

Compute First Networking (CFN) Scenarios and Requirements

draft-geng-rtgwg-cfn-req-00

L. Geng, China Mobile

P. Willis, BT

Background

Service providers are exploring the edge computing to achieve better response times and transfer rate by moving the computing towards the edge of the network. The service nodes from multiple edges normally have two key features:

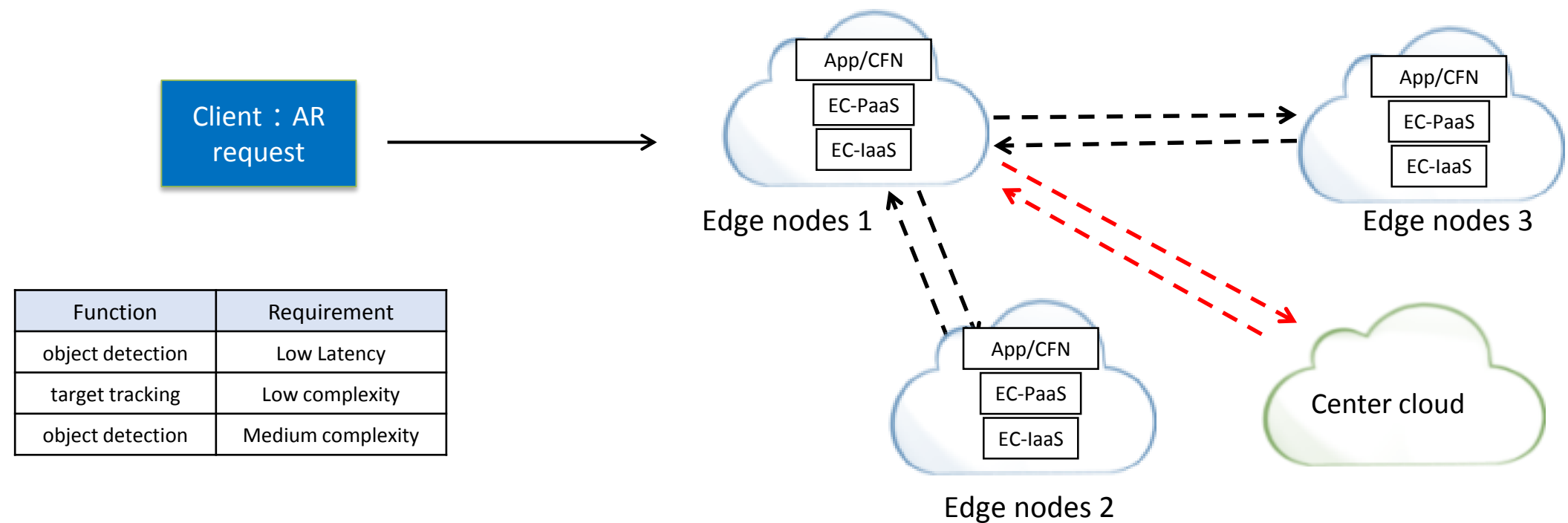
- service equivalency
- service dynamics

Compute First Networking is proposed to meet the above needs, and the service request to edge computing needs to be dispatched to and served by the most suitable edge, considering:

- available computing resources
- network conditions

- This draft describes **scenarios** and **requirements** of CFN.

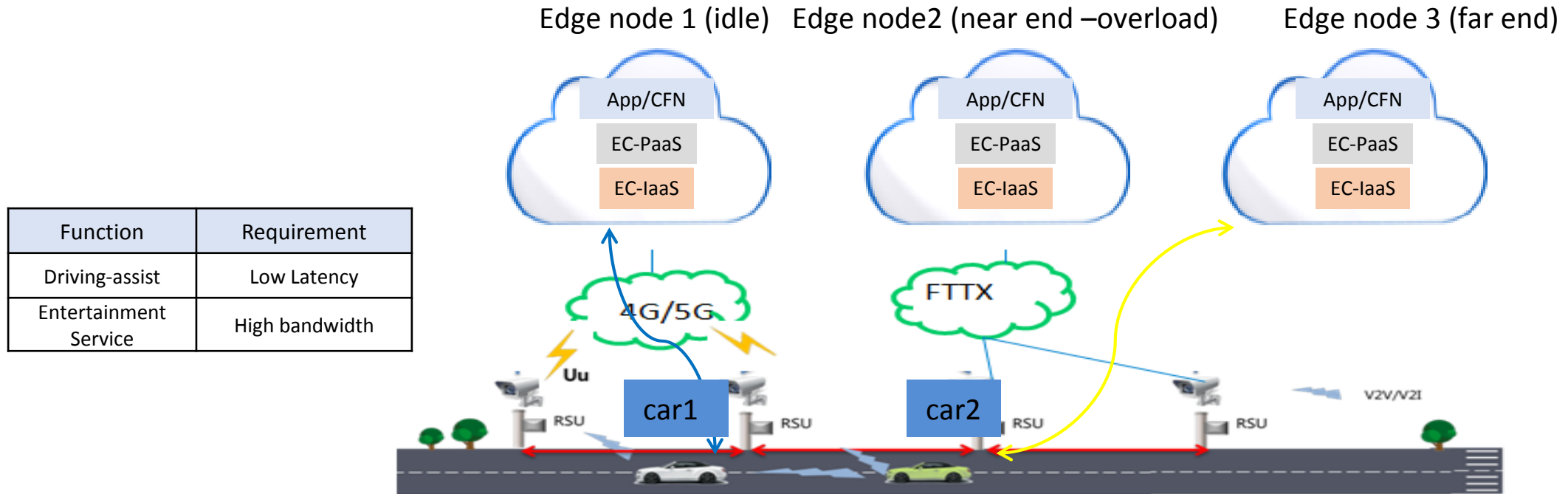
Usage Scenarios 1 - Cloud Based Recognition in Augmented Reality (AR)



In AR scenario, the computing resource for a specific service at the edge can be instantiated on-demand. Once the task is completed, the resource can be released.

- Cloud service nodes: tasks with the most intensive computational complexity like object recognition or latency non-sensitive tasks like AI based model training.
- Edge service nodes: tasks with medium computational complexity or low latency requirement like object detection, feature extraction and template matching;
- End device: only handles the tasks like target tracking and image display.

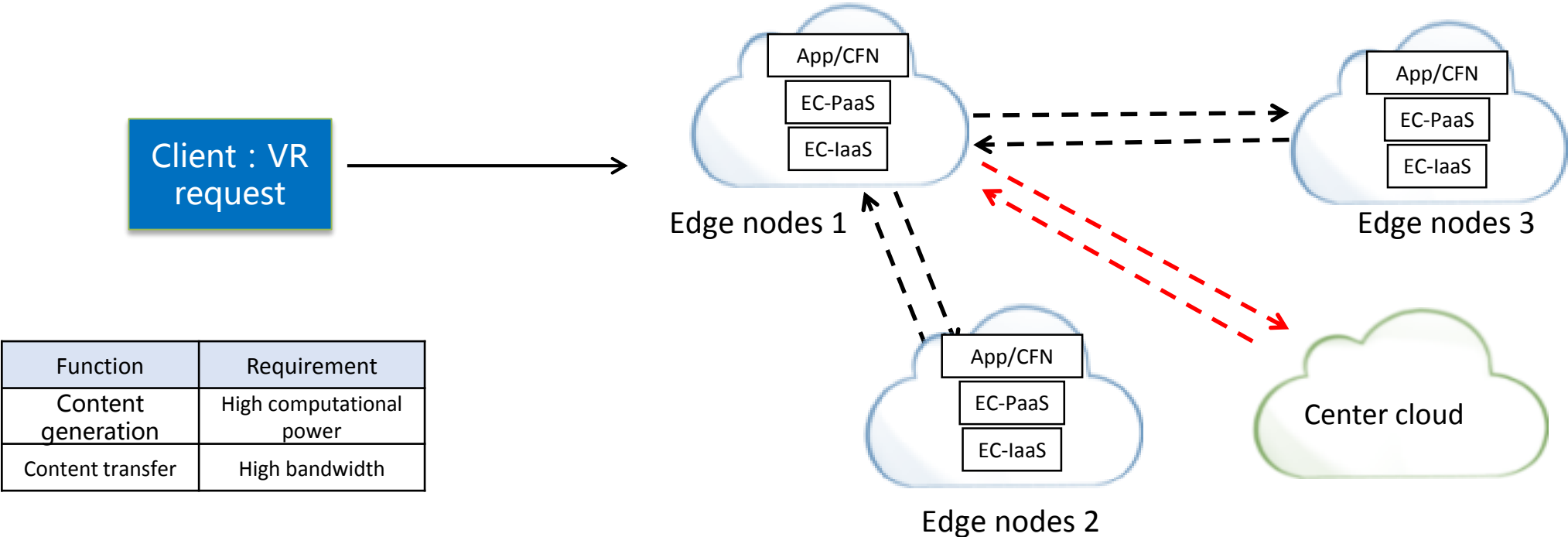
Usage Scenarios 2 - Connected Car



In auxiliary driving scenarios, when the local edge site is overload, the delay-insensitive services should be dispatched to other light loaded nodes instead of local edge nodes.

- The edge node can collect the comprehensive road and traffic information and perform data processing, so that the delay-sensitive service is preferentially processed locally to ensure the service availability and user experience.
 - high security risk of the vehicles can be signale.
 - warnings can be sent to the cars driving too fast or under other invisible dangers.

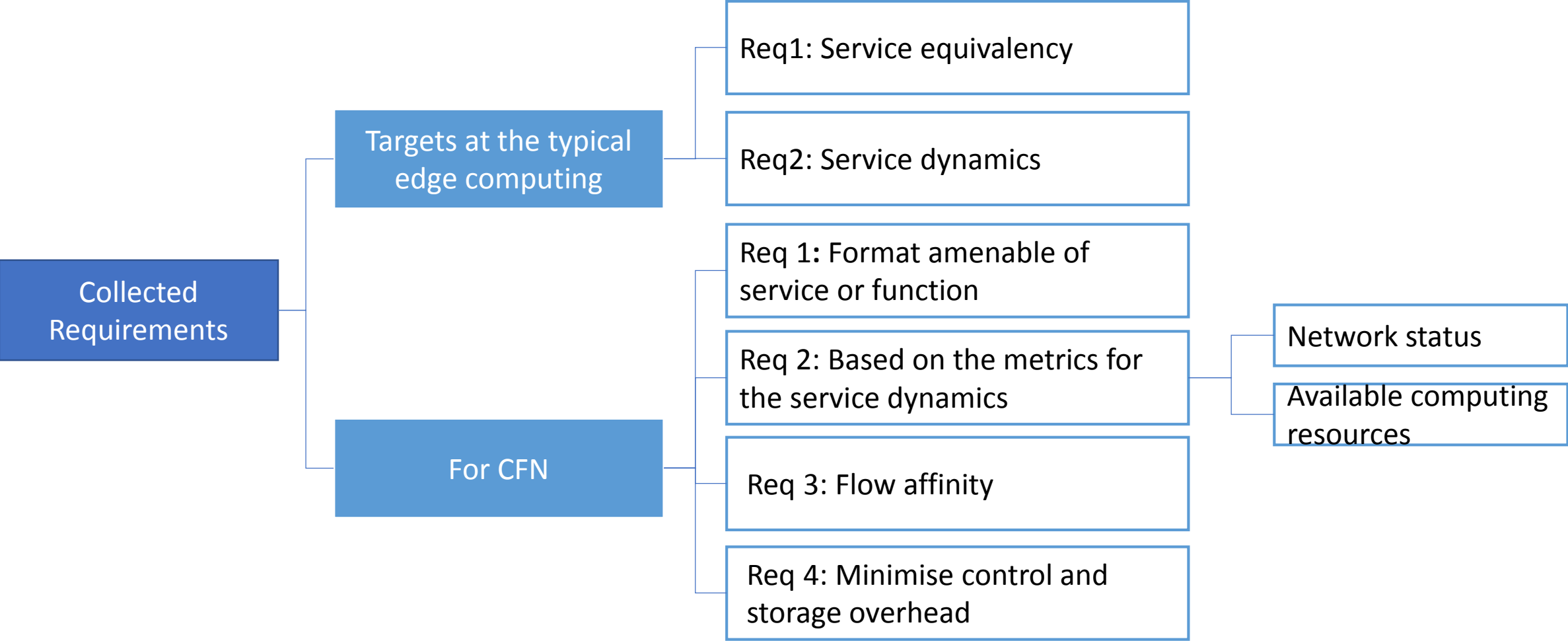
Usage Scenarios 3 - Cloud Virtual Reality (VR)



Cloud VR services have high requirements on both network and computing so that the edge node to serve the request has to be carefully selected to avoid the overloading.

- The end device usually only uploads the posture or control information to the cloud
- The VR contents are rendered in the edge cloud.
- The video and audio outputs generated from the cloud are encoded, compressed, and transmitted back to the end device via high bandwidth network.

Collected Requirements



Requirements- targets at the typical edge computing

Requirement 1- Service equivalency: Equivalent service is provided by multiple edges to ensure better scalability and availability.

Requirement 2- Service dynamics: A single edge has very dynamic resources over time to serve a request. Its dynamics are affected by computing resource of service node, network path congestion, failover and others.

Requirements- for CFN

Requirement 1- Should be identified in a format amenable to processing in the network of what the service provided or function called

Requirement 2- Should be based on the metrics for the service dynamics, including both network status and available computing resources. Service request assignment should not be static.

Requirement 3- Must be possible to have a method to signal flow affinity requirements and handle flows on the same edge based on flow affinity of applications.

Requirement 4- Must minimised control and storage overhead, due to the limitation of edge nodes' compute resources

In Summary

- CFN tries to leverage the network distributed nature by considering both network status and computing resources among multiple edges. This document illustrate some usage scenarios and requirements for CFN.
- CFN architecture should addresses how to distribute the computing resource information at the network layer, how the data plane adapts when the edge to handle the first service request is not known in advance, and how to assure flow affinity.

Comments?

Thank you!