cio.com
http://www.cio.com/article/2979997/data-center/microsoft-researcher-why-micro-datacenters-really-matter-to-mobiles-future.html?utm_campaign=sflow_tweet#tk.rss_all

Microsoft researcher: Why Micro Datacenters really matter to mobile's future

Bob Brown

Sep 3, 2015 4:07 AM PT

Microsoft Research distinguished scientist Victor Bahl has been spreading the word about Micro Datacenters, also known by the adorable name "cloudlets," as a key concept for optimizing the performance and usefulness of mobile and other networked devices via the cloud. Service providers have embraced this vision most strongly from the start, but it won't be long before enterprise IT pros will likely do the same, Bahl says.

Here's a more in-depth look at the What, Why and When of mDCs:

I notice that a lot of the research you're involved in includes not just mobility, but the cloud. Are the two inextricably linked going forward?

Yes, absolutely. We expect more and more from our mobile devices but despite major advances in technology, resource poverty continues to limit the type of applications we are able to run on these devices. Constraints such as battery life and bandwidth are fundamental and not simply a temporary limitation of current technology. To break free of such perennial problems, we must build technology that enables mobile users to seamlessly use nearby computing resources, which have a persistent high-quality connection to the cloud. This way, users can get the benefits of cloud computing without incurring the wide area network delays and jitter that are common in today's Internet. If we do this, crisp interactive response from immersive mobile applications that augment human cognition will be easier to achieve.

You delivered a keynote address at IEEE WCNC earlier this year titled "Cloud 2020: The Emergence of Micro Datacenters for Mobile Computing." 2020's not that far off: What are Micro Datacenters and why should enterprise IT pros care about them?

At the core a micro data center, or cloudlet or mDC, is a rack of servers available at thousands of locations around the world, never more than a few milliseconds away from the client devices. The mDC is connected to the "classical" mega data center i.e. the cloud via a low-latency, high-bandwidth Internet connection. The software running in these mDCs supports multi-tenancy, which means different services from different providers can be supported simultaneously. Client devices including smartphones, wearable computers, and other IoT devices can use an mDC both as a computing resource as well as a caching resource. This end-to-end arrangement allows developers to build new applications and enhance the performance for existing applications.

As mDCs become popular, enterprise IT pros will most likely have to deploy them on premises as cloud accelerators for the services their users depend on. As users come to expect and rely on the high performance and new applications enabled by mDCs, IT pros will be expected to monitor and manage these servers 24x7x365.

What sorts of applications do you envision would be enabled with Micro Datacenters that are impossible with today's mobile architectures?

Well-known services and applications such as Bing search, Office 365, Azure services etc., will all benefit from these cloud accelerators. In addition, any application that requires heavy use of computation (CPU, GPU, memory) and battery will also benefit. For example vision-based applications, video and sensor analytics, speaker recognition, etc. When wireless bandwidth cost is an issue, offloading computation to mDCs will reduce spectrum usage. By

offloading computation, battery life on the end devices will improve as they will do less work, assuming that the energy cost of computation is more than the communications cost.

How drastically do you expect our current mobile tools — smartphones, tablets, early smart watches — will change by 2020?

Wearable computers will be the rage. Applications that benefit from real-time data analytics will be pervasive and vision applications that augment human cognitive abilities will be on the rise. Generally, I believe mDCs will open the door to a new world of disaggregated cloud computing, which will improve the performance of new IoT services and current cloud services.

What role would Microsoft play in supporting such Micro Datacenters? Would the company essentially be getting into the Content Delivery Network provider game, or working with such companies?

In Microsoft Research our goal is to invent technologies that help ensure Microsoft's future. Microsoft product groups make decisions of what to ship heavily based on their customer needs and their relentless pursuit towards making Azure the best cloud platform in the world. While I cannot comment on what Microsoft product groups will do in the future, CDNs, mDCs and other similar technologies are all within the scope of what Microsoft does.

Where does the most engineering and research effort need to go in order for us to realize this improved mobile computing world?

There is a lot to be done - for example, packaging mDCs to make it easy to deploy them in places where there may not be much IT support; physical security to protect the data that resides on mDC servers and disks; programing framework, to make it easy for software developers to deploy services and applications on a global-scale mDC infra-structure; also management and support for geo-distributed analytics.

There is already a lot of research, which documents the virtues of deploying mDCs close to the user. In recent years executives of several large multi-national IT and telco companies have picked up on it and are describing the benefits of mDCs / cloudlets in their talks. There is even a new ETSI standardization effort on mobile edge computing and new conferences (including the Mobile Edge Computing Conference and The First IEEE Symposium on Edge Computing) dedicated to mobile edge computing. So this is happening.

How do you anticipate enterprise IT pros' jobs changing as a result of where mobile computing is headed? Are there things you would advise them to do, skills-wise, so as not to fall behind?

The Internet is wonderful at connecting people and services. But, it is best-effort i.e. it is not fast, nor does it provide predictable performance. We are becoming more mobile, are doing more, and are context switching more. IT customers are demanding faster access, faster response time, and immediate analysis of large quantities of data which requires more computation and more storage than ever before. Unfortunately, because of the way our network protocols are designed even a few 10s of milliseconds delay can kill user satisfaction. Unless we "fix" the Internet, of which cellular networking is a big part, the only way I see us solving this problem is by deploying thousands of mDCs or cloudlets globally.

In general management of mDCs will be challenging as company executives expect their IT pros to think though the issues and provide smooth accessibility and operation across the different geographic regions their company operates in, even in places that do not have convenient access to the servers. Also, as mentioned previously, IT pros will have to think about how these servers are physically secured. Getting ahead by understanding what mDCs are capable of, their role in emerging cloud and IoT space; asking the right questions and providing honest-to-goodness requirements to mDC providers will lead to competitive advantage. I expect most IT professionals will embrace cloudlets as part of their portfolio of enterprise services they support. The good news is, many engineers understand the outstanding issues and are working on them tirelessly.

Microsoft Research

Microsoft

emergence of micro datacenter (cloudlets/edges) for mobile computing

Victor Bahl

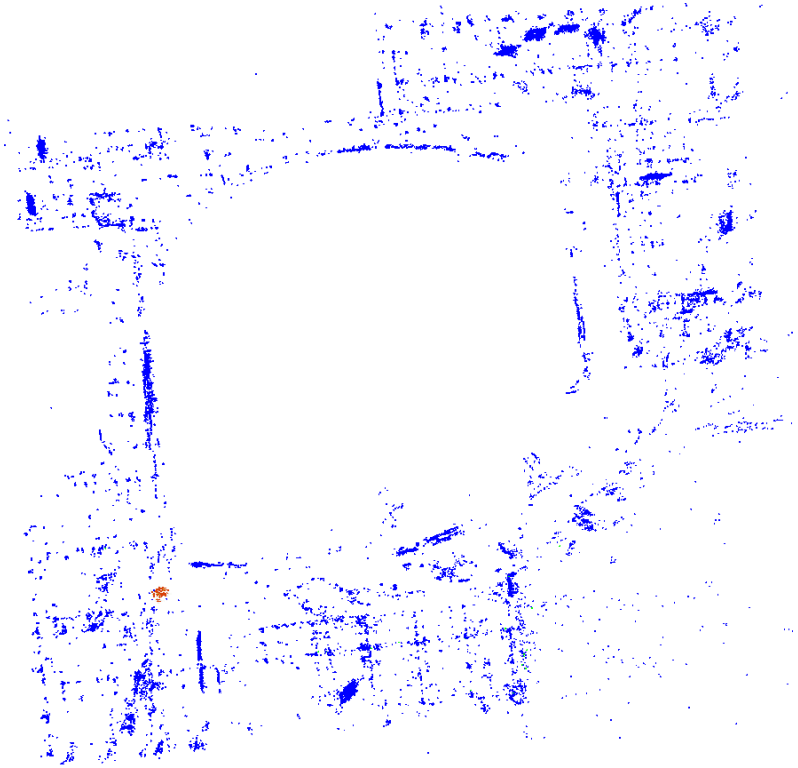
Wednesday, May 13, 2015

what if our computers could see?

Microsoft's's HoloLens

who?

where?



what?

Video credits: Matthai Philipose Microsoft Research

Microsoft

seeing is for real

BBC NEWS LONDON updated at 11:11 ET

Metropolitan Police officers start wearing body cameras

The New York Times

New York Police Officers to Start Using Body Cameras in a Pilot Program

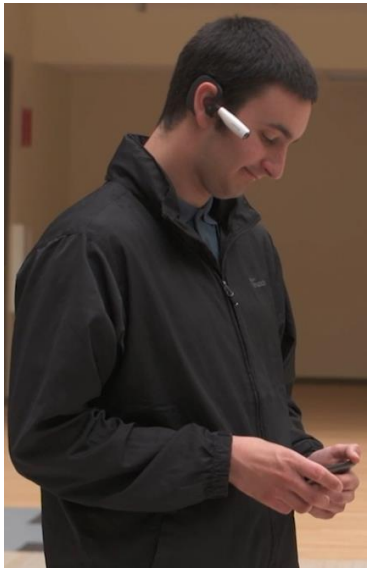
By J. DAVID GOODMAN SEPT. 4, 2014

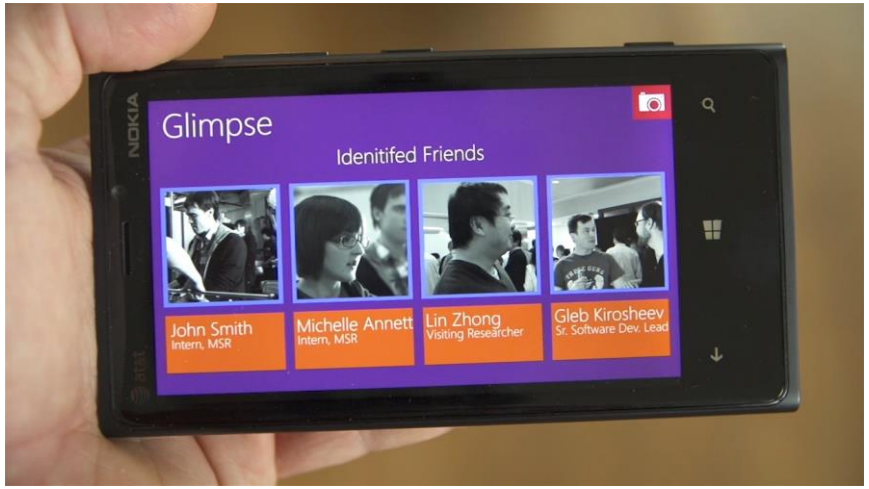
The Washington Post

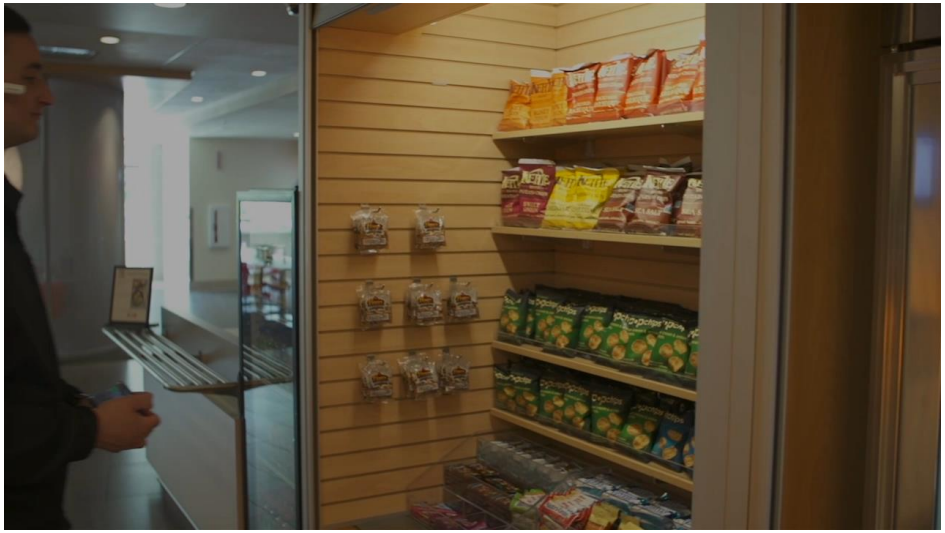
D.C. police will wear body cameras as part of pilot program

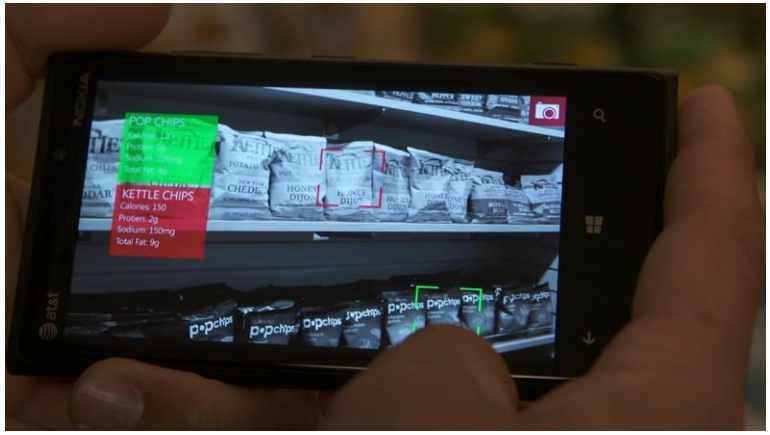
Microsoft

MSR's Glimpse project









Microsoft

vision is demanding

recognition using *deep neural networks*

	face[1] [1]	scene [2]	object[2][3]
memory (floats)	103M	76M	138M
compute	1.00 GFLOPs	2.54 GFLOPs	30.9 GFLOPs
accuracy	97%	51%	94% (top 5)

1: 4000 people; 2: 1000 objects from *ImageNet*, *top 5:* one of your top 5 matches

human-level accuracy, heavy resource demands
… offloading computation is highly desirable

[1] Y. Taigman *et al*. ***DeepFace: Closing the Gap to Human-Level Performance in Face Verification***. In CVPR 2014. **(Facebook)**

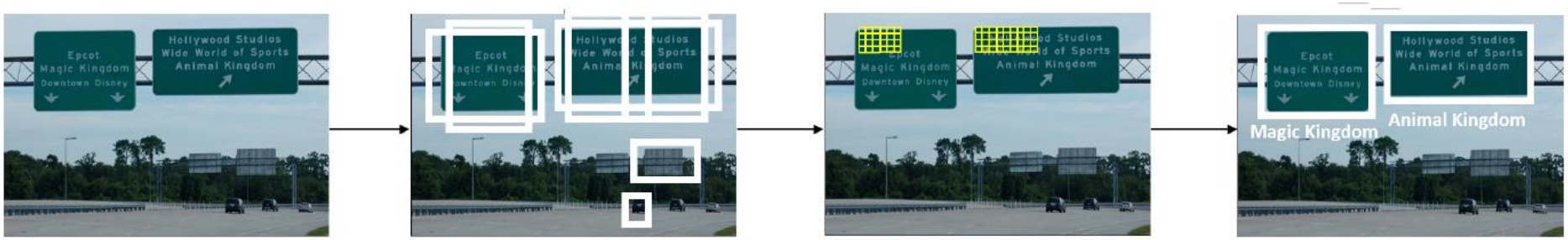
[2] B. Zhou *et al*. ***Learning deep features for scene recognition using places database***. In NIPS, 2014. **[MIT, Princeton, ..]**

[3] K. Simonyan & A. Zisserman. ***Very deep convolutional networks for large-scale image recognition***. 2014 **[Google, Oxford]**

Microsoft

under review

recognition: server versus mobile



Object Detection | Feature Extraction | Object Recognition/Labeling

road sign recognition[1]

stage	Mobile (Samsung Galaxy Nexus)	server (i7, 3.6GHz, 4-core)	Spedup (server:mobile)
detection	2353 +/- 242.4 ms	110 +/- 32.1 ms	**~15-16X**
feature extraction	1327.7 +/- 102.4 ms	69 +/- 15.2 ms	**~18X**
recognition[2]	162.1 +/- 73.2 ms	11 +/- 1.6 ms	**~14X**
Energy used	11.32 Joules	0.54 Joules	**~21X**

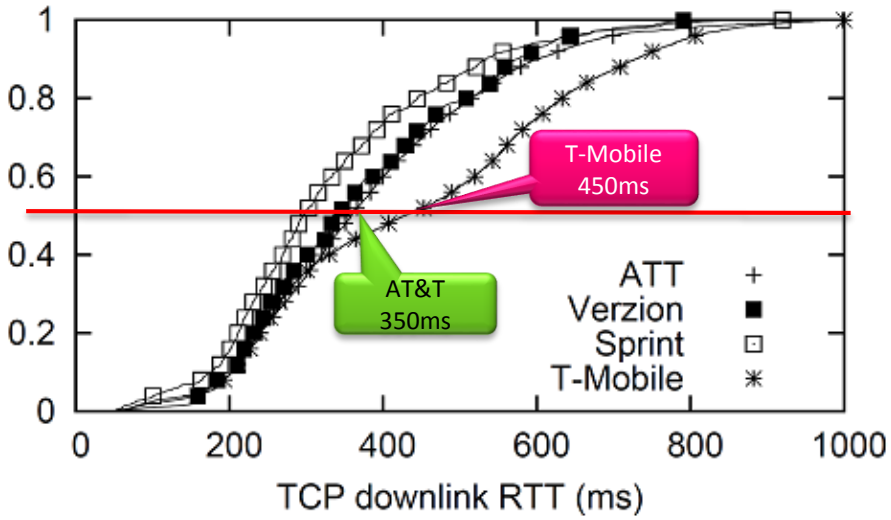
[1]convolution neural networks

[2]classifying 1000 objects with 4096 features using a linear SVM

Microsoft

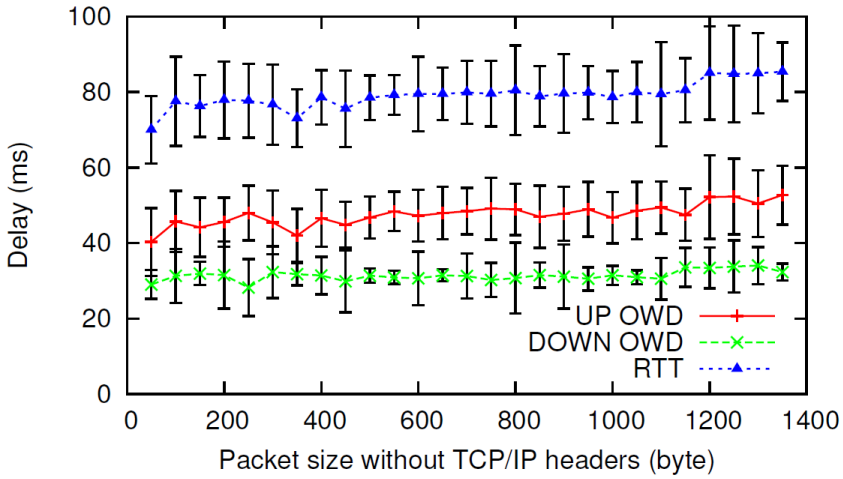
how long does it take to reach the cloud?

3g networks



MobiSys 2010

4g-lte networks



MobiSys 2013

Microsoft

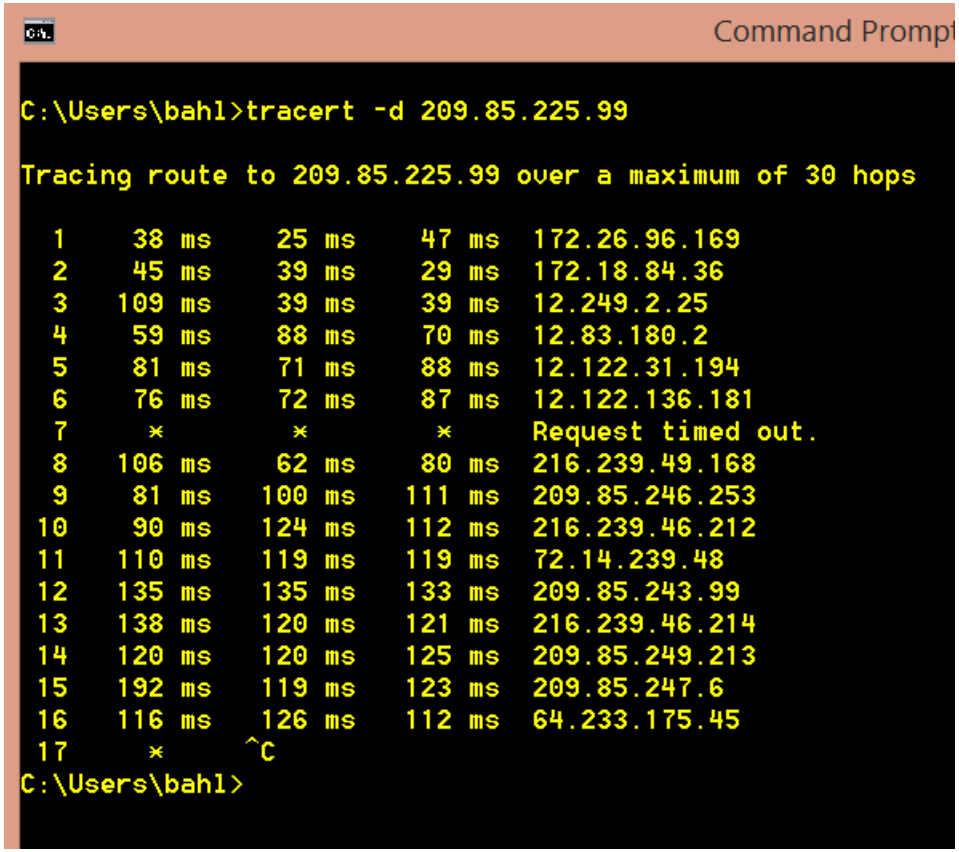
2 years later, we still have latency issues

(May 9, 2015)

major cloud provider A

Data Center	Average Latency
West US	115ms
South Central US	131ms
East US	155ms
North Central US	171ms
North Europe	222ms
West Europe	223ms
Japan West	251ms
East Asia	251ms
Japan East	253ms
Southeast Asia	253ms
Central US	276ms
Content Delivery Network *	276ms
East US 2	287ms
Brazil South	371ms
Australia Southeast	398ms
Australia East	441ms

major cloud provider B

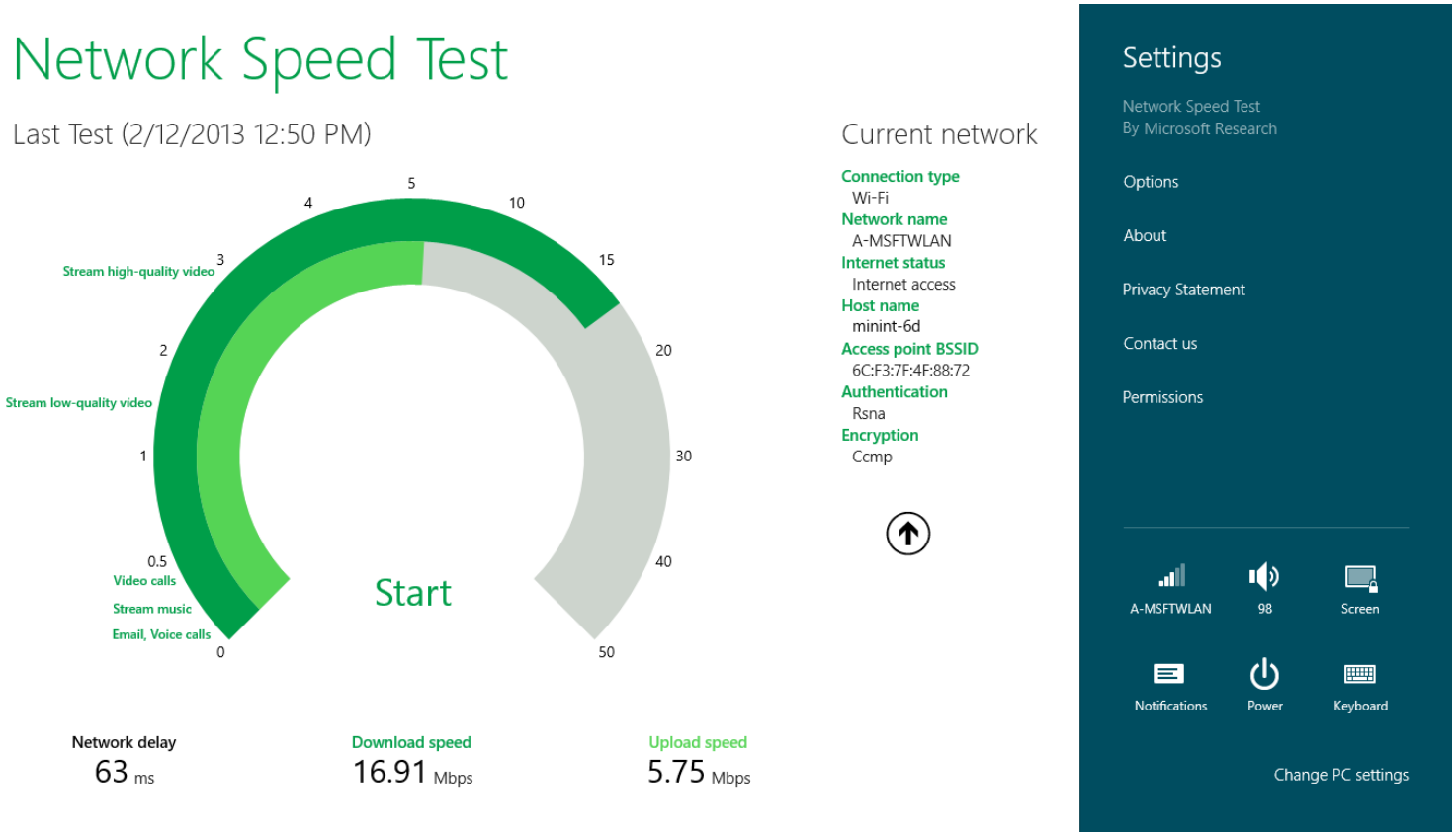


also, http://claudit.feld.cvut.cz/claudit/rtdata.php

Microsoft

Windows MarketPlace

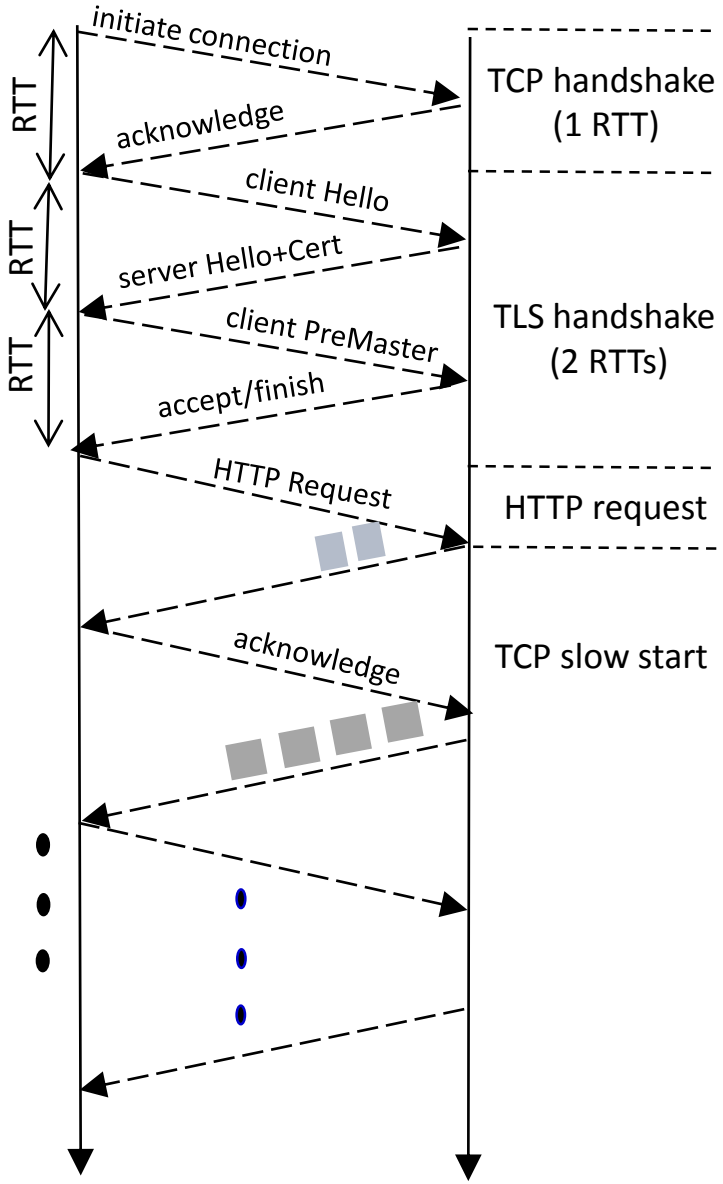
try it out – download Microsoft's Network Speed Test

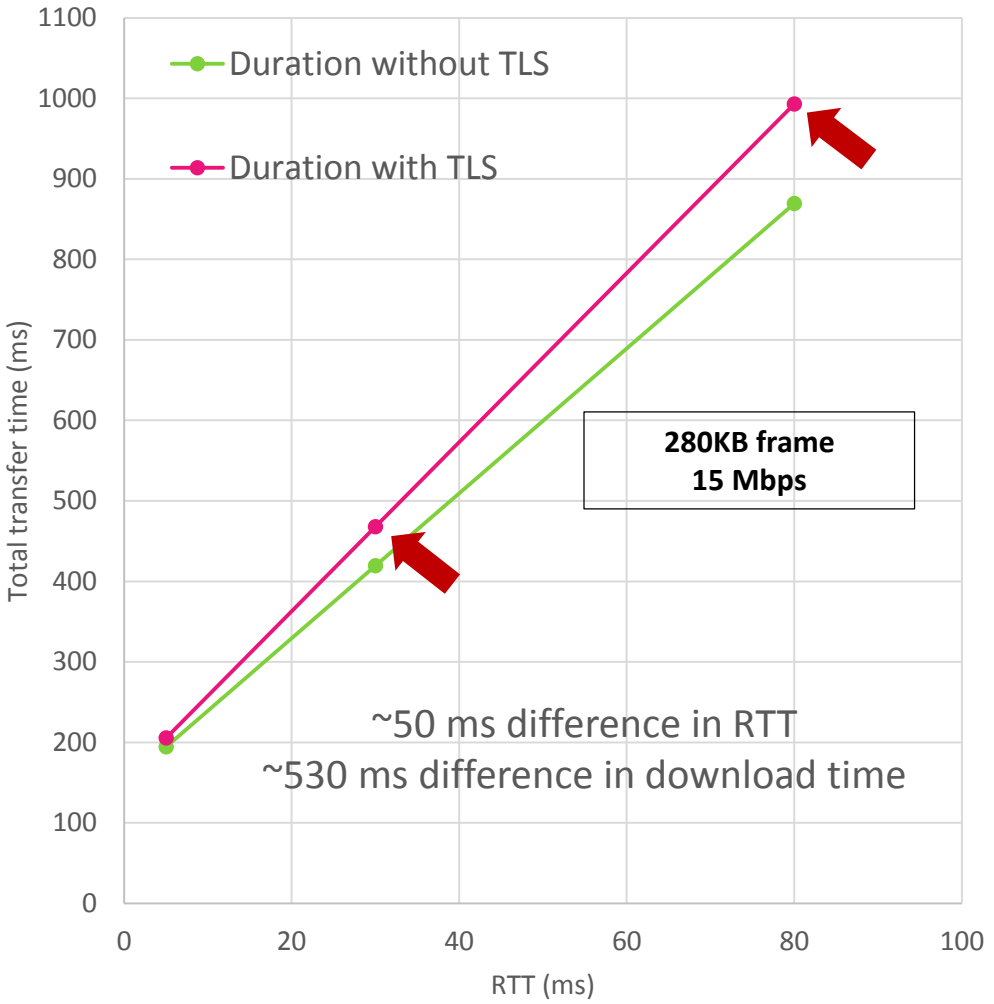


Available on Windows Phone and Windows 8

Microsoft

popular protocols make things worse!

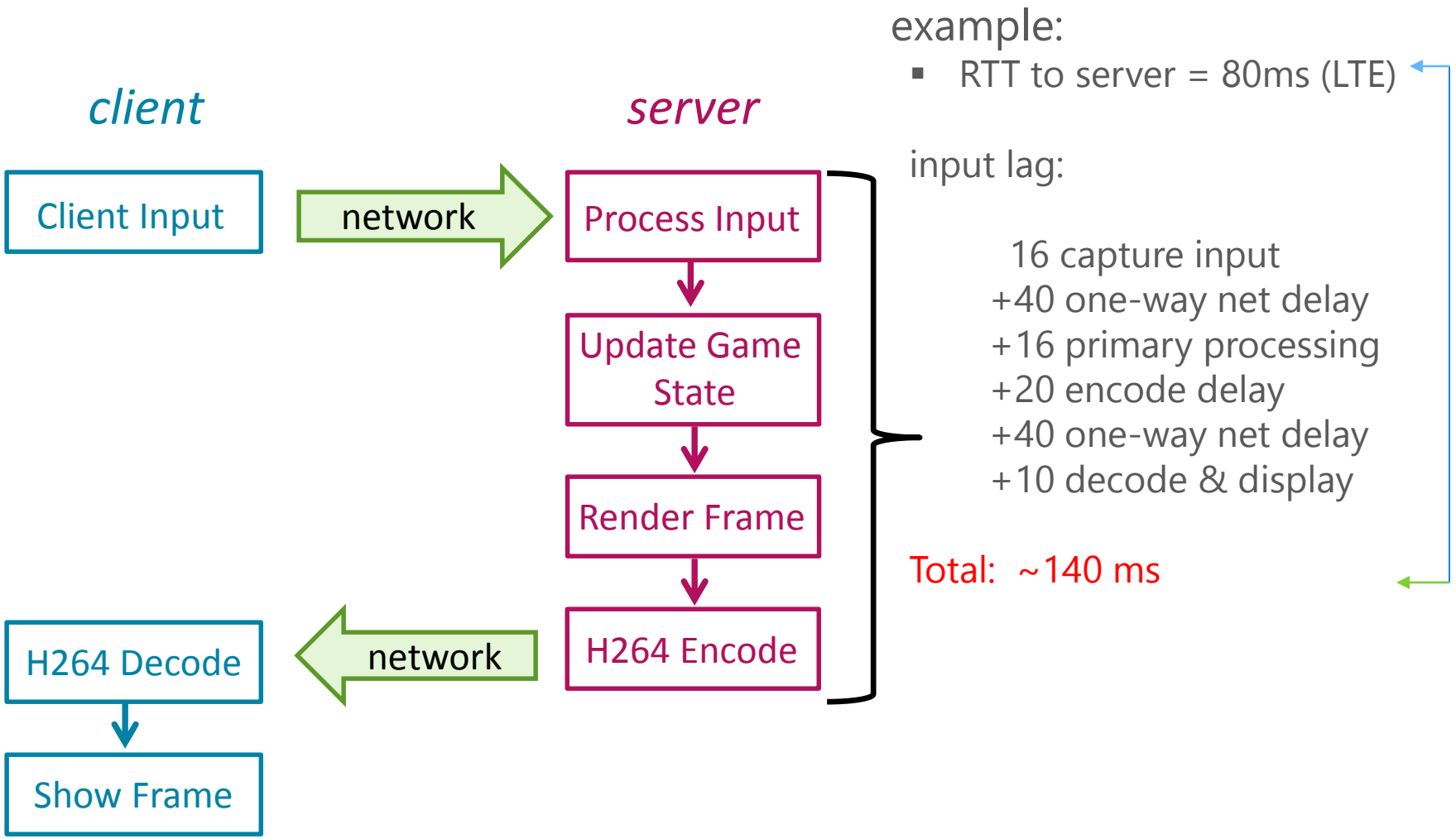




Microsoft

even with UDP - end user impact

fast action cloud gaming



impact of 5, 30 & 80 msec latency

(fast action gaming)

Microsoft

latency matters!

"being fast really matters…half a second delay caused a 20% drop in traffic. and it killed user satisfaction"

- Marissa Mayer @ Web 2.0 (2008)

"…a 400 millisecond delay resulted in a -0.59% change in searches/user",

[i.e. Google *would lose 8 million searches per day - they'd serve up many millions fewer online adverts*]

- Jake Brutlag, Google Search (2009)

"…for Amazon every 100 ms increase in load times decreased sales with 1%"

- Andy King, book author

"…when 50% of traffic was redirected to our edges preliminary results showed a 5.9% increase in click-thru rates"

- Andy Lientz, Partner GPM, BingEdge (2013)

the fact of the matter is …

offloading computation to a resource-rich cloud brings the true power of CS into your hands

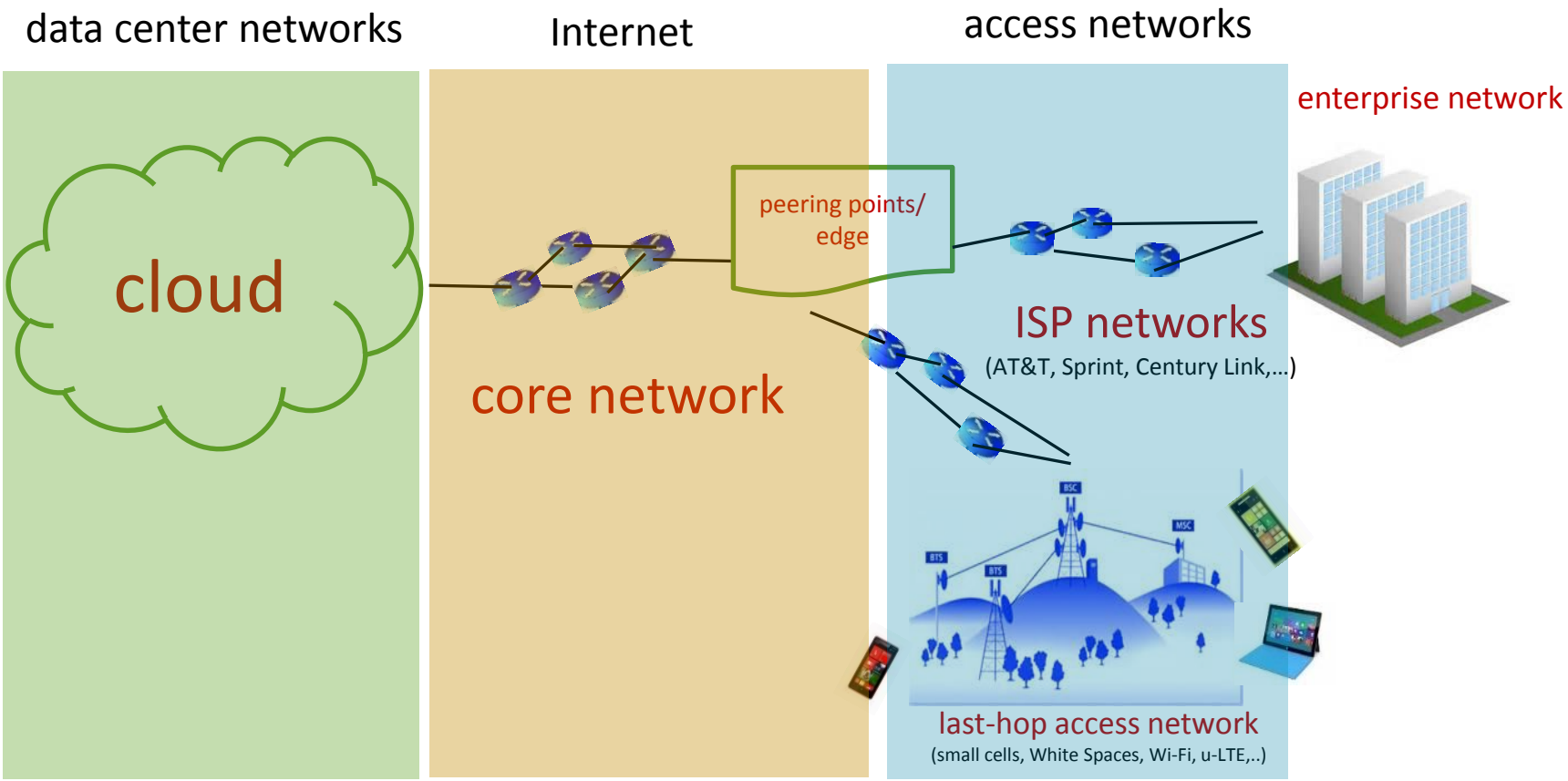
high latency & jitter to the cloud can make cloud services unusable

poor performance impacts revenue and turns users away

… and we have a latency problem

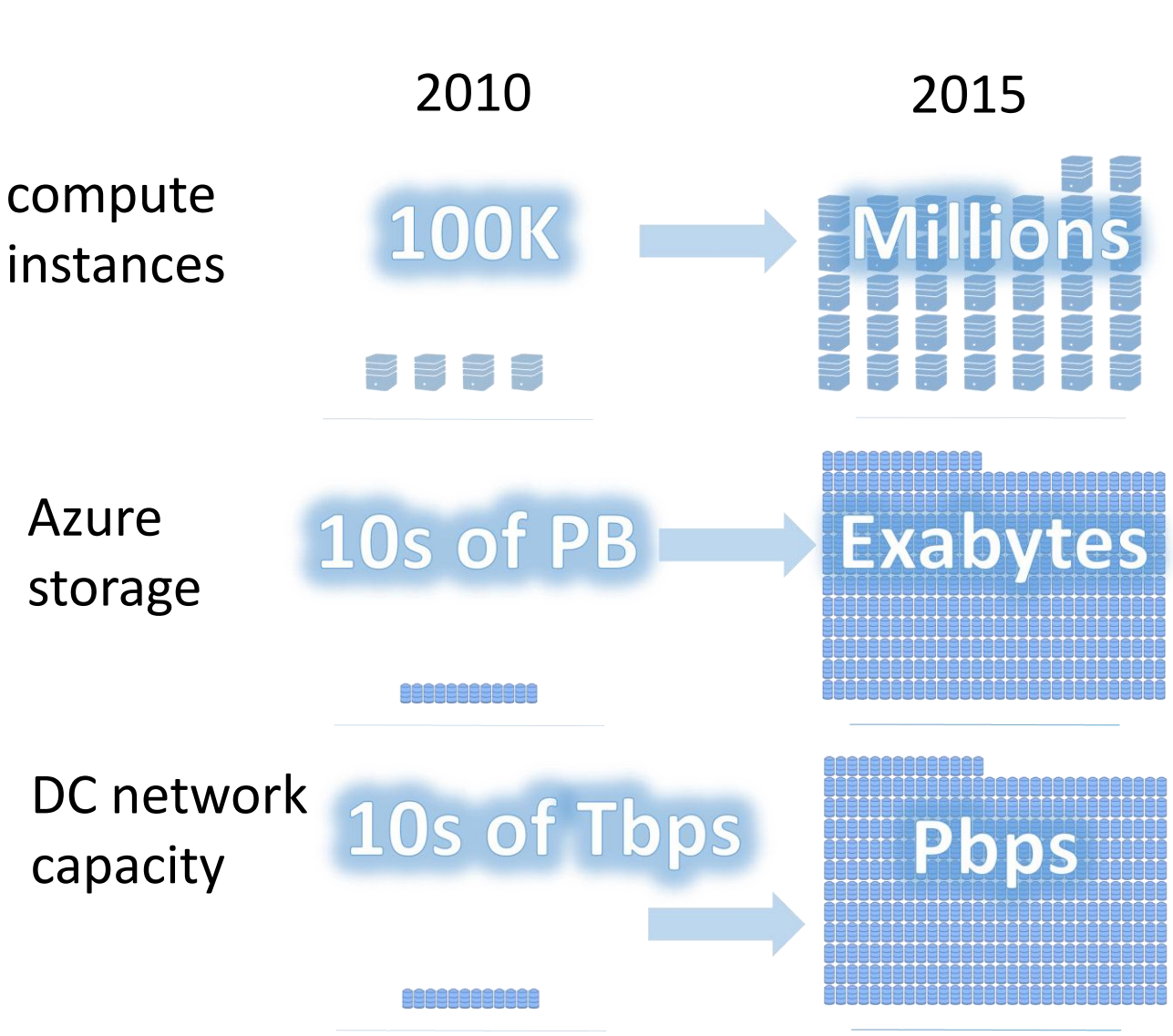
reducing latency

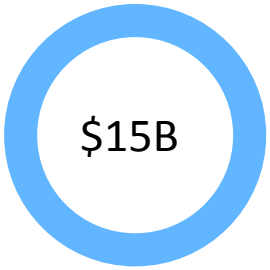
contributors to latency



Microsoft

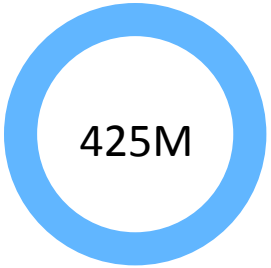
Microsoft's hyper-scale cloud





MICROSOFT CLOUD INVESTMENT

STORAGE TRANSACTIONS IN MARCH 2015

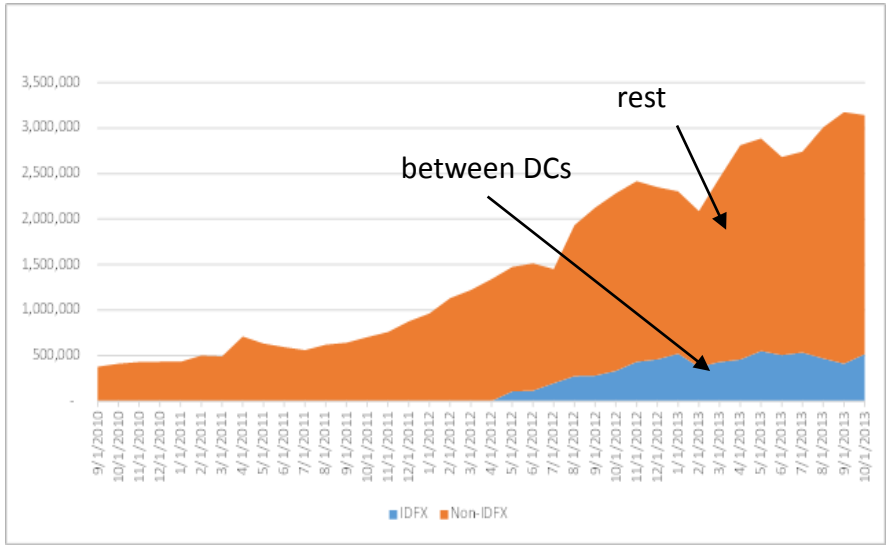


ACTIVE DIRECTORY USERS

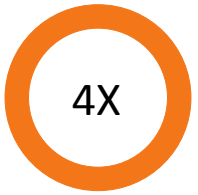
Microsoft

Microsoft's hyper-scale network

Microsoft's network is one of the largest in the world



MILES OF FIBER
(DC & WAN)



WRAP THE EARTH IN
NORTH AMERICAN FIBER

massive traffic growth is stressing the underlying core networks

areas MSR researchers are working on: SIGCOM 2014

performance: significant number of circuits sit idle while others are oversubscribed (latency increases)

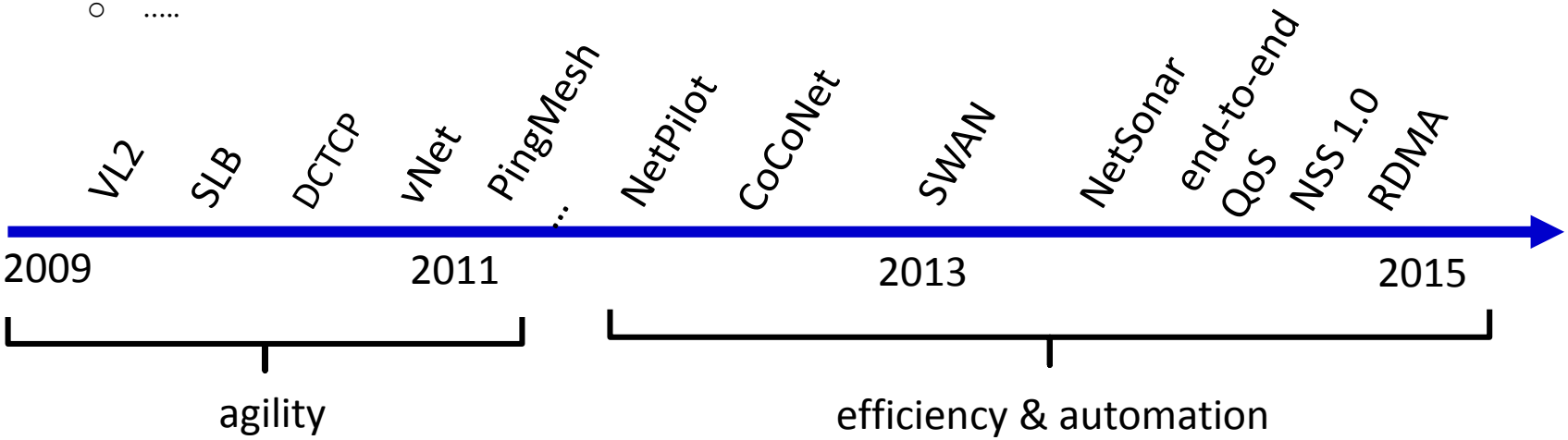
failures: long convergence time during network topology changes with planned and unplanned network events

SIGCOMM, NSDI, CACM, …

MSR's contributions to Microsoft cloud networking & to academia

reseachers worked hand-in-hand with Azure, Bing, Windows, ….

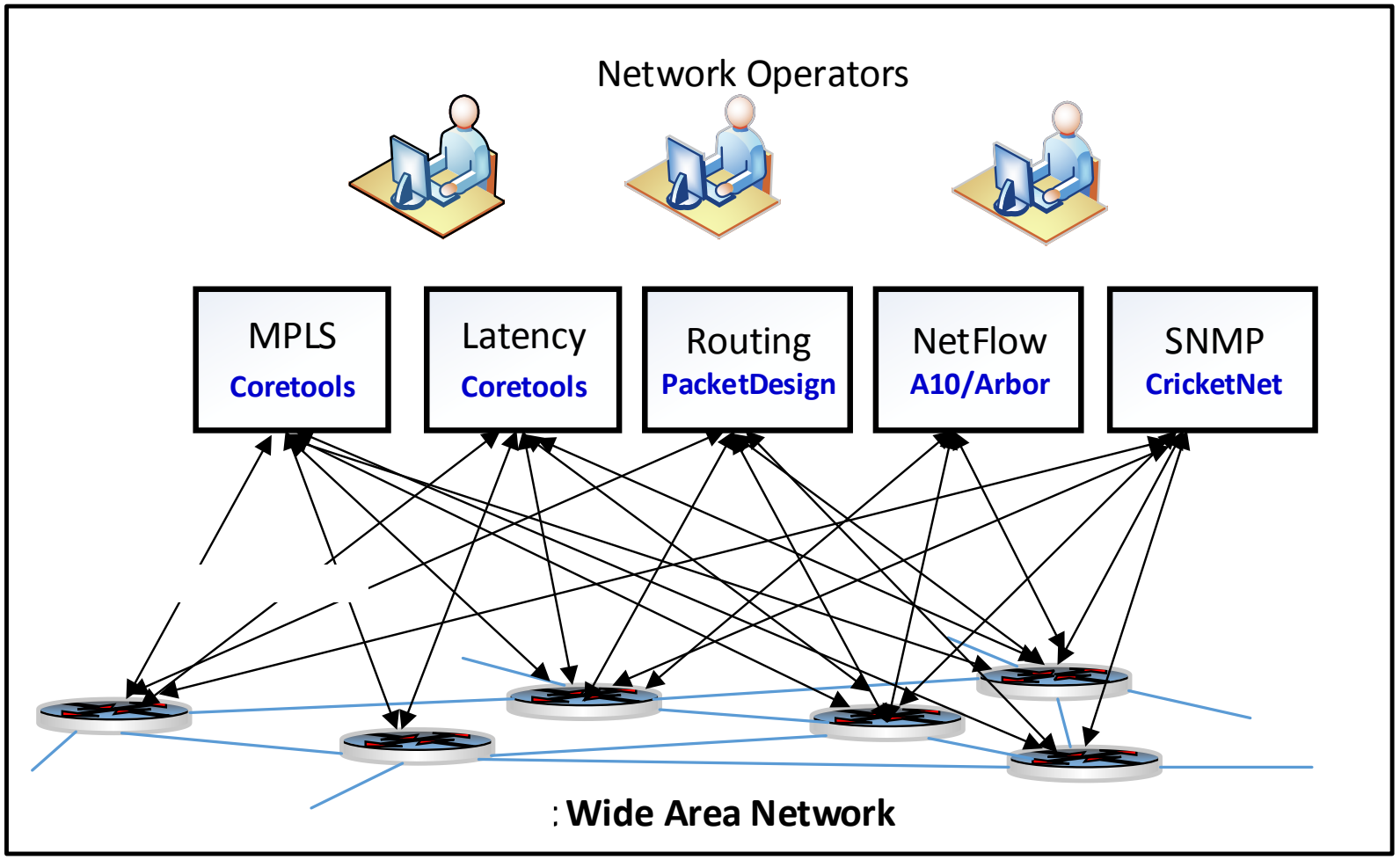
- steady stream of significant tech transfers
  - full –bisection bandwidth (Q10): 80x cost reduction, 20x outage reduction, in all Azure DCs
  - software load balancer (SLB): 15x cost reduction, carries all Azure traffic
  - software-defined WAN: increased inter-DC bw utilization from ~40% to ~95%,
  - virtual networking: enabled MSFT hybrid cloud offering via HyperV virtual network product
  - …..



- plenty of research accolades as well
  - papers recognized as "Research Highlight" by ACM

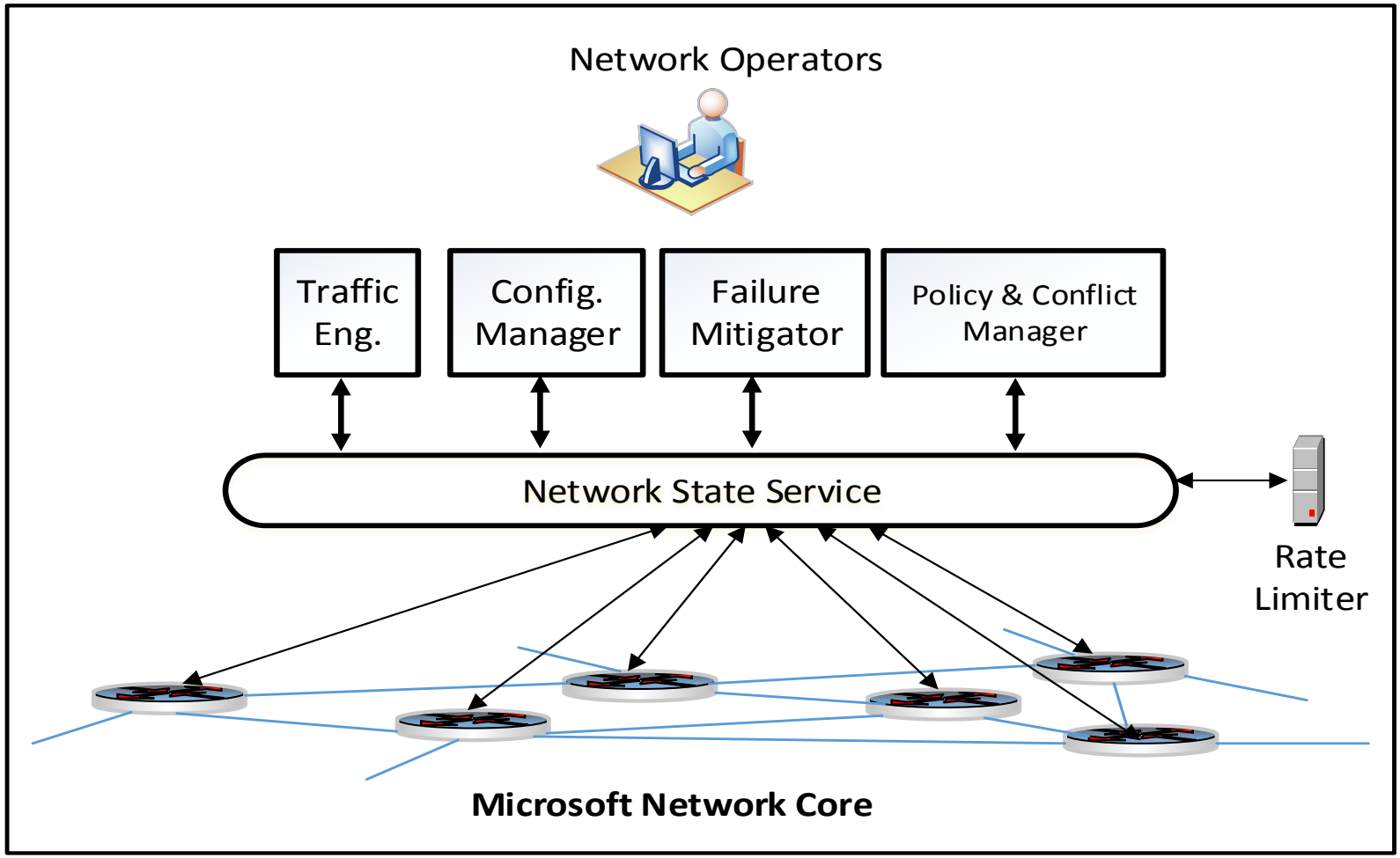
Microsoft

improving efficiency of wide area network



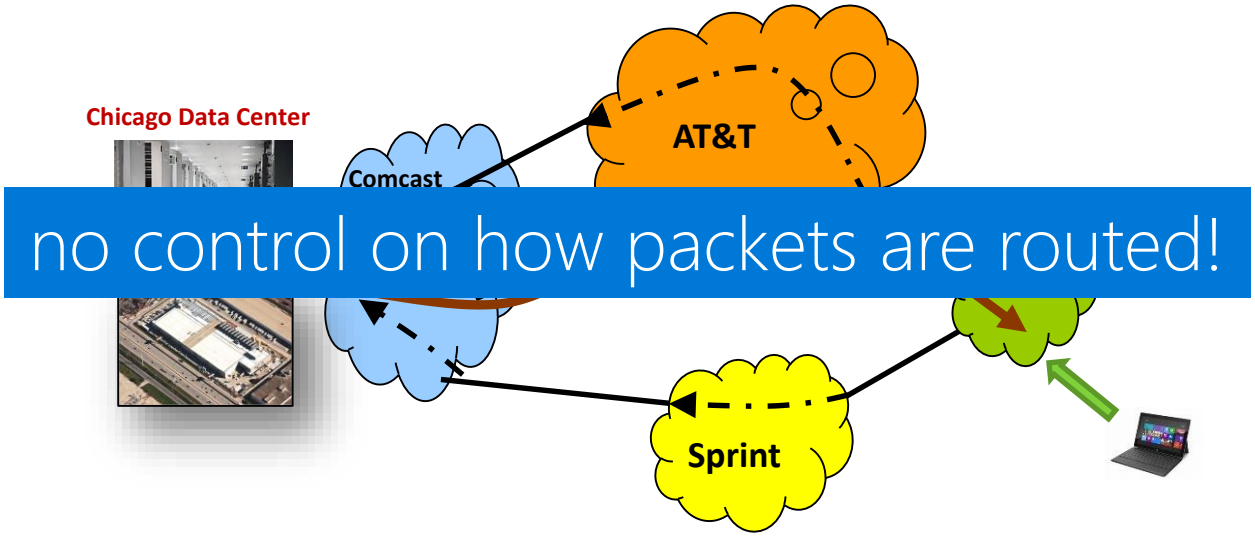
improving efficiency of wide area network

with MSR's network state service



Internet: a network of networks of networks

a collection of many autonomous systems (AS) managed by many ISPs with complex peering relationships



as of March 6, 2013 (source: PEER 1)

- 22,961 AS numbers (AS numbers uniquely identify networks on the Internet, e.g. 8075 for Microsoft)
- 50,519 peering connections

… but we can reduce latency further

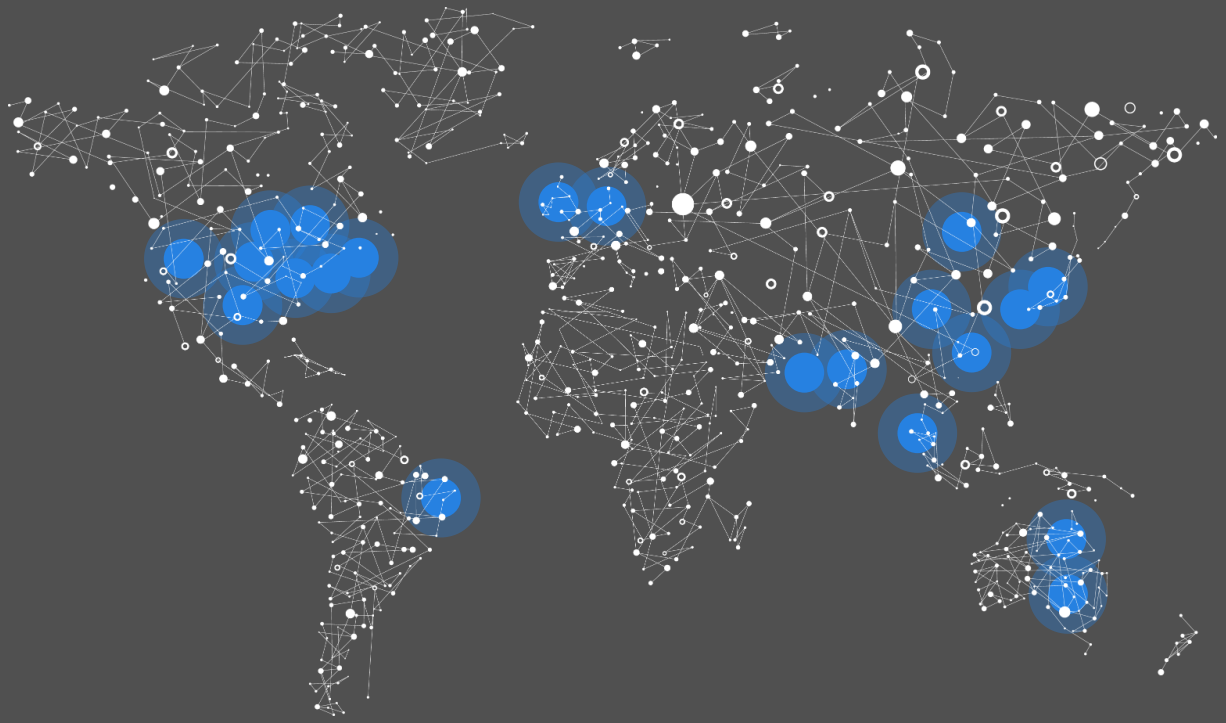
get the packets under our control as soon as possible

how?

- bring the cloud closer to the end-user
  - ✓ build lots of DCs around the world & place them in strategic locations

bringing the cloud closer

build lots of hyper-scale data centers around the world



Azure compute regions

19

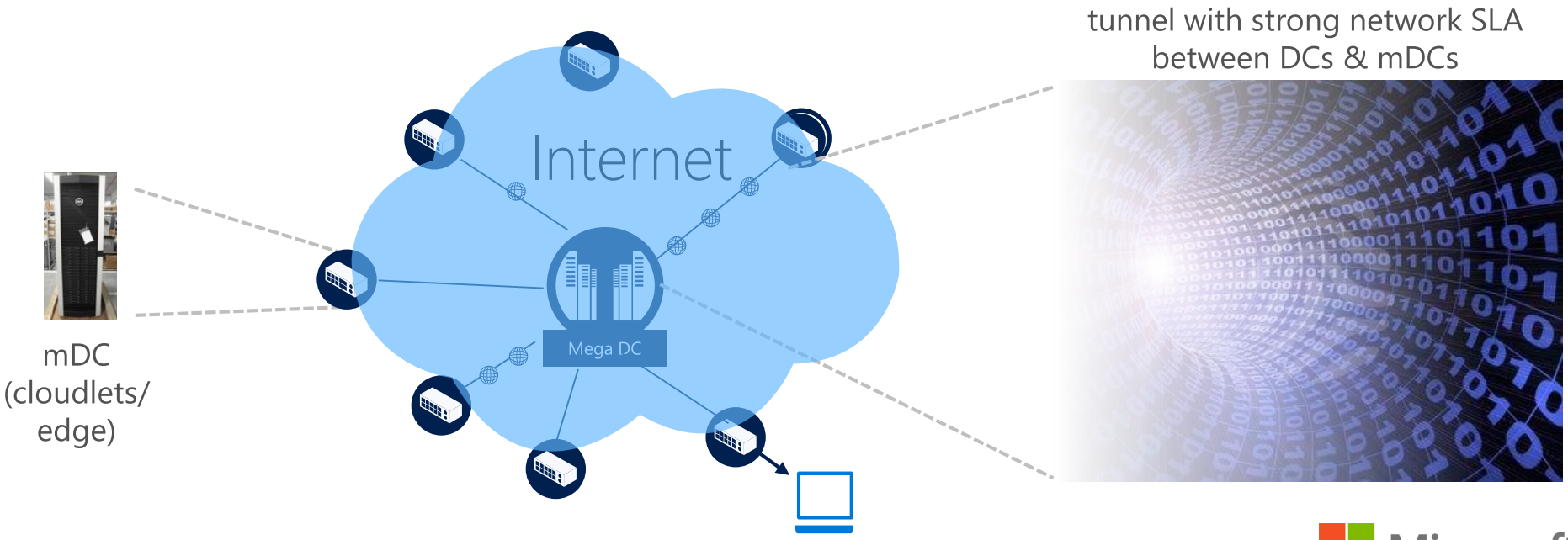
Azure compute regions open today

more than AWS and Google cloud combined

is building hyper-scale data centers enough?

no, it's capital intensive and expensive to operate

smarter approach: build an extensive infrastructure of micro DCs (1-10s of servers with several TBs of storage, $20K-$200K/mDC) and place them everywhere



Microsoft

micro DCs

site acceleration (classic)

content caching

- Xbox videos, NetFlix videos, Windows updates,…

split TCP connections

- from Bing data, on avg. can reduce latencies by ~30 msec
  - predictive search query responses improved ~25-35% based on random sampling before and after deploying edge serves in a couple of US cities

Akamai
Limelight
CloudFront
Level 3
EdgeCast
Rackspace
⋮

mDCs are “classic” CDNs nodes, that can improve the performance of search engines, office productivity tools, video and audio conferencing & future cloud services

Microsoft

additional benefits of mDCs

latency reduction

- ✓ serve static content immediately
- ✓ SSL termination / split TCP
- edge to DC protocol enhancements

bandwidth saving

- ✓ compression
- procrastination
- edge analytics

service & internet monitoring

reliable connectivity

- overlay networking
- path diversity

battery saving

- computation offloads
- client proxying

high-end game streaming

- lower device cost
- reduce developer fragmentation

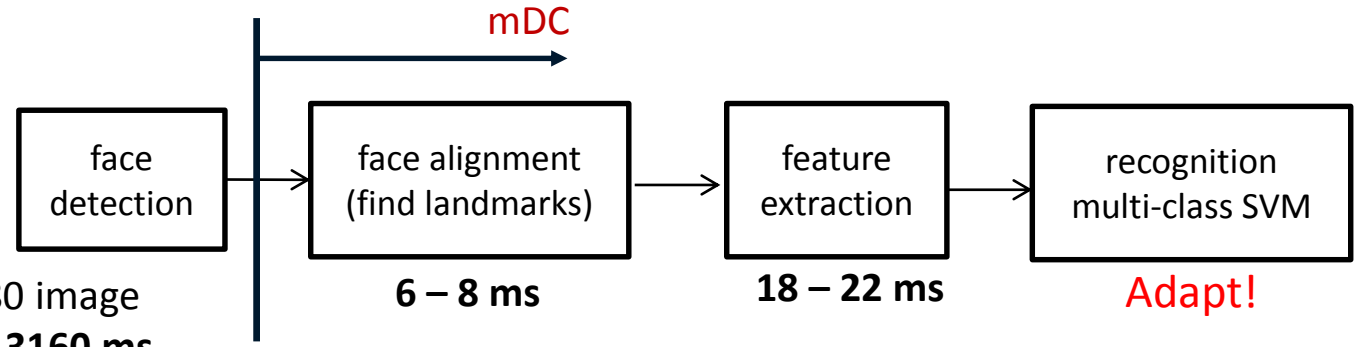
new services

protection against DoS

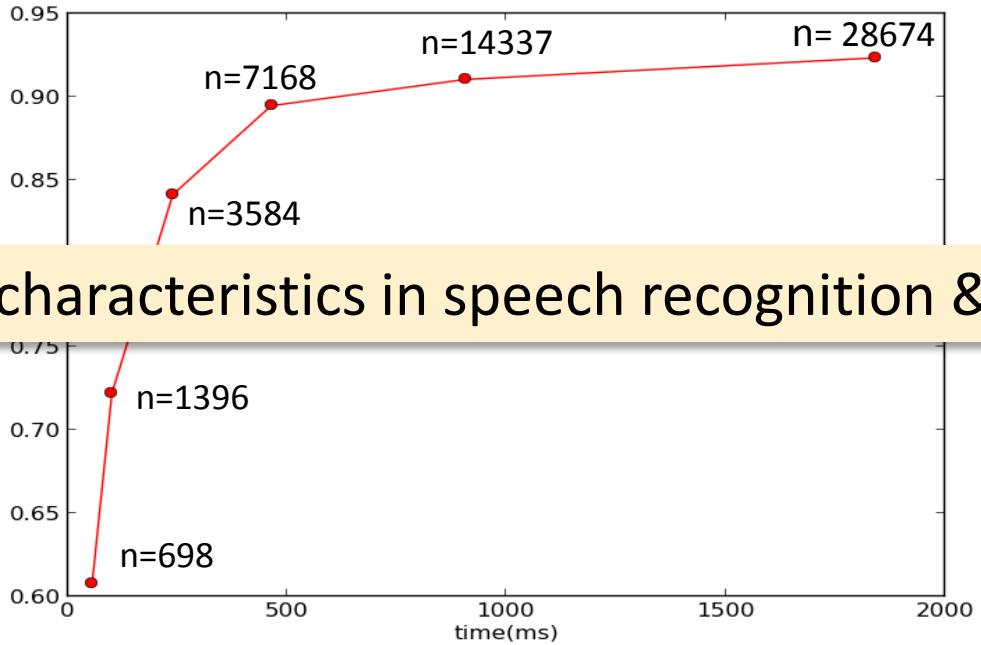
reduced load on DCs

new services: object recognition

the lower the latency, the better the results



For a 640x480 image
client: 890 – 3160 ms
server: 72 – 115 ms



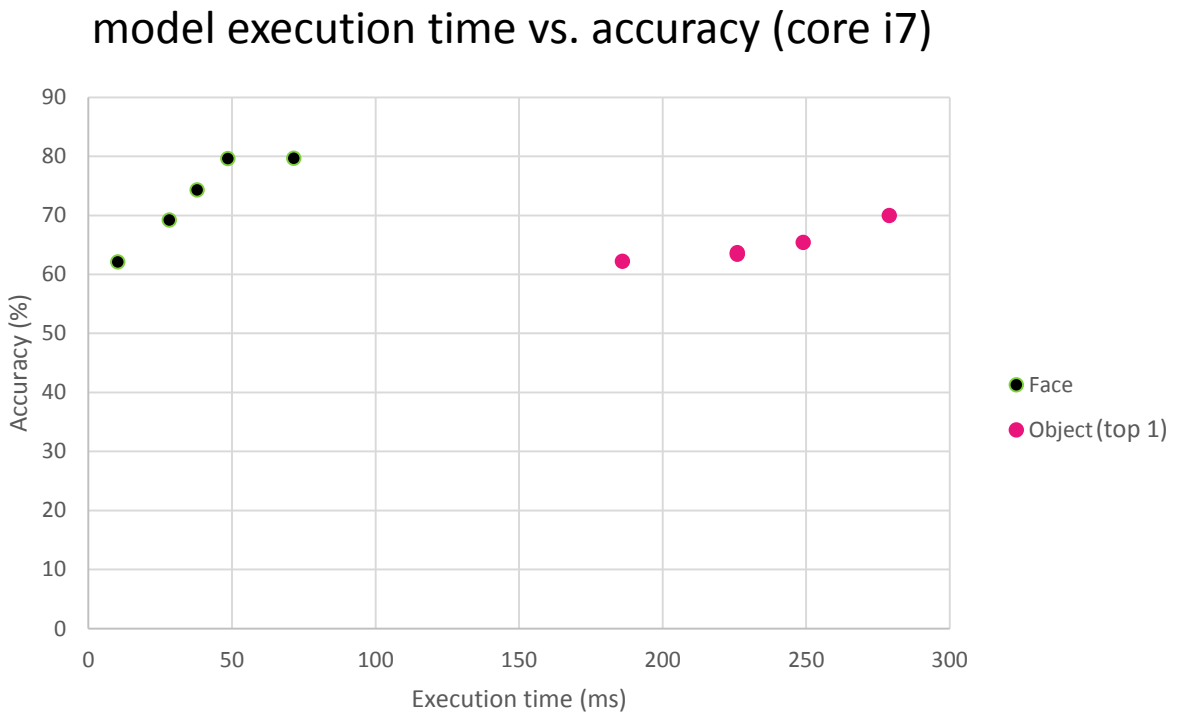
similar characteristics in speech recognition & search

Face prediction Time

Microsoft

under submission

using DNNs - similar results - lower transport latency helps

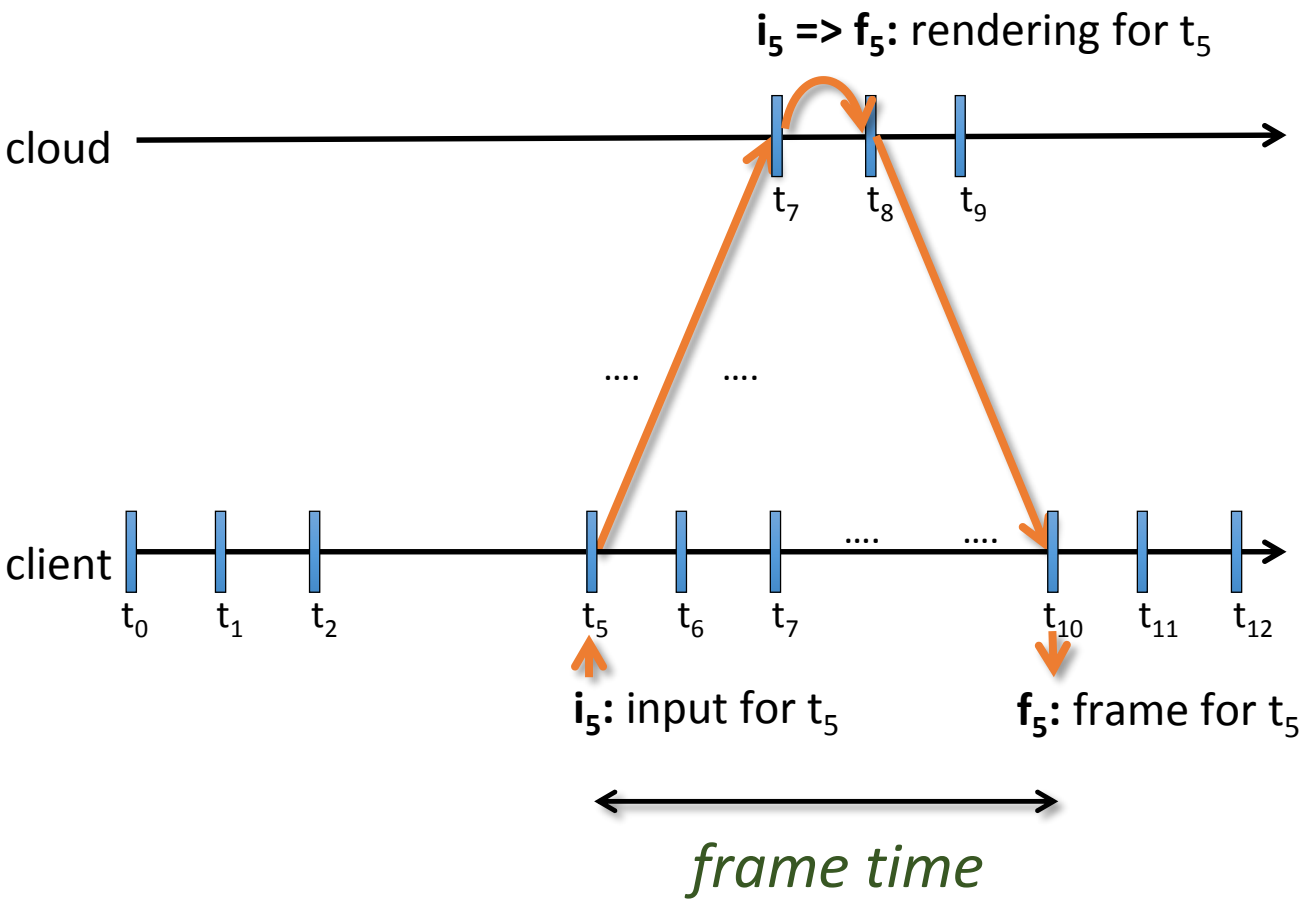


50-100ms can allows ~10-20% more accurate model

Microsoft

face recognition with mDCs

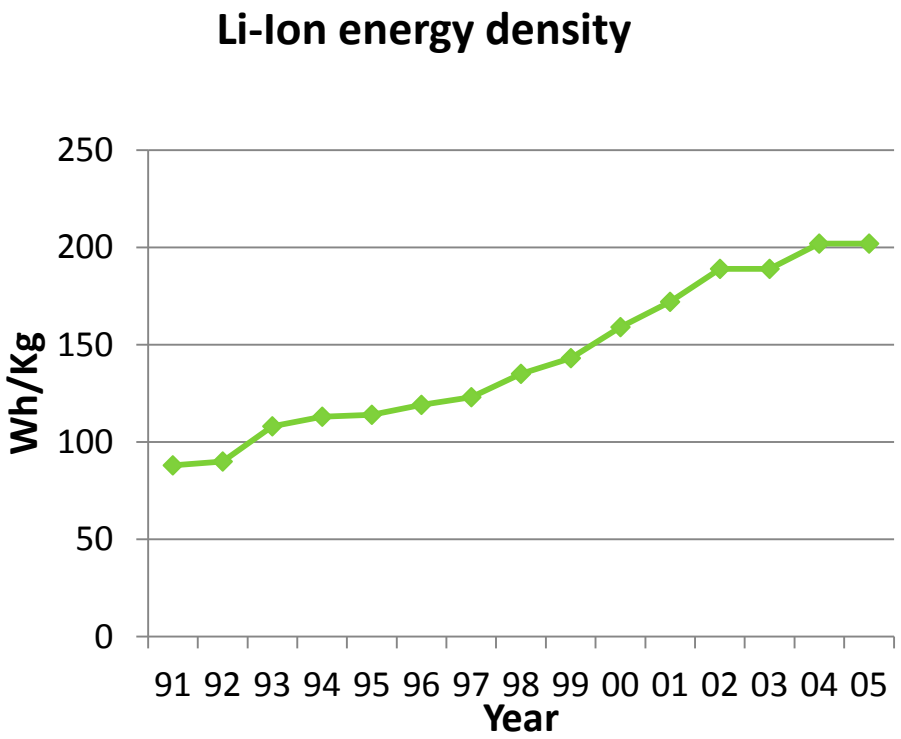
(new) service: cloud gaming



cloud gaming
(with speculative execution)

battery life…

silver bullet seems unlikely



lagged behind

- higher voltage batteries (4.35 V vs. 4.2V) – 8% improvement
- silicon anode adoption (vs. graphite) – 30% improvement

trade-offs

- fast charging = lower capacity
- slow charging = higher capacity

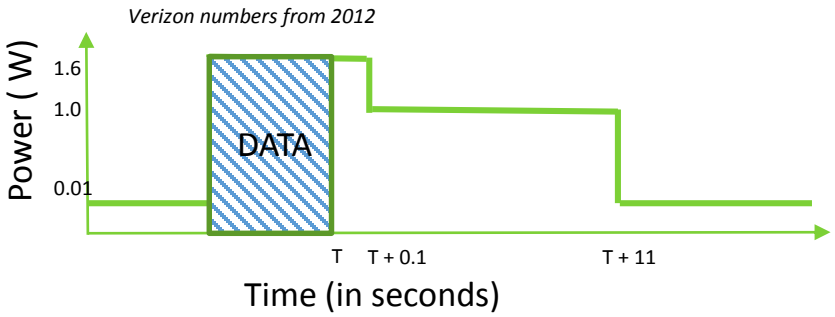
contrast with

CPU performance improvement during same period: 246x

Microsoft

battery use in SmartPhones…

LTE consumes > 1.5W when active
LTE chip active for ~10 secs of extra tail time (1W power)



….but how did we get here

Microsoft

a bit of context/history… 6 years ago

The New York Times

Customers Angered as iPhones Overload AT&T

By JENNA WORTHAM
Published: September 2, 2009

The New York Times

DIGITAL DOMAIN

AT&T Takes the Blame, Even for the iPhone's Faults

By RANDALL STROSS
Published: December 12, 2009

PCWorld

Report: AT&T Reputation Tarnished by iPhone Flaws

By Tony Bradley, PCWorld Dec 14, 2009 2:01 PM

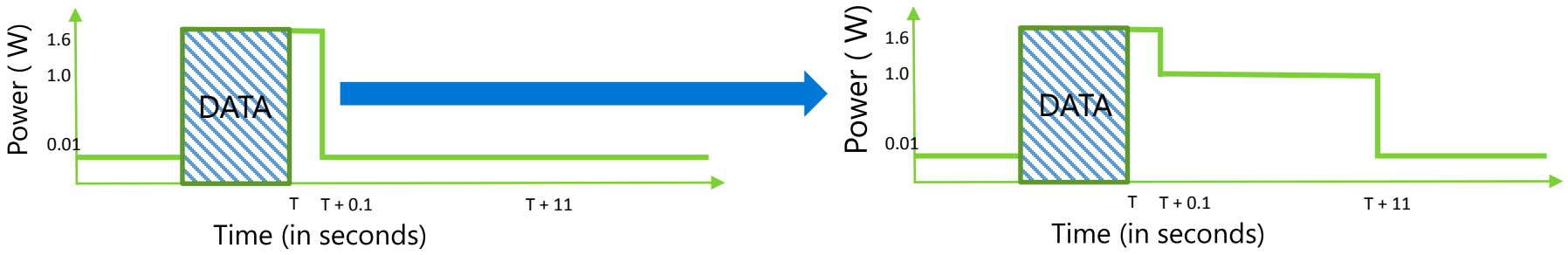
original design:

bring radio to low power state immediately

mobile operator requirement:

keep LTE chip **active for ~10 sec**. of extra tail time
(to reduce the signaling load)

Verizon numbers from 2012

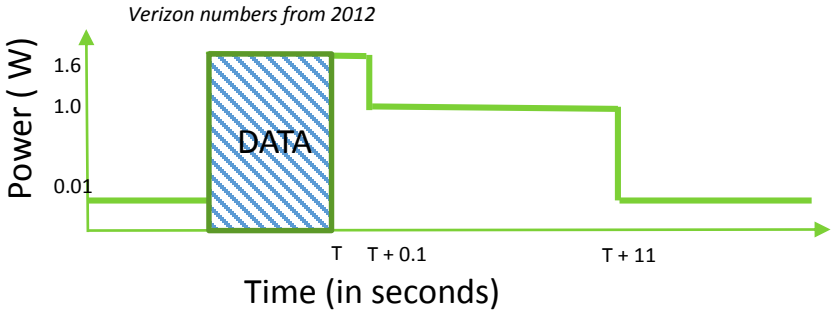


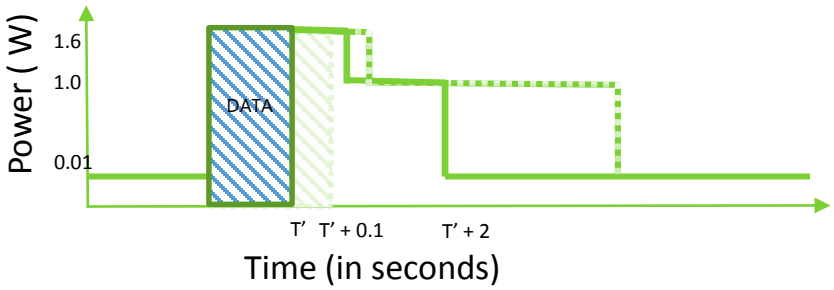
Microsoft

mDCs can increase use time

LTE consumes > 1.5W when active
LTE chip active for ~10 secs of extra tail time (1W power)

with mDCs:
faster transfers => less time in high power state
aggressively enter lowest power state





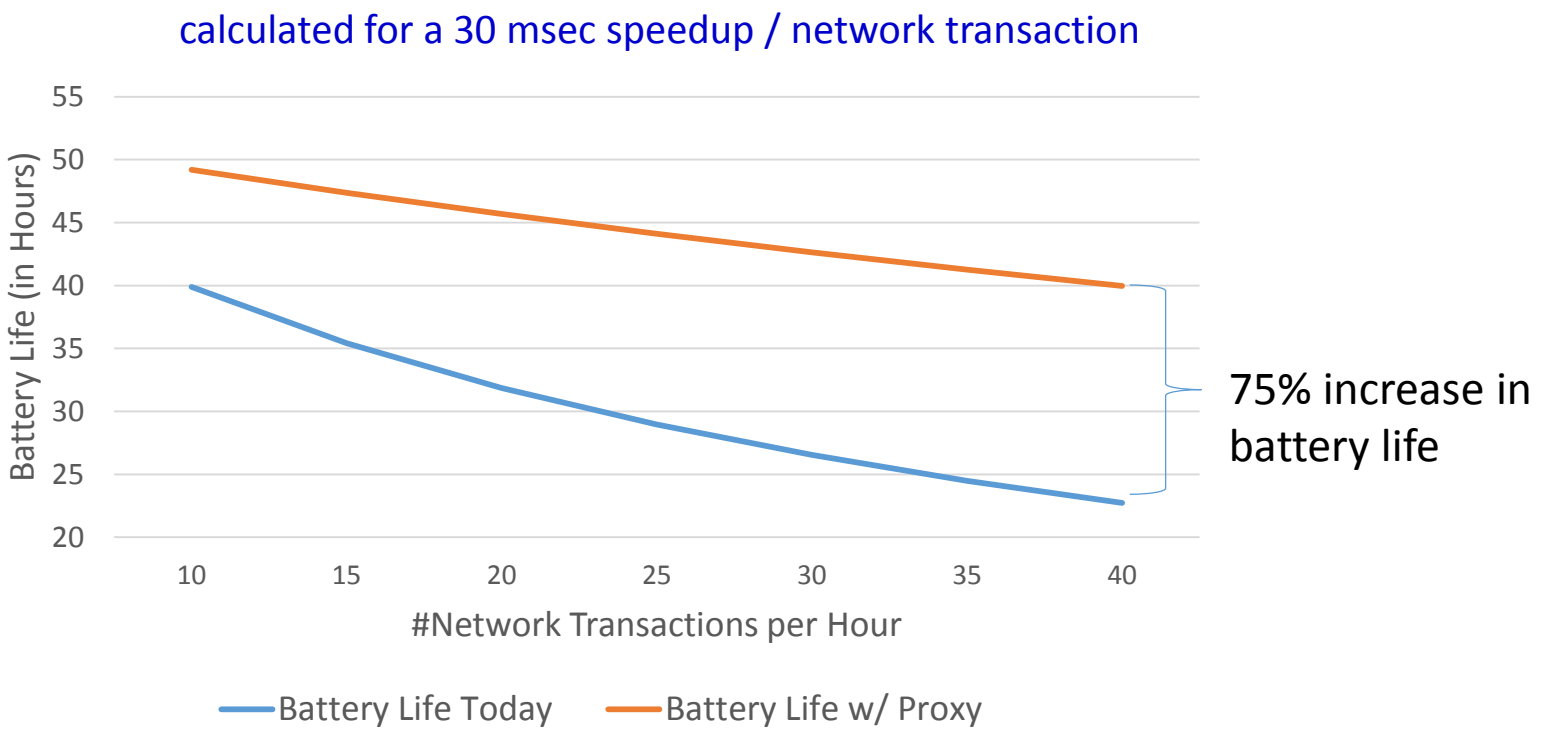
Energy savings / transfer: 1.6W*speedup + 1W*9sec = 10.6J (assuming speedup of 1 second)

for 20 network transfers/hour (notifications, email, etc.), with 1 sec speedup
total energy savings per 24 hr. = 6624 J
→ Saving of **26%** in a 1500 mAH cell phone battery*

* Samsung Standard LI-ION battery with rating of 1500mAh/3.7Vdc

Microsoft

especially good for mobile
battery life improvement



these types of saving occur across the board for all battery types and all types of mobile devices

* Samsung Standard LI-ION battery with rating of 1500mAh/3.7Vdc

Microsoft

saving bandwidth….

security, traffic, tracking

locating objects of interest

customer queue analytics

current approach

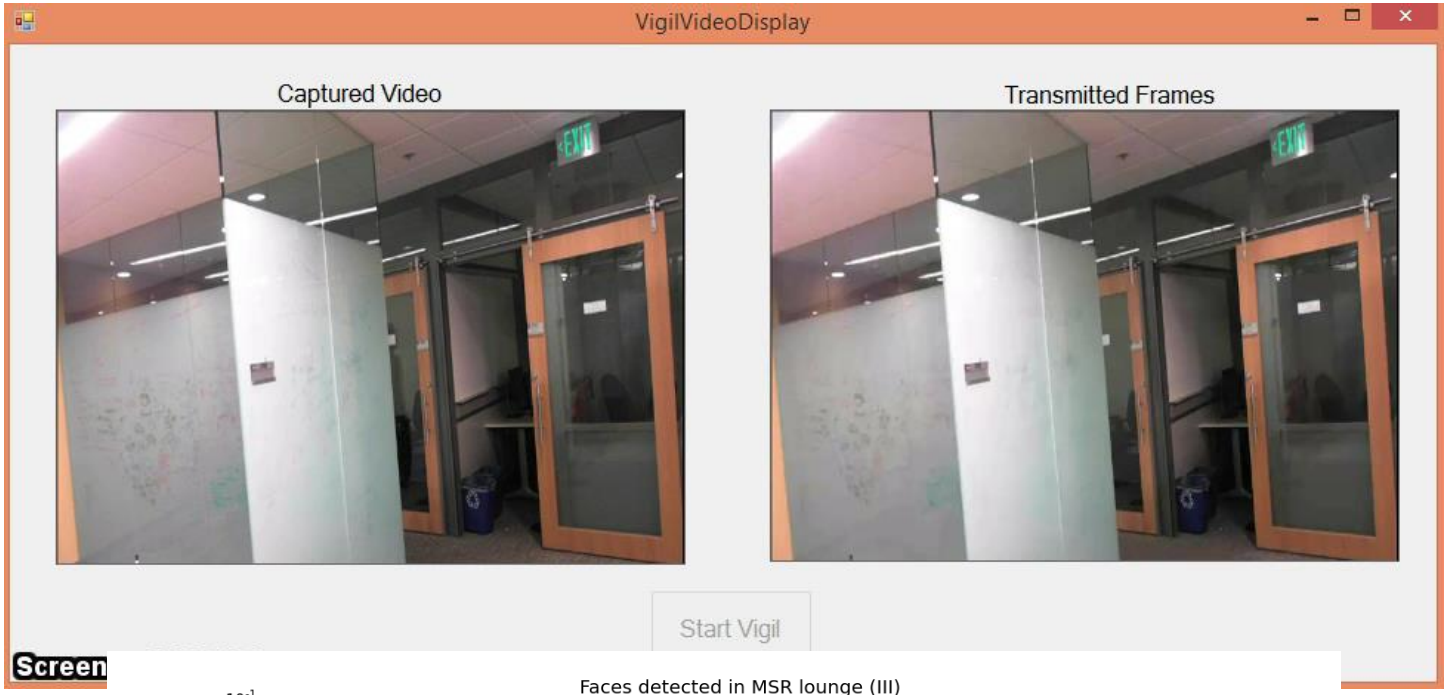
- upload the captured video to the cloud for remote analysis

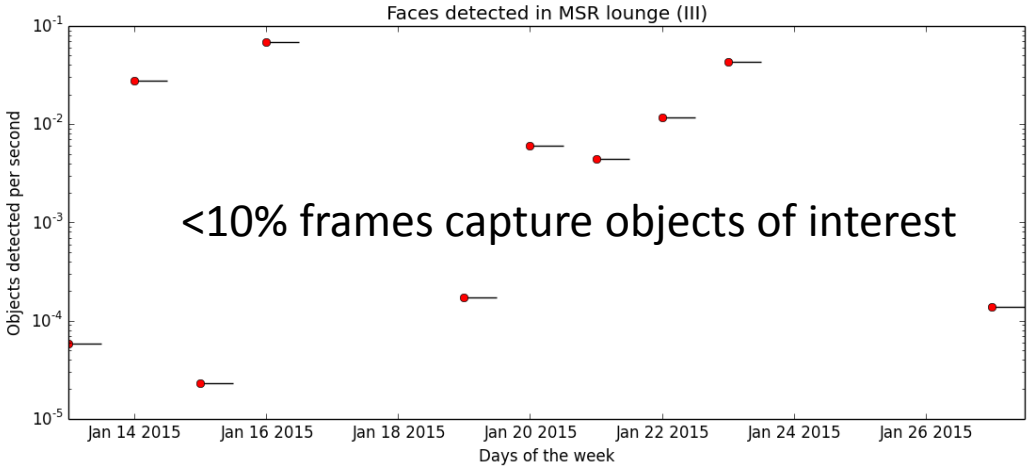
observations

- too much data captured per hour (>10GB/hour)
- bandwidth limits scale and use of system
- unable to support near real-time tracking & security

Microsoft

saving network bandwidth
(wireless video surveillance)



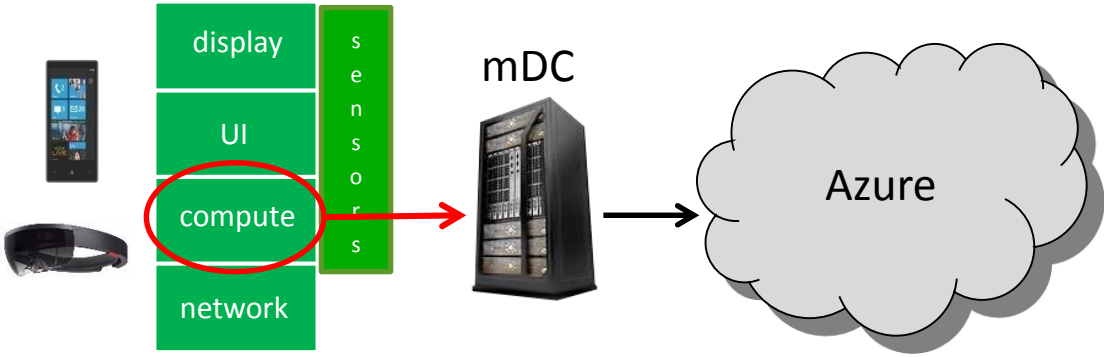


saving network bandwidth
(parking spot detector)

a couple of on-going problems

offloading computation

remote execution reduces energy consumption and improves performance



challenges

- what to offload?
- how to dynamically decide when to offload?
- how to do so with minimum programmer effort?
- how to support multi-tenancy with bullet-proof privacy?

Microsoft

programming frameworks for cloud offloading

	Microsoft's MAUI	Intel's CloneCloud	USC's Odessa
remote execution unit	methods	threads	tasks

- MAUI exploits .NET framework to dynamically partitioning & offload method execution [MobiSys'10]
- CloneCloud supports existing applications, but requires tight synchronization between cloud and phone [EuroSys 2011]
- Odessa creates a data-flow graph to exploit parallelism [MobiSys 2011]

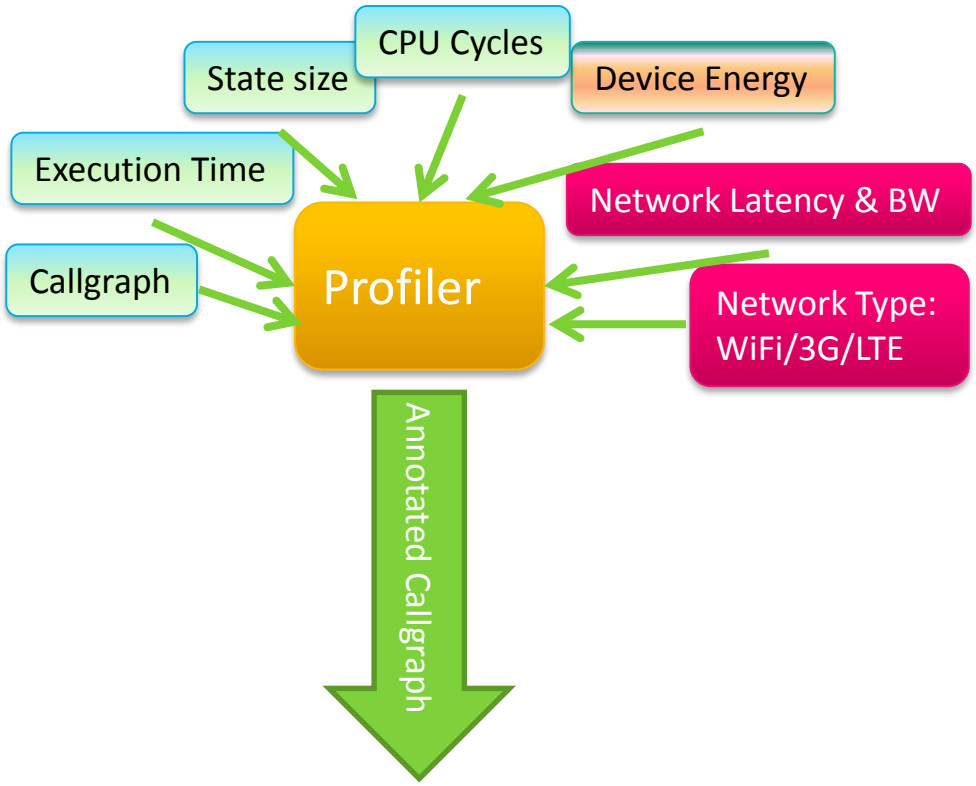
all have a profiler & a solver

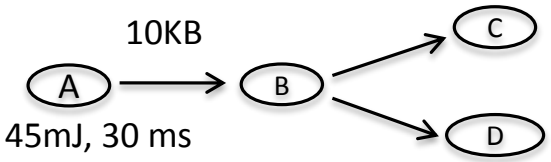
also see: *http://elijah.cs.cmu.edu/*

Microsoft

MAUI's profiler and decision engine

profiler





decision engine:

partition a running app – use Integer Linear Program

Example – Maximize:

$$\sum_{v \epsilon V} (I_v \times E_v) \quad - \quad \sum_{(u,v) \epsilon E} (|I_u - I_v| \times C_{u,v})$$

energy saved — cost of offload

Such that:

$$\sum_{v \epsilon V} (I_v \times T_v) \quad + \quad \sum_{(u,v) \epsilon E} (|I_u - I_v| \times B_{u,v}) <= \text{Lat.}$$

execution time — time to offload

and

$I_v <= R_v$ for all $v \epsilon V$

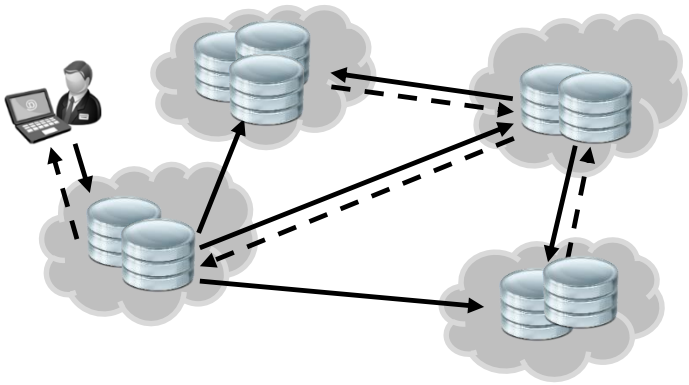
Microsoft

geo-distributed analytics

lots of data being generated at the edges, need support for sophisticated analysis

possible solution(s)

- pull all data into a central data center; answer queries from there
- leave data where it is collected; fetch on demand per query



costly and wasteful; not realtime
very long latency; can't run Hive or Spark on WAN

Observations

- connectivity is expensive, low bw & high latency
- need to support near real-time triggers (e.g.. faults/ fire)
- some of the data is infrequently accessed

geo-distributed analytics

allow data & query tasks to be placed at any site

- some datasets remain at the edge; others move to resource-rich DCs
- make job schedulers' robust to high latency by pipelining

mimic *optimal* data & task placement

- minimize average query latency
  - E.g.., move data iff the cumulative *shuffle volume* of its queries exceeds data size
  - Eg., place network-heavy tasks on a site where there is more data to be read

recapping benefits of mDCs

latency reduction

- serve static content immediately
- SSL termination / split TCP
- edge to DC protocol enhancements

bandwidth saving

- compression
- procrastination
- edge analytics

service & internet monitoring

reliable connectivity

- overlay networking
- path diversity

battery saving

- computation offloads
- client proxying

high-end game streaming

- lower device cost
- reduce developer fragmentation

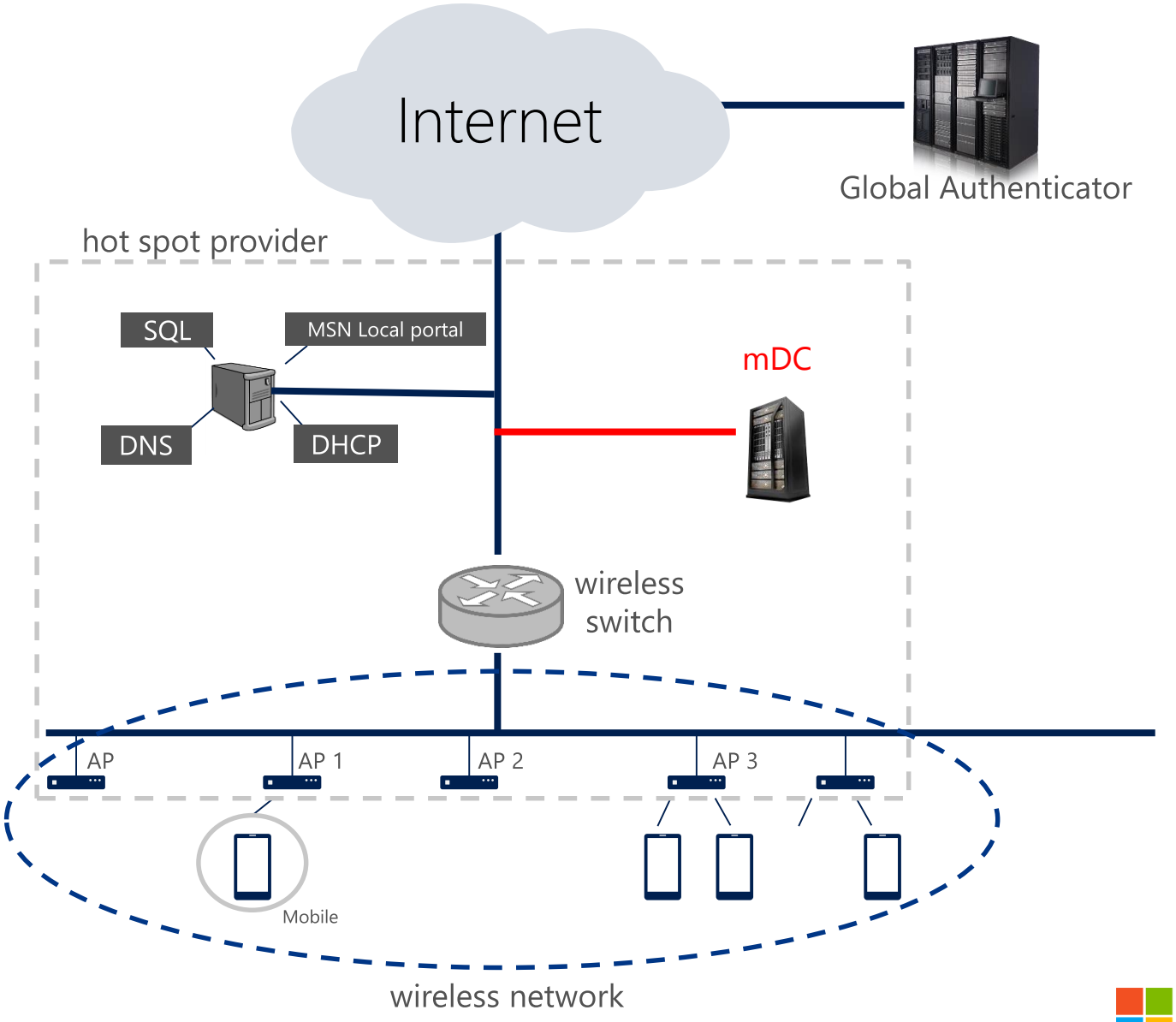
new services

protection against DoS

reduced load on DCs

deployment

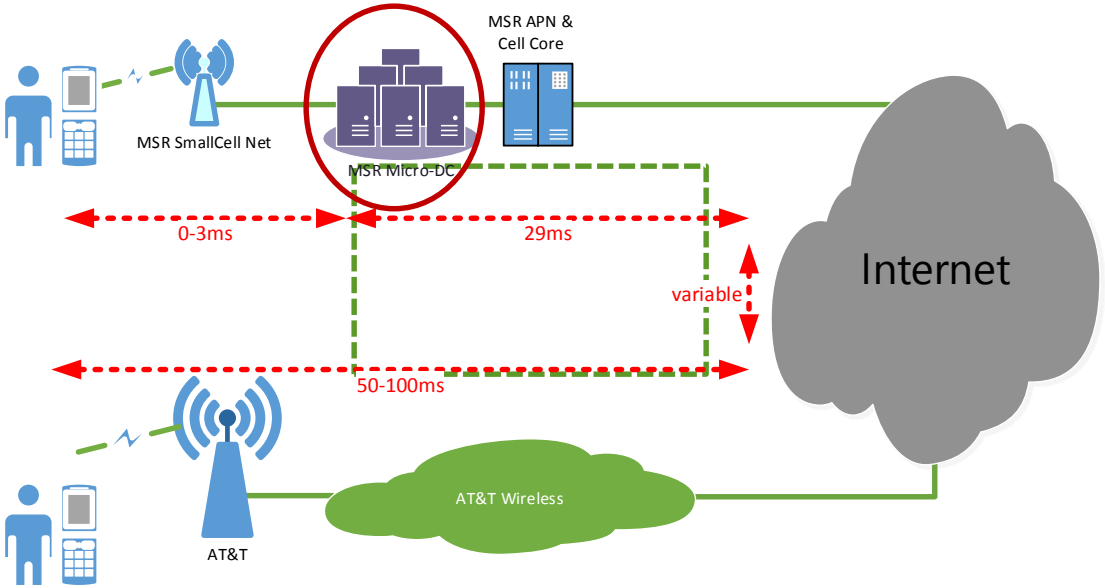
mDCs with Wi-Fi or White-Fi



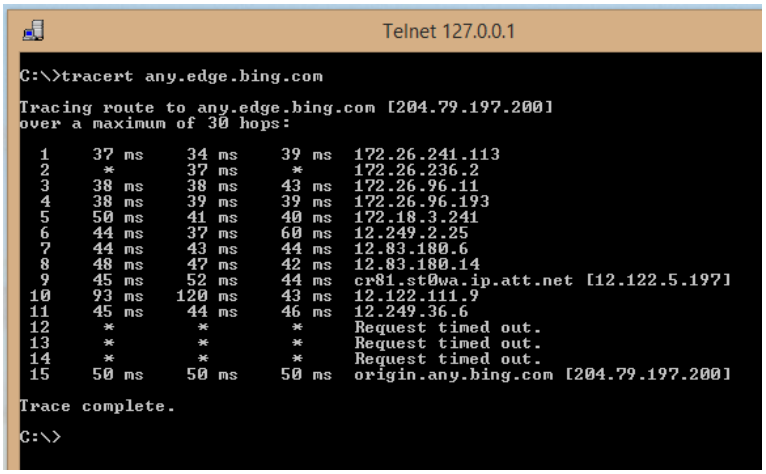
Microsoft

mDC with small cells

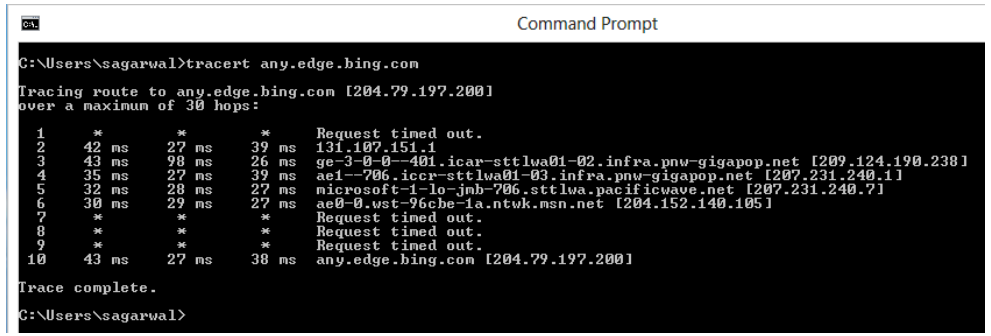
QCOM's Small Cell



Downlink	~110 Mbps
Uplink	~15 Mbps
RTT	~10 msec



tracert from AT&T LTE to *any.edge.bing.com* *(15 hops)*



tracert from SC to *any.edge.bing.com* *(10 hops)*

Microsoft

the wave is coming …

Akamai Cloudlets

- **Dynamically manipulate images in the Cloud for responsive web design**
- **Fast and Easy Access Control for your Web Applications**

intel

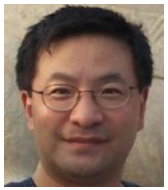
Increasing Mobile Operators' Value Proposition With Edge Computing

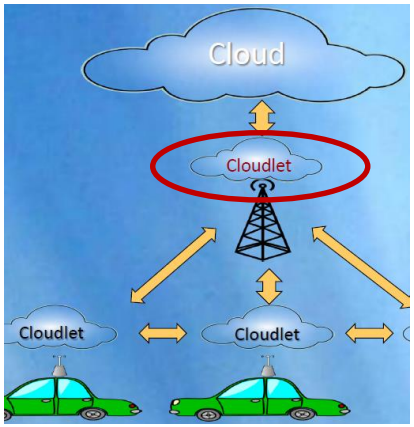
Turn bit pipes into smart pipes with an Intel® architecture-based server embedded into a Nokia Siemens Networks* base station

5G with **Undelay Networks** and **Local Cloud**

"local cloud are essential for backbone and core network scalability"

Dr. Geng Wu, Chief Scientist, **Intel** (Wireless World Research Forum, Vancouver, BC, Oct. 22, 2013)





"cloudlets for reducing latency, security and reliability"

- Dr. David Soldani, VP **Huawei**
- (IEEE ICC, June 12, 2013)

Microsoft

…and it's becoming bigger

MOs moving towards edge services

Liquid Net

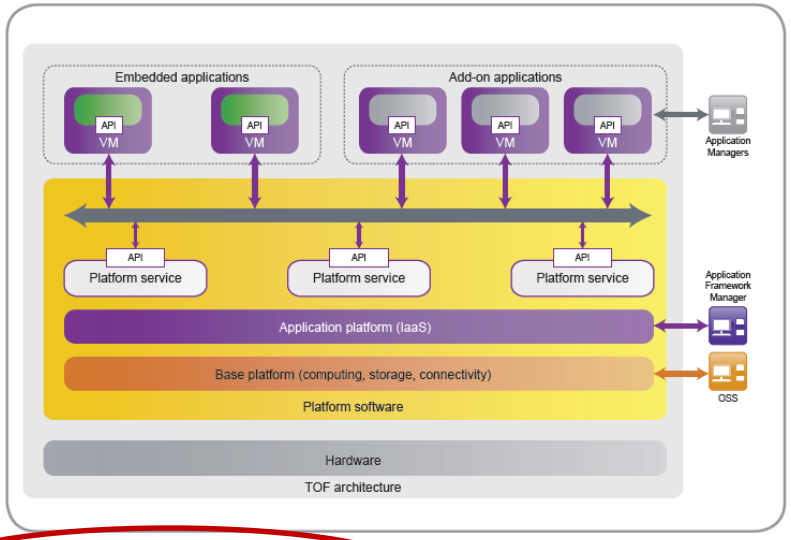


Figure 4: Base station application architecture

Nokia Networks

NETWORKWORLD

News | Blogs | Newsletters | Videos | Events | INSIDER | More

Security | LANs & WANs | UC / VoIP | Cloud | Infrastructure Mgmt | Wireless | Software | Data Center | SMB

Broadband | Collaboration | Equipment | Mobile | Networks | Security | Storage | White Papers | Webcasts | Tests

News

Nokia Siemens to merge cloud, base-station computing to boost performance

The company's Liquid Applications platform will use computing power in the cloud and in base stations, based on conditions

By Stephen Lawson, IDG News Service
February 24, 2013 04:06 PM ET

Add a comment | Print | Share | +1 | Like | 0 | More

IDG News Service - Nokia Siemens Networks will expand the role of cellular base stations with a new platform that will store and deliver some application data locally, while tapping into information about subscribers and traffic to improve the process.

The company announced the system, called Liquid Applications, at an event in Barcelona on the eve of Mobile World Congress. Liquid Applications can improve consumers' mobile experience but cutting delays as well as delivering more relevant content, CEO Rajeev Suri

ETSI World Class Standards

Nokia Networks reveals ETSI mobile edge computing collaboration

IBM, Intel, Vodafone, and Huawei all on board

October 20, 2014 | By Michael Carroll

IBM

United States [change]

News room | Search

Home | Solutions | Services | Products | Support & downloads | My IBM | Welcome [IBM Sign in] [Register]

News room
News releases
Press kits
Image gallery
Biographies
Background
News room feeds
Global news rooms
News room search
Media contacts

News room > News releases >

IBM and Nokia Siemens Networks Announce World's First Mobile Edge Computing Platform

News release | Related XML feeds | Contact(s) information

Barcelona, Spain - 25 Feb 2013: Mobile World Congress: Nokia Siemens Networks and IBM (NYSE: IBM) announced today a collaboration to deliver the world's first mobile edge computing platform that can run applications directly within a mobile base station. This new platform allows mobile operators to create a truly unique mobile experience, relieve the ever increasing strain on network infrastructure and bring completely new solutions to market.

The new platform can accelerate the delivery of media-rich services by delivering content directly from the base station, ensuring enhanced quality of experience for consumers in the face of ever increasing data traffic growth. The platform also enables a new generation of

Contact us
Contact a media relations representative
Site feedback

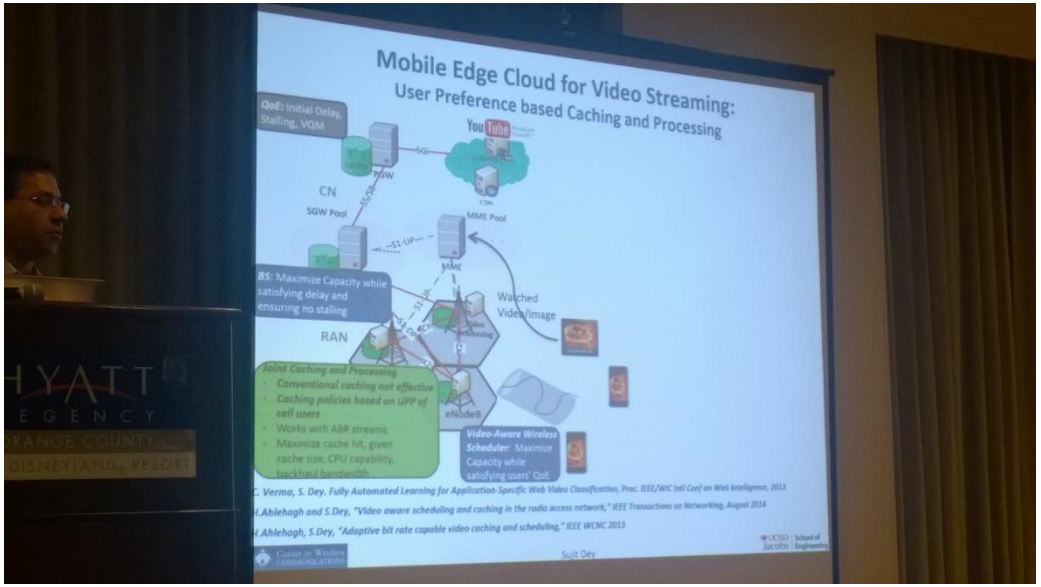
Share
Facebook
Twitter

Microsoft

overheard at a recent conference (IEEE ICNC 2015)

John Apostolopoulos
CTO &VP, Cisco, USA

“fog computing”



Sujit Dey, Professor/ Director
Center for Wireless CommunicationsUCSD

Microsoft

it's hot in the research community as well…

there is plenty of research literature (incl. MSR's) that shows edge computing significantly enhances mobile experience

first paper →

Satya (CMU), Bahl (Microsoft), Caceres (AT&T), Davies (Lancaster)
The Case for VM-based Cloudlets in Mobile Computing
IEEE Pervasive Computing, October 2009

~ 900 citations

Cuervo (Duke), Balasubramanian (UMASS), Wolman, Saroiu, Chandra, Bahl (Microsoft)
MAUI: making smartphones last longer with code offload
ACM MobiSys conference, June 2010

~ 825 citations

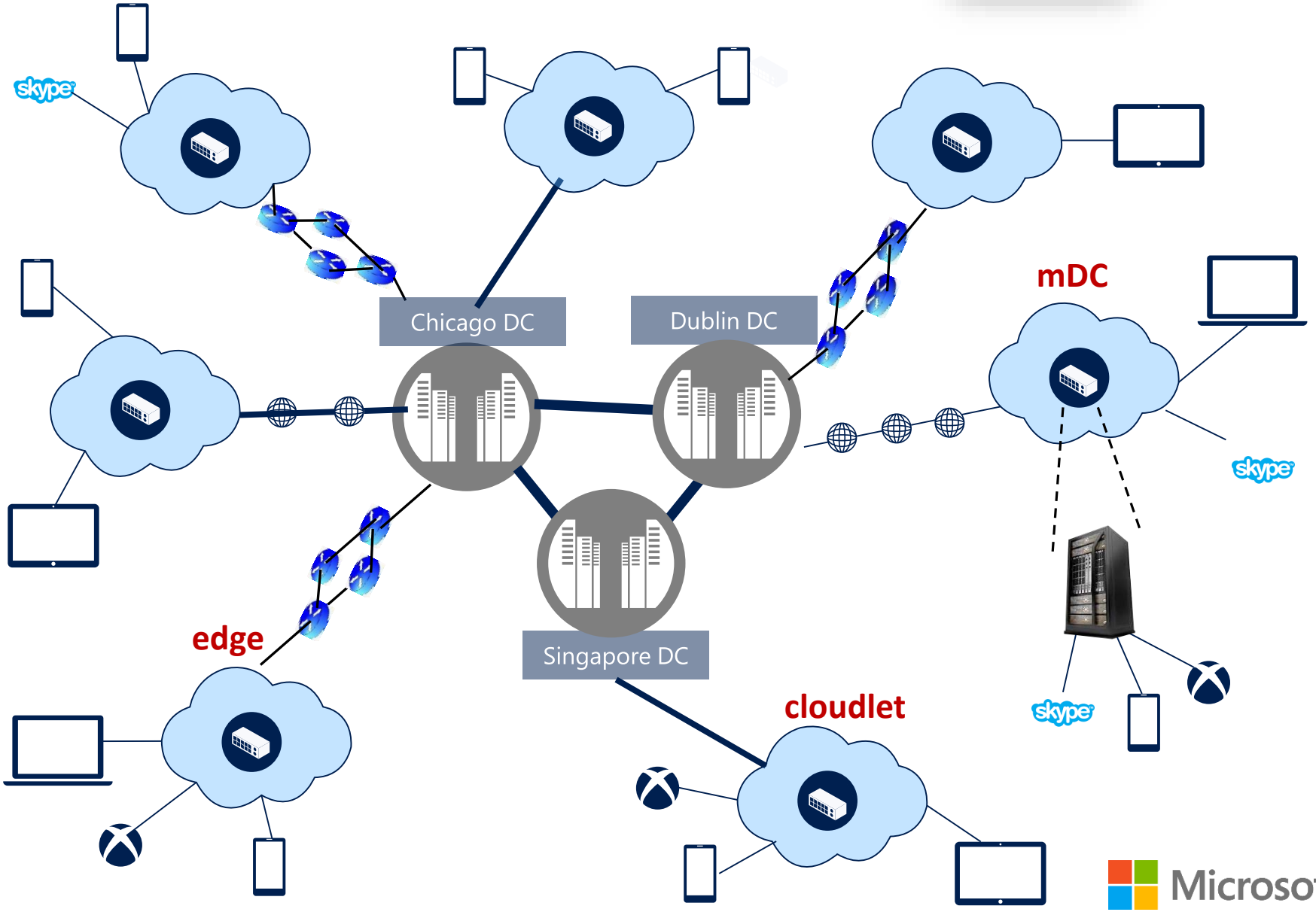
Why a Cloudlet Beats the Cloud for Mobile Apps

Posted on December 13, 2009 by lewisshepherd

Microsoft

cloud computing 2020

=  hyper scale data center



Microsoft

with mDCs (cloudlets) you can…

- develop new *(latency sensitive, CPU & battery intensive) (IoT)* applications, which *(dynamically)* partition themselves
- pursue infrastructure research in an emerging cloud platform, which promises to be pervasive
- deploy your own mDCs & connect them to Azure

Microsoft

merci!

mDC benefits -
app & game streaming

run any ecosystem's apps on resourced-starved devices by streaming them from the cloud

- circumvent client-side compatibility complexities

- with mDCs, reduce
  - latency -- keeping users engaged
  - jitter & packet loss – reduce user frustrating in highly interactive sessions
  - backbone bandwidth so both MOs and we pay less to other ISPs

note: standard proxy + split TCP insufficient for interactive traffic

mDCs can reduce dependency on cellular networks

offload to Wi-Fi aggressively

compress aggressively

already doing this

procrastinate instead of prefetch

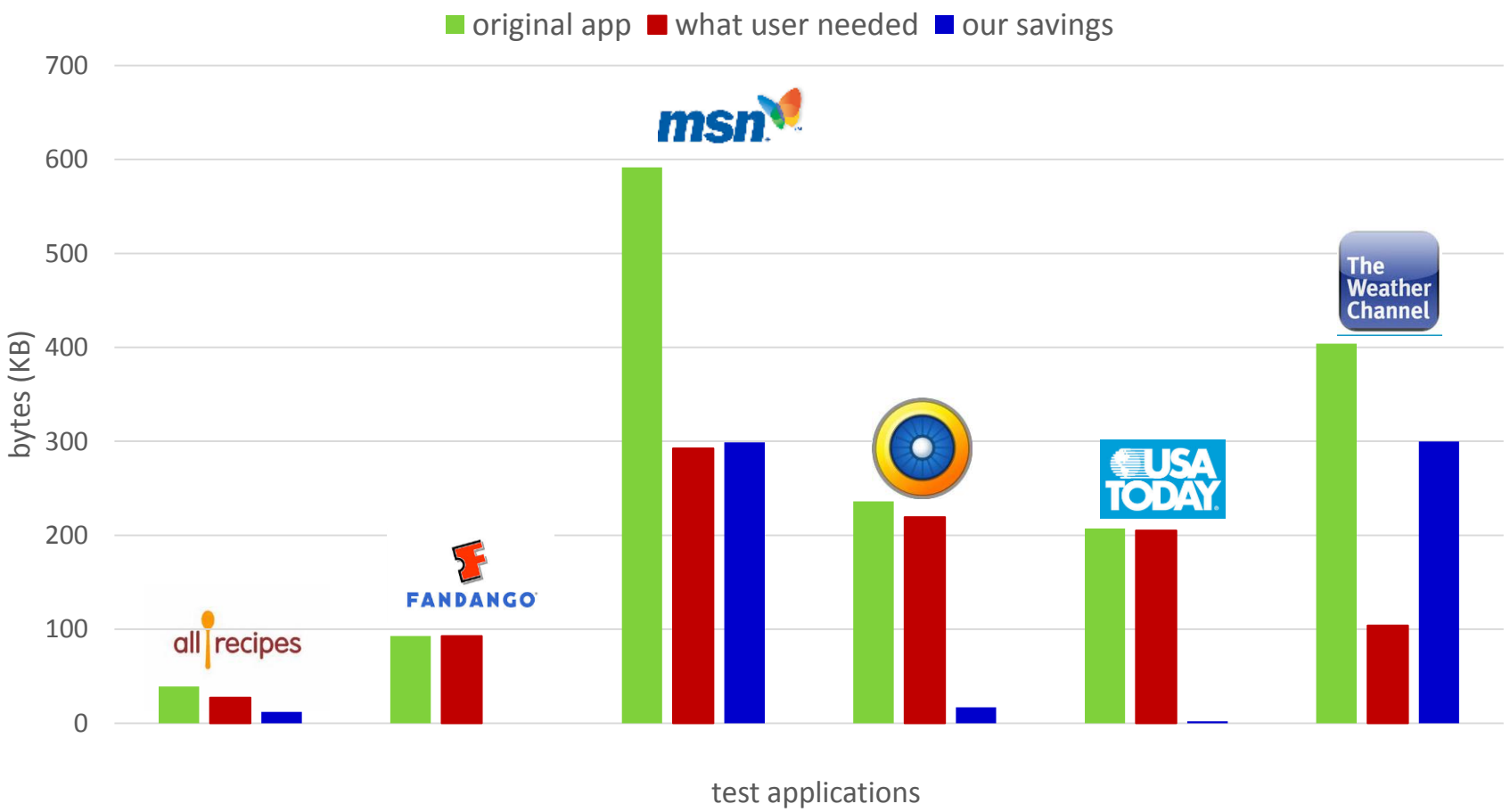
MobiSys 2014

- many network apps. fetch data whether or not it is consumed
- idea: mDC fetches the data but holds on to it until user explicitly needs it
  - ✓ save cellular bandwidth without the latency penalty

procrastinate & save

few results on bandwidth saving

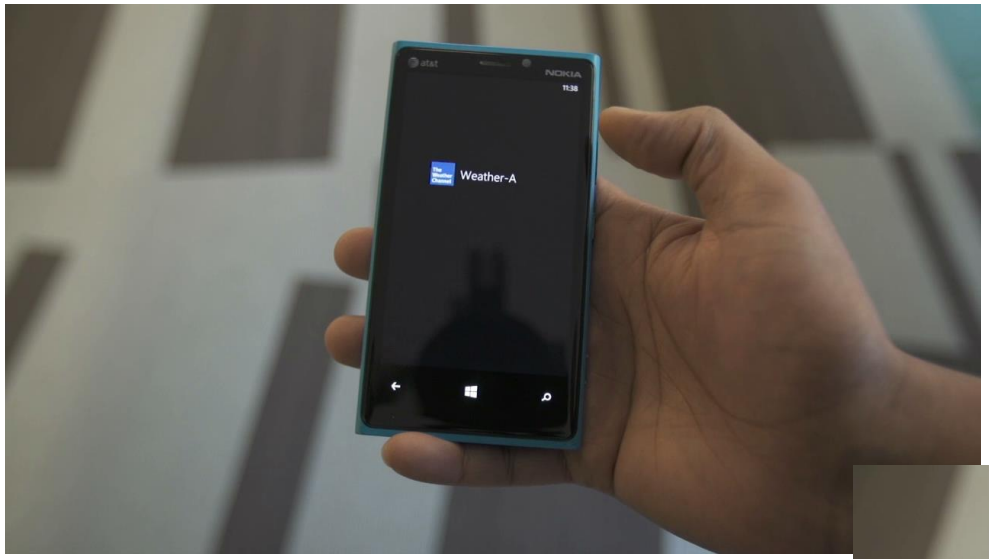
the system automatically decides what is not needed by the end-user



micro datacenter - benefits

reducing dependency on cellular networks (with procrastination)

get data only when needed (without mDC)



get data only when needed (with mDC)