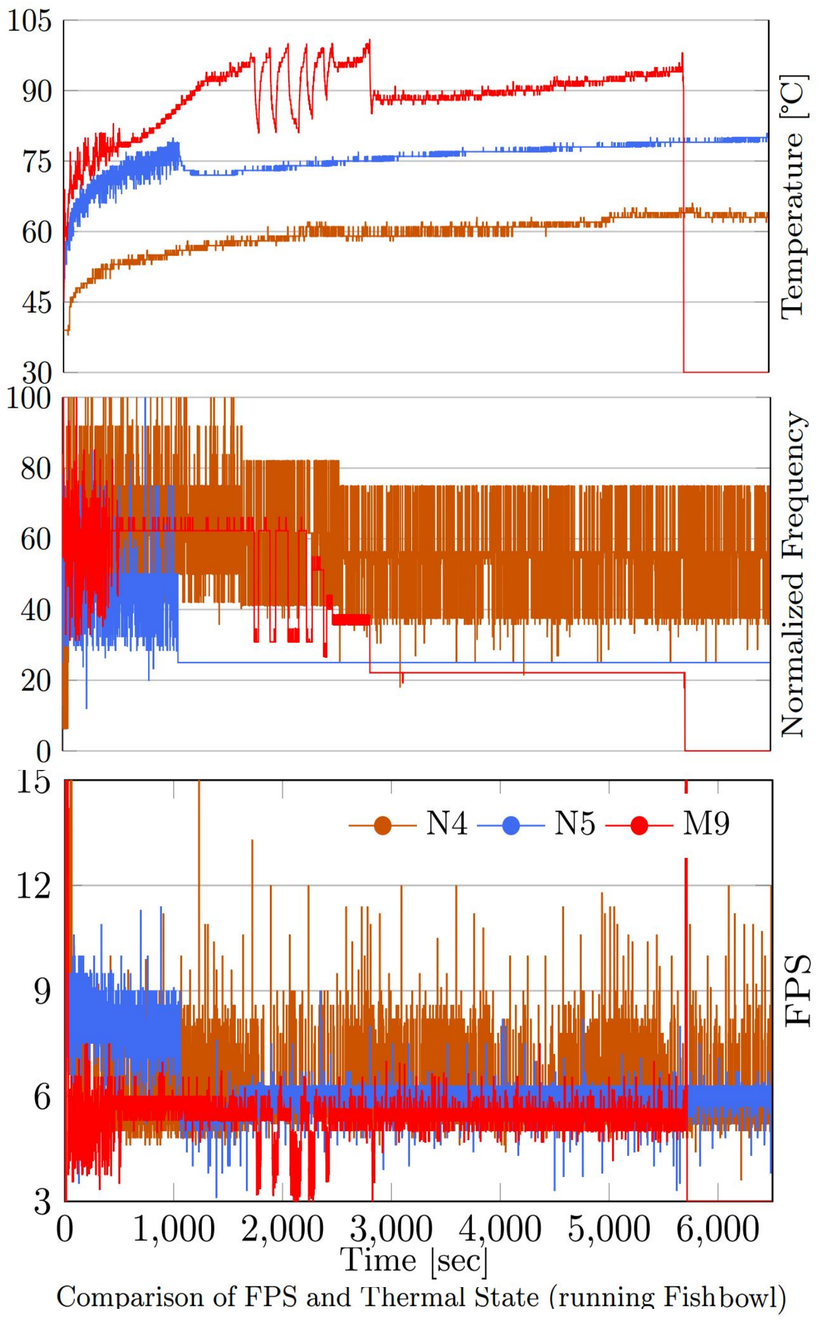
TeSLA: Thermal Service Level Agreement for Mobile Devices

HyunJong Joseph Lee, Minsung Jang*, William Agnew, Karsten Schwan, Ada Gavrilovska
{josephlee, wagnew, karsten, ada}@gatech.edu, minsung@research.att.com*



Comparison of FPS and Thermal State (running Fishbowl)

Motivation

- Rapid advances in mobile processors show promise in running compute-intensive applications (apps) on resource-constrained mobile devices. The mobile applications, however, face challenges in obtaining the full potential of those processors, especially when sustained performance is crucial for the apps.
- The need of sustained performance sharply increases temperature on the processors. Then a throttling mechanism kicks in to cool them off and consequently makes the apps seriously slow.

Problem Statement

- Processor's protection mechanism, thermal engine prohibits the applications from gaining the full processing performance for the necessary duration of time.
- Degradation of core performance by thermal engine results in worsen user-experience.
- A new resource management abstraction, and argue for its importance in allowing mobile applications to gain sustained and predictable performance.
- Thermal-state-aware resource management permits mobile devices to be maintained at the optimal thermal level that avoids the thermal engine intrusion.

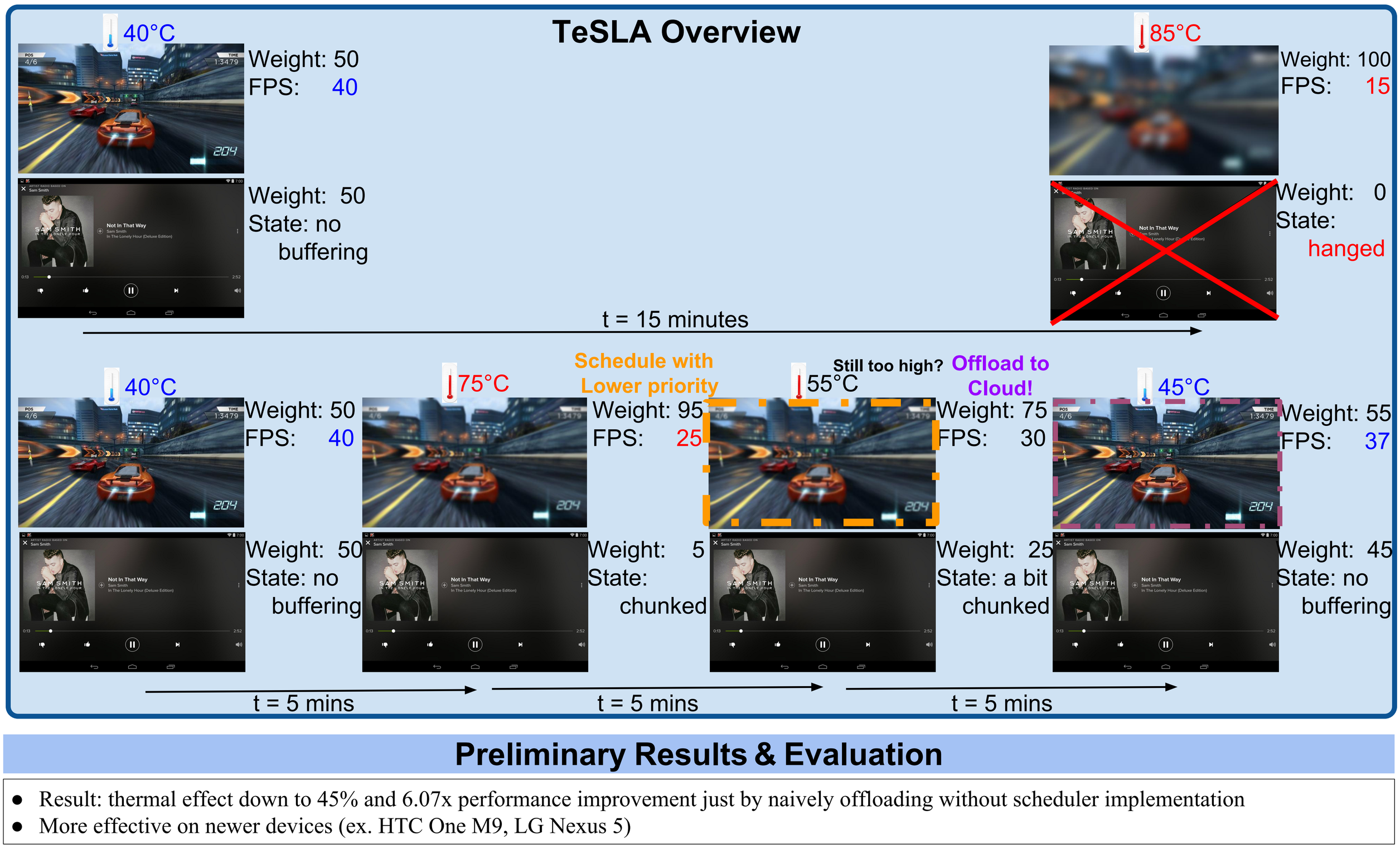
Proposed Solution(s)

- Thermal Engine by vendors, using Dynamic Voltage Frequency Scaling (DVFS) that can adapt core frequencies to four levels.
  - DVFS is limited to stepwise frequencies at hardware support.
- Leveraging off-the-shelf Big.little core technology to migrate a CPU workload of Big to little ones.
  - Heat dissipation among cores in tiny dies weakens the impact of migration.
- Mobile offloading to cloud or nearby Cloudlet to offload compute-intensive workload from fanless hot-device to spacious workstation.

Our Approach & Contributions

- Our approach: TeSLA, ***T**he**r**mal **S**ervice **L**evel **A**greement* between underlying mobile system and apps, to guarantee sustained and predictable performance for apps in mobile devices.
- Contributions:
  - Identifying thermal-related system performance issue & its magnitude: Our preliminary study shows that performance-need from even a single app causes unfettered system-wide thermal problem, which results in performance drop affecting **all** apps in the device.
  - Providing a mechanism to relieve the thermal-oriented performance issue: TeSLA proposes a mechanism to stay the mobile processors in the Goldilocks temperature for sustained and uniform performance by suppressing temperature under throttling threshold so that the apps run with long-lasting good performance.
  - Implementing a TeSLA prototype: The prototype works on real Android devices and apps from the market as intended.
- Motivating example: TeSLA successfully manages temperature below the thermal threshold of the processors and guarantees the availability of desirable performance for a compute-intensive Photoshop-like app.

TeSLA Overview



Preliminary Results & Evaluation

- Result: thermal effect down to 45% and 6.07x performance improvement just by naively offloading without scheduler implementation
- More effective on newer devices (ex. HTC One M9, LG Nexus 5)