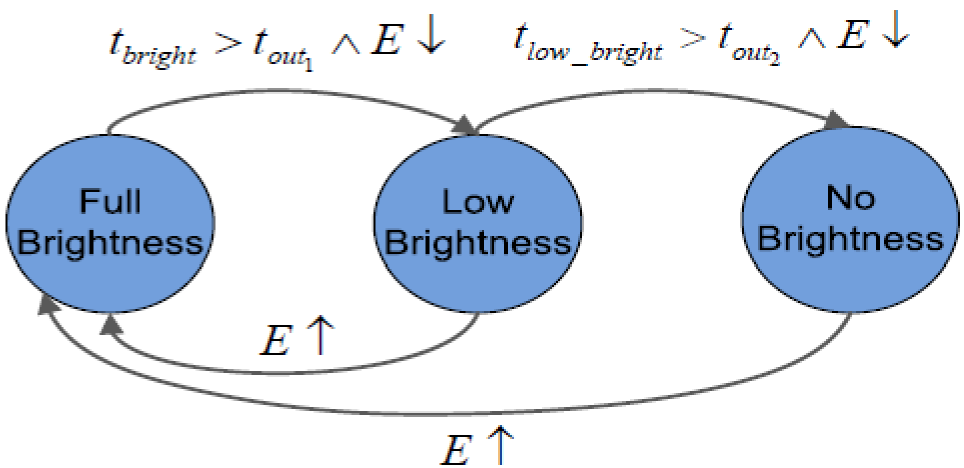
MobiTrick (4/4)

- Challenges faced
  - Communication among heterogeneous sensors
  - Low-power design with CUDA
  - Low-power design vs. Performance
  - Synchronized operation of heterogeneous sensors
  - Identifying the right parameters for stereo vision
    - Focal lengths
    - resolutions
    - FOV
    - sensor types
    - shutters types
    - frame rates
    - Interfaces
    - Mount type
    - etc, ...

MobiTrick's sensing platform

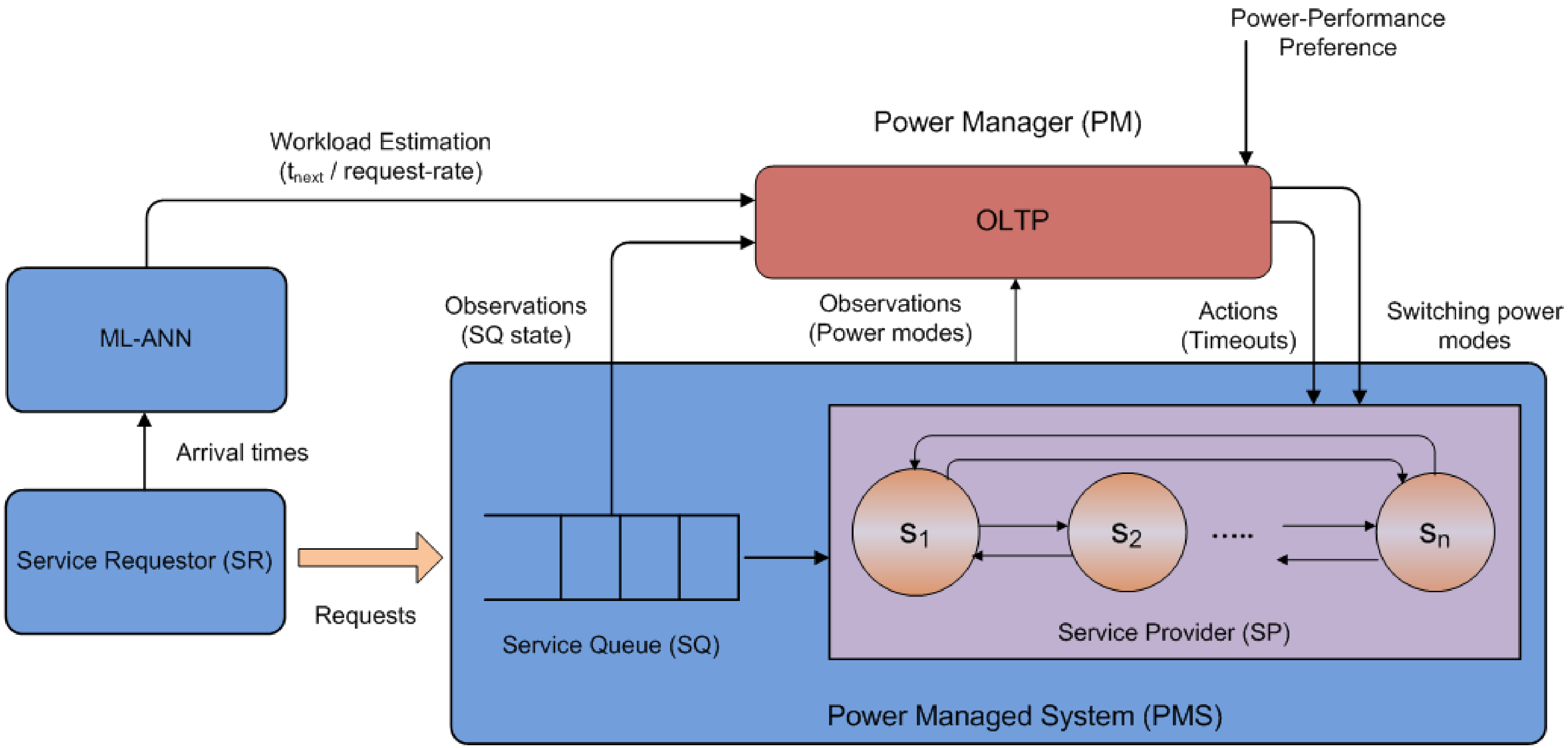
Online Learning of Timeout Policies (OLTP) - 1

- Dynamic timeout values in each state
  - Changing timeout decisions
  - Estimating the workload
  - Adaptation to the workload
- Multi-objective optimization
  - Reducing latency, reducing power consumption



State-transition diagram of screen power management

Online Learning of Timeout Policies (OLTP) - 1

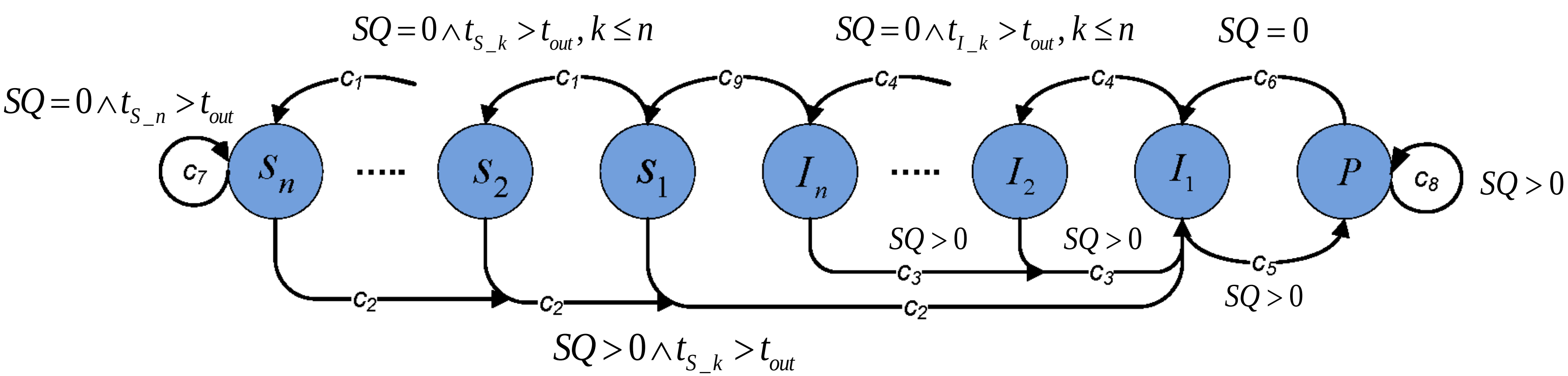


Depiction of the system under power management with OLTP

Online Learning of Timeout Policies (OLTP) - 2

Dynamics of the OLTP

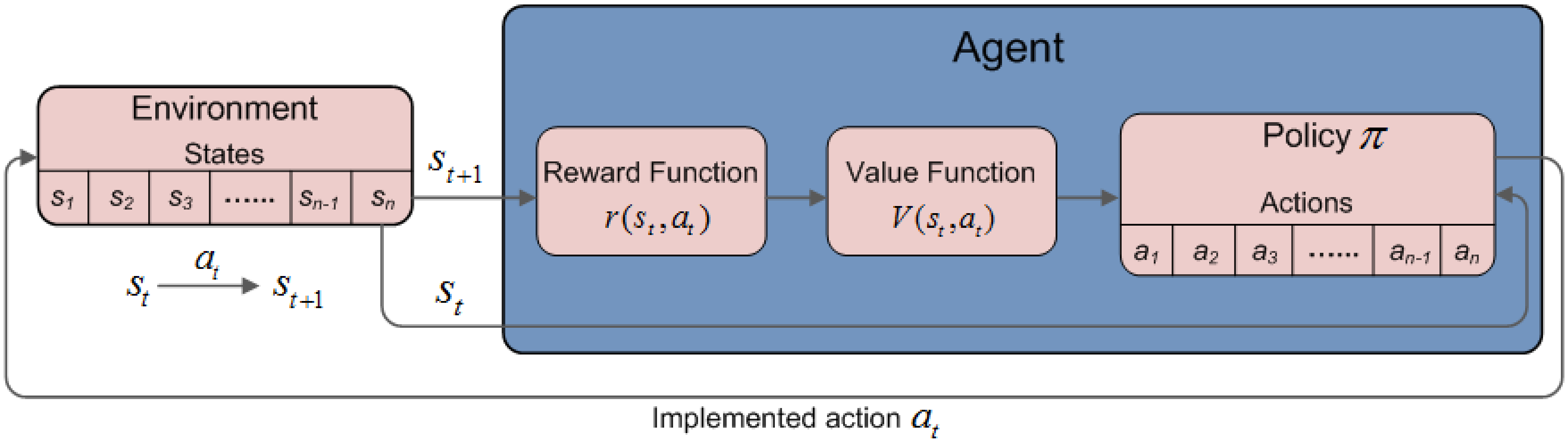
- Learning timeouts based on:
  - Power-performance preference
  - Workload estimation



State-transition mechanism of the OLTP for a generic PMS

Online Learning of Timeout Policies (OLTP) - 3

Generic RL Framework



Online Learning of Timeout Policies (OLTP) - 4

RL Framework for OLTP

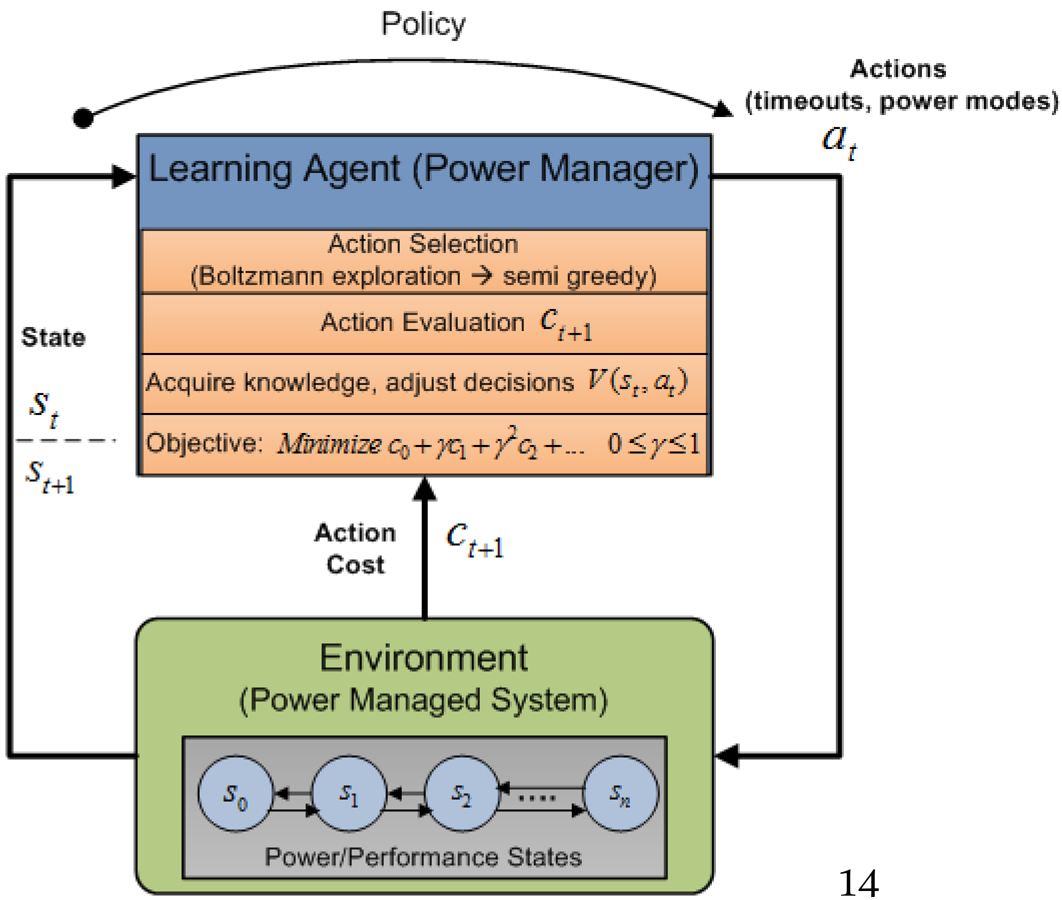
State Space:

$$\begin{cases} S=(WL, SQ, SP) \\ A=\{t_{out}^{k}\}=\{\varepsilon_k T_{thr}\},\ \ \varepsilon_k \in R^{+},\ \ k=1,2,\ldots,n \end{cases}$$

Cost Function:

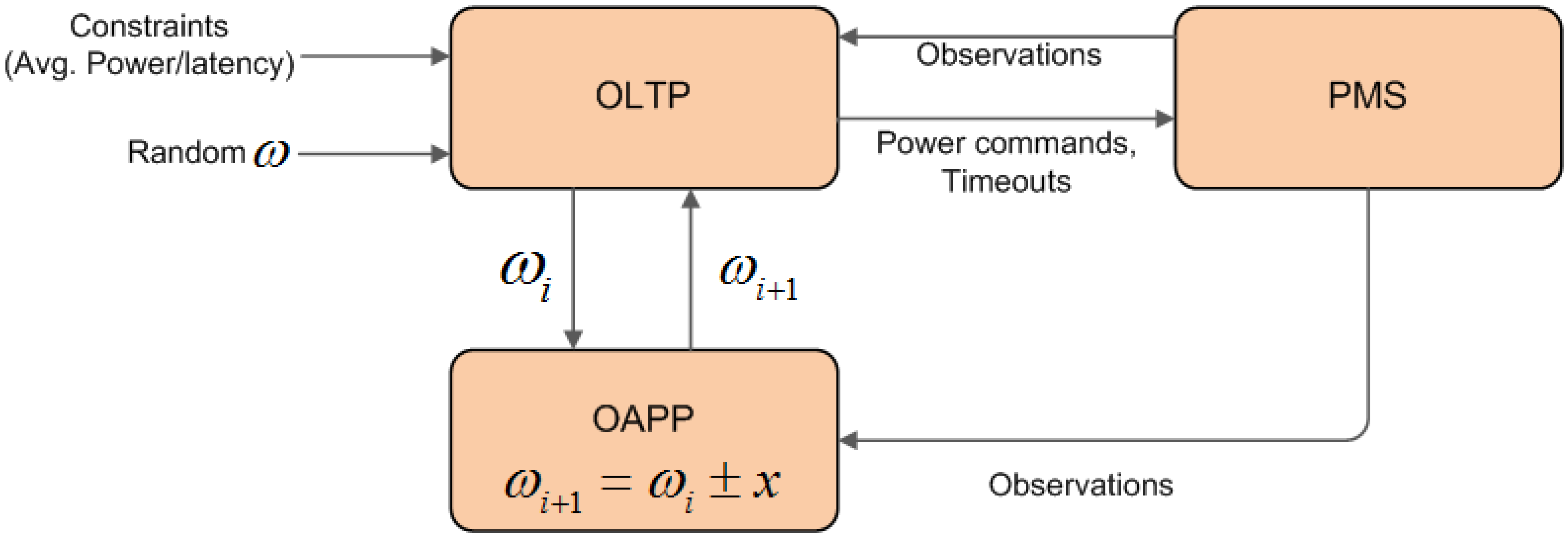
$$\forall s,a \in S \times A:$$

$$c_{t+1}(s,a,\omega)=(1-\omega)p_t(s,a)+\omega l_t(s,a)$$

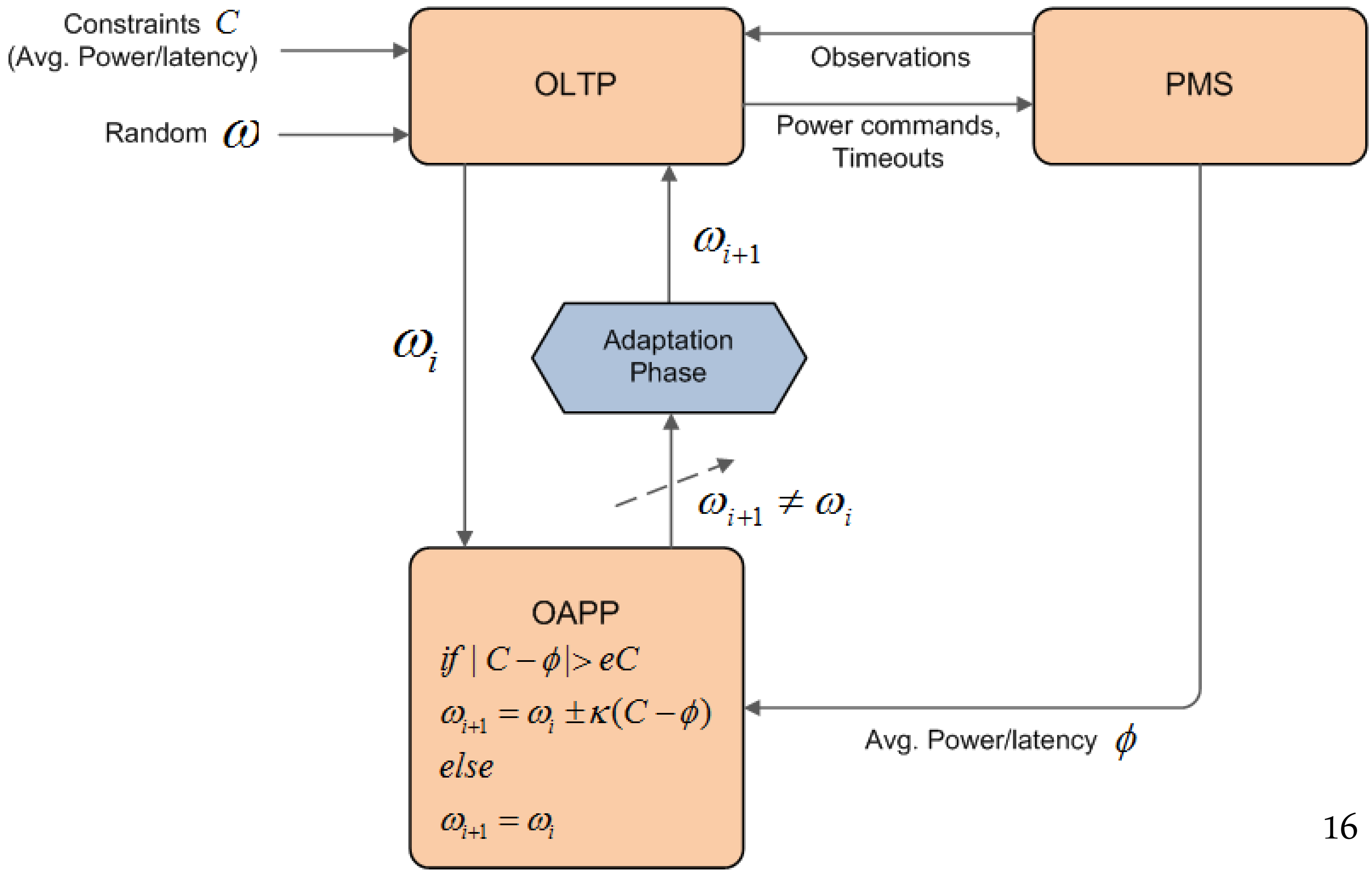


Online Adaptation of Power/Performance (OAPP) - 1

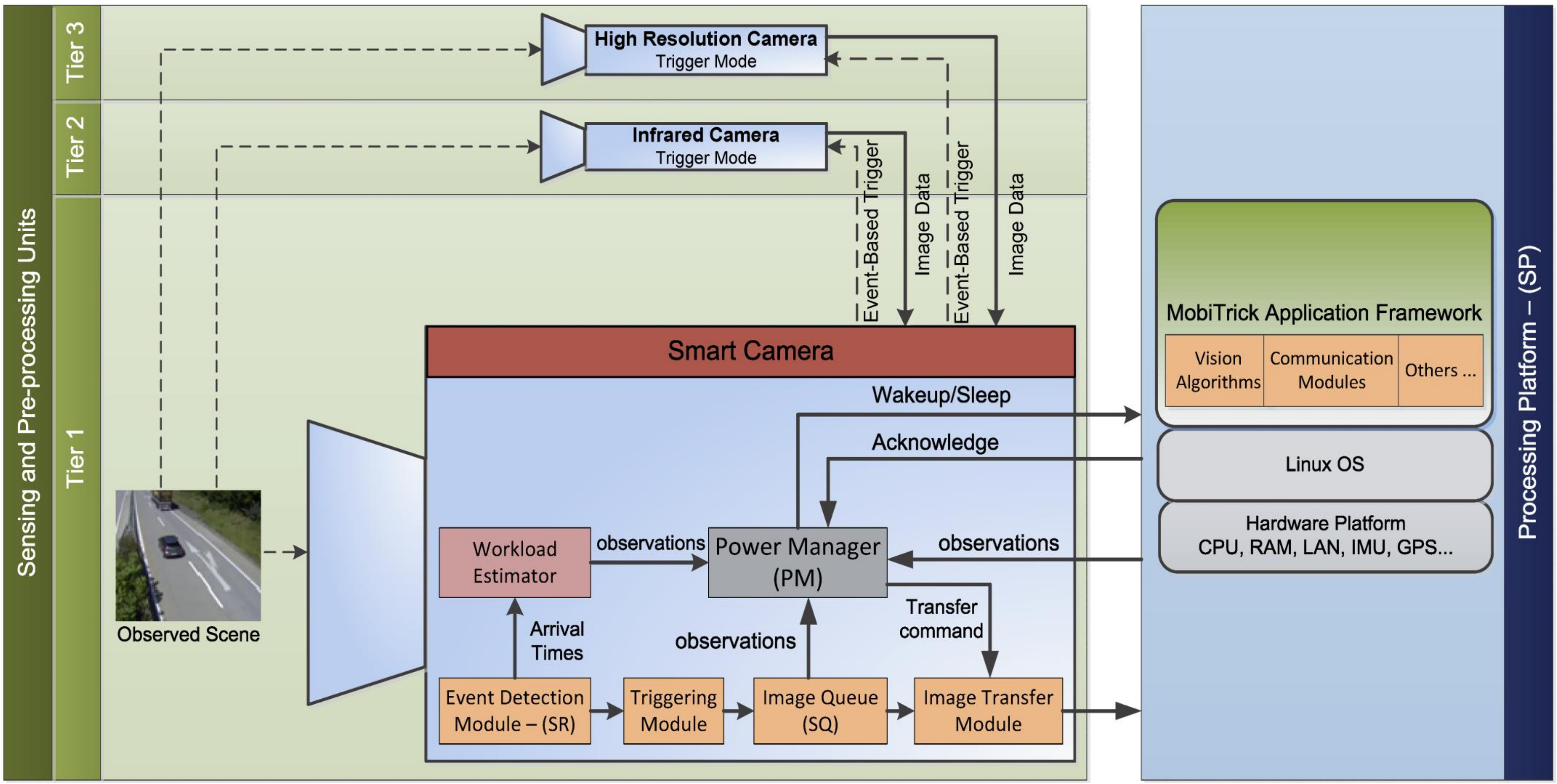
- Online constraints adaptation
- Adjusting objectives weights
- Online, discrete OAPP controller



Online Adaptation of Power/Performance (OAPP) - 2



The Implementation Platform (MobiTrick) - 1



The Implementation Platform (MobiTrick) - 2

- Processing platform
  - Intel ATOM 1.6 GHz
  - Nvidia ION GPU
  - 4 GB RAM
  - Power modes: idle, sleep, processing
- Sensing Components
  i) TMDXIPCAM8127 (5MP, RGB)
    - 1x Arm Cortex A8
    - 2x Arm Cortex M3
    - 1x C674x DSP
    - Encoding co-processor
    - 512 MB DDR

  ii) PhotonFocus (1.4 MP, Infrared)

  iii) AV10005DN (10 MP, RGB)

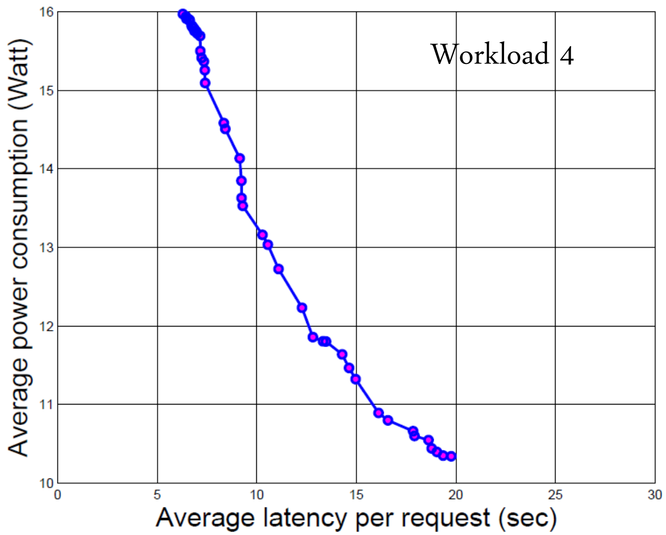
  iv) Accelerometers, gyroscopes, magnetometer, GPS receiver

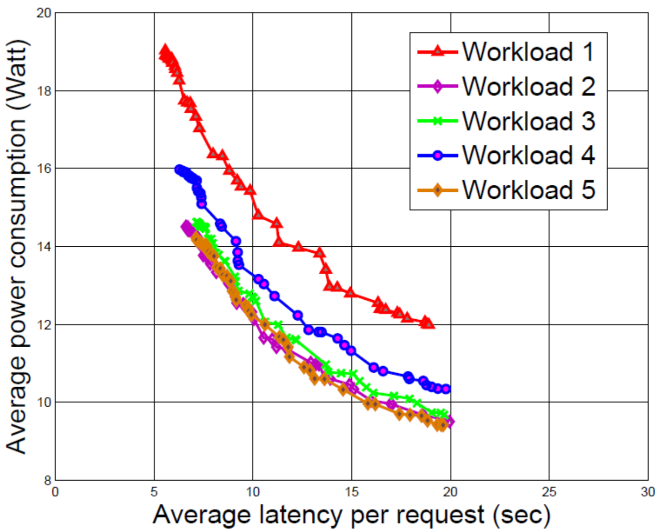
Evaluation Results – (1)

OLTP: Power-Performance Pareto Front (MobiTrick's PMS)

Workload	Mean Inter-Arrival Time	No. of Requests	Duration
Workload 1	6.79 sec	11649	22 hours
Workload 2	11.13 sec	7762	24 hours
Workload 3	11.07 sec	7803	24 hours
Workload 4	9.06 sec	9502	24 hours
Workload 5	12.05 sec	7155	24 hours

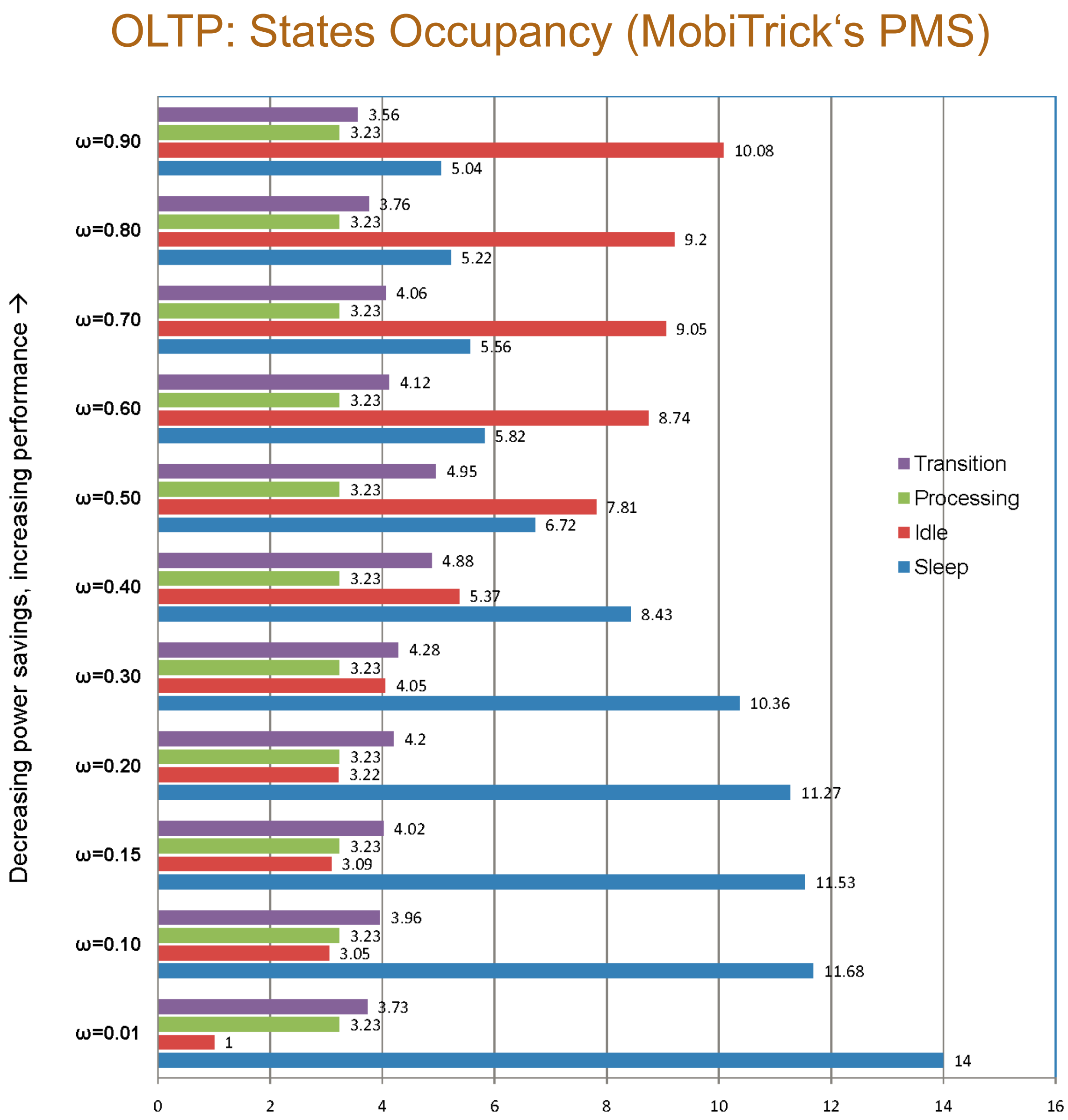
Characteristics of different workloads





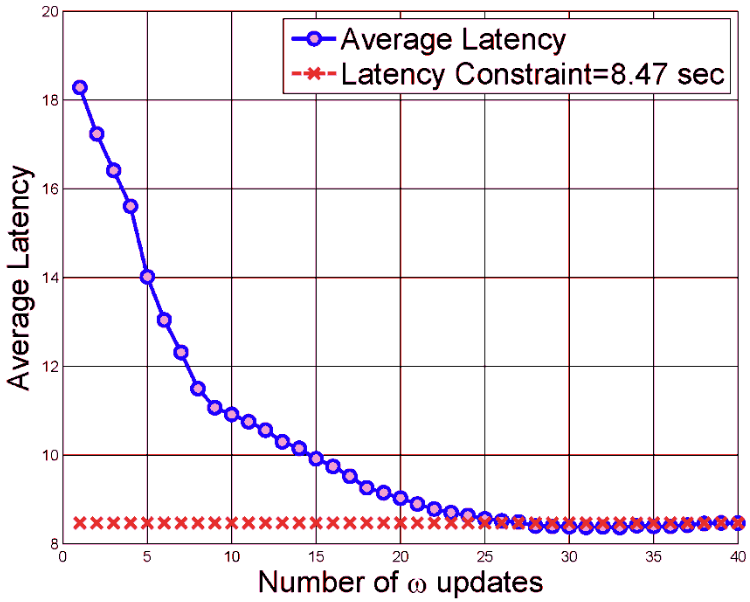
MobiTrick's PMS: 1 idle, 1 sleep and 1 processing state

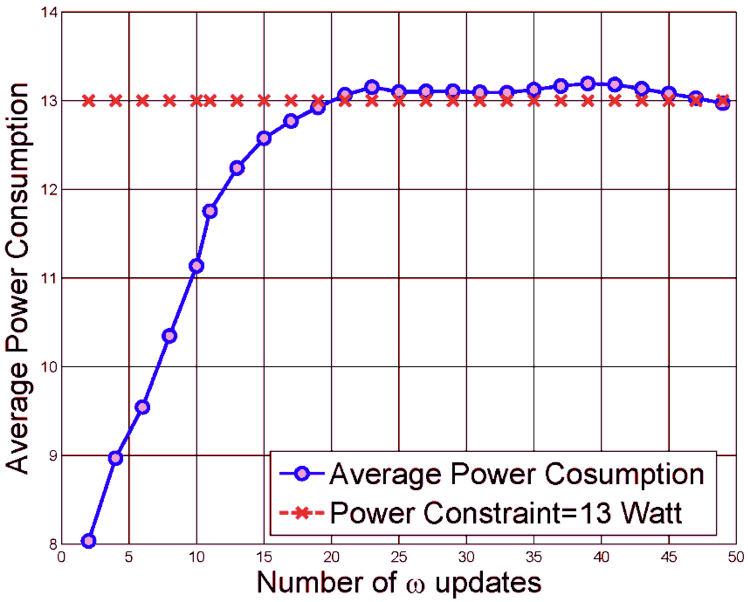
Evaluation Results – (2)



Evaluation Results – (3)

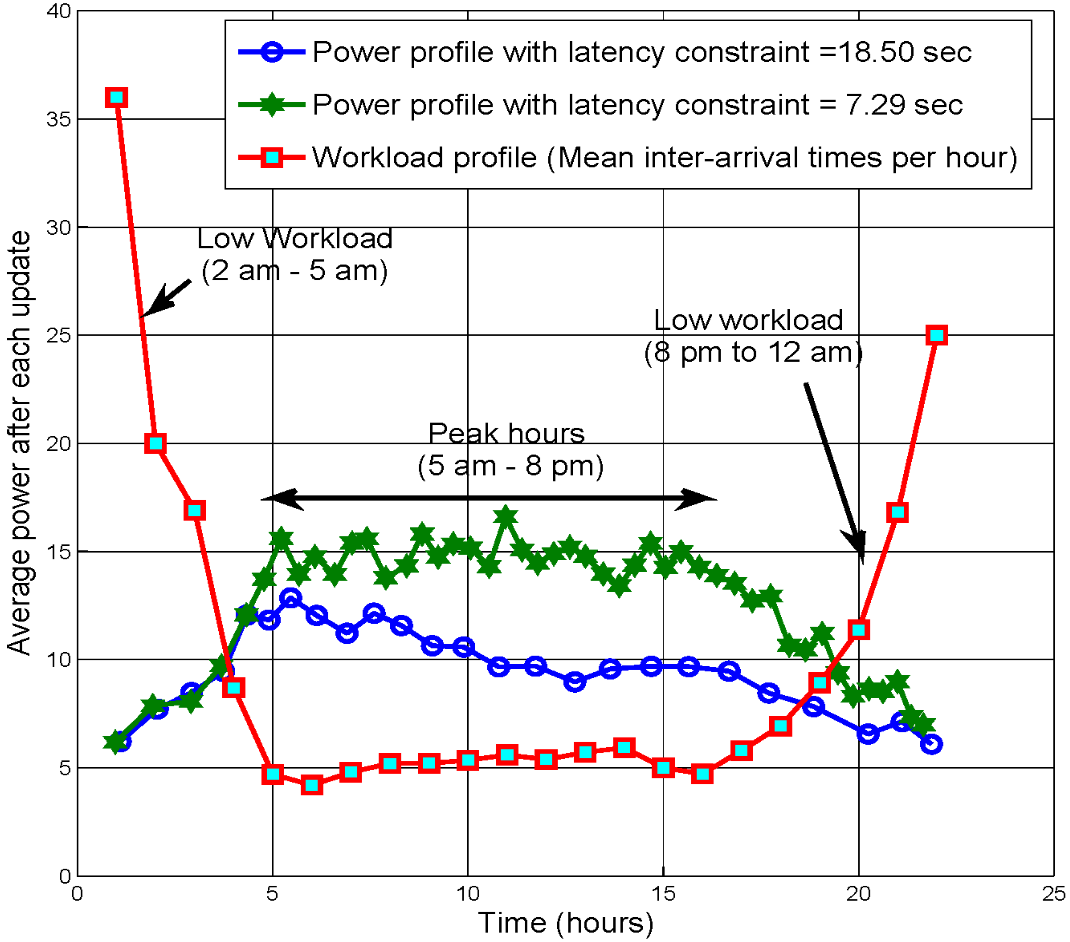
OAPP - MobiTrick‘s PMS





Evaluation Results – (4)

OLTP/OAPP: Power Profile With Changing Workload (MobiTrick's PMS)



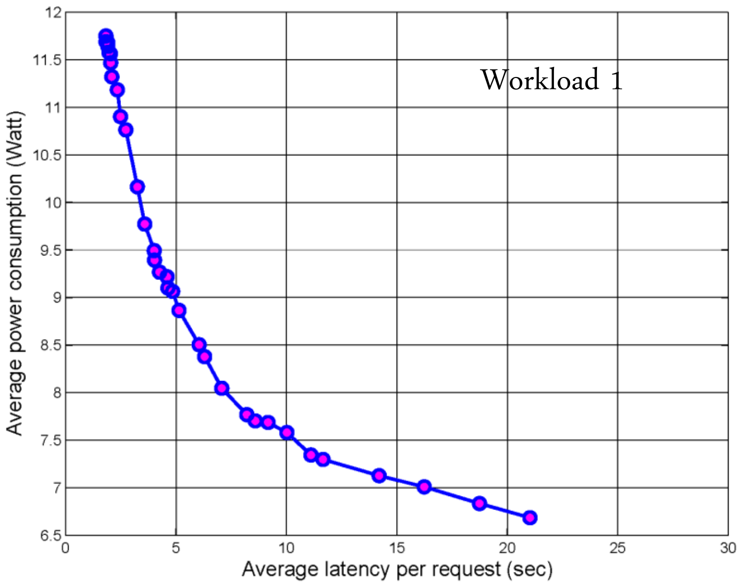
Power profile with changing workload for two different latency constraints

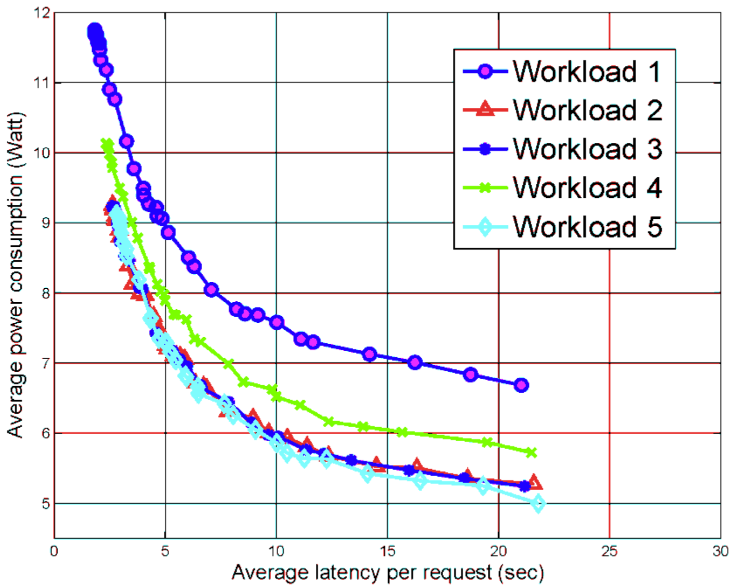
Evaluation Results – (5)

OLTP: Power-Performance Pareto Front (Synthetic PMS)
Multiple idle and sleep states

Workload	Mean Inter-Arrival Time	No. of Requests	Duration
Workload 1	6.79 sec	11649	22 hours
Workload 2	11.13 sec	7762	24 hours
Workload 3	11.07 sec	7803	24 hours
Workload 4	9.06 sec	9502	24 hours
Workload 5	12.05 sec	7155	24 hours

Characteristics of different workloads



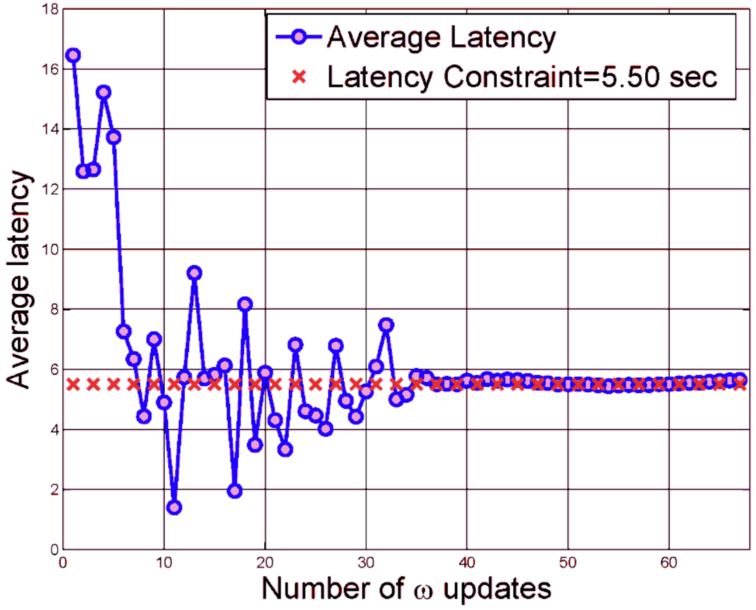


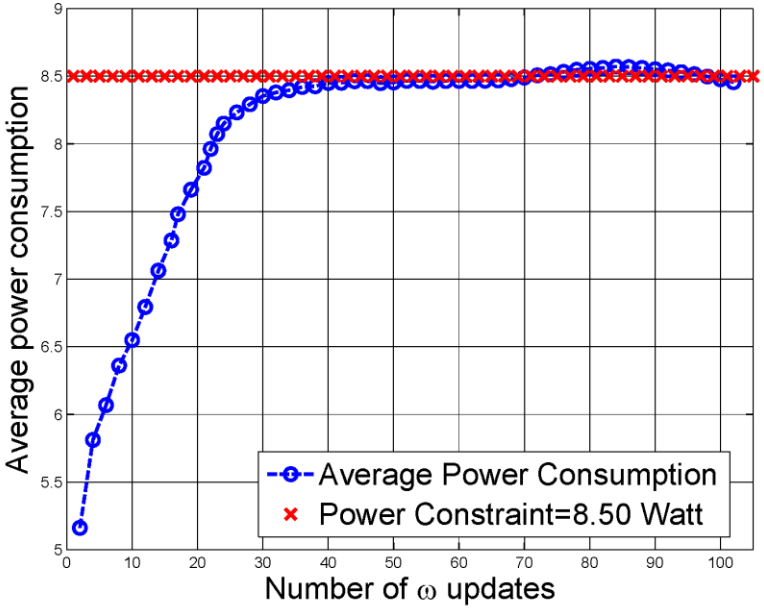
Synthetic PMS: 2 idle, 2 sleep and 1 processing state

Evaluation Results – (6)

OAPP - Synthetic PMS

(Multiple idle and sleep states)





Conclusion

- Online, model-free DPM approach
- Computation & memory efficiency
- Controllable power-performance tradeoff
- Adaptation to non-stationary workloads
- Online constraints adaptation
- Compatibility with larger systems

Future Directives

- Migration to OS level
- Variable timeout values
- Continuous timeout values
- Multiprocessor DPM
- Dynamic frequency scaling