Towards Deploying Decommissioned Mobile Devices as Cheap Energy-Efficient Compute Nodes

Mohammad Shahrad and David Wentzlaff

Monday, July 10, 2017
Growing Dominance of Smartphones

[Graph showing the growth of smartphone, PC, and tablet sales from 2009 to 2014. The graph indicates a significant increase in smartphone sales compared to PCs and tablets.]

[Bar chart illustrating smartphone subscriptions per region for 2015 and 2021. The chart shows a total of 6.3 billion subscriptions across different regions.]

[Source: Gartner, IDC, Zuberbühler Associates AG, Ericsson Mobility Report 2016]
Decommissioned Smartphones

- Physical Damage
- Expanding OSs
- New Features
- Fashion

- 4K

- Bulk Recycling
- Electronic Waste

- Refurbish
Using the computational capacity of decommissioned devices

Reusing smartphones as compute nodes?
Outline

Why using mobile SoCs now?

How to integrate mobile devices into data center fabric?

What to do with such mobile devices in data centers?

Does such deployment make economic sense?
Trend 1: Commodity vs. Mobile Performance

[Rajovic et al., Supercomputing with commodity CPUs: Are mobile SoCs ready for HPC?, SC ’13]
Trend 2: Diminishing Performance Growth

End of Moore’s Law is starting to kick in for mobile SoCs.

The performance gap between a new vs. a 3-year old phone is decreasing.
Extending Effective Lifetime to Improve Carbon Footprint

[Erkan et al., Life cycle assessment of a smartphone, ICT4S ‘16]
Proposed Integration Fabric

Power Supply
Router
Fans
Mobile Device Cage (180 x 80 x 9 mm)

2U Height

84 x 470 ≈ 470

SoC TDP = 3.5 W
Networking

1. USB Tree + Master Node
   - Establish virtual Ethernet through master node
   - Master becomes single point of failure
   - No Need for a Master Node

2. USB On-The-Go (OTG)
   - High BW per Device
   - No Need for a Master Node
   - Ethernet Switch Cost (price+space)

3. Intra-server WiFi
   - Low SNR and high congestion.
   - Low BW
   - High Energy
   - High Latency

Master Node

USB OTG Ethernet

Ethernet Switch

Upstream Network

Supply
Mobile Batteries as Distributed UPSs

Mobile batteries are designed for high-density energy storage.

Average mobile battery: > 3100mAh
After 3 years (15% dep/yr): >1900mAh
84 device per server: ~160Ah
(4x-8x the typical UPS density)

Batteries could be used to perform power capping in data centers.
Targeted Applications?

- Lower CPU Performance
- High Relative I/O Bandwidth
- Non-CPU Elements of Mobile SoC
- Lower Reliability
I/O-intensive Application

Prior examples:

**Fast Array of Wimpy Nodes** [Andersen et al., SOSP ’09]
- Using low-end embedded CPUs in a key-value store system.
- Two orders of magnitude more queries per Joule.

**Web Search Using Mobile Cores** [Reddi et al., ISCA ’10]
- Comparing Atom to Xeon processors or perform web search.
- Better energy efficiency
- Worse QoS

Smaller cores generally have better I/O to compute ratio.

Mobile cores have better performance-per-Watt.
Low-end VM Provisioning

<table>
<thead>
<tr>
<th>instance</th>
<th>vCPU</th>
<th>Mem (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t2.nano</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>t2.micro</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>f1-micro</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>g1-small</td>
<td>0.5</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Each mobile device (on average):
- > 2GB of memory
- > 5 cores

Multiple tiny burstable instances can be provisioned on each mobile device.
GPU-accelerated Dwarfs

• Current GPU-accelerated IaaS instances use beefy GPUs (e.g. EC2 P2 instances)
• Mobile SoCs are equipped with smaller GPUs
• Enabling low-end GPU acceleration for small VMs

Lower-reliability Nodes

• Dynamic reliability platforms use resource heterogeneity for cost saving while meeting SLAs
• Mobile devices can be used as low-cost low-reliability nodes
Does it make economic sense?
TCO Comparison with a Similar-density Server

Two simplifying assumptions:

1. Parallelizable workload (using Geekbench 4 cross-platform benchmark)
2. Considered eBay retail price for used phones (could be much cheaper)

<table>
<thead>
<tr>
<th>System</th>
<th>Server</th>
<th>CPU/SoC</th>
<th>Server Height</th>
<th>CPU per Server</th>
<th>Grade per CPU</th>
<th>Total RAM (GB)</th>
<th>Total Storage (TB)</th>
<th>Per Rack Unit (1U)</th>
<th>Grade</th>
<th>Memory</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Lenovo Flex System x880 X6 [21]</td>
<td>Intel Xeon E7-8890 v3</td>
<td>1.2U</td>
<td>2</td>
<td>74482 [14]</td>
<td>144 GB (18×8)</td>
<td>1.56 TB (2×800GB)</td>
<td>124,137</td>
<td>120 GB</td>
<td>1.30 TB</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Proposed Server</td>
<td>Snapdragon 805 (Samsung Galaxy Note 4)</td>
<td>2U</td>
<td>84</td>
<td>3060 [14]</td>
<td>252 GB (84×3)</td>
<td>2.63 TB (84×32GB)</td>
<td>128,520</td>
<td>126 GB</td>
<td>1.32 TB</td>
<td></td>
</tr>
</tbody>
</table>

TCO: Total Cost of Ownership
TCO comparison with a Similar-density Server

$\delta$: annual depreciation rate  
(determines the residual value)

A: Lenovo Flex server  
B: Our proposed server
Thanks for your attention!

Next slide: Discussion Topics
Some Discussion Topics

• Reliability of used mobile devices under data center workload

• The best management scheme (centralized, distributed, server-level, etc.)

• Using SoC accelerators for domain-specific applications. (security, neural net, etc.)

• Dealing with extreme heterogeneity

• Security concerns with used phones