Mobility-Enhanced Edge inTelligence (MEET) for 6G

Zhisheng Niu

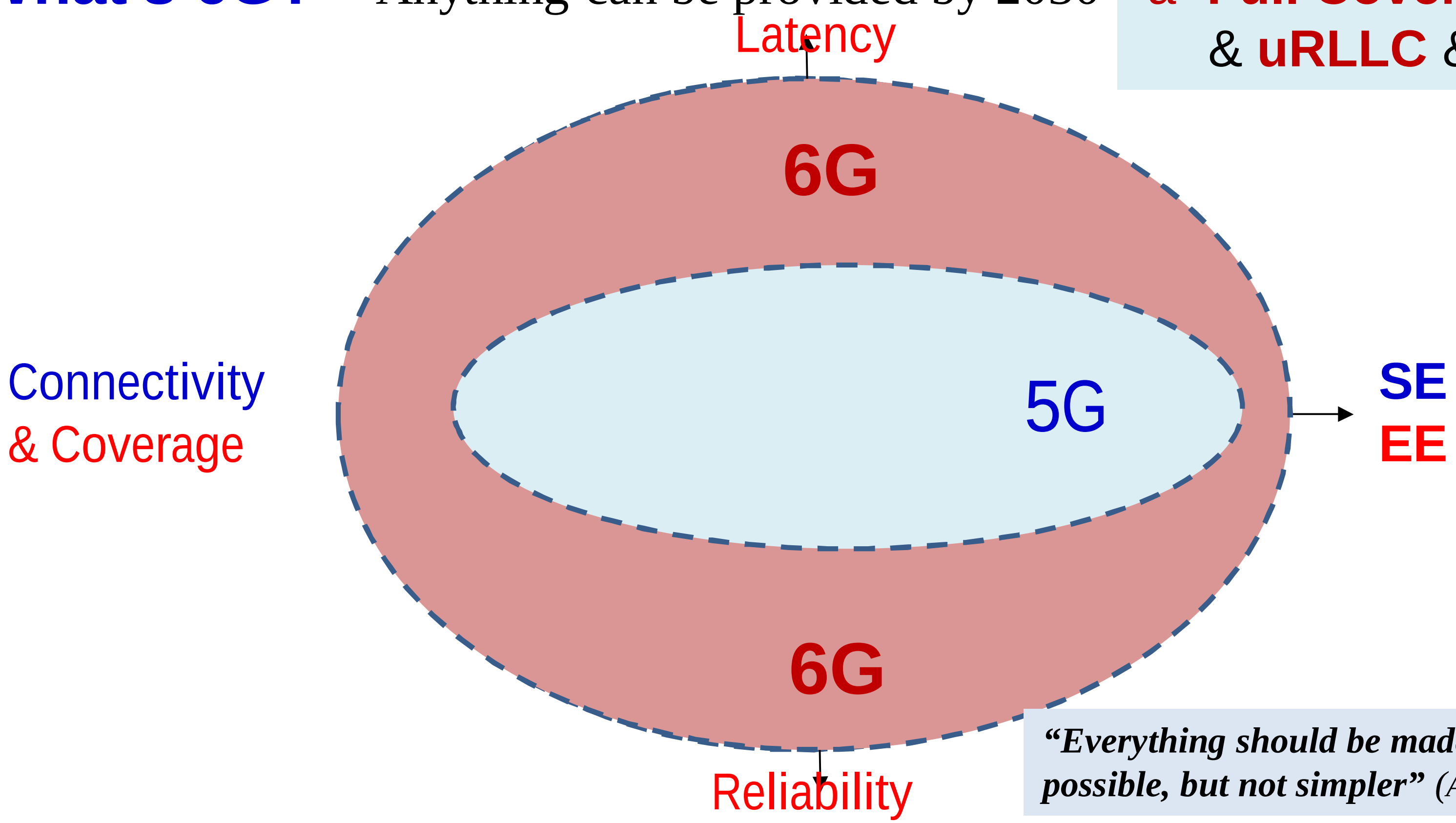
Department of Electronic Engineering
Tsinghua University

March 26, 2019

5G: What's Missing?

- **What's 5G? --** *Everything 4G couldn't provide* ***à There will be no 6G!***
  - It's EVERYTHING, like **ATM** (*Asynchronous Transfer Mode*)**!**
  - Everything is nothing & Nothing is everything ("***Duck Theorem***")
  - Realistically, anything can be provided by 2020 **à eMBB & MTC**
- **What's 6G?** -- Anything can be provided by 2030

à **Full Coverage** & **uRLLC** & **EE**



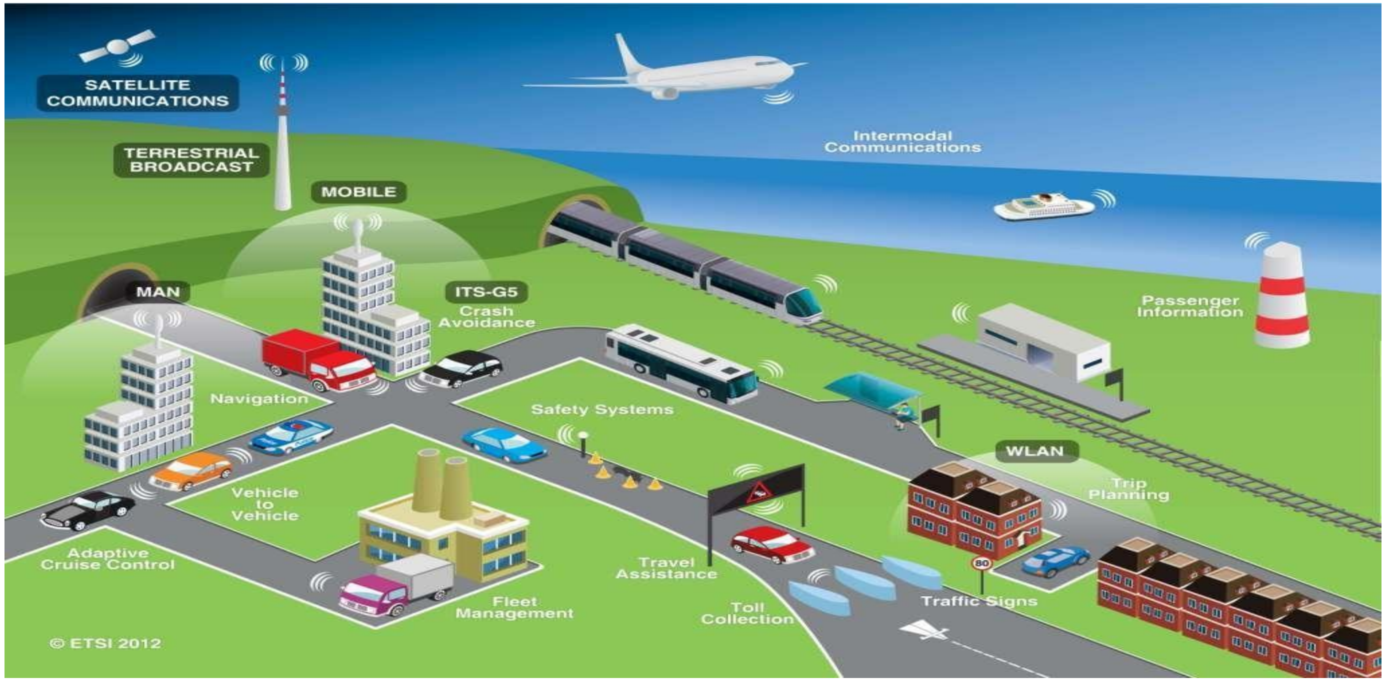
"Everything should be made as simple as possible, but not simpler" *(Albert Einstein)*

6G Vision

- **“6Genesis” by Academy of Finland, April 2018**
  - **3 strategic directions**
    - ü near-instant & unlimited wireless connectivity
    - ü distributed computing and intelligence
    - ü materials and antennas at very high frequencies
  - **First 6G Wireless Summit**
    - ü 24-26 March 2019, Levi, Lapland, Finland
- **3F’s in IMT-2020 (China)**
  - **FULL Coverage**
  - **FULL Spectrum**
  - **FULL Applications**

Space-Air-Ground Integrated Network

- **Key Challenges**
  - High-precision 3D localization **(~cm)**
  - Fusion of multi-dimensional & multi-scale sensing information
  - Near-instant context-information distribution **(~ms)**
  - Dynamic reconfiguration of heterogeneous resources & network functions



"**Space-Air-Ground Integrated Vehicular Networks for Immersive Driving Experience**" (**PI**), ***NSFC Principal Project*** (2017-2021)

uRLLC: a Grand New Challenge

- **Reliability & Latency: Inter-winded**
  - **Reliability**: Successful delivery probability within a deadline (*goodput*)
  - **Latency:** Reliable delivery as fast as possible
- **Reliability** *w/o* **latency-constraint** *or* **resource-limitation**
  - **Retransmission** (e.g., ARQ in Internet)
  - **Diversity** (e.g., multi-path routing, multi-channel transmission)
- **Low latency** *w/o (too high)* **reliability requirement**
  - **Short packet/frame** (finite blocklength)
  - **Blocking or alternative routing** (e.g., telephone network)

ultra Reliable AND Low-Latency with Limited Resources!

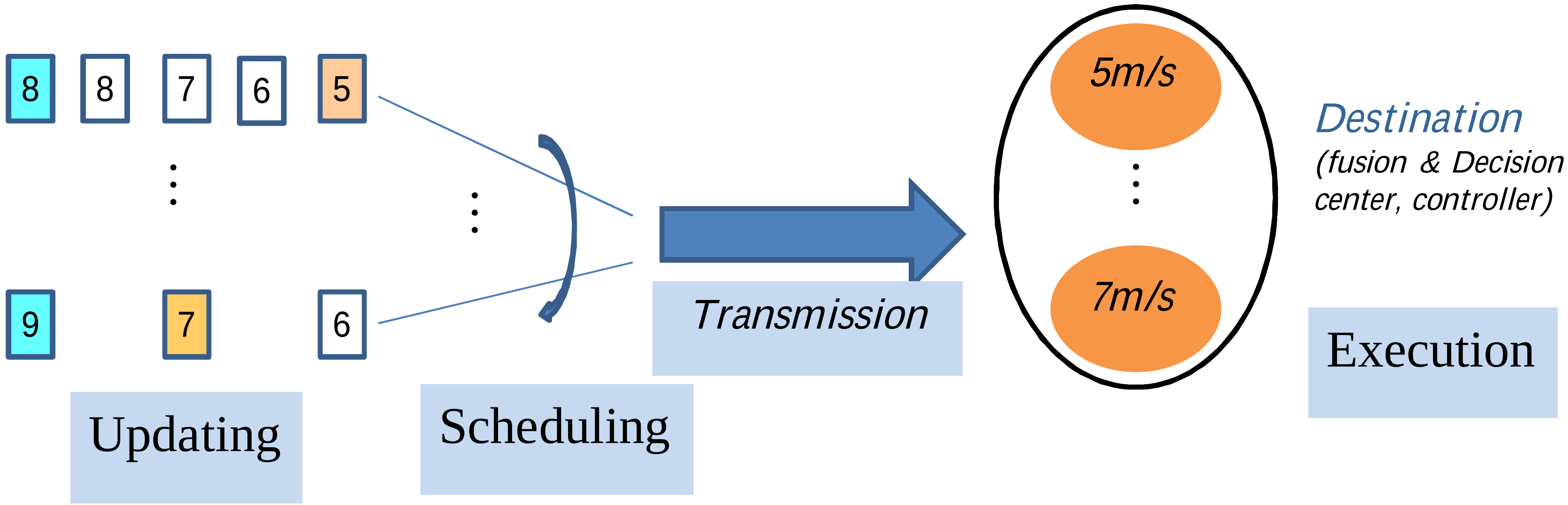
Latency: a Puzzling Word

- **Air Interface vs E2E?**
  - E2E ***communication*** delay: packetizing, coding/modulating, transmitting, decoding/demodulating, de-packetizing, fronthaul/backhaul, …..
  - You can't just move the bottleneck to others
- **Communication vs Information?**
  - E2E ***information*** delay: sensing/learning, updating, scheduling, transmitting, execution, …
  - **Freshness** (**Age**) of Information
- **Mean vs Bound?**
  - Delay bound guarantee is much harder than mean delay guarantee
  - Delay bound violation probability: $P\{D>D_{max}\}<\varepsilon$ leads to reliability

Freshness of Information

Age of Context/Status Information

- Timely update on context/status information à ***crucial in decision-making systems***
- Acquisition à Scheduling à Transmission à Execution à Feedback



Z. Jiang, B. Krishnamachari, S. Zhou, Z. Niu, "**Decentralized status update for age-of-information optimization in wireless multiaccess channels**," *IEEE International Symposium on Information Theory (ISIT)*, 2018

X. Zheng, S. Zhou, Z. Jiang, Z. Niu, "**Closed-Form Analysis of Non-Linear Age-of-Information in Status Updates with an Energy Harvesting Transmitter**", ***IEEE Trans. Wire.ess Commun***., 2019 (*under revision*)

uRLLC: Context-awareness

- **Traditionally, networks designed for the worst-case**
- **In reality, the worst case is very rare**

Poisson (λ) → **Buffer(k)** → **Exp(μ)**

M/M/1(k) service system

$$\rho=\lambda/\mu$$

$$E[D] = \frac{\rho(1-\rho)\sum_{i=0}^{k} i\rho^{i}}{\lambda(1-\rho^{k+1})}$$

$$P\{D \geq Dmax\} = \rho e^{-\mu(1-\rho)Dmax} \leq \varepsilon$$

$$P_B = \rho^{k+1}(1-\rho)/(1-\rho^{k+2})$$

For k=2, λ=3, D_{max}=10ms

- If E[D]=10ms, then μ=18; If E[D]=1ms, then μ=56
- If $P_B \leq 10^{-3}$, then μ=28; If $P_B \leq 10^{-5}$, then μ=138
- If $\varepsilon=10^{-3}$, then μ=251; If $\varepsilon=10^{-5}$, then μ=621

"Context-aware uRLLC V2X for Connected & Automated Cars" (**PI**), ***Intel Collaborative Research Institute for Intelligent and Connected Automated Cars*** (2018-2021)

Context-aware uRLLC

- **Traditionally, networks designed for the worst-case**
- **In reality, the worst case is very rare**
- **uRLLC should be context-aware**

KPI	value	Scenario
Delay (e2e, status update packets)	1-10ms	• Automated overtaking & high density platooning • status updates for collaborative collision avoidance
	50ms	• See through (10 Mbit/s) & Bird's Eye view (40 Mbit/s)
	100ms	• Trajectory handshake
Reliability	10^{-5} 10^{-3}	• Automated overtake &High density platooning • Trajectory handshake
Positioning	10cm	• Vulnerable road user discovery
	30cm	• Automated overtake & High density platooning & Trajectory handshake

"Context-aware uRLLC V2X for Connected & Automated Cars" (**PI**), ***Intel Collaborative Research Institute for Intelligent and Connected Automated Cars*** (2018-2021)

uRLLC: Solutions *(There is no free lunch)*

- **Bring extra resources *(redundancy)* closer to UEs**
  - **Communication**: mmWave à coverage cost & blockage
  - **Compute**: cloud computing, ***edge computing***, traffic offloading, …
  - **Caching/Push**: mobile caching, ***edge caching,*** content push, …

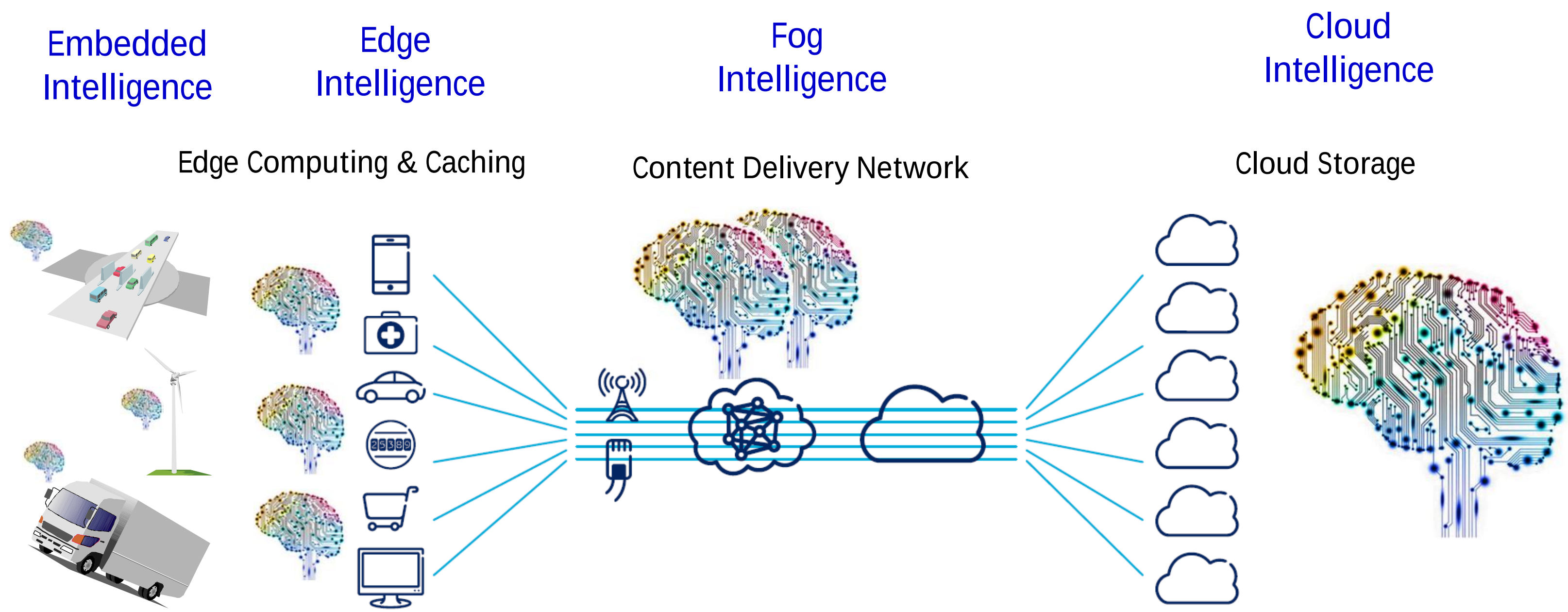
SDN/NFV enables 3C convergence

- **Bring extra information *(intelligence)* across network**
  - Traffic characteristics & QoS requirements
  - Network topology and conditions
  - Mobility information

BigData/AI enables distributed intelligence

Connected Intelligence via AI

- **Embed *intelligence* across whole network** *(access, routers, gateways, servers)* **to provide greater level of automation and adaptiveness** *(agility, resiliency, security, etc.)*



Make networks full of AI（愛）

Challenge: Who responsible for deployment/operation of edge clouds?

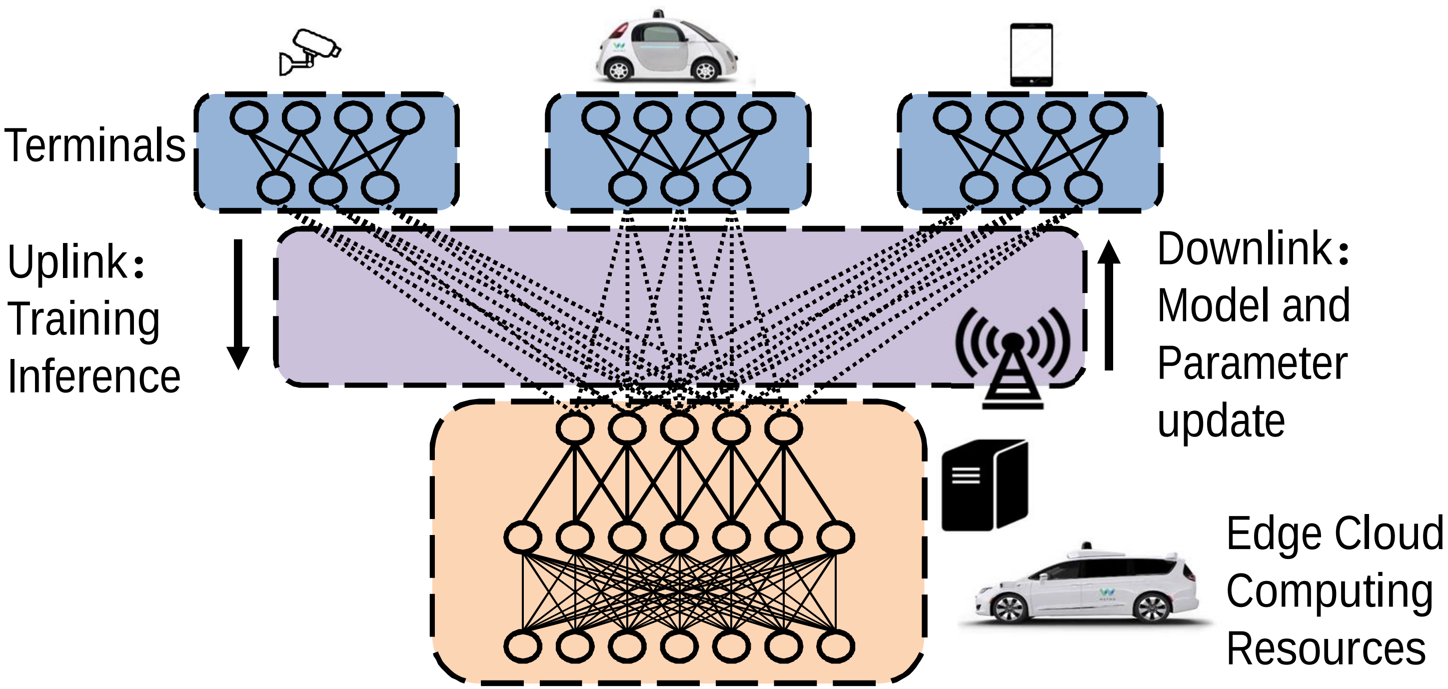
"Smart Networking in the Era of Artificial Intelligence" (**Co-PI**), ***NSFC–Scientific Foundation of Ireland*** (2018-2022)

Intelligent Vehicles for Smart Networking & City

- Autonomous vehicles with rich sensing, communication, computing, and caching capability, and, power supply, enabling them to be moving sensors NW, moving edge clouds, and even moving BSs
- Moving vehicles can bring Matters (people, goods), Energy, and Information (intelligence) to every corners of the city

Distributed Learning via Moving Vehicles

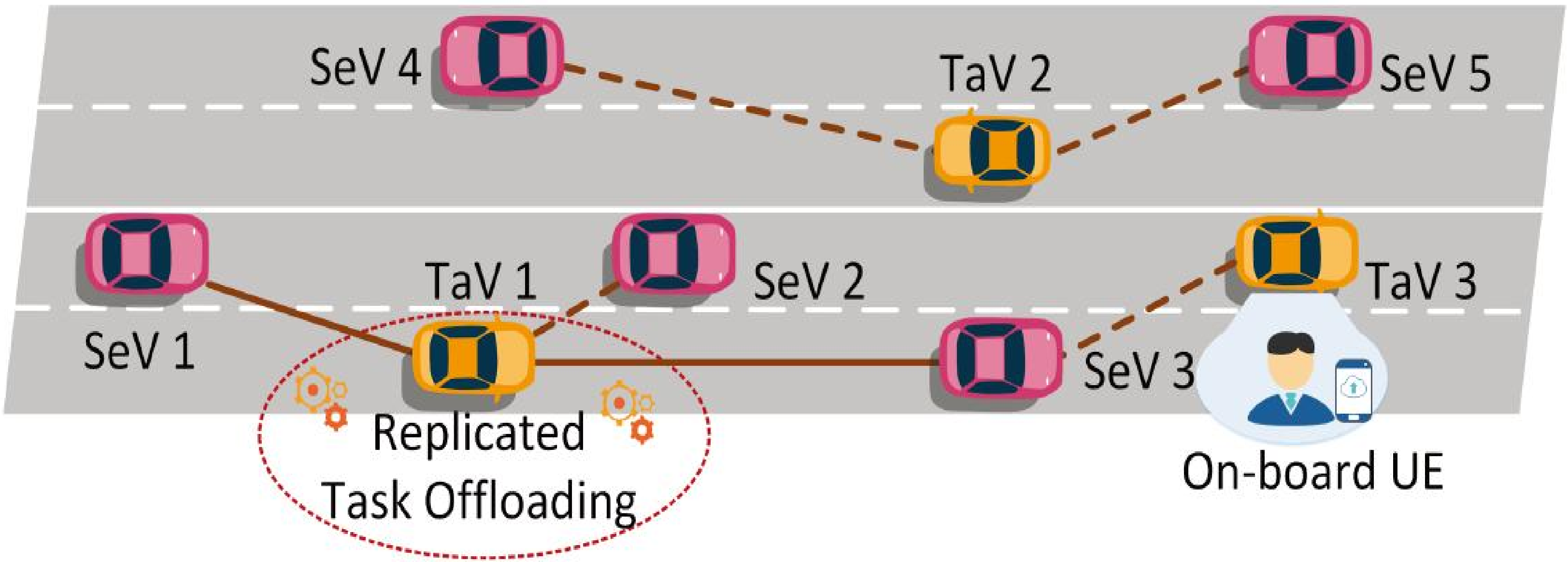
- **How can networking help AI? à Learning over the Air**
  - Cannot send **ALL** data to clouds
  - Limited **compute** and **storage** on embedded nodes
  - Leverage edge caching and computing to improve efficiency of deep learning models via wireless networks



W. Shi, Y. Hou, S. Zhou, Z. Niu, Y. Zhang, and L. Geng, "**Improving Device-Edge Cooperative Inference of Deep Learning via 2-Step Pruning**," ***IEEE INFOCOM'19 Workshop***, April 2019.

Augment Intelligence via Moving Vehicles

- Moving intelligence for *opportunistic* access & *swarm* intelligence
- **Mobility-Enhanced Edge inTelligence (MEET)**



When/where to offload/cache/push ***(when/where are the opportunities)*?**

How much gain can be achieved by moving clouds?

S. Zhou, Y. Sun, Z. Jiang, Z. Niu, "**Exploiting Moving Intelligence: Delay-Optimized Computation Offloading in Vehicular Fog Networks**", ***IEEE Commun. Mag***, 2019 *(accepted)*

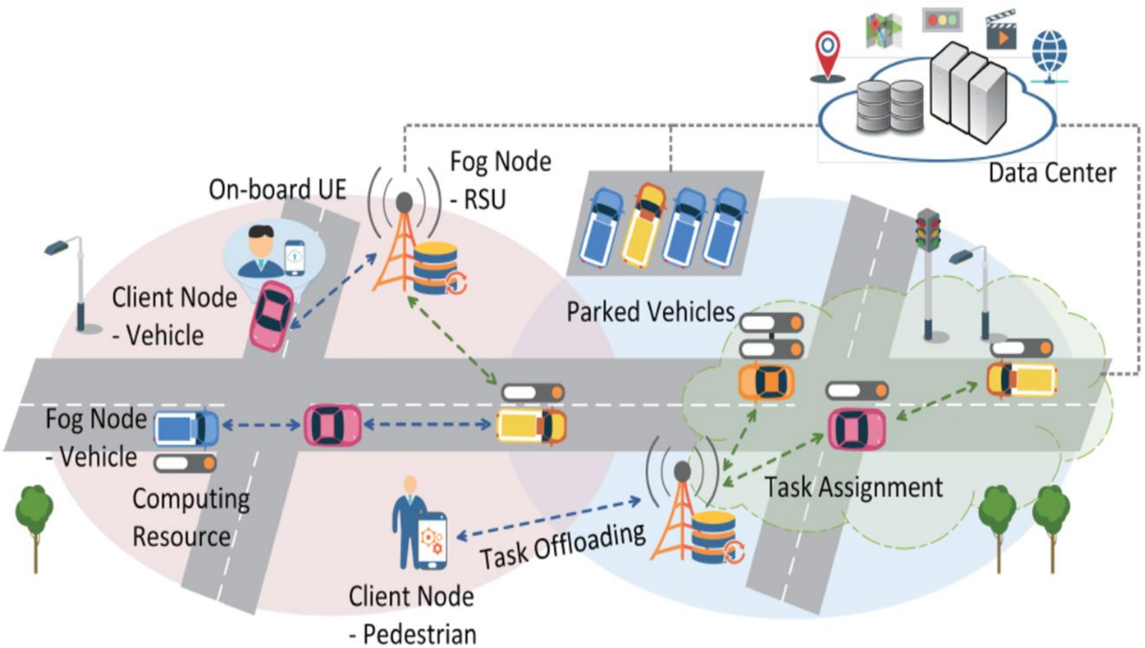
Mobility as the Opportunity

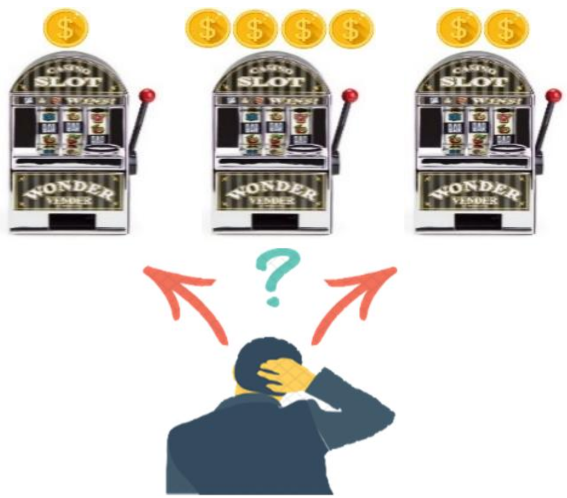
- "Mobility increases the capacity of ad hoc wireless networks" **(Grossglauser/Tse, Infocom, 2001)**
  - Mobility causes fast fading, bursty and non-uniform traffic
  - Mobility brings opportunities for good channel condition
- "Generalized Pollaczek-Khinchin Formula for Markov Channels" **(Huang/Lee, IEEE TCom, 2013)**
  - Fast fading channels improve performance
- "A Dynamic Programming Approach for Base Station Sleeping in Cellular Networks" **(Gong/Zhou/Niu, IEICE TCom, 2011) &** "Base-station sleeping control and power matching for energy-delay tradeoffs with bursty traffic" **(Wu/Niu, IEEE TVT, 2016)**
  - Non-uniform & Bursty traffic increases energy saving gain

Offloading while Learning

Adaptive & Volatile Multi-Armed Bandit

- **Explore** more when **load** is light and **opportunity** is rich
- **Exploit** more when **load** is heavy or **opportunity** is rare





Multi-armed bandit (MAB)

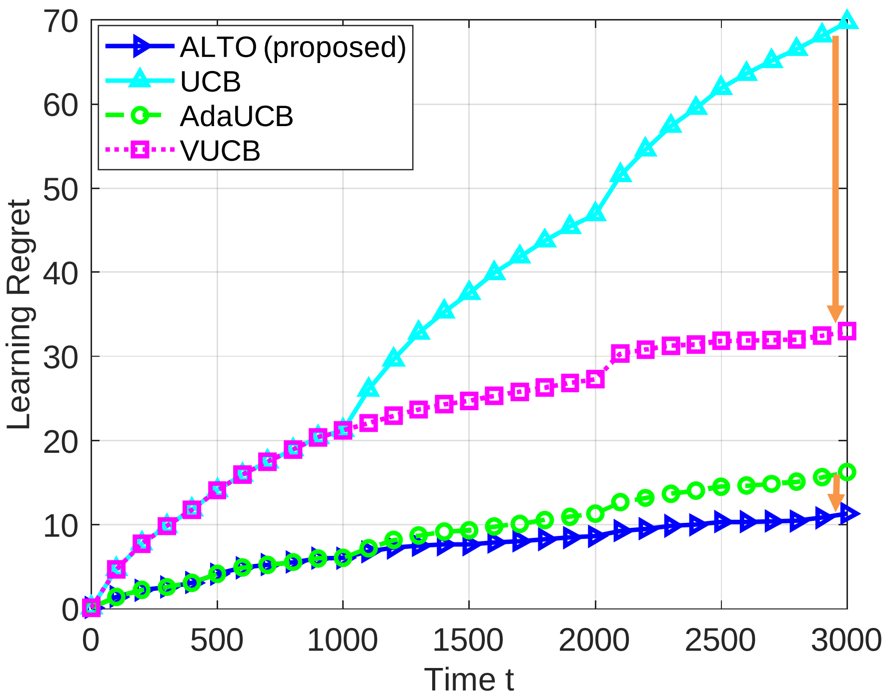
Y. Sun, X. Guo, S. Zhou, Z. Jiang, X. Liu, and Z. Niu, "**Learning-Based Task Offloading for Vehicular Cloud Computing Systems**", ***IEEE ICC'18***. May 2018

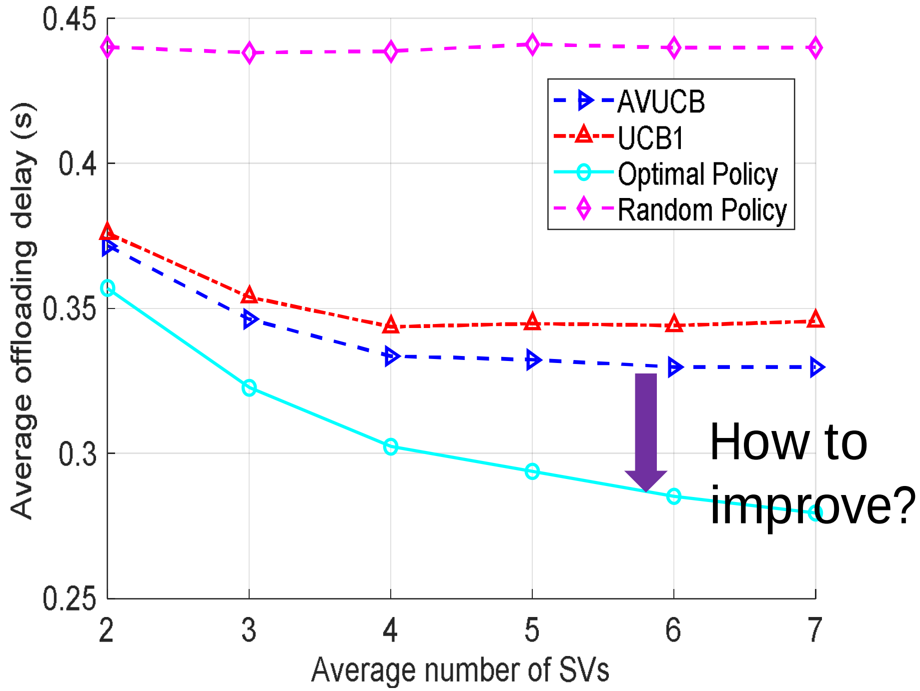
Y. Sun, X. Guo, J. Song, S. Zhou, Z. Jiang, X. Liu, and Z. Niu, ''**Adaptive learning-based task offloading for vehicular edge computing systems**," ***IEEE Trans. Veh. Technol.***, 2019 (accepted)

Single Task Offloading in MEC

Adaptive Learning-based Task Offloading (ALTO)

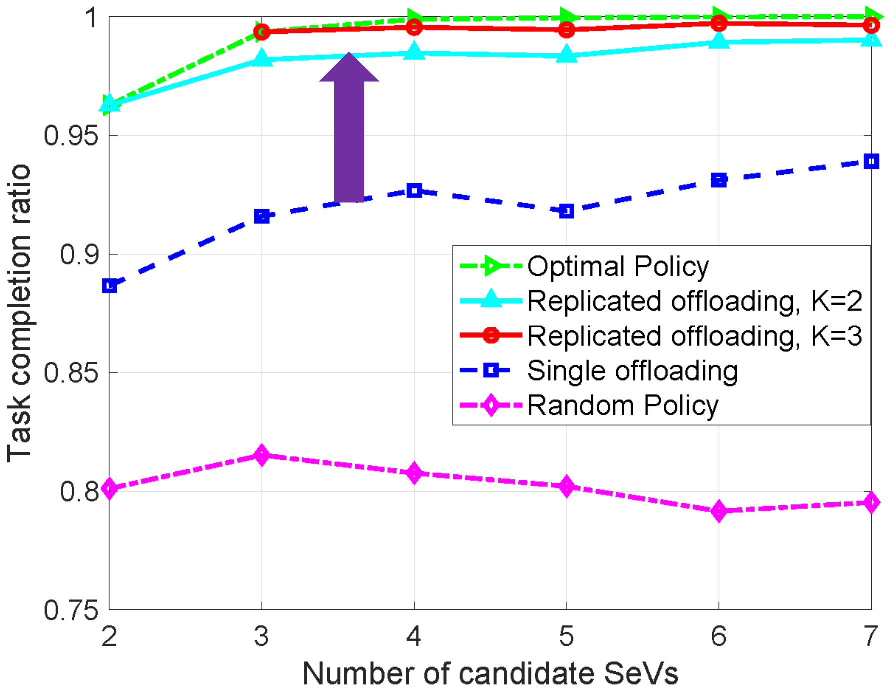
- Benchmark: Upper Confidence Bound (UCB) algorithm
- Opportunistic: Volatile UCB
- load-aware: AdaUCB
- Load-aware & Opportunistic: ALTO

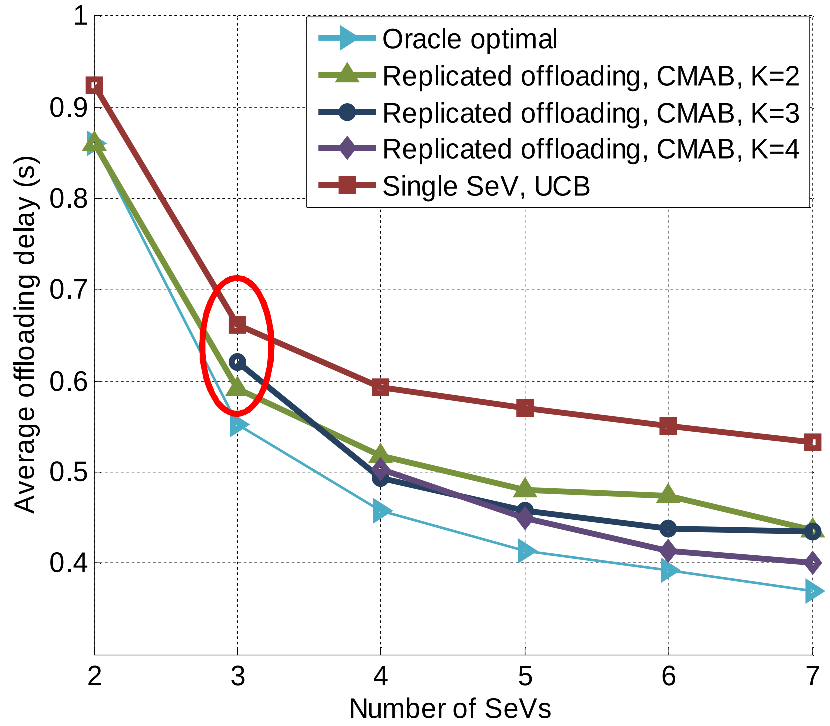




Replica Offloading with VCC

- **Replica offloading gets close-to-optimal performance**
- **Too many replicas not efficient if SeVs not enough**

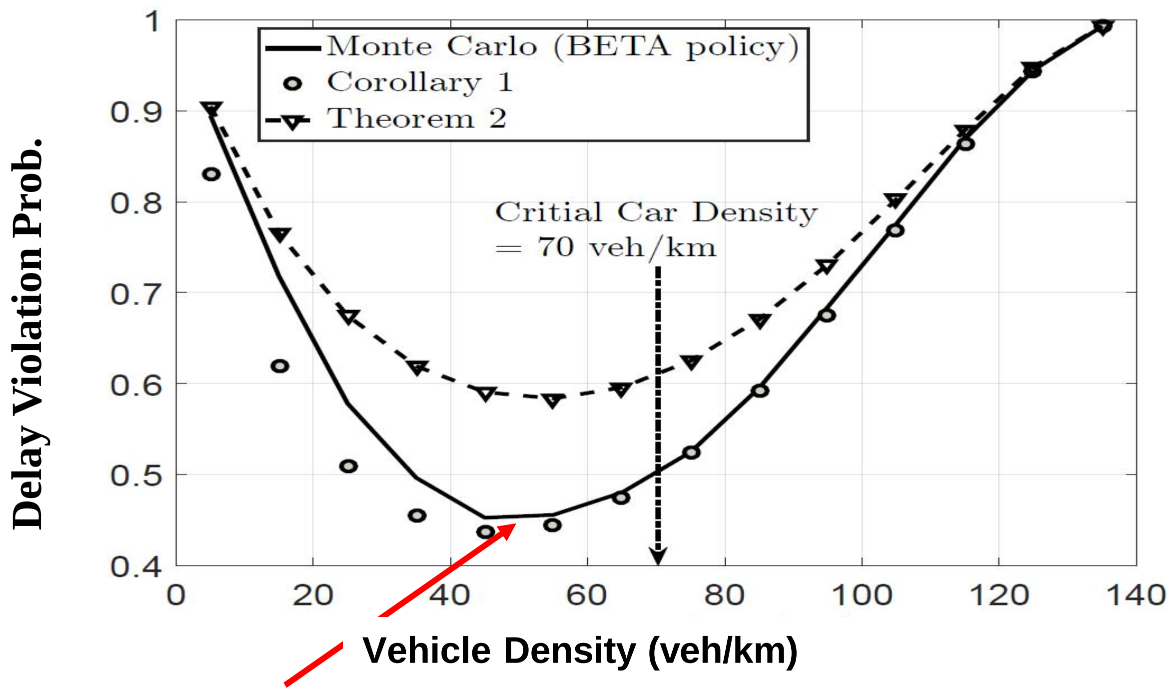




Y. Sun, X. Guo, S. Zhou, Z. Jiang, X. Liu, Z. Niu, "**Learning-based task replication for vehicular cloud computing systems**", ***IEEE Globecom, 2018***.

Smart Mobility

- **Optimal SeV density?**
  - Higher SeV density à more computing opportunities à traffic jam à Less computing opportunities

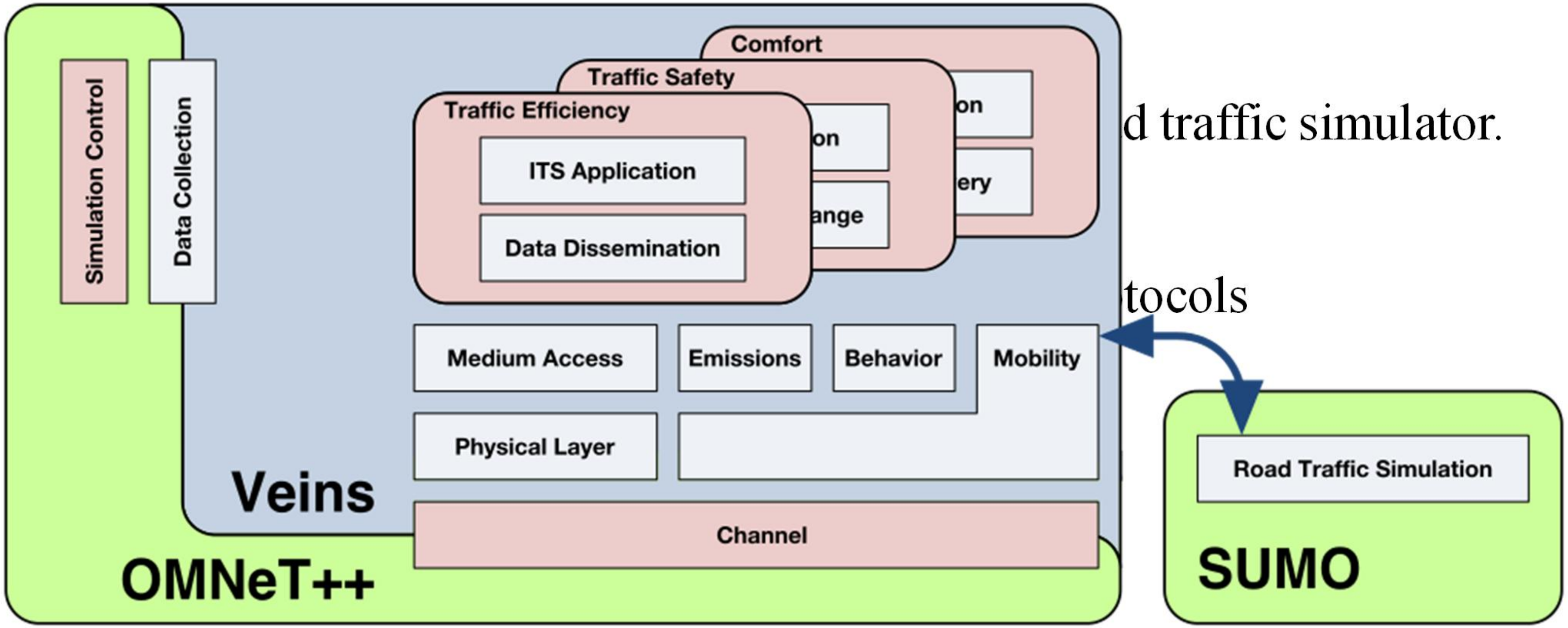


Optimum vehicle density for VCC is slightly smaller than traditional critical vehicle density to maximize traffic efficiency!

Z. Jiang, S. Zhou, X. Guo, Z. Niu, "**Task Replication for Deadline-Constraint Vehicular Cloud Computing: Optimal Policy, Performance Analysis and Implications on Road Traffic,**" ***IEEE IoT J***. Feb. 2018

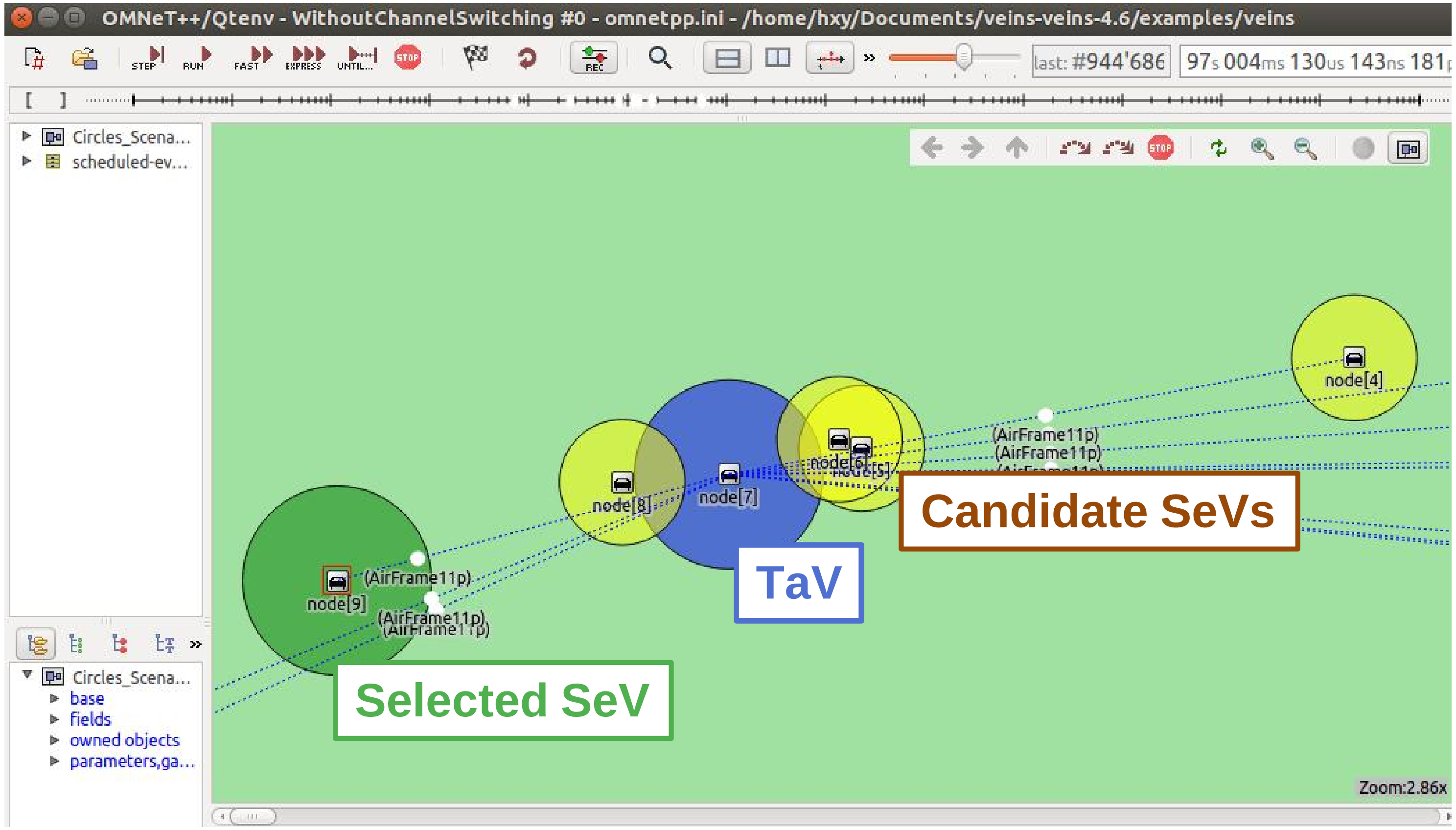
System Level Simulator

- **Veins (Vehicles in Network Simulations)**
  - An open source framework for running vehicular network simulations.
  - Obtain map and traffic information from SUMO.
  - Build PHY, MAC layer for simulation in OMNeT++.
- d traffic simulator.
- tocols



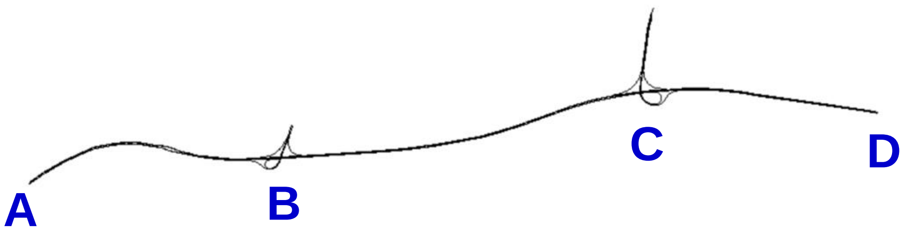
System Level Simulator

- 12 km stretch of G6 Highway in Beijing from Open Street Map.

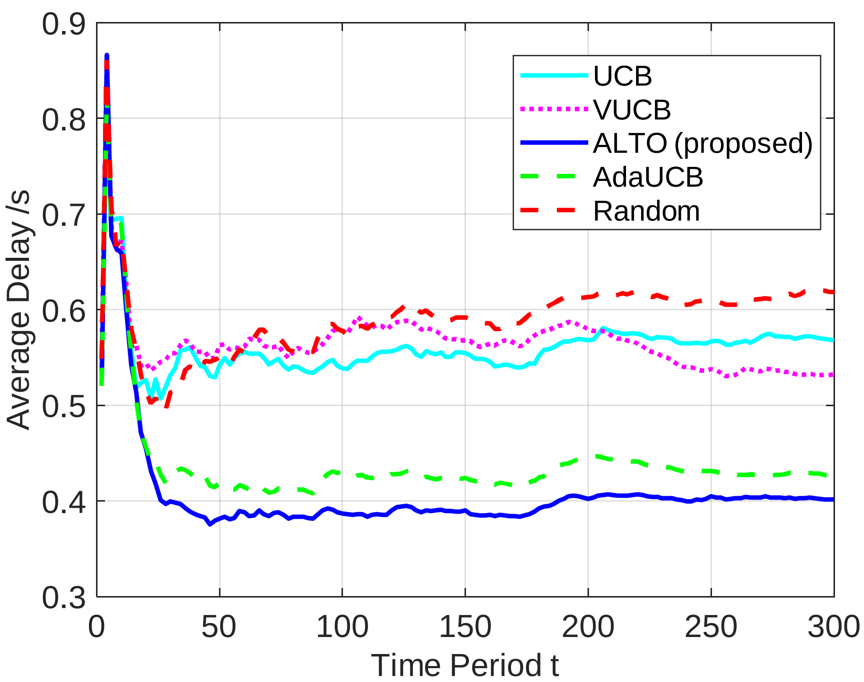


Simulation Results

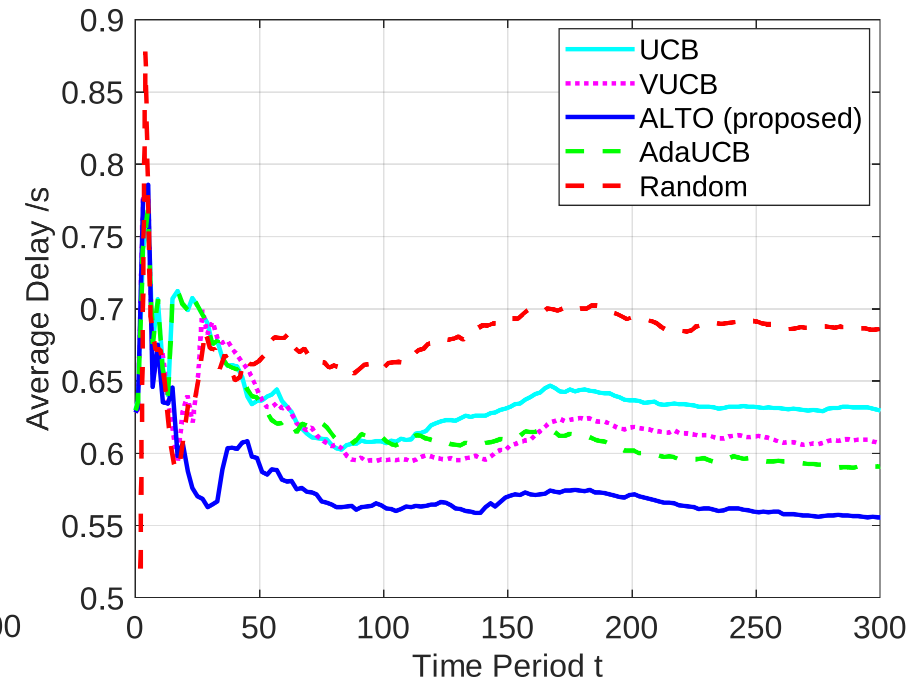
- **Algorithms comparison**



Routes	SeV arrival rate /s
A -> D	0.1
A -> C	0.05
B -> D	0.05



1 TaV



10 TaVs, inter-arrival time=10s

Smart Mobility for Intelligence-on-Demand

- **Coverage-oriented navigation & Coverage-on-demand**
- **Service-oriented navigation & Service-on-demand**

Today:	Shortest distance;	Min. travel time;	Highway-first;	Avoid congestion
	⇓	⇓	⇓	⇓
Future:	Max. throughput;	Min. Latency;	Coverage-first;	Avoid hotspots

Summary

- **Space-Air-Ground Integrated Network (SAGIN) with moving intelligence will fulfill 6G**
- **Future 5G/6G network**
  - **Software-defined**
  - **Cloud/Edge-based**
  - **AI-enabled**
  - **Mobility-enhanced**

Mobility-Enhanced Edge inTellegence (MEET)