ITU KALEIDOSCOPE
SANTA FE 2018

*Machine learning for a 5G future*

# AI as a Microservice (AIMS) over 5G Networks

**Tai-Won Um, Gyu Myoung Lee, Jun Kyun Choi**
Chosun University
twum@chosun.ac.kr

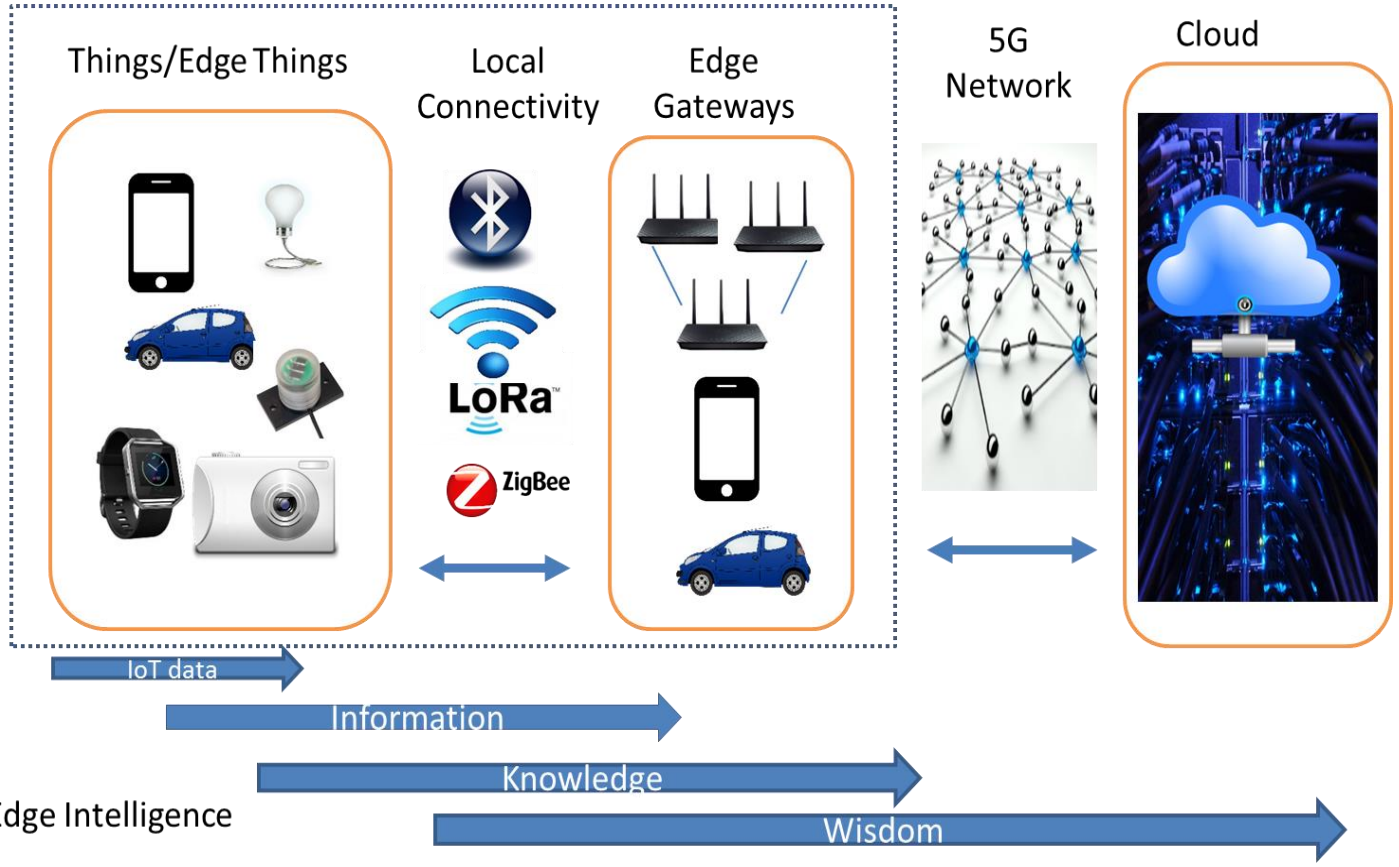
**26-28 November**
**Santa Fe, Argentina**

# Table of Contents

# Background

- As data-driven decision-making services are being infused into IoT applications at the 5G networks, AI algorithms are being deployed as monolithic application services for autonomous decision processes.
- How do we design, model and expose intelligent services for decision making at the Edge of Things to address latency related problems of AI services across the 5G networks?
- Intelligent decision making at the Edge of Things introduces new AI dimension to IoT services such as real-time local processing of IoT data for quick decision making without transporting the heavy data through the expensive 5G networks.
- The challenge here is how do we develop data-centric IoT Services in which AI is a first-class design element?

# Smart Edge of Things for Edge Intelligence

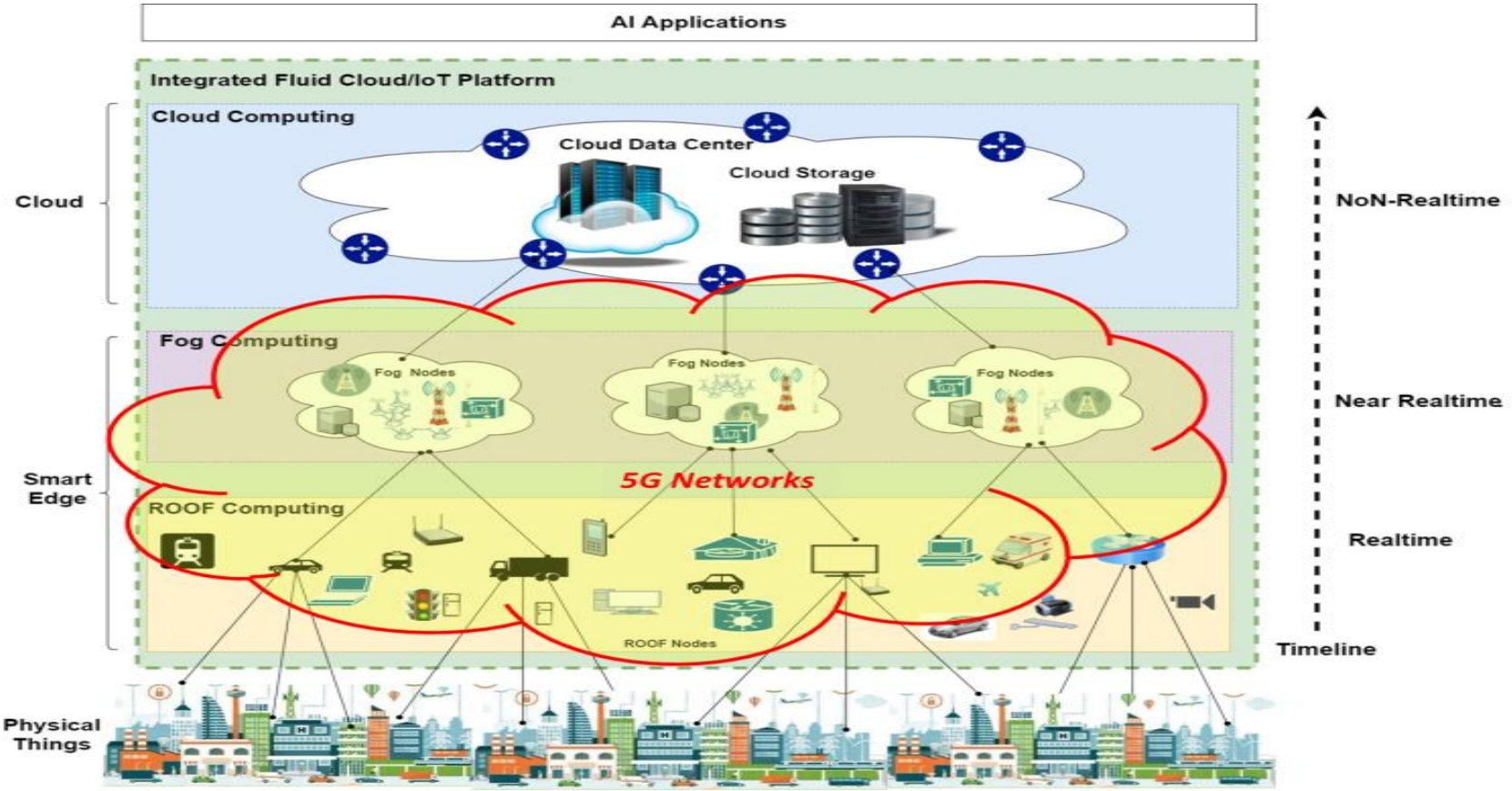



- Aggregate Data
- Filter Data and elimina outliers
- Extract Features from data
- Find patterns in the data
- Make decision (Knowledge)
- Get feedback and improve future decision (Wisdom)

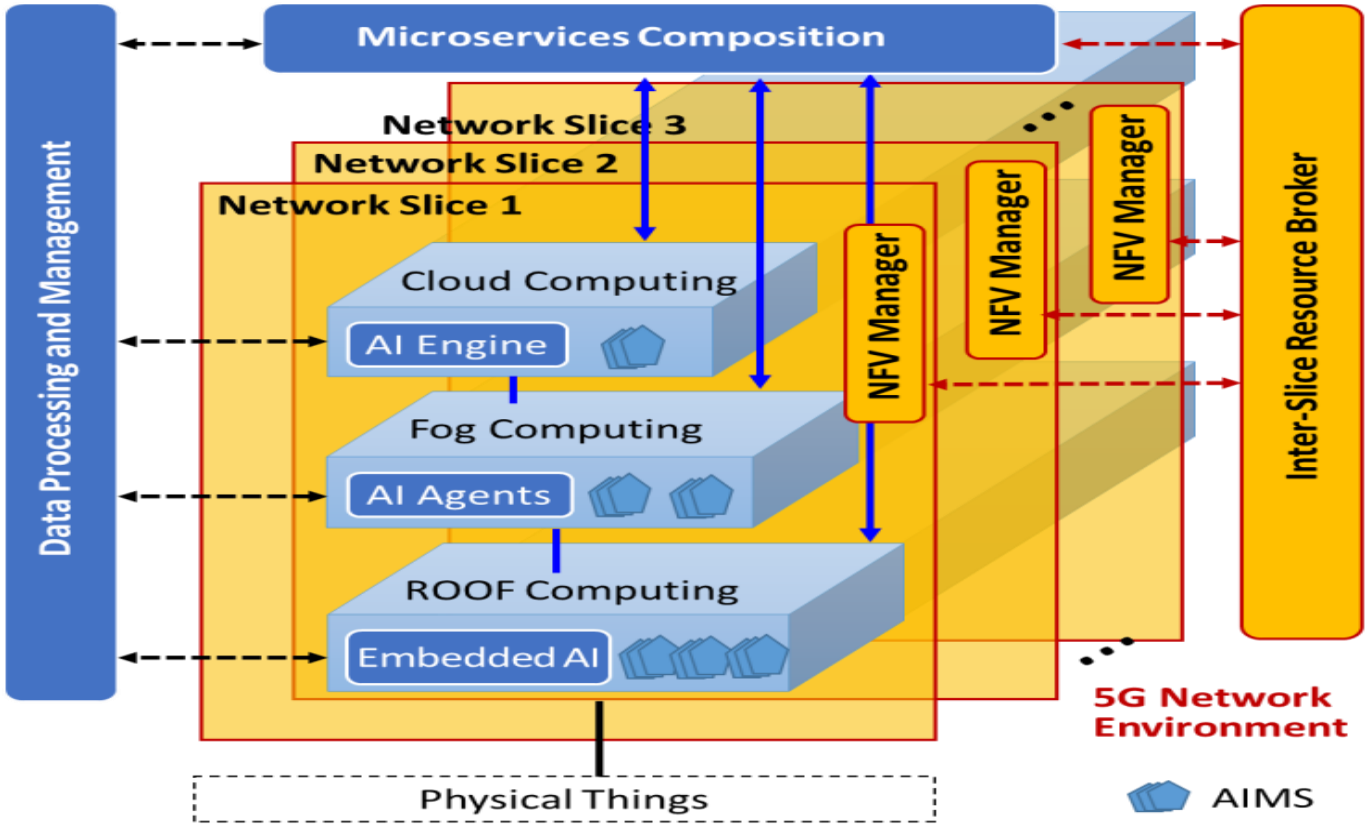
# AI as Microservices (AIMS) at the Edge of Things

- We proposes a hierarchically integrated infrastructure spanning the ROOF, Fog and Cloud computing platforms, to exploit resources at the Edge of Things and Cloud data centers, as well as microservice concepts to incorporate AI capabilities into IoT applications.
- The microservice concept allows the decomposition or factoring of the current monolithic AI services into smaller functions deployed as AI microservices.
- The microservices can then be deployed closer to the data sources and users allowing seamless composition of AI services across the ROOF-Fog-Cloud layer.
- The AI features and functions are composed from the distributed microservices AIMS integrated platform, allowing AI functionality as intelligence services to be implemented and deployed close to the data sources and users.
- The composition of the microservices can be realized sequentially based on NFV management and inter-slice resource brokering process.

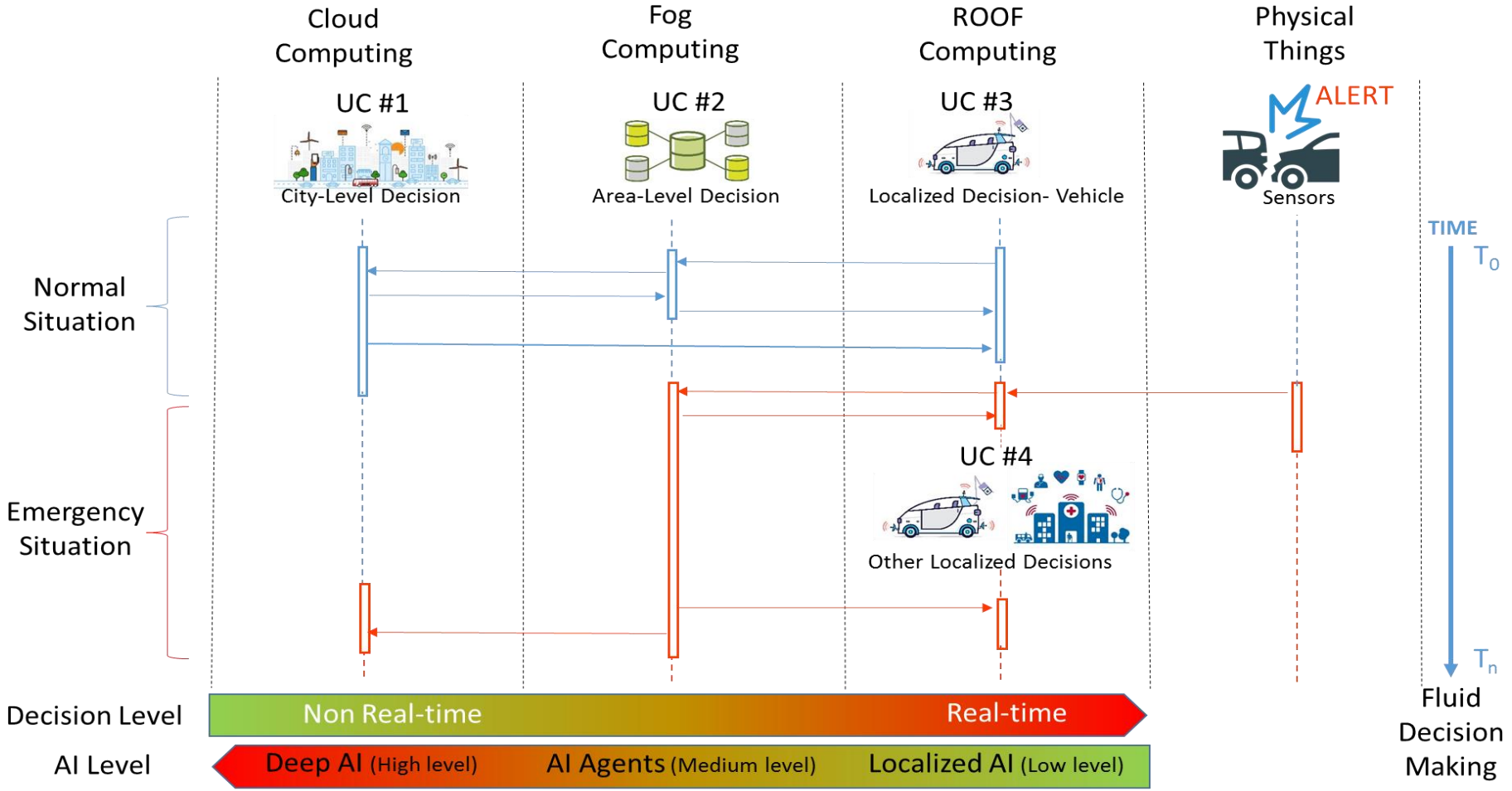
# Edge-Cloud high-level integrated architecture over 5G networks for various AI enabled IoT applications



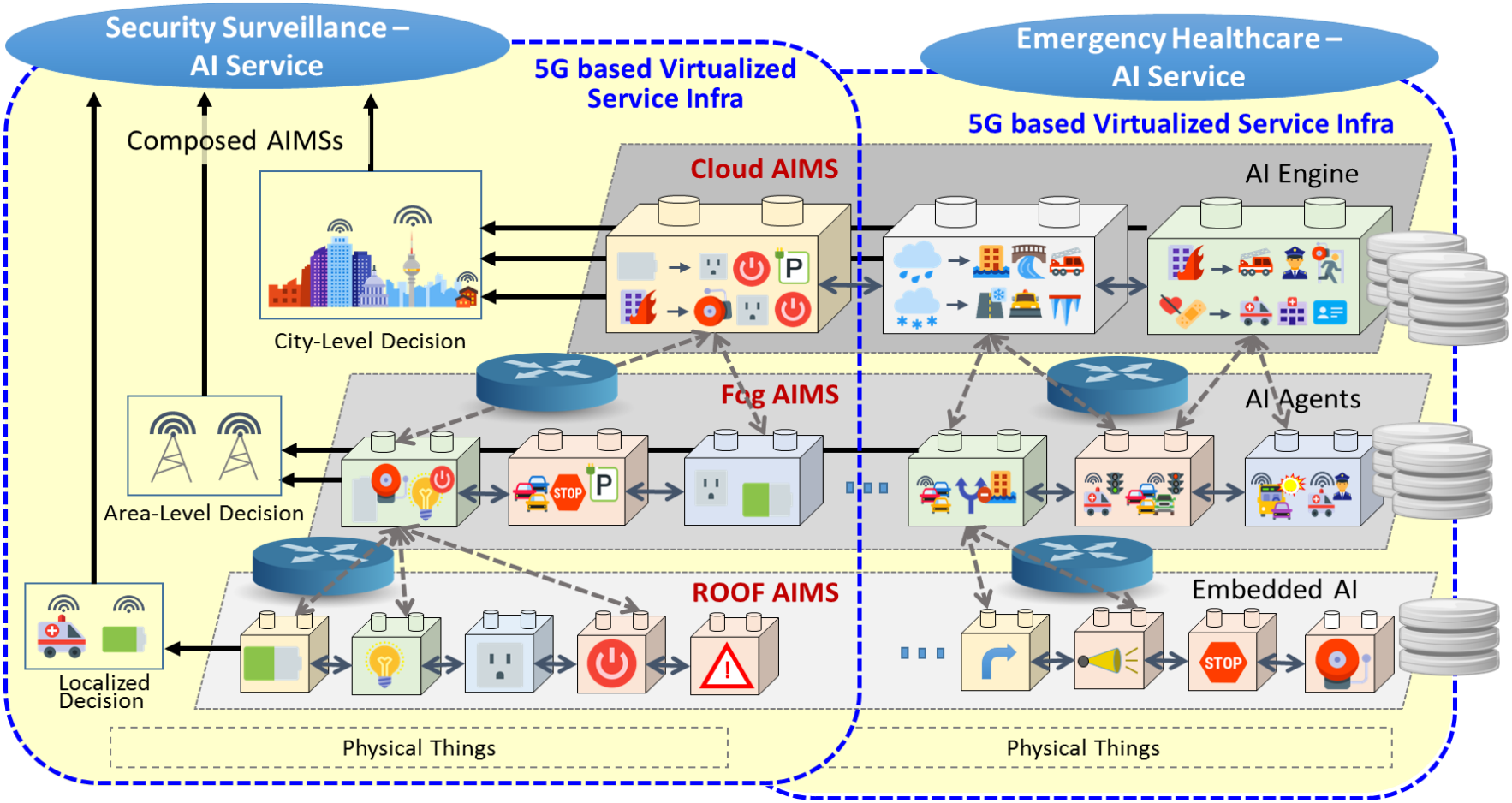
# The ecosystem of microservices distributed across the ROOF-Fog-Cloud systems over 5G networks



# AIMS Use Cases



# AIMS Use Cases



# Challenges

- Defining an integrated platform architecture for Cloud, AI and 5G
  - The AIMS aims to define and develop an integrated platform architecture for the incorporation of multi-clouds systems and IoT for AI based services.
- Specifying essential components and interfaces to support data-driven AI services
  - Various components and interfaces for communication across a federation of ROOF, Fog and Cloud platform would be specified.
- Supporting the harmonious management of computing resources
  - New mechanisms and strategies for dynamic and fluid resource allocation and scheduling would be investigated to reduce response time for task execution across the 5G integrated AIMS platform.
- Applying new mechanisms using intelligence in data lifecycle
  - How to determine based on the available resources, what tasks or functions should be executed and at what layer of the infrastructure are important technical challenges.
- Supporting trusted AI services
  - How can trust management be used to provide security, dependability and reliability for AIMS and associated data at various layers of the ROOF, Fog and Cloud integrated platform?

# References


- [1] A. Krylovskiy, M. Jahn and E. Patti, "Designing a Smart City Internet of Things Platform with Microservice Architecture," 2015 3rd International Conference on Future Internet of Things and Cloud, pp. 25-30, Aug. 2015.
- [2] "Microservices," https://en.wikipedia.org/wiki/ Microservices.
- [3] P1931.1, "Standard for an Architectural Framework for Real-time Onsite Operations Facilitation (ROOF) for the Internet of Things," https://standards.ieee.org/project/1931_1.html.
- [4] N. Mohan and J. Kangasharju, "Edge-Fog cloud: A distributed cloud for Internet of Things computations," 2016 Cloudification of the Internet of Things (CIoT), pp.1-6. 2016.
- [5] Marco Gramaglia, Albert Banchs, "Definition and specification of connectivity and QoE/QoS management mechanisms," 5G NORMA Deliverable D5.2, Jul. 2017.
- [6] P. Patel, M. Intizar Ali and A. Sheth, "On Using the Intelligent Edge for IoT Analytics," in IEEE Intelligent Systems, vol. 32, no. 5, pp. 64-69, Sep. 2017.
- [7] M. Fazio, A. Celesti, R. Ranjan, C. Liu, L. Chen and M. Villari, "Open Issues in Scheduling Microservices in the Cloud," in IEEE Cloud Computing, vol. 3, no. 5, pp. 81-88, Sep. 2016.
- [8] H. Liu, F. Eldarrat, H. Alqahtani, A. Reznik, X. de Foy and Y. Zhang, "Mobile Edge Cloud System: Architectures, Challenges, and Approaches," IEEE Systems Journal, vol. 12, no. 99, pp. 1-14, Feb. 2017.
- [9] A. Celesti, M. Fazio, M. Giacobbe, A. Puliafito and M. Villari, "Characterizing Cloud Federation in IoT," 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 93-98, Mar. 2016.
- [10] C. Long, Y. Cao, T. Jiang and Q. Zhang, "Edge Computing Framework for Cooperative Video Processing in Multimedia IoT Systems," in IEEE Transactions on Multimedia, vol. 20, no. 99, pp. 1126-1139, May 2018.
- [11] Y. Ai, M. Peng, and K. Zhang, "Edge cloud computing technologies for internet of things: A primer," Digital Communications Networks, vol. 4, iss. 2, pp. 77-86, Apr. 2017.
- [12] A. Taherkordi and F. Eliassen, "Poster Abstract: Data-centric IoT Services Provisioning in Fog-Cloud Computing Systems," 2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI), pp. 317-318. 2017.
- [13] James Kobiolus, "Building AI Microservices for Cloud-Native Deployment," https://wikibon.com /building-ai-microservices-for-cloud-native-deployments/.
- [14] M. Villari, M. Fazio, S. Dustdar, O. Rana, and R. Ranjan. "Osmotic Computing: A New Paradigm for Edge/Cloud Integration," IEEE Cloud Computing, pp. 76–83, 2016.
- [15] Qi Wang, et. al., "SliceNet: End-to-End Cognitive Network Slicing and Slice Management Framework in Virtualised Multi-Domain, Multi-Tenant 5G Networks," 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1-5, 2018.
- [16] "Edge X Foundry," https://wiki.edgexfoundry.org/ display/FA/EdgeX+Foundry+Project+Wiki.
- [17] A. Sill, "The Design and Architecture of Microservices," IEEE Cloud Computing, vol. 3, no. 5, pp. 76-80, Sept. 2016.
- [18] "BigClouT: Big data meeting Cloud and IoT for empowering the citizen clout in smart cities," http://bigclout.eu/.

# ITU KALEIDOSCOPE
## SANTA FE 2018

**Thank you**