Infrastructure for edge computing

Gopika Premsankar

Ivey Business School, University of Western Ontario

March 22, 2021

About me

Western
UNIVERSITY · CANADA

Finnish Cultural Foundation

About me

Western UNIVERSITY · CANADA

Finnish Cultural Foundation

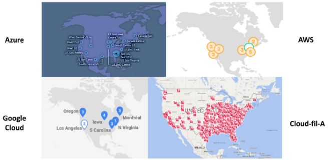
Aalto University
School of Science

Agenda

- Edge computing: where can it be deployed

Agenda

- Edge computing: where can it be deployed
  - Dependent on provider / use case



https://medium.com/@cfatechblog/edge-computing-at-chick-fil-a-7d67242675e2

Agenda

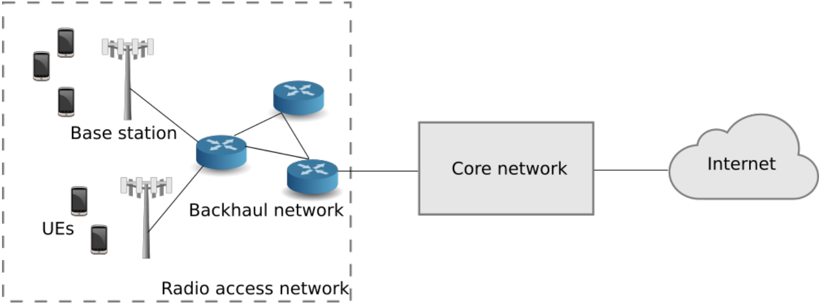
- Edge computing: where can it be deployed
  - Dependent on provider / use case
  - Focus on 5G perspective

Agenda

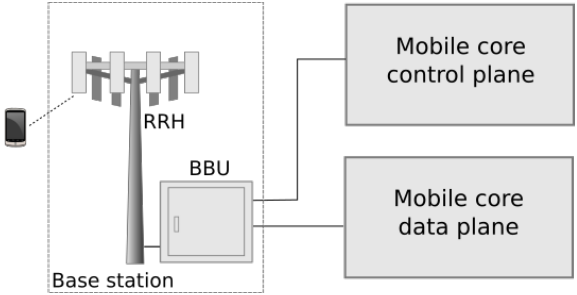
- Edge computing: where can it be deployed
  - Dependent on provider / use case
  - Focus on 5G perspective
- Optimal placement of edge computing devices for connected cars

Edge computing: a 5G perspective

Overview of cellular network architecture



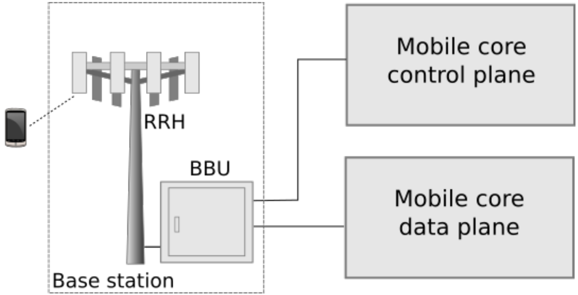
Control and data planes



- Base stations comprise Remote Radio Heads and Baseband Units

https://5g.systemsapproach.org/arch.html

Control and data planes

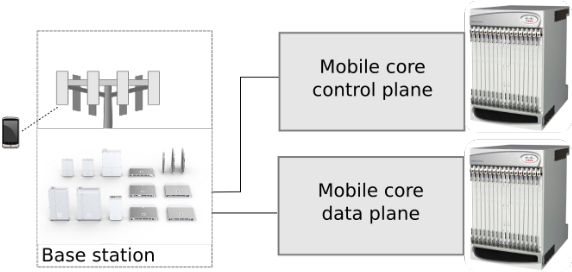


- Base stations comprise Remote Radio Heads and Baseband Units
- Signaling traffic over control plane
- User data over data (or user) plane

---

https://5g.systemsapproach.org/arch.html

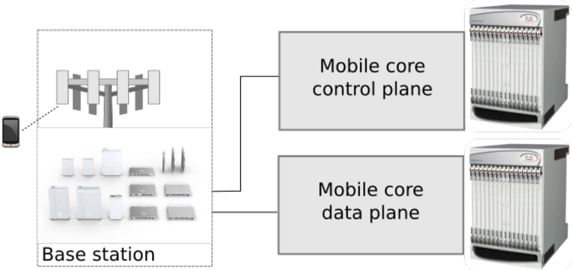
Cellular networks: hardware



- Specialized hardware

https://www.ericsson.com/en/portfolio/networks/ericsson-radio-system/ran-compute
https://www.cisco.com/c/en/us/products/wireless/asr-5000-series/

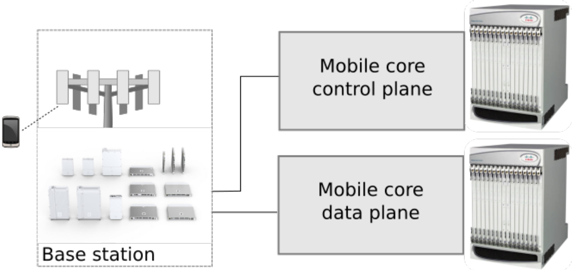
Cellular networks: hardware



- Specialized hardware
- 3GPP-standardized interfaces

https://www.ericsson.com/en/portfolio/networks/ericsson-radio-system/ran-compute
https://www.cisco.com/c/en/us/products/wireless/asr-5000-series/

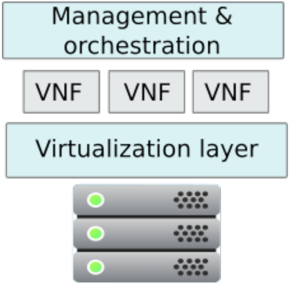
Cellular networks: hardware



- Specialized hardware
- 3GPP-standardized interfaces
- Closed, vendor-specific and proprietary software

https://www.ericsson.com/en/portfolio/networks/ericsson-radio-system/ran-compute
https://www.cisco.com/c/en/us/products/wireless/asr-5000-series/

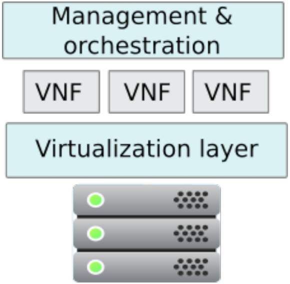
Softwarization of core network



- NFV introduced in 2013

G. Premsankar, K. Ahokas, S. Luukkainen. **“Design and implementation of a distributed mobility management entity on OpenStack”**, IEEE CloudCom 2015.

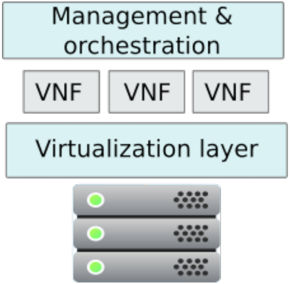
Softwarization of core network



- NFV introduced in 2013
- Virtualized network functions (VNFs) that run on commodity hardware

G. Premsankar, K. Ahokas, S. Luukkainen. **"Design and implementation of a distributed mobility management entity on OpenStack"**, IEEE CloudCom 2015.

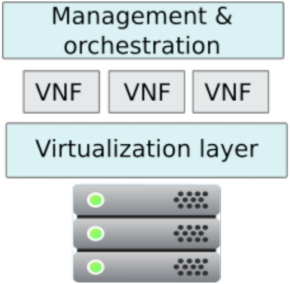
Softwarization of core network



- NFV introduced in 2013
- Virtualized network functions (VNFs) that run on commodity hardware
- Benefits: flexibility and scalability

G. Premsankar, K. Ahokas, S. Luukkainen. **"Design and implementation of a distributed mobility management entity on OpenStack"**, IEEE CloudCom 2015.

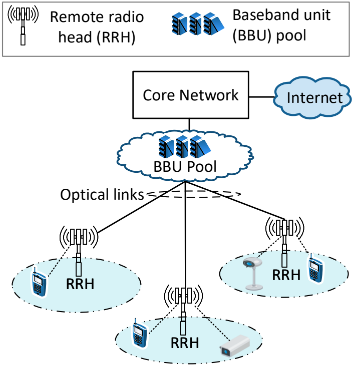
Softwarization of core network



- NFV introduced in 2013
- Virtualized network functions (VNFs) that run on commodity hardware
- Benefits: flexibility and scalability
- 5G architecture: cloud-native functions that can be joined as service chains

G. Premsankar, K. Ahokas, S. Luukkainen. "**Design and implementation of a distributed mobility management entity on OpenStack**", IEEE CloudCom 2015.

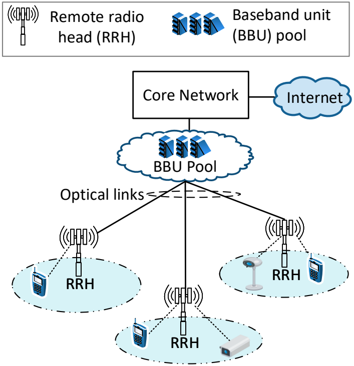
Softwarization of the access network



- RRHs and centralized BBU pools

B. Jedari, G. Premsankar, G. Illahi, M. Di Francesco, A. Mehrabi, A. Ylä-Jääski. **"Video Caching, Analytics and Delivery at the Wireless Edge: A Survey and Future Directions."** IEEE Communications Surveys & Tutorials (2020).

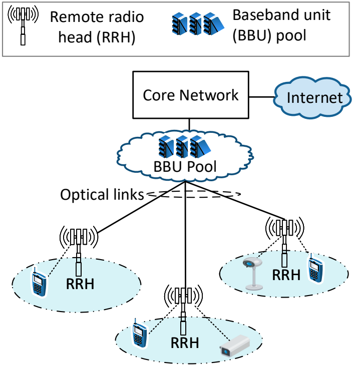
Softwarization of the access network



- RRHs and centralized BBU pools
- 5G: further split BBU into lower (distributed units) and higher protocol layers (central units)

B. Jedari, G. Premsankar, G. Illahi, M. Di Francesco, A. Mehrabi, A. Ylä-Jääski. **"Video Caching, Analytics and Delivery at the Wireless Edge: A Survey and Future Directions."** IEEE Communications Surveys & Tutorials (2020).

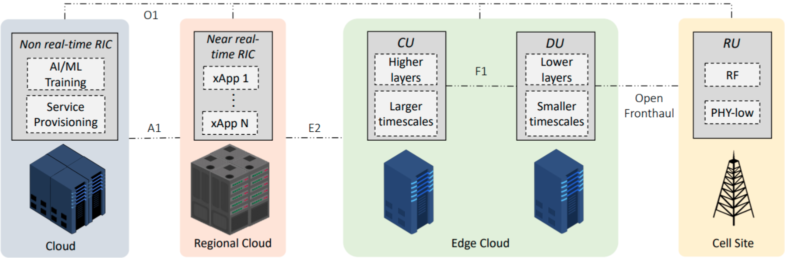
Softwarization of the access network



- RRHs and centralized BBU pools
- 5G: further split BBU into lower (distributed units) and higher protocol layers (central units)
- Vendor-specific implementation

B. Jedari, G. Premsankar, G. Illahi, M. Di Francesco, A. Mehrabi, A. Ylä-Jääski. **"Video Caching, Analytics and Delivery at the Wireless Edge: A Survey and Future Directions."** IEEE Communications Surveys & Tutorials (2020).

Open and inter-operable RAN



- O-RAN: open interfaces in radio access network
- Support for AI / ML models at different levels depending on the time-scale required

L. Bonati, S. D'Oro, M. Polese, S. Basagni, T. Melodia. "Intelligence and Learning in O-RAN for Data-driven NextG Cellular Networks." arXiv preprint arXiv:2012.01263 (2020).

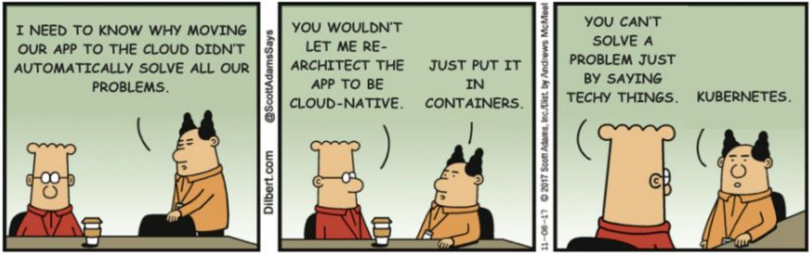
How does it all fit together

- Network slicing

How does it all fit together

- Network slicing
- Virtualization supports dynamic and flexible operation
- End-to-end management of both radio and core network resources

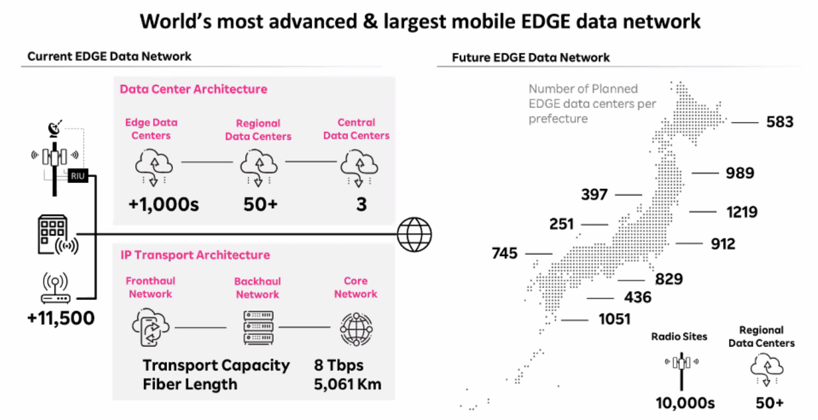
How does it all fit together



Where will cellular operators deploy edge

- Placement depends on use case, latency, resource constraints
- Mobile operators can leverage existing infrastructure: central offices, distributed antenna hub
- More distributed approach followed by non-incumbent operators

Where will cellular operators deploy edge



Source: `https://www.lightreading.com/the-edge/rakuten-dish-network-and-akamai-chart-future-of-edge-computing/d/d-id/767811?`

Summary: Edge computing in 5G

- Blurring of networking, hardware and software
- Disaggregation, virtualization and commoditization
- Access to radio and low-level information in edge and control applications

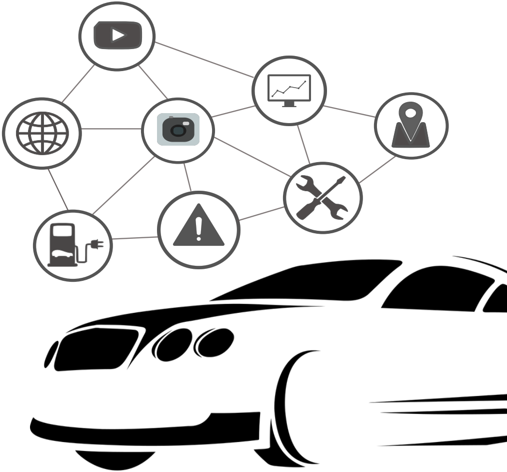
Resources

- 5G Mobile Networks: A Systems Approach by Larry Peterson and Oguz Sunay
  https://5g.systemsapproach.org/
- Linux Foundation's open glossary of edge computing
  https://github.com/State-of-the-Edge/glossary
- B. Jedari, G. Premsankar, G. Illahi, M. Di Francesco, A. Mehrabi, A. Ylä-Jääski. "Video Caching, Analytics and Delivery at the Wireless Edge: A Survey and Future Directions." IEEE Communications Surveys & Tutorials (2020)

Edge computing:
optimal placement for connected cars

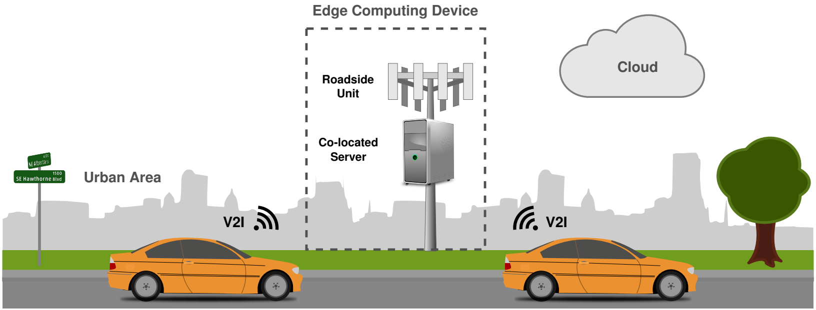
Connected vehicles in smart cities

- Autonomous cars
- Collision avoidance
- Dynamic real-time routing of vehicles
- Real-time computer vision applications



Edge computing for connected cars

- Smart road side units (RSUs) equipped with servers



G. Premsankar, B. Ghaddar, M. Di Francesco, R. Verago. “**Efficient Placement of Edge Computing Devices for Vehicular Applications in Smart Cities**” IEEE/IFIP NOMS 2018. Best student paper award.

Edge computing for connected vehicles: challenges

- Mobile vehicles need continuous connectivity

Edge computing for connected vehicles: challenges

- Mobile vehicles need continuous connectivity
- Low latency communication

Edge computing for connected vehicles: challenges

- Mobile vehicles need continuous connectivity
- Low latency communication
- Increasing amount of data to be processed

Objective

How to deploy a network of edge computing devices in a city for connected cars?

Constraints:

- Mobile vehicles need continuous connectivity
- Dense and built-up environment
- Limited computing resources at edge

Urban area deployments

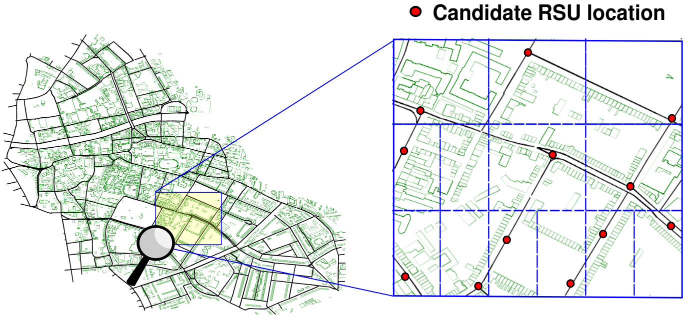
Optimal placement

Propose a mixed integer linear programming formulation for efficient deployment of edge computing devices

Features:

- Specify target network and computational demand coverage
- Accurately characterizes the effect of built-up environment on communications
- Uses open public data

System model



- Area modelled as a set of cells $\mathcal{C}$
- Each cell has a candidate location for installing an RSU
- RSUs can be deployed with different power levels $\mathcal{P}$
- Wireless communication over IEEE 802.11p

Optimization problem: RSU-OPT

A mixed integer linear programming model to minimize cost of deploying RSUs while ensuring:

- target network coverage γ
- target computational demand α

Decision variables:

- y_i Place RSU in cell i
- $x_{i,k}$ Transmit power level k to assign to RSU in cell i

RSU-OPT: objective

$$\textit{minimize} \quad \boxed{\sum_{i=1}^{|\mathcal{C}|} c y_i} + \boxed{\sum_{i=1}^{|\mathcal{C}|} \sum_{j=1}^{|\mathcal{C}|} a_{i,j} z_{i,j}} + \boxed{\sum_{k=1}^{|\mathcal{P}|} \sum_{i=1}^{|\mathcal{C}|} b_k x_{i,k}}$$

- Fixed cost for each RSU
- Cost based on distance between RSU and covered cell
- Cost based on transmit power level

RSU-OPT: constraints for network coverage

Pre-computed adjacency matrix

$$\sum_{k=1}^{|\mathcal{P}|} A_{j,i,k} x_{j,k} \geq z_{i,j} \quad \forall i, j$$

$$\sum_{j=1}^{|\mathcal{C}|} z_{i,j} \geq h_i \quad \forall i$$

$$\sum_{i=1}^{|\mathcal{C}|} h_i l_i \geq \gamma \sum_{i=1}^{|\mathcal{C}|} l_i$$

Meet network coverage requirement for gamma percent of roads

$$\sum_{k=1}^{|\mathcal{P}|} x_{i,k} = y_i \quad \forall i$$

RSU-OPT: constraints for computational demand

$$\sum_{i=1}^{|\mathcal{C}|} h_i d_i \geq \alpha \sum_{i=1}^{|\mathcal{C}|} d_i$$

Meet computational demand for alpha percentage of area

$$\sum_{i=1}^{|\mathcal{C}|} z_{i,j} d_i \leq m y_j \quad \forall j$$

Do not exceed available CPU cycles at each RSU

Evaluation: simulation settings

- Area: city center of Dublin (3.2 by 3.1 square kilometers)
- Simulations using moderate and high traffic conditions
- Optimization problem solved using CPLEX

Evaluation: comparison with other approaches

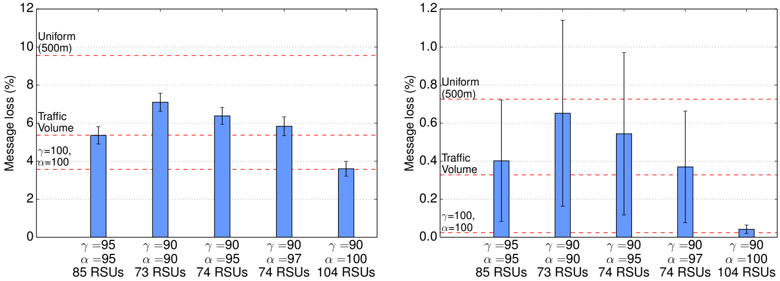
Baseline approaches:

- Every 500 m apart (primarily at intersections)
- Every 1 km apart
- Heuristic based on traffic volume:
  Deploy RSUs in cells with high traffic volume

Metrics:

- Number of RSUs deployed
- Message loss due to lack of network connectivity
- Message loss due to CPU capacity (cycles) exceeded at RSUs

Evaluation: RSU-OPT settings



- RSU-OPT (α = γ = 95) similar to traffic volume heuristic
- RSU-OPT (γ = 90) has lower loss than Uniform (500 m)
- Indicates number of RSUs required for target coverage

Summary: RSU placement

- Optimal placement of edge devices crucial for meeting network guarantees
- Even limited compute at the edge results in benefits for edge applications
- General model that can be extended to other wireless communication technologies

Energy-efficient edge computing

Optimization goals

- Switch off under-utilized resources

Optimization goals

- Switch off under-utilized resources
- Trade-off between energy and latency

Optimization goals

- Switch off under-utilized resources
- Trade-off between energy and latency
- Characterize requirements for AI applications at the edge

Orchestration frameworks

- Incorporate scheduling and placement decisions in state-of-the-art orchestration frameworks
- Improve expressiveness of orchestration policies

Resources

- Johannes Bisschop. AIMMS optimization modeling, 2006
- Ed Klotz and Alexandra M. Newman. "Practical guidelines for solving difficult mixed integer linear programs." Surveys in Operations Research and Management Science 18.1-2 (2013): 18-32.
- IBM academic initiative program `https://www.research.ibm.com/university/`
- NEOS Server `https://neos-server.org/neos/`