Socnoc XPU: Dual-protocol Switch SoC for Integrated Edge Cloud

Leo Li
li@socnoc.ai
Socnoc AI Inc.
合肥边界智芯科技有限公司
https://www.socnoc.ai
**Differences between Datacenter Cloud and Edge Cloud**

<table>
<thead>
<tr>
<th>Datacenter Cloud</th>
<th>Edge Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>• massive scale (&gt;10k servers connected)</td>
<td>• small scale (less than 50 servers connected)</td>
</tr>
<tr>
<td>• highly organized (storage pool, computing pool, networking pool, GPU pool)</td>
<td>• less organized (mix and match)</td>
</tr>
<tr>
<td>• management focused (datacenter OS, network OS, xOS)</td>
<td>• hard to managed</td>
</tr>
<tr>
<td>• constant environment (professional building)</td>
<td>• poor environment</td>
</tr>
<tr>
<td>• devops friendly (in-field engineers)</td>
<td>• no engineers</td>
</tr>
<tr>
<td>• countable sites (8000 datacenter globally)</td>
<td>• massive sites (2.8 million edge rooms)</td>
</tr>
<tr>
<td></td>
<td>• very low cost!</td>
</tr>
</tbody>
</table>
Goal: Low cost Switch SoC for Integrated Edge Cloud

Basic Feature:
• High outbound Bandwidth: 4x10Gbps/2x25Gbps/1x100Gbps
• Cluster ready: 8/16/32/64 servers
• High Intra-connection Bandwidth: 10-20Gbps
• Low Latency: Host-to-host latency less than 10us
• Dual network topology for data path and management path
• **Low networking cost: less than $5 per Gbps for host-to-host networking**
• Support x86, arm CPUs, GPUs and AI sticks

Advanced Feature:
• Protocol offload: TCP/IP offload, RDMA accelerated
• Device pooling ready: GPU pool and SSD pool
• Plug-n-play maintenance and remote resource management
Solution: Dual-protocol Switch SoC for Integrated Edge Cloud

- Socnoc XPU: dual protocol switch SoC supporting multichannel PCIe bus and ethernet protocol to connect CPU, GPU, SSD in the same plug-n-play fashion.

- Feature
  - PCIe interface, no special chips
  - Low cost, $5 per Gbps per connection
  - Low latency
  - Shared system BUS (manageable)
  - Low TDP

### Product specification

<table>
<thead>
<tr>
<th>Product</th>
<th>Intra-Bandwidth</th>
<th>Outbound Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>X100(2021Q4)</td>
<td>20Gbps</td>
<td>40/80Gbps</td>
</tr>
<tr>
<td>X200(2022Q4)</td>
<td>80Gbps</td>
<td>100/200Gbps</td>
</tr>
<tr>
<td>X300(2023Q4)</td>
<td>160Gbps</td>
<td>200/400Gbps</td>
</tr>
</tbody>
</table>

XPU-based Integrated Edge Cloud

- dual protocol (CPU and GPU)
- dual ports (data and management)
- plug-n-play
- remote management ready
Solution: SR-IOV and Mainstream Cloud and Edge OS

Data Fabric

Software Stack

Virtualization and Software Stack

- SR-IOV based networking virtualization for network performance
- TCP/IP for Cloud and Edge microservice-like Cluster OS
- RDMA and Pooling Fabric for Device and I/O Sharing
Prototype and Demo System (2021Q4 release)
Just connecting

Hefei

Socnoc: Just connecting!

li@socnoc.ai

https://www.socnoc.ai