Joint Exposure of Network and Compute Information for Infrastructure-Aware Service Deployment

<draft-rcr-opsawg-operational-compute-metrics>

Sabine Randriamasy (Nokia Bell Labs), Luis Contreras (Telefonica), Jordi Ros Giralt (Qualcomm Europe, Inc.), Roland Schott (Deutsche Telekom)

IETF 119, Brisbane, March 2024
Motivation

• Standardization of network information is quite mature but is in progress for compute information.

• There is a need to define a set of compute metrics to support various use cases being served in the IETF.

• Some ad hoc work exists in the IETF:
  • CATS (e.g., draft-du-cats-computing-modeling-description)
  • ALTO (e.g., draft-contreras-alto-service-edge)
  • OPSAWF (e.g., RFC 7666 MIB)

• Metrics are also defined in other bodies such as the Linux Foundation, DMTF, ETSI NFV, etc:
  • Raw compute infrastructure metrics (e.g., processing, memory, storage)
  • Compute virtualization resources and service quality metrics (e.g., VNF resources in VMs)
  • Service metrics including compute-related information (e.g., service delay, availability)
Problem space

Service Life Cycle:

New service → Service deployment → Service selection

<table>
<thead>
<tr>
<th>Action to take</th>
<th>Information needed</th>
<th>Who needs it</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Service placement</td>
<td>- Compute</td>
<td>- Service provider</td>
</tr>
<tr>
<td>- Communication</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action to take</th>
<th>Information needed</th>
<th>Who needs it</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Service selection</td>
<td>- Compute</td>
<td>- Network provider</td>
</tr>
<tr>
<td>- Path selection</td>
<td>- Communication</td>
<td></td>
</tr>
<tr>
<td>- Application (client or proxy)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
History and updates from IETF 118

• -01 version presented at IETF 118, collecting initial feedback
• Updates from -01 (now in -03 version)
  • Use cases documented for better illustrating the problem space
  • New section on “Production and Consumption Scenarios of Compute-related Information”
    • Reference to raw resources and allocated resources
  • New section on “Metrics Selection and Exposure”
    • Reference to how the metrics are exposed and (2) which kind of metrics need to be exposed
    • Discussion on dimensions to consider when identifying compute metrics
    • Discussion on abstraction levels and information access
    • Reference to distribution and exposure mechanisms
• Added Roland as co-author
Next steps

• Collect feedback from CATS WG
  • Some discussion now on mailing list
  • We encourage more discussion

• Prepare new version for IETF 120
  • Incorporating comments received so far from CATS, NMRG and Side Meeting on “Information Exposure for Edge Computing” during IETF 119