Computing-Aware Traffic Steering (cats)

Compute-Aware Metrics Working with ALTO

CATS Chairs

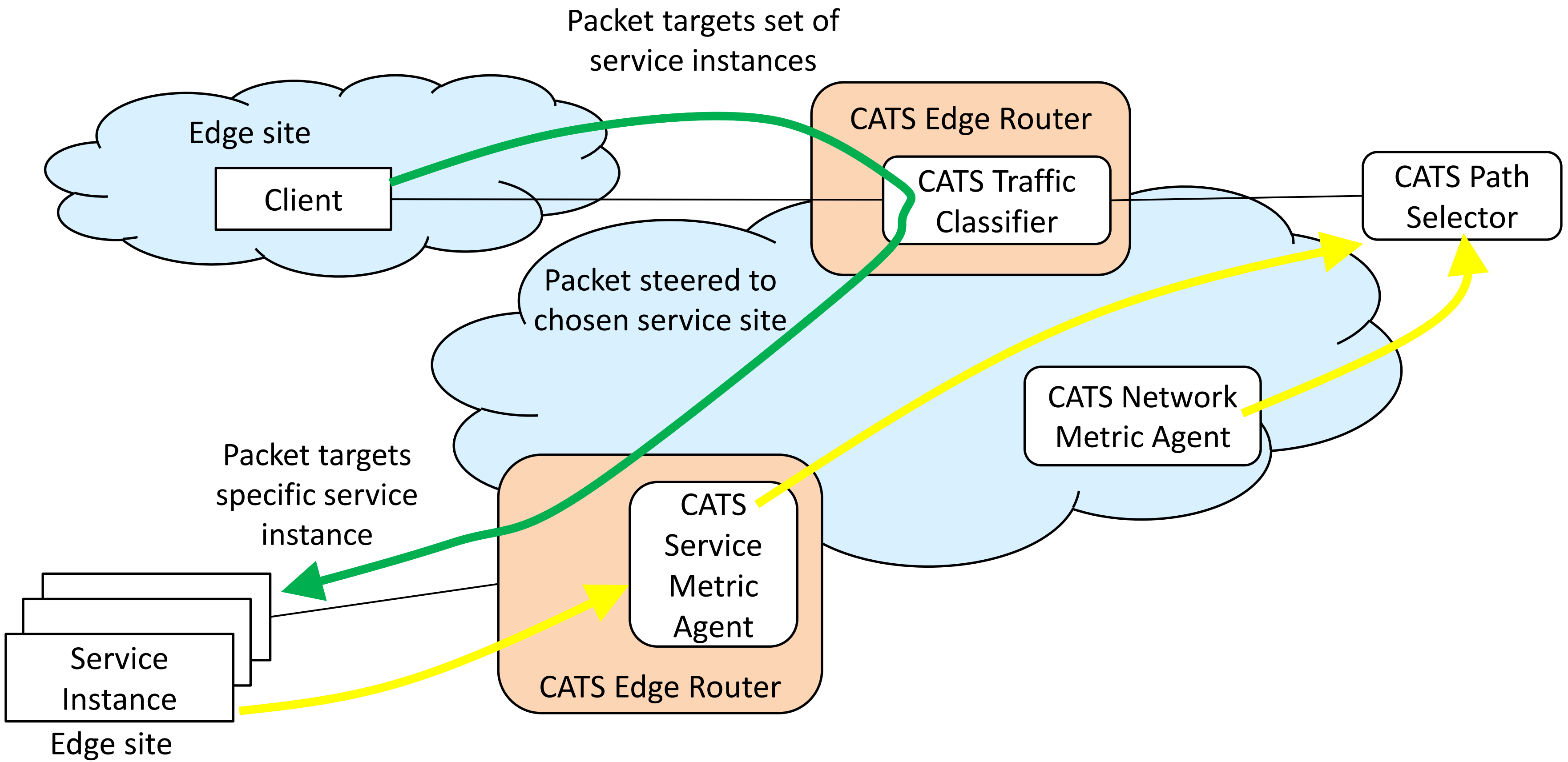
Adrian Farrel (adrian@olddog.co.uk)

Peng Liu (liupengyjy@chinamobile.com)

IETF-117 – San Francisco – July 2023

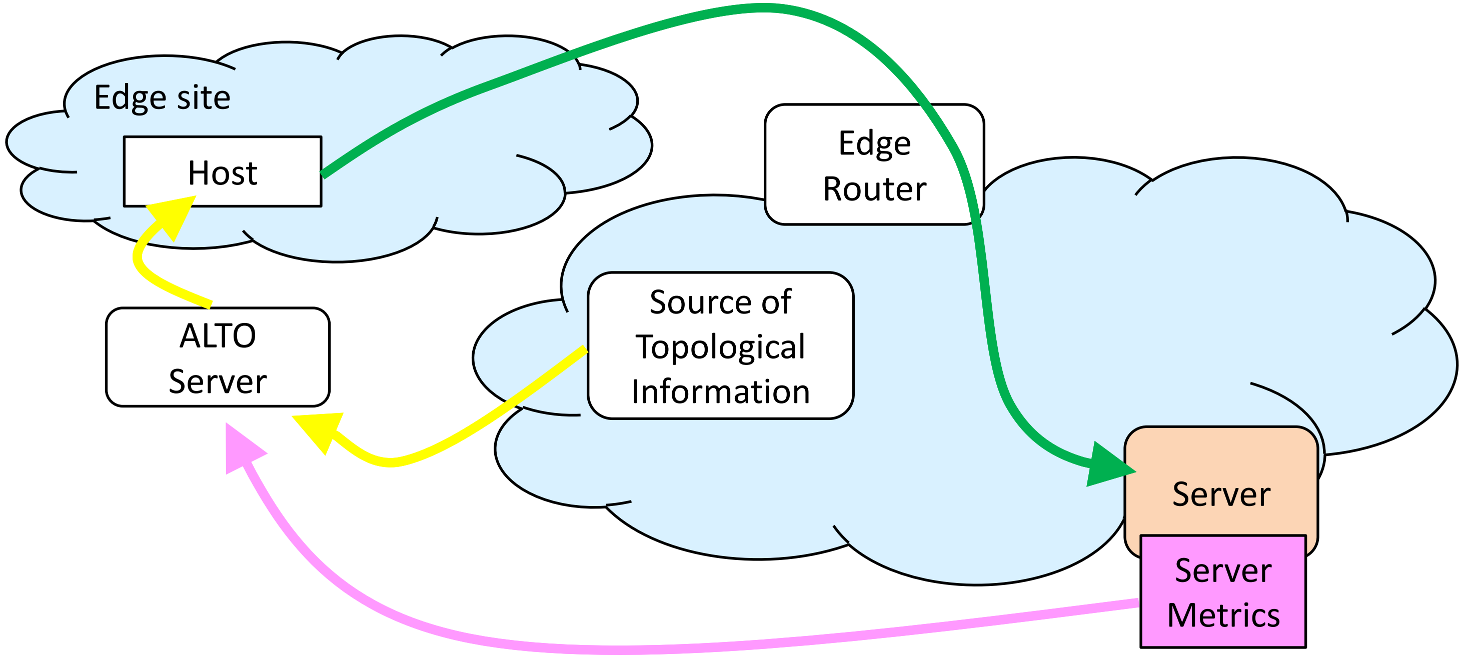
What is CATS? What's the Problem Space?

- From the charter
  - A general framework for the distribution of compute and network metrics and transport of traffic from **network edge** to service instance.
  - Integrate network and **compute conditions** in the optimization function


Packet targets set of service instances
Edge site
Client
CATS Edge Router
CATS Traffic Classifier
CATS Path Selector
Packet steered to chosen service site
CATS Network Metric Agent
Packet targets specific service instance
CATS Service Metric Agent
CATS Edge Router
Service Instance
Edge site

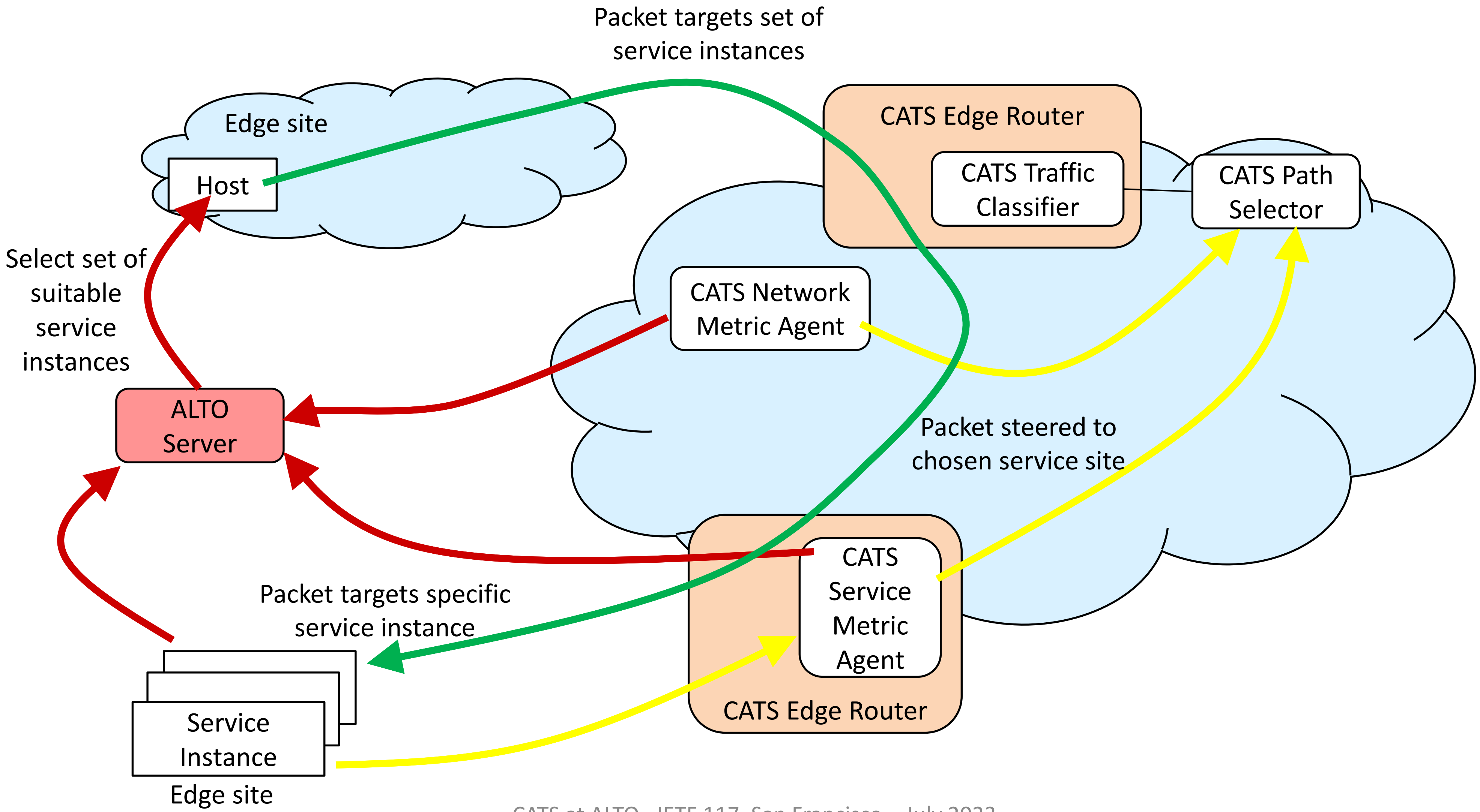
Compare with the ALTO Space

- From the charter
  - Allow a **host** to choose optimal paths
  - A server with knowledge of the network



Potential Joint Deployment

- This is just a thought for the future



Compute Metrics in CATS

- CATS uses two sets of information to select a remote compute site
  - Network topology, capabilities, and state
    - We already have most of this
    - Latency is work-in-progress in Routing Area
  - Server compute locations, capabilities, and state
    - CATS is only just starting to look at this
    - “State” means ability to process a new request in a specific time (i.e., load)
- Guiding principles
  - Need standardised metrics so everyone can understand them
  - Want to be easily able to combine the network and compute metrics
  - Simplicity is fundamental

ALTO / CATS Cooperation on Metrics

- It seems that ALTO may also be interested in compute metrics
  - The deployment scenario is different, but the concept is the same
  - ALTO may be interested in a wider set of server metrics
- Is it possible for the metrics were consistent?
  - Different usage, but similar methodology and concepts
- Could we pool our understanding of compute metrics?
- If we want to share ideas, how should we progress?
  - Cross-review of drafts?
  - Joint (virtual) interim meeting?