edge computing
a historical perspective & direction
10 years & counting

Victor Bahl
Distinguished Scientist
Director, Mobility & Networking Research
Microsoft Research

Monday, August 20, 2018
Microsoft’s big bet: Azure
millions of servers

54 Azure regions
2M miles intra-DC fiber
150+ data centers
80+ Tb data on backbone
Microsoft’s big bet: Azure

FY18: Intelligent Cloud: $23 billion revenue (Azure grew 89% in Q4)

FY17 growth numbers:

- **15%** YoY Microsoft server products and cloud services revenue growth
- **97%** YoY Azure Revenue Growth
- **250+** FY17 Azure announcements
- **2x** YoY Azure compute usage

>90% of Fortune 500 use Microsoft Cloud
Microsoft’s data centers

- Columbia river, hydro-electric power
- Each facility is 8 MW in size, total of 64 MW
- Expanding rapidly, powered by wind farms
The Case for VM-Based Cloudlets in Mobile Computing

A new vision of mobile computing liberates mobile devices from severe resource constraints by enabling resource-intensive applications to leverage cloud computing free of WAN delay, congestion, and failures.

Why a Cloudlet Beats the Cloud for Mobile Apps

Posted on December 13, 2009 by lewisshepherd
offloading & programming the edge (2009-10)

July 12 –14, 2009

MAUI: Making Smartphones Last Longer with Code Offload

Eduardo Cuervo¹, Aruna Balasubramanian¹, Dae-ki Cho⁰,
Alec Wolman¹, Stefan Saroiu¹, Ranveer Chandra¹, Paramvir Bahl¹
¹Duke University, ⁰University of Massachusetts Amherst, ¹UCLA, ¹Microsoft Research

ABSTRACT
This paper presents MAUI, a system that enables fine-grained energy-aware offload of mobile code to the infrastructure. Previous approaches to these problems either relied heavily on programmer support to partition an application, or they were coarse-grained requiring full process (or full VM) migration. MAUI uses the benefits of a managed code environment to offer the best of both worlds:

Given the t
the energy foremost cl
One pop

services is remote execution: applications can take advantage of the resource-rich infrastructure by delegating code execution to remote servers. For the last two decades, there have been many attempts to make mobile devices use remote execution to improve perfor-

MobiSys 2010

Overcoming Language Barriers Using MAUI

citation 1996 (as of 8/20/18)
opportunistic use of infrastructure for dynamic offloading

approach
- developers build standalone apps with simple annotations but **no changes to program logic**
- system uses nearby and cloud-server resources in **opportunistic manner**

properties
- apps. always work, even when disconnected
- simple programming model (lowers barrier to widespread adoption)
impact of latency on recognition performance
impact of latency on recognition performance
led to research, papers, keynotes, & a prediction

the disaggregated cloud!

cloud computing 2020

Dec. 12, 2013
prediction was based on

the virtues of edge computing

- latency reduction
  - serve content immediately
  - SSL termination with split TCP
- 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 & applications
- protection against DoS
- reduced load on DCs

there is plenty of research literature (incl. MSR’s) that shows edge computing significantly enhances mobile experience
several developments since then

press articles

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

research projects

Elijah  Cloudlet-based Mobile Computing
Gabriel  Wearable Cognitive Assistance using cloudlets

Government initiatives

standards

conferences

industry initiatives

NSF Workshop on Grand Challenges in Edge Computing

SEC 2016 The First IEEE/ACM Symposium on Edge Computing
October 27-28, 2016, Washington DC, USA
... but we needed a killer app
Vision: Cloud-Powered Sight for All
Showing the Cloud What You See

Paramvir Bahl     Matthai Philipose     Lin Zhong
Microsoft Research
Redmond, WA 98052
{bahl,matthaip,t-lizhon}@microsoft.com

ABSTRACT
We argue that for computers to do more for us, we need to show the cloud what we see and embrace cloud-powered sight for mobile users. We present sample applications that will be empowered by this vision, discuss why the timing is right to tackle it, and offer our initial thoughts on some of the important research challenges.

Categories and Subject Descriptors
A.1 [General Literature]: introductory and survey

General Terms
Algorithms, Design, Human Factors, Languages, Performance, Security

and share it with the cloud. Inspiring applications and services have been demonstrated that analyze continuously collected accelerometer and microphone data, e.g., [LPL-109], and occasionally data from phone camera. However, information transfer is greater than limited.

In this paper, we assert that mobile applications should be configured to analyze data to them at a much higher rate to unleash the creativity of developers. Applications that make use of this concept today, we let the computer know where and how we move, but our computer see what we see.
MSR’s Glimpse project
Vision: Cloud-Powered Sight for All
Showing the Cloud What You See

Paramvir Bahl  Matthai Philipose  Lin Zhong

Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision

Robert LiKamWa†, Bodhi Priyantha†, Matthai Philipose†, Lin Zhong†, and Paramvir Bahl†

Glimpse: Continuous, Real-Time Object Recognition on Mobile Devices

Tiffany Yu-Han Chen
MIT CSAIL
yuhang@csail.mit.edu

Lenin Ravindranath
Microsoft Research
lenin@microsoft.com

Shuo Deng
MIT CSAIL
shuodeng@csail.mit.edu

Paramvir Bahl
Microsoft Research
bahl@microsoft.com

Hari Balakrishnan
MIT CSAIL
hari@csail.mit.edu

ABSTRACT
Glimpse is a continuous, real-time object recognition system for camera-equipped mobile devices. Glimpse captures full-motion video, locates objects of interest, recognizes and labels them, and tracks them from frame to frame for the user. Because the algorithms for object recognition entail significant computation, Glimpse runs them on server machines. When the latency between the server and mobile device is higher than a frame-time, this approach lowers object-recognition accuracy. To regain accuracy, Glimpse uses an active cache of video frames on the mobile device. A subset of the frames in the active cache are used to track objects on the mobile, using (small) hints about objects that arrive from

Frame 1  (t = 0 ms)
Frame 20  (t = 650 ms)

Figure 1: Offloading every frame to a server reduces trackability (right): the stop sign’s location is wrong.
canonical example for edge computing
the connected car
in-vehicle video analytics for detecting open parking spaces in urban environment
Vision: Cloud-Powered Sight for All
Showing the Cloud What You See
Paramvir Bahl, Matthai Philipose, Lin Zhong

Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision
Robert LiKamWa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl

Glimpse: Continuous, Real-Time Object Recognition on Mobile Devices
Tiffany Yu-Han Chen, Lenin Bavingrathan, Shuo Peng

ParkMaster: An in-vehicle, edge-based video analysis service for detecting open parking spaces in urban environments
Giulio Grassi, Sorbonne Universités, UPMC, LIP6
Kyle Jamieson, Princeton University; University College London
Paramvir Bahl, Microsoft Research, Redmond
Giovanni Pau, Sorbonne Universités, UPMC, LIP6

ABSTRACT
We argue that leveraging the cloud can improve the performance of mobile users. We extend this idea by this new model, which offers our superior performance in a cloud-based environment.

Though mobile platforms such as smartphones are powerful, they are still far from offering the same level of performance as desktop computers. The mobile cloud model has emerged as a promising solution for improving the performance of mobile users. However, the challenge of integrating different mobile platforms and providing a smooth user experience remains a significant challenge.

Glimpse is an innovative system that addresses this challenge by leveraging the cloud to provide real-time object recognition on mobile devices. It enables users to identify objects in real-time, such as detecting people, cars, and other objects, and provides a seamless experience for users.

ABSTRACT
Glimpse is a system for full-mobile object recognition that allows users to significantly improve the performance of vision-based services on mobile devices. It provides real-time object recognition in real-time, significantly improving the user experience for mobile devices.
aha moment!

for every 8 people in the US & for every 29 people worldwide!

→ live video streams are being generated from factory floors, traffic intersections, camera mounted on cars, police vehicles, & retail shops

extract value from video streams in-context, in-the-moment to generate actions & workflows

with cloud computing, it’s the golden era for computer vision, AI & machine learning

potential to impact science, society & business
first attempt: public security

prevailing approach (at the time):
upload video to the cloud for remote (offline) analysis

limitations
• large quantities of data (>10GB/hour)
• bandwidth availability limited coverage & accuracy
• human availability limited the systems usefulness
  – no automatic real-time tracking or alerts
saving network bandwidth
(increasing coverage & accuracy)

<10% frames capture objects of interest
fun project: securing corporate buildings
some disturbing local news

Hit-and-run driver nearly kills woman on bike in Bellevue

Car strikes, kills toddler in stroller in Bellevue

77-year-old pedestrian killed by teen driver in Bellevue

traffic safety: a world-wide movement

- 1.2 million people die on the world’s roads every year.
- 20-50 million suffer non-fatal injuries.
- In the US, 19,000 people were killed in the first 6 months of 2016 (up 9% compared to 2015).

**Vision Zero** is a multi-national road traffic safety project that aims to achieve a highway system with no fatalities or serious injuries in road traffic. It started in Sweden and was approved by their parliament in October 1997.
cities all over North America are embracing it
city planners care about -

- how often are vehicles speeding & failing to yield to ?
- are pedestrians disregarding traffic signals?
- are bicyclists ignoring or are they running ?
- any trends that hint at the reasons why certain are broken in certain places?
- did a countermeasure have the desired effect?

courtesy: Franz Loewenherz, Senior Transportation Planner, City of Bellevue, WA
city planners need data & analytics to perform corrective measures

In 2013, WSDOT built a new roundabout at the intersection

2005 - 2010 60 collisions recorded by the Bellevue Police Department
...we got going, we had a “killer” application and it was about saving lives
Vision Zero: eliminate pedestrian/biker deaths
Use widely deployed traffic cameras
  • Car/bike/ped counts, near-collisions, anomalies

next-generation traffic control

Amy Carlson,
Vice President & Area Office Manager, CH2M Hill
picked up by local media

Microsoft looks to stop bike crashes before they happen, testing Minority Report-style predictive intelligence

BY LISA STIFFLER on October 14, 2015 at 1:00 pm

declined interview but...

“Microsoft, Bellevue team up to prevent crashes”
video query: pipeline of transforms

vision algorithms ("transforms") chained together
transforms implement specified interfaces

example: count the number of moving cars on a road segment

transform 1 (decoder)  transform 2 (object detector)  transform 3 (object tracker)  transform 3 (classifier& counter)
many implementation choices

40+ detector implementations
- motion-based: background subtraction
- DNN-based: Yolo detection
- exhaustive search

60+ tracker implementations
- moving pattern
- color histogram
- key-point features: SURF, SIFT

which implementation will you select?
which implementation is better?

DNN + histogram (0.17 fps)

BGS + movement (42.3 fps)
each implementation’s performance is impacted by the selection of “knob” positions

- **frame rate**: 30 fps for HD cameras
- **resolution**: 1080p, 720p, 480p...
- **window size**: region of interest

**Licence Plate Reader**

- Frame rate 3
- Resolution 720p
  - Accuracy: 0.93, CPU: 0.54 cores

- Frame rate 1
- Resolution 480p
  - Accuracy: 0.27, CPU: 0.09 cores
knobs/parameters impact quality & resource demands

frame rate

resolution

window size
impact of knobs/parameters on quality & resource demands

orders of magnitude cheaper resource demand for little quality drop

no analytical models to predict resource-quality tradeoff
no one plan is uniformly the best...

differ by \(46x\) in their accuracy, \(250x\) in speed!

best plan is dependent on the camera, lighting, track direction, object color, ...

---

processing thousands of live streams
to support different types of queries at scale:
• must reduce processing cost of a query
• must schedule resources efficiently across queries

lag: time difference between frame arrival and frame processing

<table>
<thead>
<tr>
<th>accuracy lag</th>
<th>high hours</th>
<th>moderate seconds</th>
<th>high seconds</th>
</tr>
</thead>
</table>

TOLL-BY-PLATE AMBER ALERT
what is the best implementations for a video analytics query?

the configuration & resource allocation that maximizes quality & minimizes lag within the given resource capacity is the best implementation
system design

query -> profiler

resource-quality tradeoff

utility (quality & lag)

offline

scheduler

trades off quality & lag across queries

online
• operational traffic cameras in Bellevue and Seattle
• 101 machine Azure cluster
• license plate reader, car counter, DNN classifier, object tracker
results
details in our NSDI 2017 paper

compared to a fair scheduler with varying burst duration:
• quality improvement: up to 80%
• lag reduction: up to 7x
Vision: Cloud-Powered Sight for All
Showing the Cloud What You See
Paramvir Bahl, Matthai Philipose, Lin Zhong

Energy Characterization and Optimization of Image Sensing Toward Continuous Mobile Vision
Robert LiKamWa†, Bodhi Priyantha†, Matthai Philipose†, Lin Zhong†‡, and Paramvir Bahl†‡

Glimpse: Continuous, Real-Time Object Recognition Mobile Devices
Tiffany Yu-Han Chen‡, Lenin Ravindranath§, Shuo Deng§

ParkMaster: An in–vehicle, edge–based video analytics service for detecting open parking spaces in urban environments
Giulio Grassi, Paramvir Bahl

The Design and Implementation of a Wireless Video Surveillance System
Tangi Zheng, Ashokkhe Chaubhogy, Paramvir Bahl, Kiel Jamieson, Sumit Rangan

Live Video Analytics at Scale with Approximation and Delay-Tolerant
Haoyu Zhang†‡, Ganesh Ananthanarayanan*, Peter Bodik*, Matthai Philipose†‡, Michael J. Freedman†‡
*Microsoft †Princeton University

Abstract
Video cameras are pervasively deployed for security and smart city scenarios, with millions of them in large cities worldwide. Achieving the potential of these cameras requires efficiently analyzing the live videos in real-time. We describe VideoStorm, a video analytics system that processes thousands of video analytics queries on live video streams over large clusters. Given the high costs of vision processing, resource management is crucial. We consider two key characteristics of video analytics: resource-quality tradeoff with multi-dimensional configurations, and variety in quality and lag goals. VideoStorm’s offline profiler generates query resource-
...and we have been deploying & learning
(Cambridge, U.K)

![Image of a road scene with a vehicle and a classification diagram]

<table>
<thead>
<tr>
<th>Classified Truth</th>
<th>Vehicles</th>
<th>Bikes</th>
<th>Peds</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vehicle</strong></td>
<td>0.95</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Bike</strong></td>
<td>0.08</td>
<td>0.67</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td><strong>Pedestrian</strong></td>
<td>0.15</td>
<td>0.15</td>
<td><strong>0.73</strong></td>
<td>0.05</td>
</tr>
<tr>
<td><strong>None</strong></td>
<td>0.09</td>
<td>0.03</td>
<td>0.11</td>
<td><strong>0.81</strong></td>
</tr>
</tbody>
</table>

We recognized it as
multi-tenancy

can a existing network of cameras be used by more than a single customer?
steerable cameras
servicing multiple applications simultaneously

- foliage monitoring
- pedestrian monitoring
- weather monitoring
- parking spot monitoring
- car counting / license plate detection
break one-to-one binding between camera & application

- traffic volume monitoring
- state of art
- fixed view camera

azure

- traffic volume monitoring
- accident detector
- amber alert

steerable PTZ camera

our system
camera management system

```
applications

vCamera  vCamera  vCamera

camera virtualization layer

pCamera

camera control
controls \( \{p, t, z\} \)

mobility-aware scheduler

predictor

Per app. SLA

<table>
<thead>
<tr>
<th>app. 1</th>
<th>( (p_1, t_1, z_1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>app. 2</td>
<td>( (p_2, t_2, z_2) )</td>
</tr>
<tr>
<td>app. n</td>
<td>( (p_n, t_n, z_n) )</td>
</tr>
</tbody>
</table>
```

Shubham
deployment

accuracy • latency • bandwidth • cost
the system we built

vision modules

tracker
counter

optimization

selector
profiler

query

geo-distributed execution layer

execution engine
execution engine

resource-quality tradeoff

query plan

resource manager

scheduling, placement ...

camera manager

early discard

early discard

training phase

neural network

model generator
crowd sourced labeled data

analytics

alert
UI

video storage

event DB

analytics output
the stack we built: MSR’s Rocket

- public safety
- traffic planning & safety
- consumer live videos
- home security
- retail surveillance
- ...

video pipeline optimizer

- resource manager
- (geo-) distributed executor
- camera manager
- GPU manager

- vision modules & neural networks
- crowd-sourced labeled data

video store

apps systems ML / vision
deployment: hybrid edge-cloud architecture

Azure US-West

Bellevue, WA edge

micro DC

factory floor edge

Azure US-East

Washington DC edge

micro DC

factory floor edge

deployment: hybrid edge-cloud architecture

Azure US-West

Bellevue, WA edge

micro DC

factory floor edge

Azure US-East

Washington DC edge

micro DC

factory floor edge
multi-camera implementation in Bellevue
live dashboard

http://vavz.azurewebsites.net/
direction counting accuracy

12 directions (lane-wise counts) occlusions due to 3D $\rightarrow$ 2D projection on lanes

95% count accuracy compared to crowdsourced ground truth
training neural networks

labeled data
national initiative to train NN
(launched July 1, 2017)
http://www.ite.org/visionzero/videoanalytics/

Volunteers needed to improve traffic safety using high-tech

by: Alison Grande Updated: Jun 1, 2017 - 8:53 PM

Video Analytics towards Vision Zero

Worldwide problems demands bold action

Make a difference, teach computers to learn

Participate Today

Partners

To help the video analytics system learn to detect real conflicts, Microsoft is collaborating with the following partners to promote this revolutionary technology:

- Microsoft
- Siemens
- KDDI
- INRIA
- Viaz
- Unity

Partners

More info
can we solve all problems?
(can humans do better?)
actively reducing accidents
making self-driving cars safer

(10 million self-driving cars by 2020 – Forbes, March 2017)
live demonstration in Hannover Messe

schematic view
some nice memories
mayor’s challenge award to Bellevue

Bellevue, WA, pursued a range of data collection activities during the Mayor’s Challenge to identify barriers to bicycling and walking, prioritize improvements, and guide investments. In February 2015, the Bellevue City Council established the Pedestrian and Bicycle Implementation Initiative (PBII) to improve safety for people of all ages and abilities who walk and bike in Bellevue. Using data collected from online surveys, impact mapping at public meetings, automated bicycle and pedestrian counters, and traffic camera video, the PBII team identified barriers to walking and bicycling and developed a $4.8M Bicycle Rapid Implementation Project (BRRP) budget proposal to guide citywide investments through 2016. The BRRP aims to expand the city’s bicycle network from 42 miles to more than 70 miles of conventional bike lanes, separated bike lanes, or off-street paths, and to complete four continuous, multi-use bicycle corridors.

**Demonstrated Successes**

- Innovative data collection techniques gathered real-time and long-term data with public input throughout the PBII process. Bellevue has emphasized understanding long-term trends and gathering feedback from people who walk and bike.
- Conducted a longitudinal assessment from 2004-2015 of non-motorized collisions using the USDOT’s Pedestrian and Bicycle Crash Analysis Tool (PB-CAT) system.
- Gathered input using key-painted polling and comment cases at 26 public meetings and an open house that attracted 1,600 attendees.
- Used citizen surveys to solicit public input at two stages in the BRRP development process.
  - Over 700 people placed more than 1,600 points on the first online map to identify locations that they felt were unsafe for walking and bicycling.
  - Over 120 people submitted more than 400 comments on conceptual designs for 50 proposed projects to make the pedestrian and bicycle systems safer.

**Winner!**
Bellevue, Washington

Bellevue, WA, pursued a range of data collection activities during the Mayor’s Challenge to identify barriers to bicycling and walking, prioritize improvements, and guide investments. In February 2015, the Bellevue City Council established the Pedestrian and Bicycle Implementation Initiative (PBII) to improve safety for people of all ages and abilities who walk and bike in Bellevue. Using data collected from online surveys, impact mapping at public meetings, automated bicycle and pedestrian counters, and traffic camera video, the PBII team identified barriers to walking and bicycling and developed a $4.8M Bicycle Rapid Implementation Project (BRRP) budget proposal to guide citywide investments through 2016. The BRRP aims to expand the city’s bicycle network from 42 miles to more than 70 miles of conventional bike lanes, separated bike lanes, or off-street paths, and to complete four continuous, multi-use bicycle corridors.

**Demonstrated Successes**

- Innovative data collection techniques gathered real-time and long-term data with public input throughout the PBII process. Bellevue has
- Conducted a longitudinal assessment from 2004-2015 of non-motorized collisions using the USDOT’s Pedestrian and Bicycle Crash Analysis Tool (PB-CAT) system.
- Gathered input using key-painted polling and comment cases at 26 public meetings and an open house that attracted 1,600 attendees.
- Used citizen surveys to solicit public input at two stages in the BRRP development process.
  - Over 700 people placed more than 1,600 points on the first online map to identify locations that they felt were unsafe for walking and bicycling.
  - Over 120 people submitted more than 400 comments on conceptual designs for 50 proposed projects to make the pedestrian and bicycle systems safer.
Video Analytics analyzes traffic camera video footage and uses near-miss collisions to predict where future crashes are likely to occur. Traffic engineers could then take corrective action to prevent them. File photo

Bellevue video analytics project receives safety award

Fri Aug 4th, 2017 3:44pm - BUSINESS
Real-Time Video Analytics: The Killer App for Edge Computing

Ganesh Ananthanarayanan, Paramvir Baldr, Peter Bodik, Krishna Chintalapudi, Methal Philipose, Lavin Revindranath, and Sudipta Sinha, Microsoft Research

Video analytics will drive a wide range of applications with great potential to impact society. A geographically distributed architecture of public clouds and edges that extend down to the cameras is the only feasible approach to meeting the strict real-time requirements of large-scale live video analytics.

According to a 2015 report by the Information Handling Services on the installed base for video surveillance equipment, there is a camera installed for every 26 people on the planet, with mature markets having a camera for every 4 people. The report predicts that the number of cameras will grow by 29 percent year over year for the next 5 years. Video analytics from these cameras are used for traffic control, surveillance, and security in both public and private venues, as well as consumer applications including digital assistants for real-time decisions.

We propose that a geographically distributed architecture of public clouds, private clusters, and edges that extend down to the cameras is the only approach that can meet the strict real-time requirements of large-scale video analytics, which must address latency, bandwidth, and provisioning challenges. First, applications require very low latency when processing video because the output of the analytics is used to interact with humans (such as in augmented reality scenarios) or to actuate some other system (such as traffic lighting). Second, high-definition video requires large bandwidths—5 Mbps or even 25 Mbps for 4K video—and streaming a large number of video feeds directly to the cloud might not be feasible. When cameras are connected wirelessly, such as inside a car, the available uplink bandwidth is very limited. Finally, using compute capacity available on the camera itself allows for correspondingly lower provisioning (or usage) in the cloud. This also means that less interesting parts of the video can quickly be filtered out, for example, using motion-detection techniques, which dramatically reduce the bandwidth that needs to be provisioned.

Aside from low latency and efficient bandwidth usage, another major consideration for continuous video analytics is video processing’s high compute cost. Because of high data volume, compute demand, and latency requirements, cameras are the most challenging “things” in the Internet of Things. Thus, large-scale video analytics could well be edge computing’s “killer app.” Tapping into the cloud’s power is fine, but there are some critical constraints that must be considered.

Video Analytics Towards Vision Zero

Franz Loewenherz, Victor Bahl, Ph.D., and Yinhe Wang, Ph.D.

For young people below the age of 35, motor vehicle crashes are the leading cause of death in the United States. In 2015, collisions resulted in 35,092 deaths and 2.4 million injuries. More than 1,100 children under the age of 15 were killed. The 7.2 percent increase in traffic fatalities from 2014 to 2015 represents the greatest percentage increase in nearly 50 years. Yet despite the massive death toll, work to prevent traffic fatalities has been woefully lacking.

Many governmental agencies continue to rely on traditional traffic safety approaches. They intervene only after enough police crash reports are filed to trigger a High Crash Corridor designation. This reactive approach in preventing crash occurrence has well-documented limitations:

- At most locations, the number of crashes is very small and subject to cyclical variations.
- Not all crashes are reported, and the level of reporting is unclear.
- The type of road users involved, the exact location, and the severity of injuries are unknown.
- Numerous “false calls” go undocumented.
- Many years of crash data are typically required to develop an understanding of the situation.

Given these trends, the crash analysis tools presently available are not effective at achieving what all of us want: fewer fatalities and serious injuries on our roadways. That’s the goal of Vision Zero, which seeks to end all traffic fatalities and serious injuries in our cities. To make this goal a reality, we must first understand the root causes of these incidents, and then we can develop innovative solutions.
opportunity for AI in video analytics

*59% of installed cameras in 2016 are IP cameras
Source: Markets&Markets, IHS market
... but then something interesting happened ...
“It’s not the mobile devices but all the other things that are out at the edge that are truly going to transform cloud computing and put an end to what we know as the cloud.

Peter Levine
Andreessen Horowitz
December 16, 2016
"we're moving from what is today's mobile-first, cloud-first world to a new world that is going to be made up of an intelligent cloud and an intelligent edge"

- Satya Nadella
  CEO Microsoft
  Build, May 10, 2017
Maverick* Research: The Edge Will Eat the Cloud

Published: 22 September 2017  ID: G00330633
Analyst(s): Thomas J. Bittman

Summary
The growth of the Internet of Things and the upcoming trend toward more immersive and interactive user interfaces will flip the center of gravity of data production and computing away from central data centers and out to the edge. (Maverick research exposes unconventional thinking and advice.)

Overview
Specific Maverick Caution

This research contradicts prevailing views on the future of cloud computing, the topology of computing architectures and the nature of applications as we move toward digital business. Instead of continued growth of mega data centers, compute and storage will move toward the edge, due to the Internet of Things and new user/machine interfaces.

“forty percent of large enterprises will be integrating edge computing principles into their 2021 projects, up from less than 1% in 2017”
2017 the year edge computing took off.

Microsoft is extending Azure IoT to the edge of the network
Posted May 16, 2017 by Frederic Lardinois ( @frederic )

Amazon Makes Foray Into Edge Computing With AWS Greengrass
Jun 7, 2017 @ 6:34 PM 8,996
Janakiram MSV, CONTRIBUTOR
I cover Cloud Computing, Machine Learning, and Internet of Things FULL BIO

AT&T wants to reinvent the cloud with low latency edge computing over 5G
With new investments in edge computing, AT&T could set the stage for business investments in IoT, VR, robotics, and more.
By Cameron Fretz | July 18, 2017 2:20 AM PDT

GE adds edge analytics, AI capabilities to Predix industrial IoT suite
At its annual Minds + Machines conference, GE Data is unveiling a slew of extensions to its Predix Edge PaaS
By Marc Ferranti
October 24, 2017

Dell Technologies Launches A New IoT Division And Partner Program Amid 'Great Boom' In Edge Computing
by Kyle Alspach on October 10, 2017, 11:26 am EDT

Deutsche Telekom targets 5G applications with edge computing testbed

ITWORLD
Microsoft will invest $5 billion in IoT. Here’s why.

April 4, 2018 | Julia White, CVP Microsoft Azure

Advancing the Opportunity in IoT and Intelligent Edge

We’ve been invested in IoT before the term was coined, when enterprises had these endpoints in their factories, buildings and other devices that were totally “dark.” Today, we’re planning to dedicate even more resources to research and innovation in IoT and what is ultimately evolving to be the new intelligent edge.
Azure’s perspective on IoT App pattern + Edge
Azure IoT edge

Cloud services at the edge
Azure ML, Azure Stream Analytics, Azure Functions, custom

Manage from the cloud
Devices and services from Azure Portal

Flexible connectivity
Intermittent, low, or no connectivity

Reduced latency and cost
Bring compute to the data, reduced bandwidth cost
project Brainwave @ the edge

an accelerated FPGA powered AI Platform:

- **Fast:** Ultra-low latency, high-throughput serving DNN models at low batch sizes
- **Flexible:** Future proof, adaptable to fast-moving AI space & evolving model types
- **Friendly:** Turnkey deployment of TensorFlow/CNTK/Caffe/etc.

![Diagram showing Pretrained DNN Model in TensorFlow, CNTK, etc. and BrainWave soft DPU on FPGA]
deploying and running a model

1. Use Azure ML to create **custom model**
2. Use Model Management Service to pull it to the Edge Device
3. Run custom model with **FPGA on Edge Device**
4. Use your **custom code** to interface with a camera or microphone
5. Use **IoTHub** to manage your Edge Module & data streams
AI at the “cutting edge

• defect detection
• video surveillance
• example
what is the edge?

Azure IoT Edge hardware requirements

- Rich OS - Windows or Linux
- flexible HW – ARM or x64
- Moby-compatible container runtime
- hardware based security – HSM or Enclave
- hardware sizing depends on workload
underwater edge

Facts and Figures
• 12.2m length, 2.8m diameter
• Available IT Space: 12 42U racks
• Max Power: 454 KW (38 KW/rack)
• Power Utilization Effectiveness of 1.07
• Payload: 864 Azure servers w/FPGA
problem space is very rich

- data security & integrity
- availability
- resource management
- edge clouds
- machine learning at the edge
- programming model
- economics
- serverless framework?
- federated edges?
- deployments: drones, automobile, retail, factory floor, homes, enterprise
- management (Kubernetes, ...)
- networking
- 5G cloudification of the telcos
- benchmarks
- geo-distributed analytics

- to edge or not to edge?
- SLA
final thoughts

- edge computing is a paradigm shift, embrace it also known as: “micro DCs” & “cloudlets”

- by 2022, video analytics market is expected to become $11.2B and that is going to change lives
  

- nation-wide deployments will create a infra-structure where the other aspects of edge computing will shine
<table>
<thead>
<tr>
<th>Build with Azure IoT Edge—it’s free</th>
<th><a href="https://aka.ms/iot-edge/">https://aka.ms/iot-edge/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>get your devices certified on Azure IoT device catalog</td>
<td><a href="https://catalog.azureiotsolutions.com/">https://catalog.azureiotsolutions.com/</a></td>
</tr>
<tr>
<td>nominate your solutions to Azure marketplace</td>
<td><a href="https://aka.ms/iot-edge-marketplace-signup">https://aka.ms/iot-edge-marketplace-signup</a></td>
</tr>
<tr>
<td>register for the Vision AI Dev Kit to build your custom vision AI solution</td>
<td><a href="https://www.visionaidevkit.com">https://www.visionaidevkit.com</a></td>
</tr>
</tbody>
</table>
thanks!
body worn cameras on the rise

Since February of 2012, Rialto, California has required all police officers to wear a camera to monitor all interactions with the public.

After this practice was instituted, the number of complaints to the department dropped by 80%. Further, the number of instances of the police using “force” on people dropped 70%.

“When you put a camera on a police officer, they tend to behave a little better, follow the rules a little better…”
- Chief William Farrar

Metropolitan Police officers start wearing body cameras

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

By J. DAVID GOODMAN SEPT. 4, 2014

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