Gap Analysis and Requirements of CAN

Gap Analysis

draft-liu-dyncast-ps-usecases-03

P. Liu, China Mobile
P. Eardley, British Telecom
D. Trossen, Huawei
M. Boucadair, Orange
LM. Contreras, Telefonica
C.Li,Huawei

Existing solutions

Here we list some ‘existing solutions’ based on the assumption of supporting the network and computing joint optimization.

DNS：

• 'early binding' to explicitly bind from the service identification to the network address

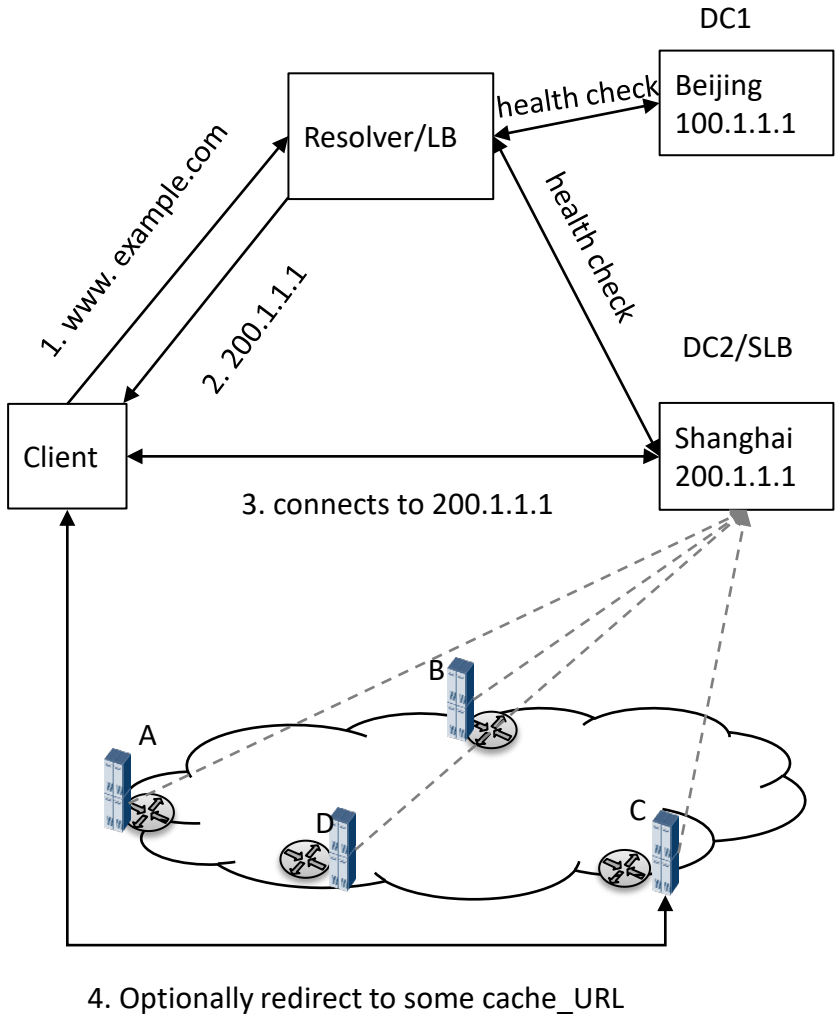
• ‘geographical location’ to pick the closest computing resource

• ‘health check’ to realize load balance and prevent single point failure

Load balancer：

• external components of network designed for computing domain to support load balance

Gap analysis of existing solutions-DNS/GSLB



- Early binding: clients resolve IP address first and then steer traffic.
  - Use the DNS entry cached at client, stale info may be used.
  - Often, resolver and LB are separate entities which incurs even more signaling overhead by needing to first resolve and then redirect to LB for final decision
  - Resolution is L7 or app-level decision making, i.e. DB lookup. Originally intended for control, NOT data plane speed!
- Health check: on an infrequent base, switch when fail-over
  - Limited computing resources on edge will change rapidly, while more frequent health check is prohibitive in cost
- Load balance over DNS: usually focused on edge server load first, then utilizing lowest latency routing to the selected server's IP address
  - Lacks the combined consideration for load & latency's for a better E2E guarantee
  - Problem of how to obtain necessary metrics for decision

Gap analysis of existing solutions - load balancer

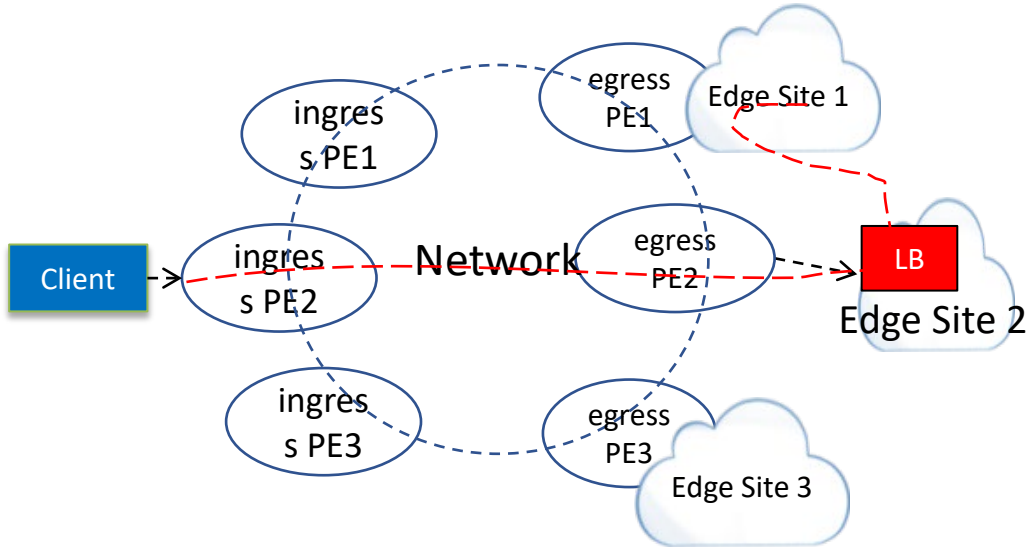
Single LB in a single site for all server instances:

Pros:

- easy deployment

Cons:

- Single point of failure at the LB
- The network path from the LB to server instances at other sites might not be optimal, e.g., the red dotted path



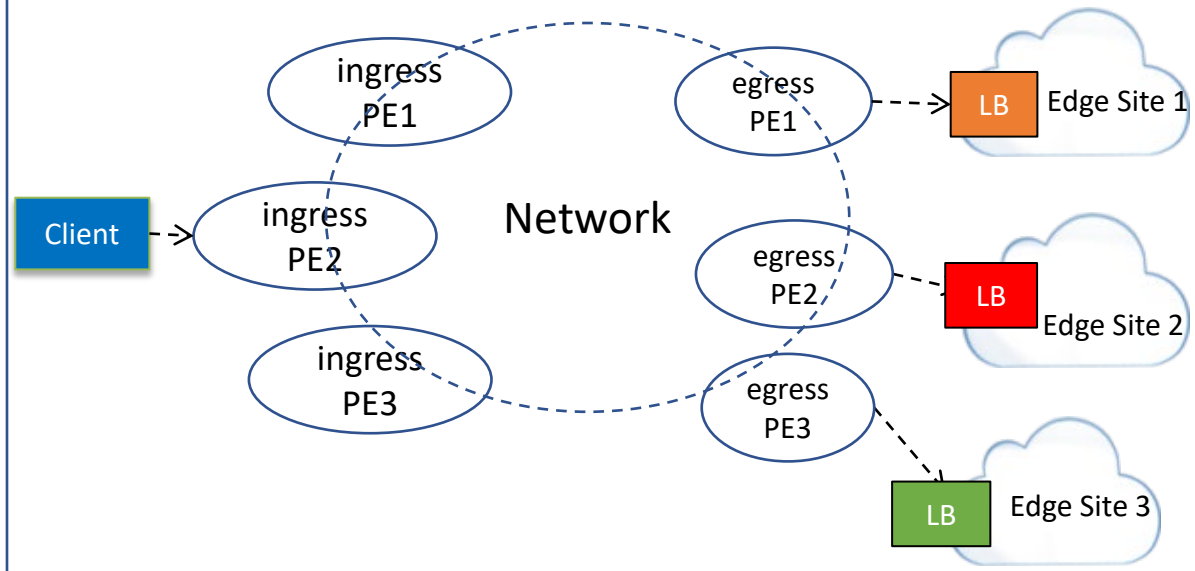
Each site has it's own LB:

Pros:

- Easy deployment.

Cons:

- No load balance among multiple sites.



Summary of gap analysis

- **Dynamicity of Relations:** Existing solutions exhibit limitations in providing dynamic instance affinity
  - E.g., DNS is not designed for this level of dynamicity (i.e., minute level originally, client needs to flushing the local DNS cache, frequent resolving may lead to overload of DNS)
- **Efficiency:** Existing solutions may introduce additional latencies and inefficiencies (e.g., more messages) in packet transmission
- **Complexity and Accuracy:** Existing solutions require careful planning for the placement of necessary control plane functions in relation to the resulting data plane traffic, which is difficult and may lead to the inaccuracy of the scheduling.
- **Metric exposure and use:** Existing solutions lack the necessary information to make the right decision on the selection of the suitable service instance due to the limited semantic or due to information not being exposed
- **Security**: Existing solutions may expose control as well as data plane to the possibility of a distributed Denial-of-Service attack on the resolution system as well as service instance.

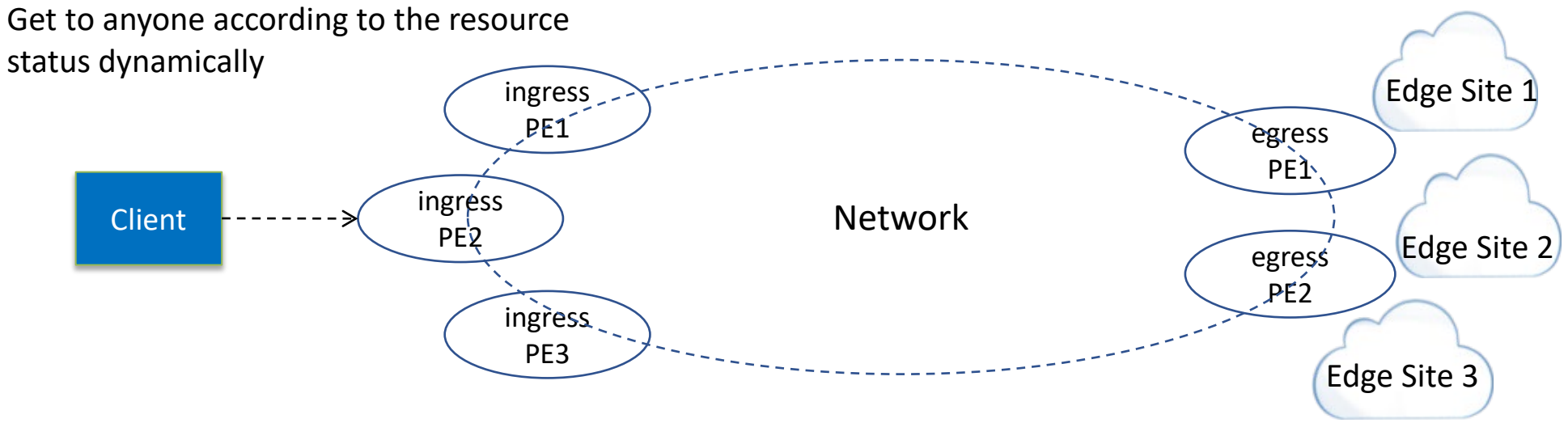
Requirements

draft-liu-dyncast-reqs-02

P. Liu, China Mobile
T. Jiang, China Mobile
P. Eardley, British Telecom
D. Trossen, Huawei
C.Li,Huawei

Main goals

- **Open choice: considering to access the location of multi-computing resource**
- **Dynamically: considering to select the appropriate computing resource dynamically**
- **Multi-metric: considering both the network and computing resource status**



Potential requirements

- **Support multi-access to the available edge sites dynamically**
  - Anycast based or other potential methods
- **Support considering and using both network and computing metrics**
  - Computing semantic model
  - Rate control signaling of metrics
  - **Support effective computing resource representation and encapsulation**
  - Single index or multi-information for specific purpose
  - **Support the interface between network and computing components**
  - Centralized and/or decentralized advertising and signaling
- **Support the session continuity and service continuity**
  - Functional equivalency in different areas
  - **Support management of network and computing resource**

Thank you!