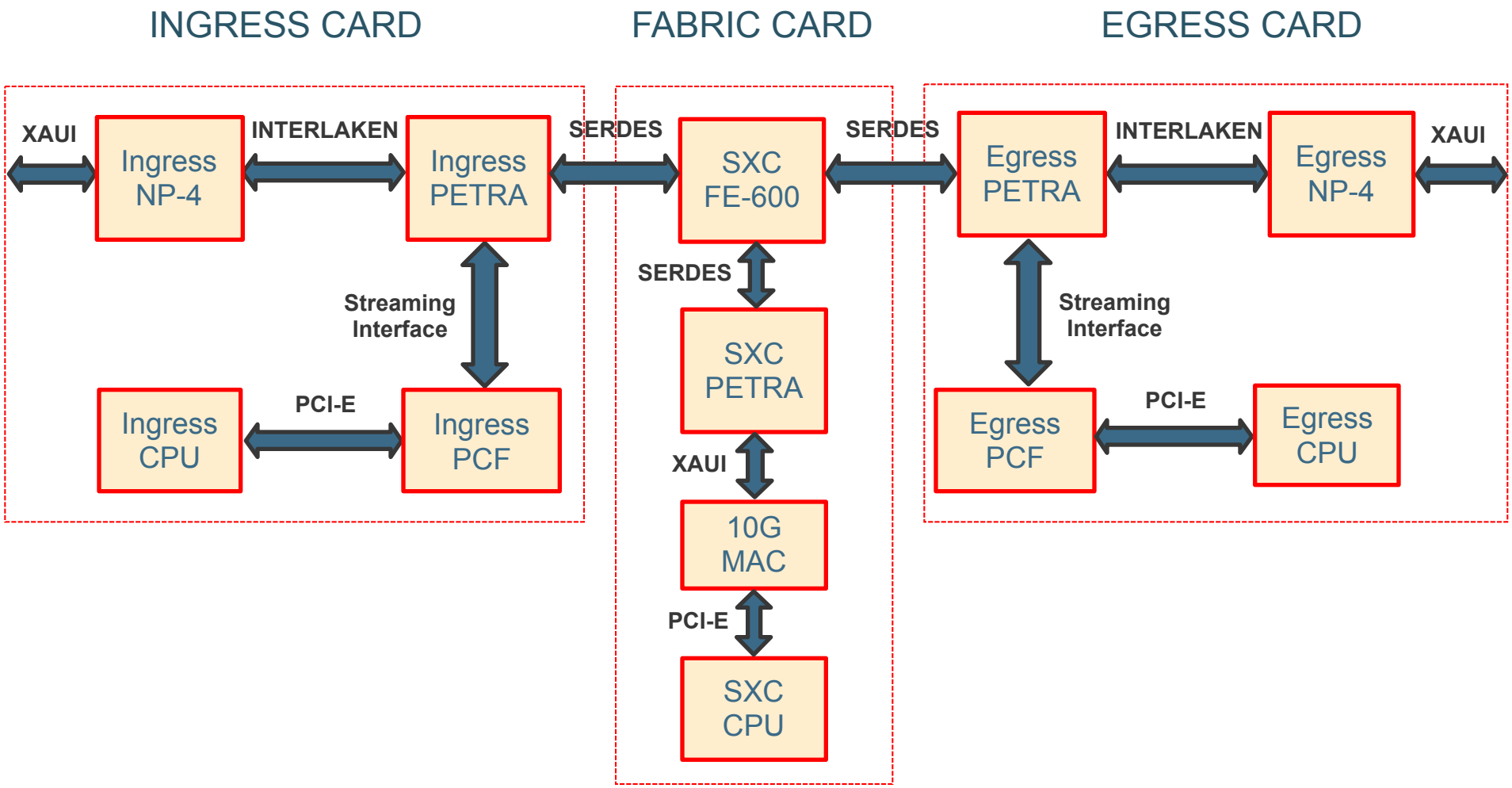
CCNx 1.0 Virtual Machine for Content Routers

Computer Science Laboratory
Networking & Distributed Systems

March 2014

parc
A Xerox Company

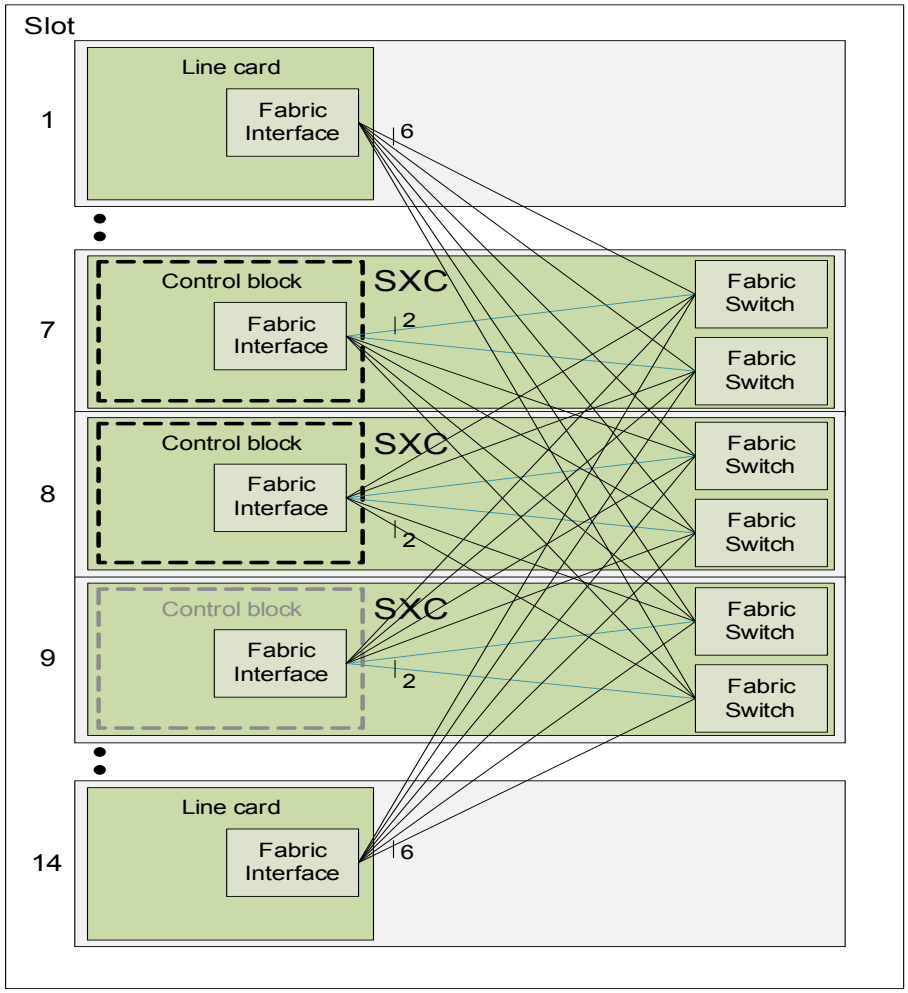
Research Platform System Architecture



parc
A Xerox Company

Research Platform Switch Fabric Details

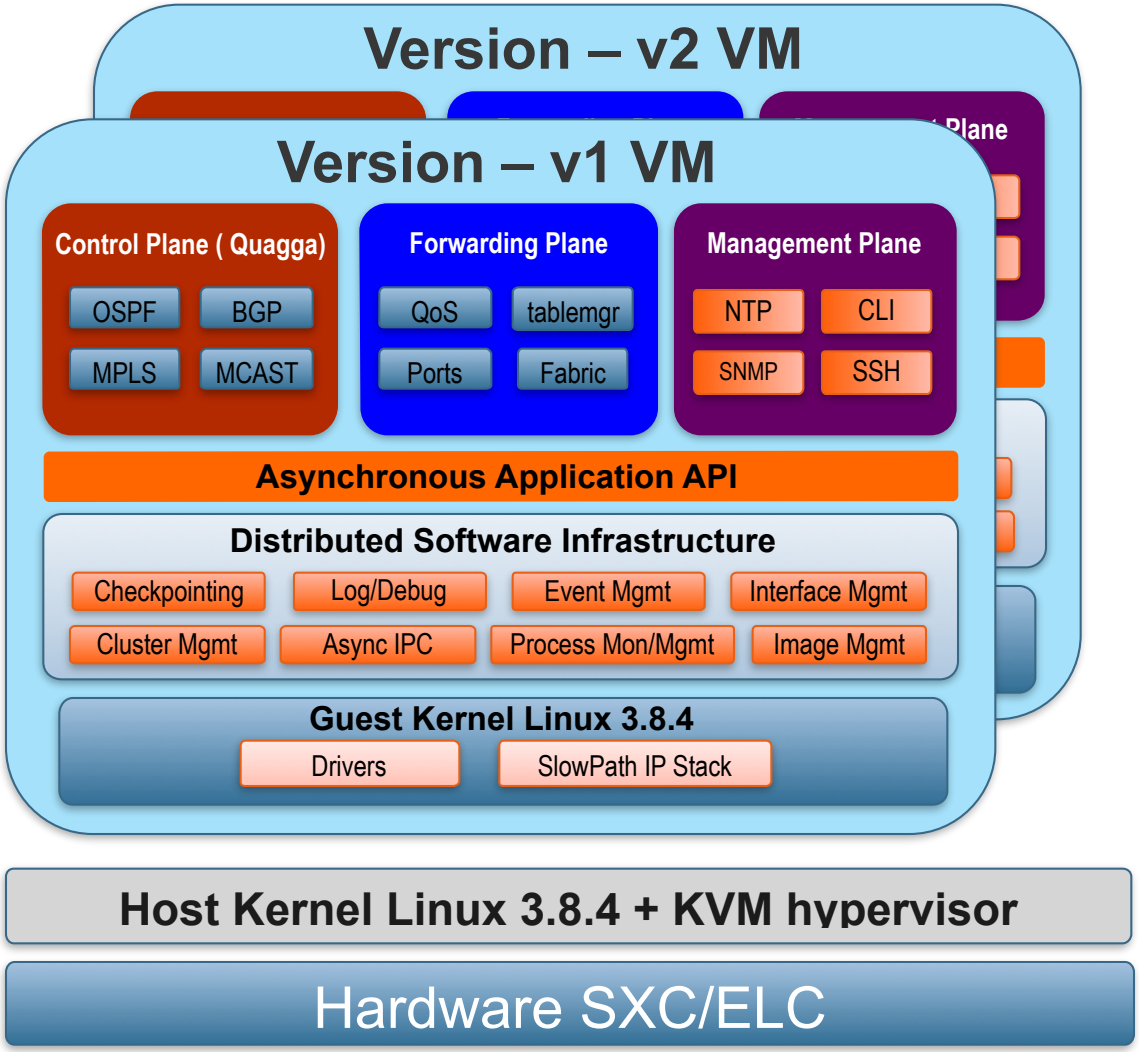
- Each SXC card contains two Fabric Switch chips, FE600A and FE600B.
- Each CFMC card has one Fabric Switch.
- All cards have a Petra Fabric Interface Chip configured with 36 serdes ports.
- Ethernet Line Cards have 6 serdes lanes routed to each available fabric switch.
- With a single SXC, 12 serdes lanes provide 12 x 6.25 Gbps = 75 Gbps available throughput. (Estimate 4.4 Gbps usable throughput per serdes lane, 52 Gbps total).
- 100 Gbps Line Cards require two SXC fabric cards to provide sufficient throughput.



parc
A Xerox Company

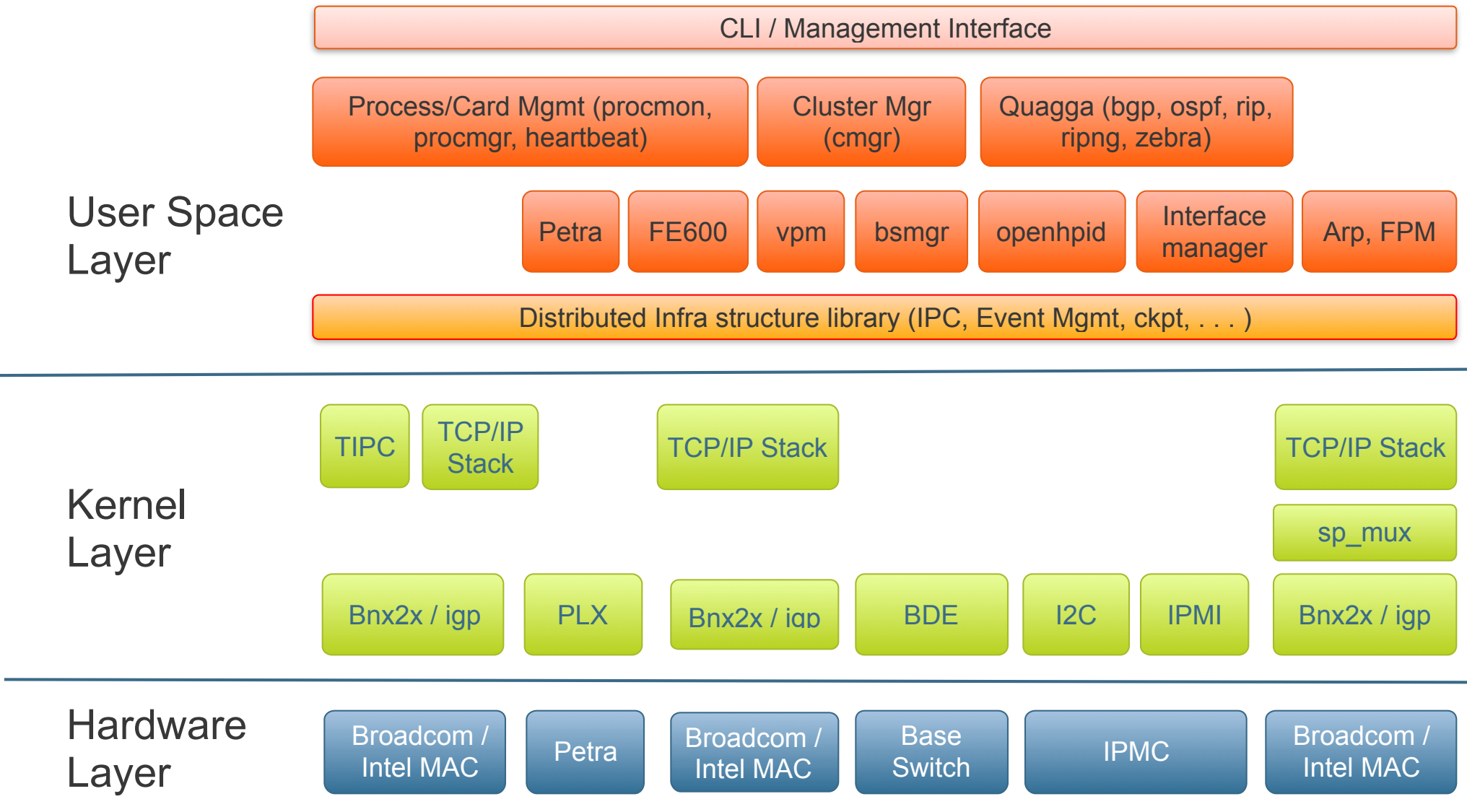
PARC-OS Software Architecture

- Built from open source linux kernel and KVM hypervisor.
- Running on a VM allows in-service software updates, multi-tenancy, and makes network slicing possible.
- Ability to restart VMs improves availability.
- VM isolation allows other services to run without impacting routing.
- Functional modules are processes, and processes are restartable.
- Fault monitoring and recovery mechanism can provide six 9's of reliability.
- Built with async processes to avoid thread switching overhead.



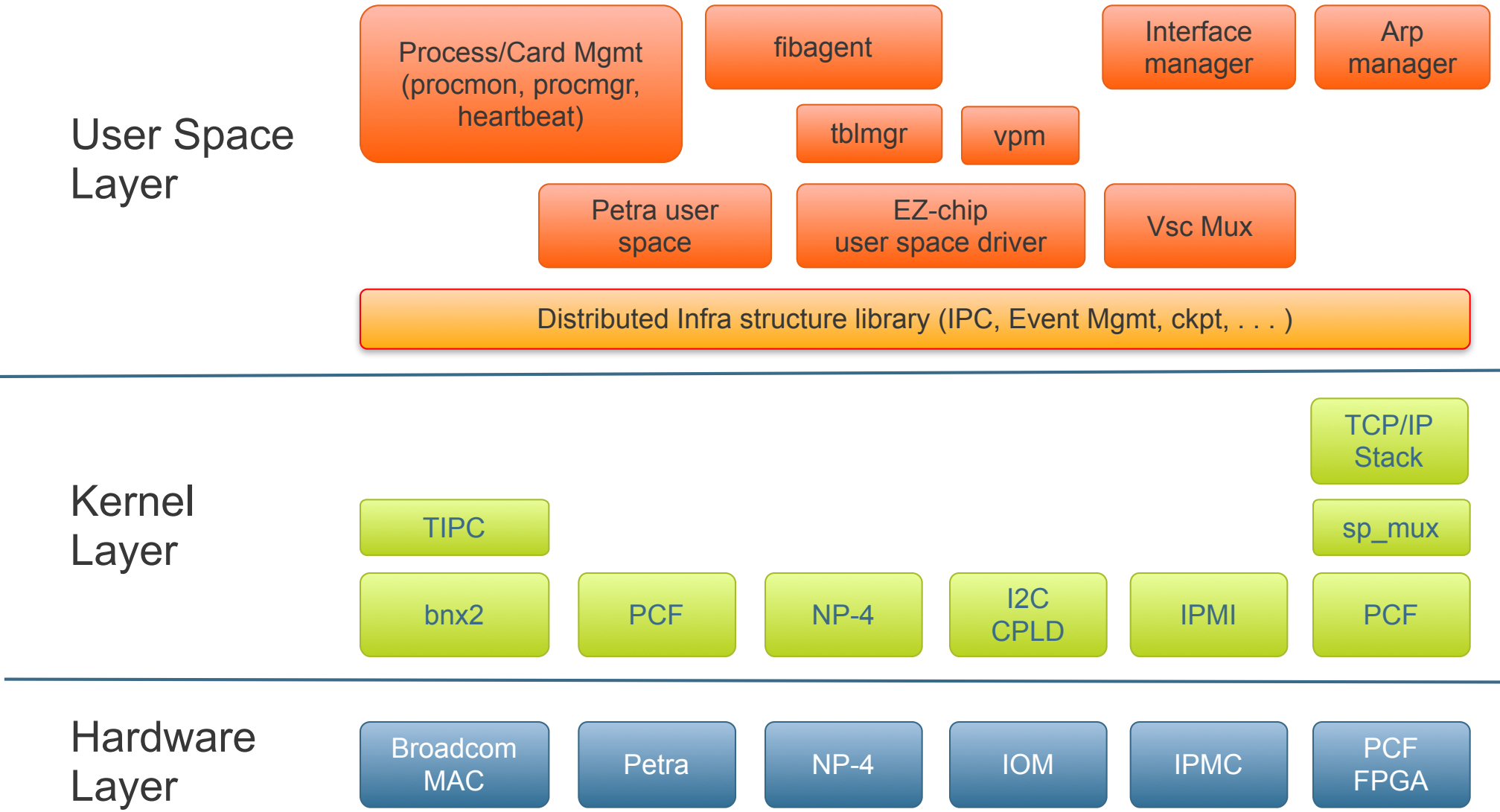
parc
A Xerox Company

Switch Controller Card VM Software Block Diagram



parc
A Xerox Company

Ethernet Line Card VM Software Block Diagram



parc
A Xerox Company

Description of PARC-OS software modules

- VPM is the Virtual Port Manager responsible for bringing the ports up by interacting with NP-4 drivers. It has a corresponding line card component running on Ethernet Line Cards (ELC).
- IM is the Interface Manager responsible for creating, deleting, and modifying IP interfaces on the Switch Controller and ELC.
- ARP Manager runs on Switch Controller and ELC and performs ARP resolution needed for packet forwarding.
- Table Manager runs only on ELC and provides an interface to its clients to program entries in NP-4.
- SP_Mux is a kernel module responsible for fabric multiplexing and demultiplexing. It runs on both Switch Controller and line cards and virtual interfaces are hosted over this module.

Quagga Control Plane

- Quagga is an open-source software control plane that is widely used in data center ethernet routers.
- It provides OSPF, IS-IS, BGP, and RIP with support for IPv6.
- On the SXC Controller Card, Quagga computes routes and installs them on the kernel route table for slow path routing.
- The Forwarding Path Manager (FPM) uses Netlink to listen for kernel route table changes.
- FPM distributes the routes to the Ethernet Line Card's fibagent process.
- Fibagent on line card installs the routes to the NP-4 using the table manager process and also installs the routes in the line card kernel.

How Interfaces are Created

- Interface CLI has been integrated into Quagga vtysh infra structure. Commands are available to create, delete, shutdown and enable an interface.
- Interface name format is **s[n]p[n]**, e.g., s3p0 specifies slot 3 and port 0.
- Example: interface <name> ip address prefix/len command creates interface.
- Sends message to SXC vpm which sends messages to the line cards:
  - To add the port to the vpm database and set port parameters.
  - To admin enable the port.
- Sends a multicast TIPC message to all interface managers (IM).
  - IM creates the virtual interface with specified interface name.
  - Assigns the IP address with the prefix.
  - When vpm detects link up, IM will 'UP' the interface.
- vpm adds the default ingress and egress QoS (Iqos and Eqos) table entries.
- IM on the ELC adds the vlan table entries.
- FPM/ARP in SXC detects the new interface created and asks the fibagent to create the Ip4Dest table entries in the ELC.

Ping sequence of messages

- First ping to a directly connected IP address will result in ARP Request sent by the kernel's IP Stack over the virtual interface (s4p0).
- The Virtual interfaces are hosted over sp_mux module. The sp_mux module will add the Petra + TH header and send it over the broadcom/igb MAC.
- The ARP response will be received by the sp_mux on the ELC thru the pcf FPGA. The received packet will have the pcf header (2 bytes) + petra (4 bytes) + TH (40 bytes).
- The ELC kernel will add the arp entry to the arp table and ARP Manager is notified of new arp neighbor by kernel's netlink layer.
- ARP manager on ELC notifies SXC ARP manager, which in turn adds the kernel entry and multicasts ADD message to all ELCs.
- ELC on receiving ADD message, adds a table manager entry completing the ARP resolution.
- After the ARP is resolved, the icmp ping packet is sent over the virtual interface.
- The sp_mux module will add the Petra + TH header and send it over the broadcom/igb MAC -> Petra -> FE600 -> NP4 -> PHY.
- The Ping response will be received by the sp_mux on the SXC thru the broadcom/igb MAC. The received packet will have the TH (40 bytes).
- The sp_mux strips the header and sends it to TCP/IP stack, completing a successful ping.