Isolated control plane for OpenRack v3 Servers

Rob Elliott, cloud native hardware platforms, HPE
Alex Fuxa, cloud native hardware platforms, HPE

2020-01-20 for Server WG call
Control planes in a typical server

<table>
<thead>
<tr>
<th>Physical interface</th>
<th>PCIe</th>
<th>SMBus (I2C)</th>
<th>NC-SI over RMII-based transport (RBT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controller</td>
<td>Host CPU, or BMC through host PCIe topology</td>
<td>BMC</td>
<td>BMC</td>
</tr>
<tr>
<td>Speed</td>
<td>1 GB/s full duplex (PCIe3 x1) or faster</td>
<td>10 to 100 kB/s (1 MB/s if upgraded to I3C Basic)</td>
<td>10 MB/s full-duplex</td>
</tr>
<tr>
<td>Definition</td>
<td>built-in</td>
<td>optional on PCIe CEM</td>
<td>defined in OCP NIC 3.0</td>
</tr>
<tr>
<td>Extra pins</td>
<td>none</td>
<td>2 (shared or point-to-point)</td>
<td>7-12 (shared or point-to-point)</td>
</tr>
<tr>
<td>Protocol</td>
<td>MCTP over VDM</td>
<td>MCTP over SMBus</td>
<td>NC-SI</td>
</tr>
<tr>
<td>Adoption</td>
<td>some devices</td>
<td>some devices</td>
<td>some NICs</td>
</tr>
</tbody>
</table>
Historically, host CPU is in charge

Device accepts all management over host PCIe interface
Software on CPU is in charge
Customer are expected to run OS software that isolates administrators from users
  • Root user can do everything
  • Download firmware
  • Configure devices
    • e.g., NIC: overlay networks (VXLANs), network policies, firewalls
    • e.g., storage: configure volumes, RAID, encryption
BMC has limited access with its slow interfaces
  • OK for things like temperature polling
Hostile tenants

In a server used for a multi-tenant cloud, the software on the host CPU is not trustworthy at all

- Tenants come and go
- Tenants may be hostile
- Might not be protected by a hypervisor (bare-metal)

Need a way for the provider/operator to control all the devices in the system
Isolated control plane

Add an isolated control plane (ICP) for operator control

- Could be the BMC or a separate system controller
- Device refuses management commands from host PCIe
- Device accepts management commands from ICP

Requires a higher speed, more reliable interface than I²C or RBT

- Large firmware images
- Large firewall rule sets for NICs
- Large machine images for storage controllers
- System partitioning (e.g., HP Superdome nPars, IBM mainframe LPARs)
- Constant monitoring and logging regardless of host CPU state
Call to Action

Help define requirements for an isolated control plane
Help define the interfaces
  • Hardware interface
    • New connectors, alternative pin assignments on existing connectors
  • Software interface
    • New Data Models/APIs for coordination inside a server
Could involve multiple organizations
  • DMTF (Platform Management Components Intercommunication WG, Redfish Forum)
  • SNIA SFF Technology Affiliation TWG
  • PCI SIG
  • OCP NIC subgroup
Will present again at the OCP Global Summit Server WG track on 5 March 2020
Open for All.

MARCH 4 & 5, 2020  |  SAN JOSE, CA
DMTF PMCI (Platform Management Components Intercommunication) WG

- DSP0218 Platform Level Data Model (PLDM) for Redfish Device Enablement, Version 1.0.0, 25 June 2019
- DSP0222 Network Controller Sideband Interface (NC-SI) Specification, Version 1.0.1, 2013-01-24,
- DSP0238 Management Component Transport Protocol (MCTP) PCIe VDM Transport Binding Specification, Version 1.1.0, 29 November 2018
- DSP0248 Platform Level Data Model (PLDM) for Platform Monitoring and Control Specification, Version 1.2.0, 9 September 2019
- DSP0261 NC-SI over MCTP Binding Specification, Version 1.2.2, 24 September 2019
- DSP0267 Platform Level Data Model (PLDM) for Firmware Update Specification, Version 1.1.0, 4 December 2019
- DSP0274 Security Protocol and Data Model (SPDM) Specification, Version 1.0.0, 22 December 2019
- DSP0275 Security Protocol and Data Model (SPDM) over MCTP Binding Specification, Version 1.0.0, 22 December 2019
- https://www.dmtf.org/standards/pmc

DMTF Redfish Forum

- DSP0266 Redfish specification, Version 1.8.0, 23 September 2019
- https://www.dmtf.org/standards/redfish
References - Other

PCI SIG
- *PCI Express Base Specification*, Revision 5.0, Version 1.0, 22 May 2019
- *PCI Express Card Electromechanical Specification*, Revision 4.0, Version 1.0.4, 7 August 2019
- [https://pcisig.com/specifications](https://pcisig.com/specifications)

MIPI Alliance
- *MIPI I3C Basic*, Version 1.0, 19 July 2018
- [https://www.mipi.org/specifications/i3c-sensor-specification](https://www.mipi.org/specifications/i3c-sensor-specification)

SNIA SFF
- SFF-TA-1021 *Specification for PCIe Enclosure Compatible Form Factor Specification (PECFF)*, Revision 0.8.3, 6 January 2020
- SFF-TA-1022 *Specification for PCIe Enclosure Compatible Form Factor Specification (PECFF) Thermal Reporting*, Revision 0.8.3, 6 January 2020
- [https://www.snia.org/sff](https://www.snia.org/sff)