ECE 590/COMPSI 590
Special Topics: Edge Computing

Edge Helping Higher-end Mobile Devices: Mobile Offloading

Wednesday January 22nd, 2018

Duke UNIVERSITY

1

Last Class Recap

- Edge and IoT devices
  - ➢ Common IoT architectures
  - ➢ Role of the gateway
- Opportunities: edge for responsive IoT applications
  - ➢ Hardware
  - ➢ Algorithms
  - ➢ Edge for system decisions

2

Duke UNIVERSITY

2

Upcoming Timelines: A Reminder

- This week:
  - ➢ Project team selection: Friday January 24th
  - ➢ Paper presentation slot sign-up: Friday January 24th

- 2.5 weeks from now:
  - ➢ Project proposal: Monday February 10th
  - ➢ Project proposal presentations: Wednesday February 12th

3

Duke UNIVERSITY

3

Does Anyone Have a Project Idea They Want to Run by the Group?

4

Duke UNIVERSITY

4

Lecture Outline

- Technology and Courage
- Higher-end mobile devices
- Cloudlets
  - ➢ Current presence
  - ➢ Challenges
- Mobile offloading
- Future directions in mobile offloading

Quiz

Technology and Courage

- What did you think?

7

Duke UNIVERSITY

7

Edge for IoT Nodes vs. Edge For High-End Mobile Nodes

Low-end IoT nodes

IoT nodes → Gateway → Cloud

Last lecture

Higher-end mobile devices

Mobile device → Cloudlet → Cloud

This lecture

8

Duke UNIVERSITY

8

Core Approaches

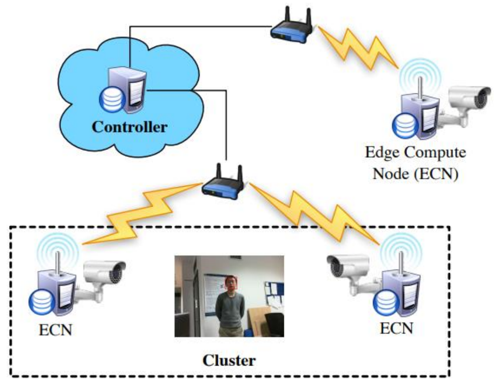
- Edge devices: **cloudlets**
- Core technique: **mobile offloading**

Higher-End Mobile Devices

- Mobile phones: prevalent use case
- AR/VR, drones, smart cars – emerging use cases

Special Case: Camera Installations

- E.g., city, campus security cameras
  - Very common
  - Of major practical importance
  - Often not **mobile** devices
  - Many video-specific mechanisms



From: The Design and Implementation of a Wireless Video Surveillance System, Zhang et al, ACM MobiCom'15

Unlike IoT devices…

- Not as resource-constrained
  - Fewer per-device customizations
  - Usually standard protocols

Unlike IoT devices…

- Complex, often high-volume, data
  - ➢ Variety of sensors – accelerometers, video, audio, …
- More complex operations
  - ➢ Thinking in full application pipelines, rather than individual tasks

13

Duke UNIVERSITY

13

Like IoT devices…

- Battery-limited
  - ➢ How long they last
  - ➢ How much heat they produce
- Usability limited by the batteries

Duke UNIVERSITY

14

Mobile Device vs. a Server

- Isn't a mobile device a desktop in your pocket?
- Server > mobile device
  - ➢ Power constraints → 500 W of power on a high-end GPU, 10 W on a mobile SoC GPU
  - ➢ Space constraints

15

Duke UNIVERSITY

15

Cloudlets

- Local mini-clouds
- Envisioned properties:
  - ➢ Powerful, well-connected, and safe
  - ➢ Close at hand
  - ➢ Build on standard cloud technology

16

Duke UNIVERSITY

16

Existing and Possible "Cloudlets"

- On-site computing
- Targeted edge installations
- Resource scavenging

17

Duke UNIVERSITY

17

On-Site Computing

- At universities
- … and other medium and large organizations
  - ➢Shrinking but not disappearing
  - ➢Usually have low utilization

18

Duke UNIVERSITY

18

Targeted Installation: Chick-fil-A (1/2)

July 2018

19

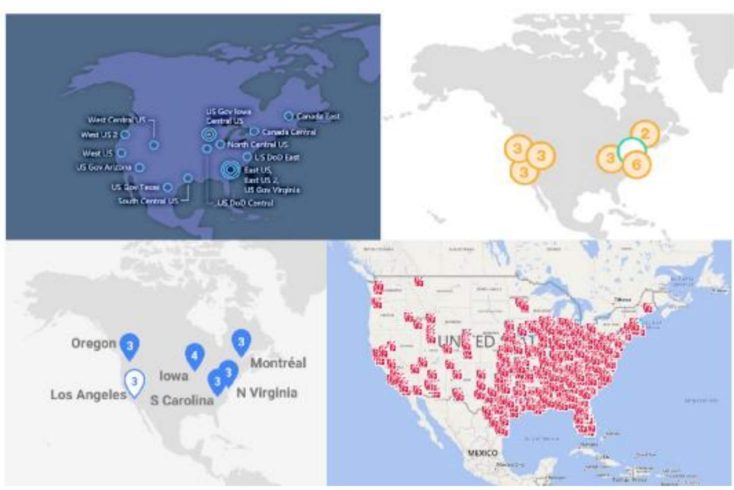
Duke UNIVERSITY

19

Targeted Installation: Chick-fil-A (2/2)

Azure

AWS



Google Cloud

Cloud-fil-A

20

Duke UNIVERSITY

20

Can Imagine Deploying More of These

- Especially for Augmented and Virtual Reality
- … and for smart cities in general

21

Duke UNIVERSITY

21

Resource Scavenging

- Finding unused devices around you
  - ➢ "Cyber foraging"
  - ➢ "*Uberization*" of computing and storage

UBER

22

Duke UNIVERSITY

22

Resource Scavenging: Open Questions

- Open questions
  - ➢ Discovery, connectivity
  - ➢ Security
  - ➢ Incentives, pricing
- Smart city computing infrastructures, e.g. Barcelona deployments, try to address these

23

Duke UNIVERSITY

23

Cloudlet Challenges

- **Mobile** devices → supporting mobility
- Cloud**let** →does not have the scale of the cloud

24

Duke UNIVERSITY

24

Cloudlets Helping Mobile Devices: Challenges: Rapid Service Provisioning

- A scenario: a student comes to Hudson Hall and needs to use our cloudlet
  - Service discovery
  - Provisioning delay
  - Do not have the scale of the cloud: do we prioritize this user over others? Shift workloads with every user?

Cloudlets Helping Mobile Devices: Challenges: Service Handoff

- A scenario: the student moves from Hudson Hall to CIEMAS
  - Do we transfer their workload state?
  - Do we de-provision their Hudson Hall services?

Cloudlets Helping Mobile Devices: Challenges

- Platform challenges
  - Challenges similar to ***wireless hand-off***
- Workload allocation and scheduling challenges

27

Duke UNIVERSITY

27

Lecture Outline

- Higher-end mobile devices
- Cloudlets
  - Current presence
  - Challenges
- Mobile offloading
- Future directions in mobile offloading
- Challenges

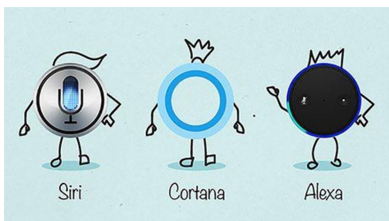
28

Duke UNIVERSITY

28

How Edge Helps: Mobile Offloading

- Executing code **not** on the mobile device
- E.g., image, video, audio, other sensor data processing
  - Face detection, person identification
  - Language translation, speaker identification
  - Activity tracking, gesture recognition



- All offload processing to the cloud

29

Duke UNIVERSITY

29

Goals: Reducing Mobile Device Energy Consumption (1/3)

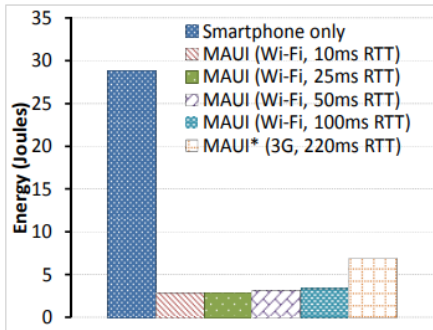
- Need to have:
  - Energy to {transmit data + receive results} < energy to {execute the operation on the mobile device}
- Design principles:
  - Pick the most compute-intensive parts of the operation
  - Reduce the size of what is transmitted: data and results
- Order-of-magnitude mobile energy savings possible

30

Duke UNIVERSITY

30

Example: Face Recognition with MAUI



ONE RUN FACE RECOGNITION

From: MAUI: Making Smartphones Last Longer with Code Offload, Cuervo et al., ACM MobiSys'10.

31

Duke UNIVERSITY

31

Goals: Reducing Mobile Device Energy Consumption (2/3)

- Not minimizing **total** energy:
  - ➢ Combined server + mobile energy spending can be higher than mobile-only energy spending
- System heterogeneity principle:
  - ➢ Server energy spending is not as important as mobile device energy spending
  - ➢ Server grade does not factor into energy minimization objective

32

Duke UNIVERSITY

32

Goals: Reducing Mobile Device Energy Consumption (3/3)

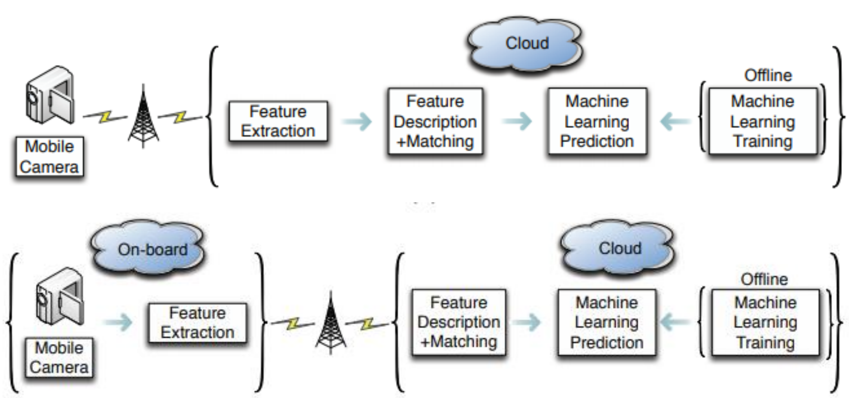
- Often: transmit partially processed, rather than raw, data
  - Energy to {extract features + transmit extracted features + receive results} < energy to {transmit data + receive results}
  - Energy to {extract features + transmit extracted features + receive results} < energy to {execute the operation on the mobile device}

33

Duke UNIVERSITY

33

Local Feature Extraction Can Reduce the Amount of Data Transmitted



From: A Hybrid Approach To Offloading Mobile Image Classification, Hauswald et al, IEEE ICASSP'14.

34

Duke UNIVERSITY

34

Goals: Minimizing Task Completion Time

- Need to have:
  - Time to {transmit data + execute operation on the server + receive the results} < time to {execute the operation locally}
- More demands on the server:
  - Need to offload to a much more capable device

35

Duke UNIVERSITY

35

Mobile Offloading: Need for Scheduling Mechanisms

- Time, energy vary with network connectivity
- Need to make decisions for different conditions
  - Different ways of placing different parts of operations
  - Offline versus online
  - Joint scheduling of different operations
  - Scheduling that takes into account different local processors and the cloud

36

Duke UNIVERSITY

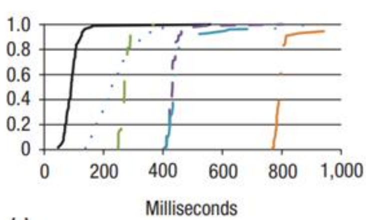
36

Role of the Edge (1/2)

- Short transmission distance helps both transmission energy and latency
  - Better performance of existing offloading scenarios
  - Offloading equations “work out” in more cases
- Potentially, additional privacy

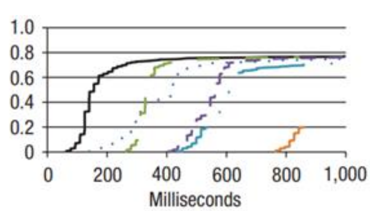
Duke UNIVERSITY

Role of the Edge (2/2)



Augmented reality

3.3 J	Mobile only	16.4 J
1.1 J	Cloudlet	5.4 J
3.1 J	AWS-East	6.6 J
5.1 J	AWS-West	8.5 J
5.2 J	AWS-EU	9.5 J
9.4 J	AWS-Asia	14.3 J



Face recognition

From: The Emergence of Edge Computing, by M. Satyanarayanan, IEEE Computer, 2017. Adapted from The Impact of Mobile Multimedia Applications on Data Center Consolidation, by Ha et al, 2013.

Duke UNIVERSITY

Lecture Outline

- Higher-end mobile devices
- Cloudlets
  - Current presence
  - Challenges
- Mobile offloading
- Future directions in mobile offloading

39

Duke UNIVERSITY

39

Future Directions: “Offload Shaping”

- Adapting operations for offloading
- A form of creative pre-processing
  - Changing application pipelines **specifically** for offloading
- Some examples from: The Case for Offload Shaping, by Hu et al, ACM HotMobile’15

40

Duke UNIVERSITY

40

Offload Shaping: Object Recognition in Video Captures (1/2)

- Object recognition works poorly on blurry frames
  - ➢ Can drop blurry frames before transmitting them to the cloud/cloudlet for processing

	Send all	Drop blurry
Bytes transferred	0.51M	0.34M
Glass energy (J)	429(2)	292(3)
Server CPU usage (normalized)	1.00(0.01)	0.81(0.01)

41

Duke UNIVERSITY

41

Offload Shaping: Object Recognition in Video Captures (2/2)

- Results from similar frames are likely to be the same
  - ➢ Discard frames that are sufficiently similar

	No shaping	Drop similar	Improve-ment
Bytes transferred	0.51M	0.23M	55%
Frames recognized	171(2)	189(1)	11%
Glass power (W)	1.82(0.01)	1.83(0.01)	-1%
E2E latency (ms)	920(8)	393(2)	57%
Glass energy (J/frame)	1.66(0.01)	0.72(0.01)	57%
Server CPU usage (normalized)	1.00(0.01)	0.27(0.01)	73%

42

Duke UNIVERSITY

42

Offload Shaping

- (+) Holistic view of the entire system
  - ➢ Fixing inefficiencies that become obvious when we think about the system beginning-to-end
- (-) Solutions likely to be application-specific
  - ➢ E.g., blur detection in one of the previous examples

Opportunities: Providing Local Context

- Information about local conditions
  - ➢ Pre-programmed
  - ➢ … or learned
- Historic data, predictions

Opportunities: Providing Local Context

- Especially when context is large
- Opportunities for behavior **specialization**

45

Duke UNIVERSITY

45

Side Note: Context Awareness in Applications is Not New

- Traces back to early 1990s
- E.g.:
  - ➢Active badge location system
  - ➢OS updates only when a phone is plugged in and is on WiFi

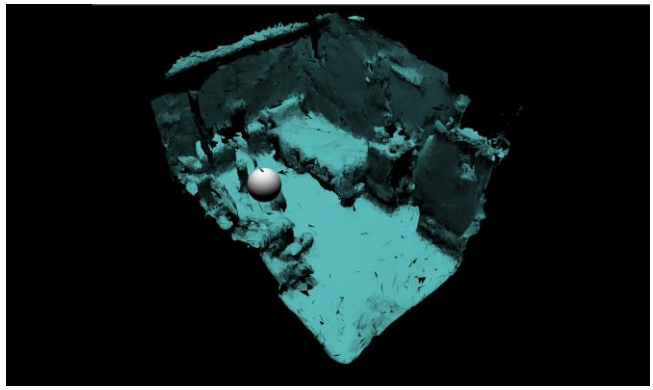
46

Duke UNIVERSITY

46

Large Local Context: 3D Maps of the Environment for AR/VR (1/2)

- Massive amounts of information and processing
  - ➢ Useful to not regenerate for all users
  - ➢ Useful to not fetch from the cloud



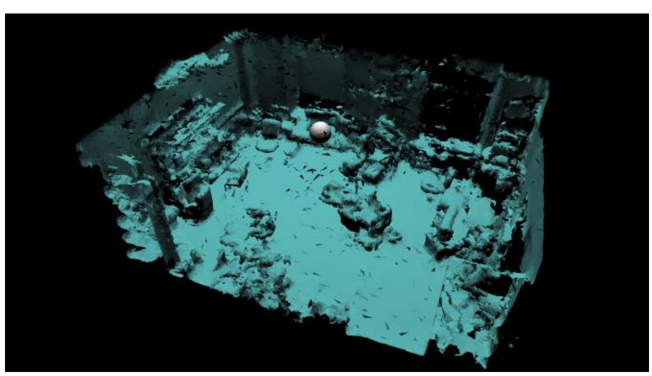
Mesh representing a student dorm room

47

Duke UNIVERSITY

47

Large Local Context: 3D Maps of the Environment for AR/VR (2/2)



Mesh representing a lab

48

Duke UNIVERSITY

48

What Could Hudson Hall and CIEMAS Cloudlets Tell Us?

- Opportunities for behavior specialization

Opportunities: Thinking Across Multiple Devices and Multiple Applications

- New paradigms

- Without the cloudlets, ***nearby devices have no exposure to each other's actions***
  - ➢ No single "choke point"

Opportunities: Thinking Across Multiple Devices and Multiple Applications

- Same application likely to be invoked on different devices served by one cloudlet

51

Duke UNIVERSITY

51

Class Recap

- Higher-end mobile devices
- Cloudlets
  - ➢ Current presence
  - ➢ Challenges
- Mobile offloading
- Future directions in mobile offloading

52

Duke UNIVERSITY

52

Next Class: Edge Helping the Cloud

- Why do **cloud computing companies** want to create edge services?
- Why do **telecom companies** want to create edge services?

53

Duke UNIVERSITY

53

Reading Material for the Next Class

- *Commoditization of the wireless industry*
- *Vodafone perspective on edge computing*

54

Duke UNIVERSITY

54

Homework

- Work on your research project
  - ➢And on your proposal specifically

Duke UNIVERSITY