Energy-Efficient Software Development for the Internet of Things

Chandra Krintz
Professor, Computer Science Dept.
UC Santa Barbara

IEE Emerging Technologies Review, May 16, 2019
What will you buy?
• When will you buy it?
• What will you pay?
The Internet of Things (IoT): Leveraging Cloud & Big Data Analytics

- Prescriptive & descriptive
- Actuation & control
- Proactive & predictive
- Performance & optimization
- Observation & surveillance
- Real-time actionable insight
- Automation & artificial intelligence

- Massive monitoring

- Food production
- Retail
- Healthcare
- Energy
- Building
- Security
- Transportation
E-Commerce & Entertainment: Data Moves From Cloud to User Devices

Download (read) dominated: streaming entertainment/content
"Cloud" and Internet Was Not Designed for IoT

75 Billion IoT Devices by 2025

Upload (write) dominated: sensing and monitoring
Cloud Tier

Edge Systems:
Small clouds,
Data centers,
Local servers,
Single board computers,
Microcontrollers, ...

The IoT Alternative:
Move the cloud (code) to the data instead of moving the data to the cloud!

Public Clouds

Resource Constrained

Resource Rich
Public Clouds

Devices

Edge Tier

Edge Systems:
Small clouds,
Data centers,
Local servers,
Single board computers,
Microcontrollers, ...

Resource Constrained

The IoT Alternative:
Move the cloud (code)
to the data instead of
moving the data to the
cloud!

Data analytics apps

External data sources

Cloud Tier

Public Clouds

Resource Rich

Amazon Web Services

Microsoft Azure

Google Cloud Platform
**Software Engineering of Multi-Tier IoT Apps**

- Each tier is very different/heterogenous
  - Capability
  - Capacity
  - Energy use
  - Availability
  - Hardware
  - Software (languages, libraries)
  - Deployment
Software Engineering of Multi-Tier IoT Apps

• Each tier is very different/heterogenous
  • Capability
  • Capacity
  • Energy use
  • Availability
  • Hardware
  • Software (languages, libraries)
  • Deployment

• Requires significant expertise to write applications that span tiers or that run on 1+ tiers
  • Required to expedite innovation
Software Engineering of Multi-Tier IoT Apps

• Each tier is very different/heterogenous
  • Capability
  • Capacity
  • Energy use
  • Availability
  • Hardware
  • Software (languages, libraries)
  • Deployment

• Requires significant expertise to write applications that span tiers or that run on 1+ tiers
  • Required to expedite innovation

• Application focus on
  • Data processing
  • Analytics including machine learning and artificial intelligence (AI)
  • Actuation, control, automation
Software Engineering of Multi-Tier IoT Apps

- Each tier is very different/heterogenous
  - Capability
  - Capacity
  - Energy use
  - Availability
  - Hardware
  - Software (languages, libraries)
  - Deployment

- Requires significant expertise to write applications that span tiers or that run on 1+ tiers
- Required to expedite innovation

Application focus on
- Data processing
- Analytics including machine learning and artificial intelligence (AI)
- Actuation, control, automation

RACELab Research:
- Portability across devices, systems, tiers
- Energy efficiency for resource-constrained systems
- Automated deployment at scale
- Write-once, Run-Anywhere for IoT
- Services & optimizations for analytics, ML, AI
Programming Complex Systems: What Would the Cloud Do?

IaaS: Cloud Infrastructure (as a Service)

- Applications
- Runtimes
- Security & Integration
- Databases
- Servers
- Virtualization
- Storage
- Networking

Managed by:
- Vendor software (public) e.g. AWS
- IaaS software (on-premises) e.g. Eucalyptus, Openstack, Kubernetes

PaaS: Cloud Platform (as a Service)

- Applications
- Runtimes
- Security & Integration
- Databases
- Servers
- Virtualization
- Server HW
- Storage
- Networking

Managed by:
- Vendor software (public) e.g. AWS Beanstalk, Google App Engine, Heroku
- PaaS software (on-premises) e.g. AppScale, Kubernetes
Write-Once Run-Anywhere for Multi-Tier IoT Apps

• Software Platform of Things (SPOT)
  • Program = Simple, **event-triggered functions**
  • Portable across IoT systems (all tiers)
  • Leverages Linux containers for portable deployment
• Open source, self-service, easy to use
Write-Once Run-Anywhere for Multi-Tier IoT Apps

- **Software Platform of Things (SPOT)**
  - Program = Simple, *event-triggered functions*
    - Portable across IoT systems (all tiers)
    - Leverages Linux containers for portable deployment
  - Open source, self-service, easy to use

- Program variables/data structures
  = Wide-area Objects of Functions (WOOFs)
    - Durable, persistent, replicated, logged

- “Cloud-like” tools and services
  - Libraries
  - Automatic deployment & placement
  - Monitoring
  - Debugging, root cause analysis
  - Replay, Pause/Resume

---

Devices-as-Services

- IoT applications (clients): collections of services implemented by device tier
  - Call for a *flipped* client-server model
  - Apps compose devices-as-services (data, actuation, control)

F. Bakir, R. Wolski, C. Krintz, and G. Sankar Ramachandran, *Devices-as-Services: Rethinking Scalable Service Architectures for the Internet of Things* USENIX HotEdge, July 2019 (to appear)
Devices-as-Services

- IoT applications (clients): collections of services implemented by device tier
  - Call for a *flipped* client-server model
  - Apps compose devices-as-services (data, actuation, control)

- Edge systems: *Edgistries*
  - Isolation/replacement, privacy, security
  - Discovery
  - Speed-matching & offloading

- Apps, edgistries, devices programmed via SPOT
  - Data processing, fusion, analytics, ML, AI, ...
Performance Results

• Software Platform of Things (SPOT)
  • Simple, portable, *event-triggered functions*
  • **Very fast and energy efficient across all tiers, even with data replication!**
    • AWS and Azure require multiple programming models

<table>
<thead>
<tr>
<th>Deployment</th>
<th>Replicas</th>
<th>Mean (ms)</th>
<th>Stdev (ms)</th>
<th>95% (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MicroWoof-&gt;Pi-&gt;Edge-&gt;Campus-&gt;AWS</td>
<td>5</td>
<td>535</td>
<td>61</td>
<td>650</td>
</tr>
<tr>
<td>Micro-&gt;Pi-&gt;Edge-&gt;Campus-&gt;AWS</td>
<td>4</td>
<td>513</td>
<td>48</td>
<td>607</td>
</tr>
<tr>
<td>MicroWoof-&gt;Edge-&gt;AWS</td>
<td>3</td>
<td>298</td>
<td>15</td>
<td>326</td>
</tr>
<tr>
<td>Micro-&gt;Edge-&gt;Campus-&gt;AWS</td>
<td>3</td>
<td>323</td>
<td>17</td>
<td>427</td>
</tr>
<tr>
<td>AWS Greegrass (Micro-&gt;Edge-&gt;AWS)</td>
<td>&gt;= 3</td>
<td>4136</td>
<td>632</td>
<td>4288</td>
</tr>
<tr>
<td>MS IoT Edge (Edge-&gt;Azure)</td>
<td>&gt;= 3</td>
<td>2621</td>
<td>1512</td>
<td>4386</td>
</tr>
</tbody>
</table>
Performance Results

- Software Platform of Things (SPOT)
  - Simple, portable, event-triggered functions
  - **Very fast and energy efficient across all tiers, even with data replication!**
    - AWS and Azure require multiple programming models

<table>
<thead>
<tr>
<th>Deployment</th>
<th>Replicas</th>
<th>Mean (ms)</th>
<th>Stdev (ms)</th>
<th>95% (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MicroWoof-&gt;Pi-&gt;Edge-&gt;Campus-&gt;AWS</td>
<td>5</td>
<td>535</td>
<td>61</td>
<td>650</td>
</tr>
<tr>
<td>Micro-&gt;Pi-&gt;Edge-&gt;Campus-&gt;AWS</td>
<td>4</td>
<td>513</td>
<td>48</td>
<td>607</td>
</tr>
<tr>
<td>MicroWoof-&gt;Edge-&gt;AWS</td>
<td>3</td>
<td>298</td>
<td>15</td>
<td>326</td>
</tr>
<tr>
<td>Micro-&gt;Edge-&gt;Campus-&gt;AWS</td>
<td>3</td>
<td>323</td>
<td>17</td>
<td>427</td>
</tr>
<tr>
<td>AWS Greegrass (Micro-&gt;Edge-&gt;AWS)</td>
<td>&gt;= 3</td>
<td>4136</td>
<td>632</td>
<td>4288</td>
</tr>
<tr>
<td>MS IoT Edge (Edge-&gt;Azure)</td>
<td>&gt;= 3</td>
<td>2621</td>
<td>1512</td>
<td>4386</td>
</tr>
</tbody>
</table>

- Devices-as-services: Capability-based security
  - Uses 3-4 orders of magnitude **less energy** for authentication than RSA or ECC based approaches
Summary

• Bringing cloud/data analytics to IoT requires new research in
  • Distributed systems with *edge intelligence*
  • Software architectures and platforms
  • Programming systems
  • Services and tools

• To facilitate (across IoT tiers)
  • Energy efficiency
  • Low latency (fast app response)
  • Privacy/security
  • Ease of programming/deployment

• We show it is possible to achieve both
  • Via new distributed software platform and programming model
  • Tailored to the IoT use case
Thanks!

• **Collaborators:** UCSB, UCSB IEE, LREC, CalPoly, Fresno State, Powwow Energy, Sedgwick Reserve, Private Growers

• **Support:** Google, Intel, IBM Research, Microsoft Research, NSF, NIH, California Energy Commission

[Logos and images]

**Students:**

Fatih Bakir, Kyle Carson, Gareth George, Nevena Golubovic, Carly Larsson, Wei-Tsun Lin, Andy Rosales Elias, Nazmus Saquib, John Thomason, Michael Zhang

**UCSB RACELab**

The Lab for Research on Adaptive Computing Environments
Computer Science Department, Harold Frank Hall (E-5), Santa Barbara, CA

ckrintz@cs.ucsb.edu, rich@cs.ucsb.edu

http://www.cs.ucsb.edu/~ckrintz/racelab.html