Using ALTO to Determine Service Edge

<draft-contreras-alto-service-edge-00>

L.M. Contreras (Telefonica)
D. Lachos, C.E. Rothenberg (UNICAMP)

Singapore, ALTO WG, November 2019
Problem Statement

• Operators are starting to deploy distributed computing environments in different parts of the network
  • Different service needs in terms of latency, bandwidth, processing capabilities, etc
• Multiple data centers of different sizes across the network (e.g., large, medium, small) with distinct dimensioning in terms of CPUs, memory and storage capabilities, as well as bandwidth capacity, and Enhanced Platform Awareness (EPA) capabilities
• Essential for a network operator to have mechanisms assisting on the decision of which DC to use
  • Decisions affect both compute and transport network substrates
• ALTO can assist on such decisión making processes
Role of ALTO

• ALTO can **assist in the selection of the more convenient edge** combining transport network information.
Expressing Computing Needs

• Computing needs are typically **structured by bundling CPU, RAM and storage units** as quotas, instances or flavors
  • E.g., Amazon Web Services, Microsoft Azure

• Same approach followed by the Common Network Function Virtualisation Infrastructure Telecom Taskforce (CNTT)

• Flavors or instances can be characterized by:
  • **Type of instance (T)**: the types of instances are characterized as B (Basic), N (Network Intensive), and C (Compute Intensive)
  • **Interface option (I)**: it refers to the interface bandwidth
  • **Compute flavor (F)**: it refers to a certain combination of virtual CPU, RAM, disk, and bandwidth for the management interface
  • **Optional storage extension (S)**: to request additional storage capacity
  • **Optional hardware acceleration characteristics (A)**: to request specific acceleration capabilities for improving the performance of the infrastructure
Association of Compute Capabilities to Network Topology

• (At least) **Three potential** solutions to be explored:
  • Leverage on (and possibly extend) [I-D.ietf-teas-sf-aware-topo-model]
  • Extend BGP-LS [RFC7752]
  • Combine information from the infrastructure profiles catalogue with topological information by leveraging on the IP prefixes allocated to the gateway providing connectivity to the NFVI PoP.
Next Steps

• Elaborate more on mapping to property maps.
• Explore the alternatives for the association of compute capabilities to network topology.
• Collect interest from ALTO Working Group.
• Document advances for IETF#107