Switch Chip
panel discussion

Moderator: Yoshihiro Nakajima (NTT)

Goal

- To share vision and direction of future switch chip for next generation networking
  - Requirements
  - Movement on whitebox switch and network OS
  - open API and SDK

Panelist

- Eli Karpilovski
  (Director, Product Marketing,
   SDN/Cloud Eco-System,
   Infrastructure and Networking/XGS, Broadcom)

- Benny Koren
  (VP of Architecture, Mellanox)

- Prem Jonnalagadda
  (Product line manager, Barefoot)

- Sachin Gandhi
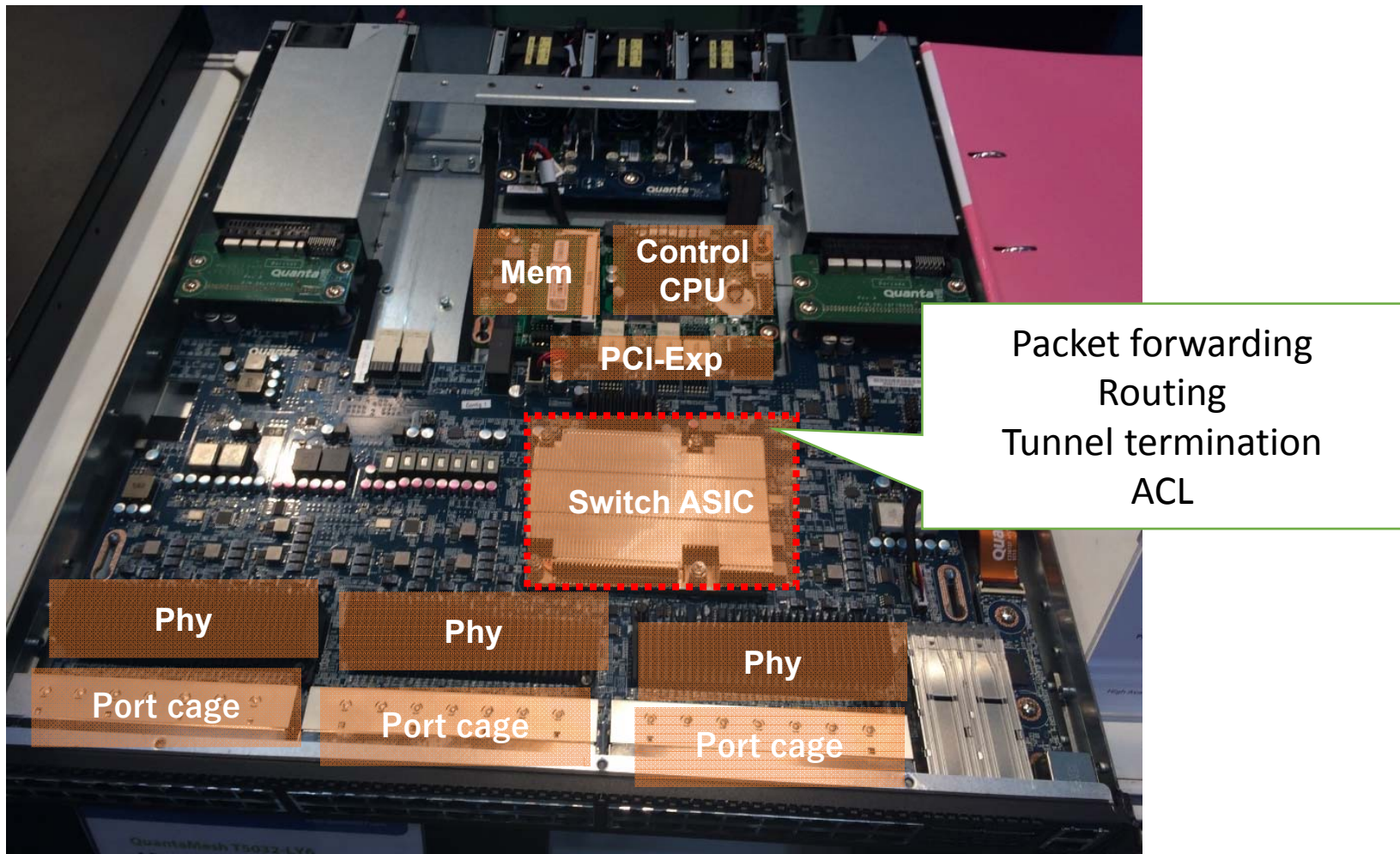 (Director of Marketing, Cavium)

Agenda

- Switch panel introduction (Yoshihiro)
  - Trend and movement
  - Market and application
- Switch chip company introduction
  - Broadcom (Eli)
  - Mellanox (Benny)
  - Barefoot (Prem)
  - Cavium (Sachin)
- Discussion
  - Requirements for next generation switch chip
  - SDN features support on switch chip
  - Supporting emerging applications
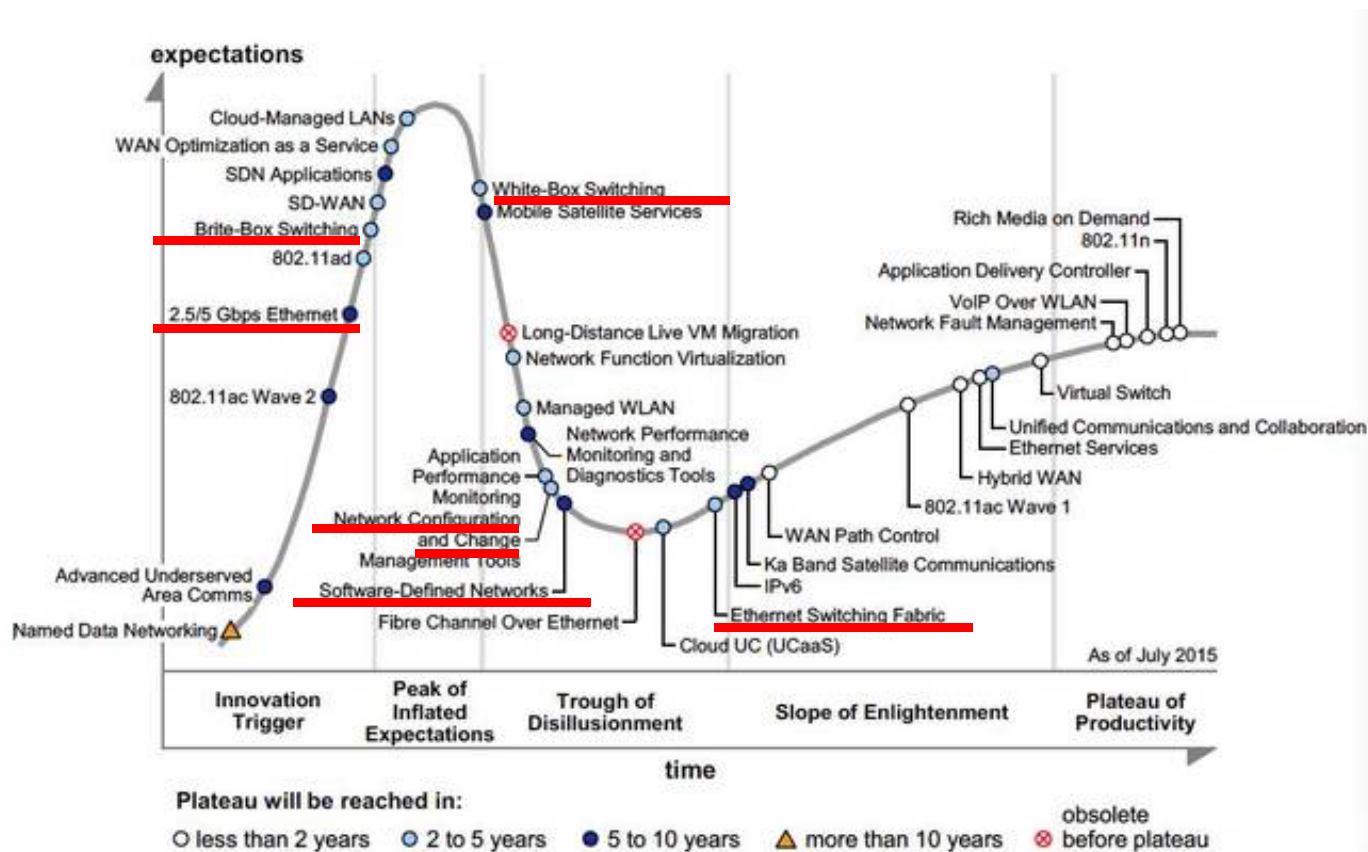 - Whitebox switch and switch APIs

Switch panel introduction

Yoshihiro Nakajima

What is switch chip?

Mem

Control CPU

PCI-Exp

Switch ASIC

Packet forwarding
Routing
Tunnel termination
ACL

Phy

Phy

Phy

Port cage

Port cage

Port cage

A whitebox ToR switch

Gartner Networking Hype Cycle 2015

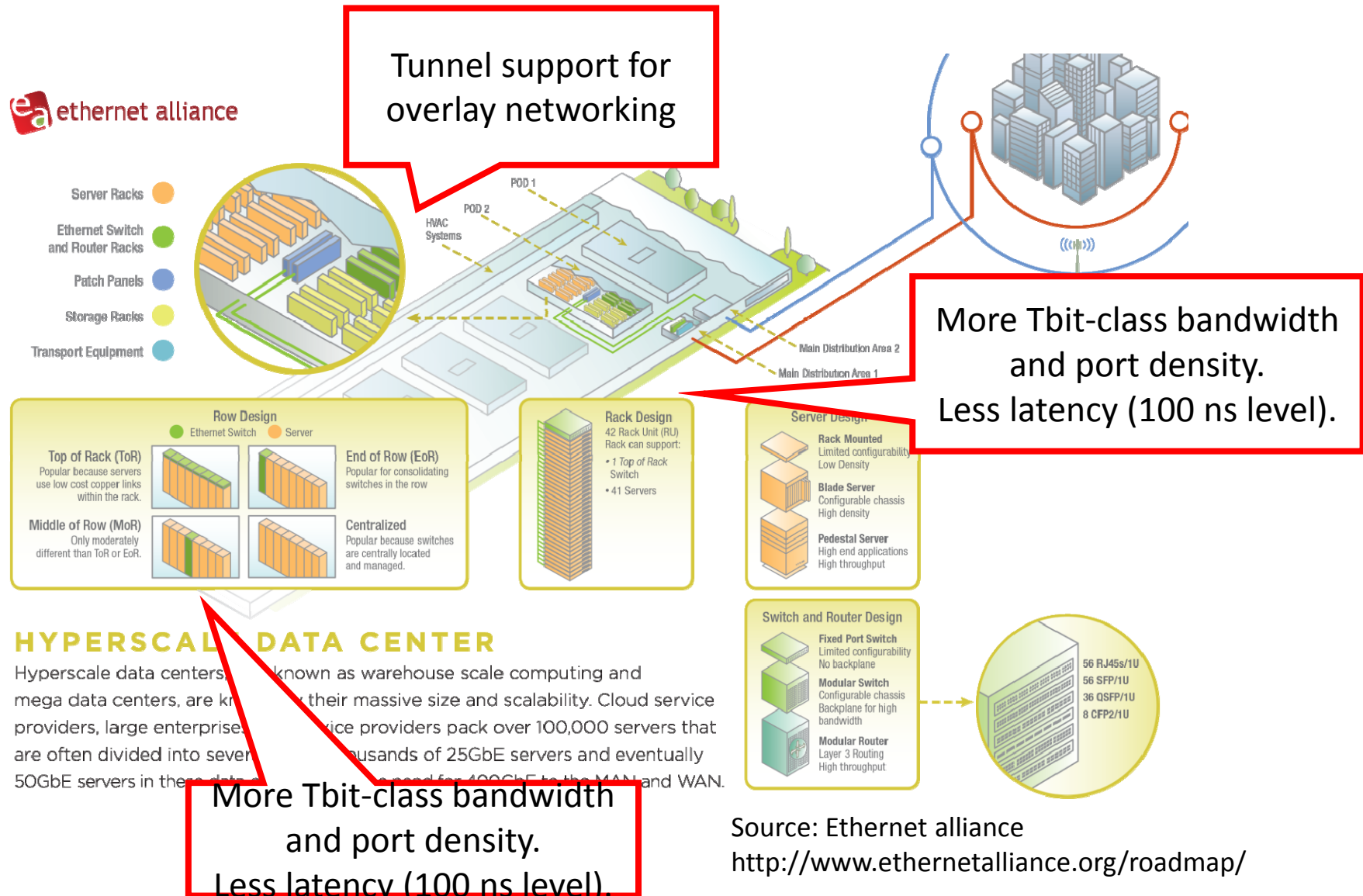Reference: Hype Cycle for Networking and Communications, 2015, http://www.gartner.com/document/3100229

Ethernet roadmap

Source: Ethernet alliance http://www.ethernetalliance.org/roadmap/
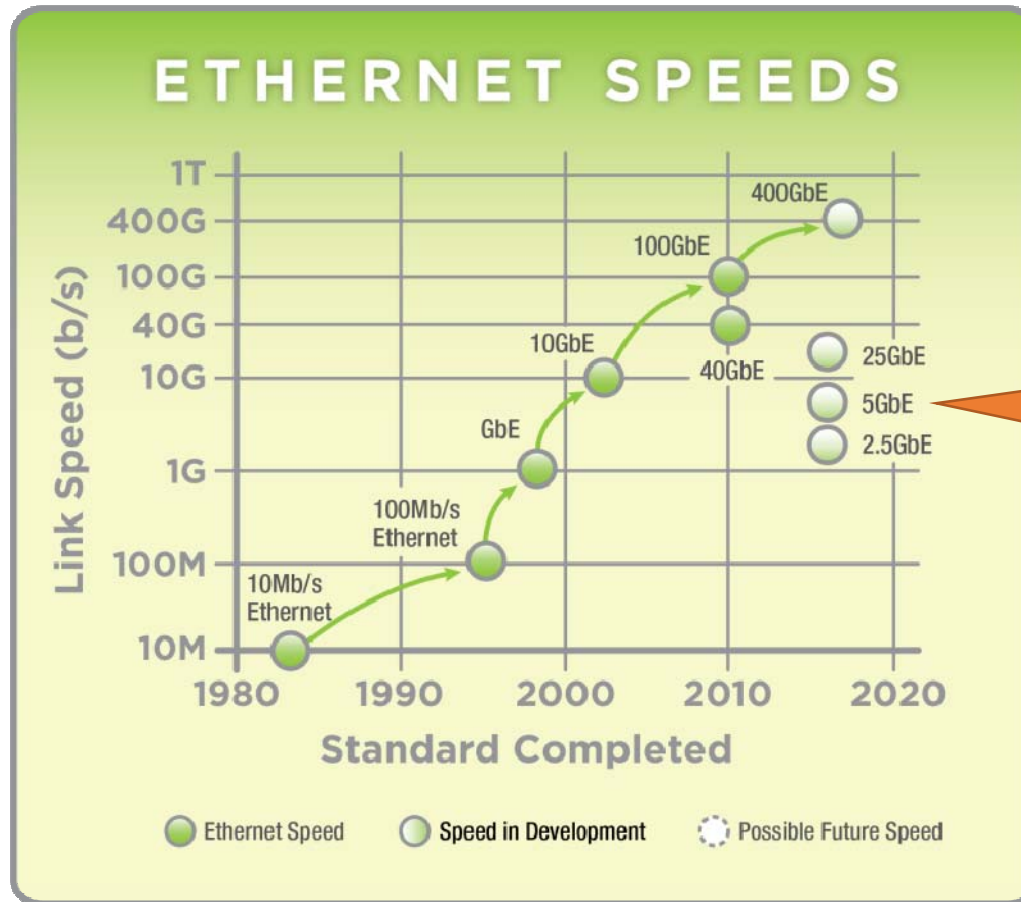
Target 1/3 (Hyperscale datacenter)

Tunnel support for overlay networking

More Tbit-class bandwidth and port density.
Less latency (100 ns level).

More Tbit-class bandwidth and port density.
Less latency (100 ns level).

Server Racks

Ethernet Switch and Router Racks

Patch Panels

Storage Racks

Transport Equipment

POD 1

POD 2

HVAC Systems

Main Distribution Area 2

Main Distribution Area 1

Row Design
Ethernet Switch   Server

Top of Rack (ToR)
Popular because servers use low cost copper links within the rack.

End of Row (EoR)
Popular for consolidating switches in the row

Middle of Row (MoR)
Only moderately different than ToR or EoR

Centralized
Popular because switches are centrally located and managed.

Rack Design
42 Rack Unit (RU)
Rack can support:
• 1 Top of Rack Switch
• 41 Servers

Server Design

Rack Mounted
Limited configurability
Low Density

Blade Server
Configurable chassis
High density

Pedestal Server
High end applications
High throughput

Switch and Router Design

Fixed Port Switch
Limited configurability
No backplane

Modular Switch
Configurable chassis
Backplane for high bandwidth

Modular Router
Layer 3 Routing
High throughput

56 RJ45s/1U
56 SFP/1U
36 QSFP/1U
8 CFP2/1U

HYPERSCALE DATA CENTER

Hyperscale data centers, also known as warehouse scale computing and mega data centers, are known for their massive size and scalability. Cloud service providers, large enterprises and service providers pack over 100,000 servers that are often divided into several to thousands of 25GbE servers and eventually 50GbE servers in the ... demand for 400GbE to the MAN and WAN.

Source: Ethernet alliance
http://www.ethernetalliance.org/roadmap/

Target 2/3 (Service Provider)

Packet Encap/Decap

Increase # of route and MAC entry

Source: Ethernet alliance
http://www.ethernetalliance.org/roadmap/

Target 3/3 (Enterprise and campus)

ENTERPRISE AND CAMPUS

Enterprises consume more Ethernet ports than the other environments by connecting desktop computers, devices and Voice over IP (VoIP) phones. The wired Ethernet networks are supplemented with wireless access points (WAPs) that are connected to Ethernet cables. 802.11ac WAPs are driving the need for 2.5 and 5GBASE T and eventually 10GBASE T. Most enterprise data centers are less than 10,000 sq ft and use Cat "x" cabling to connect to servers
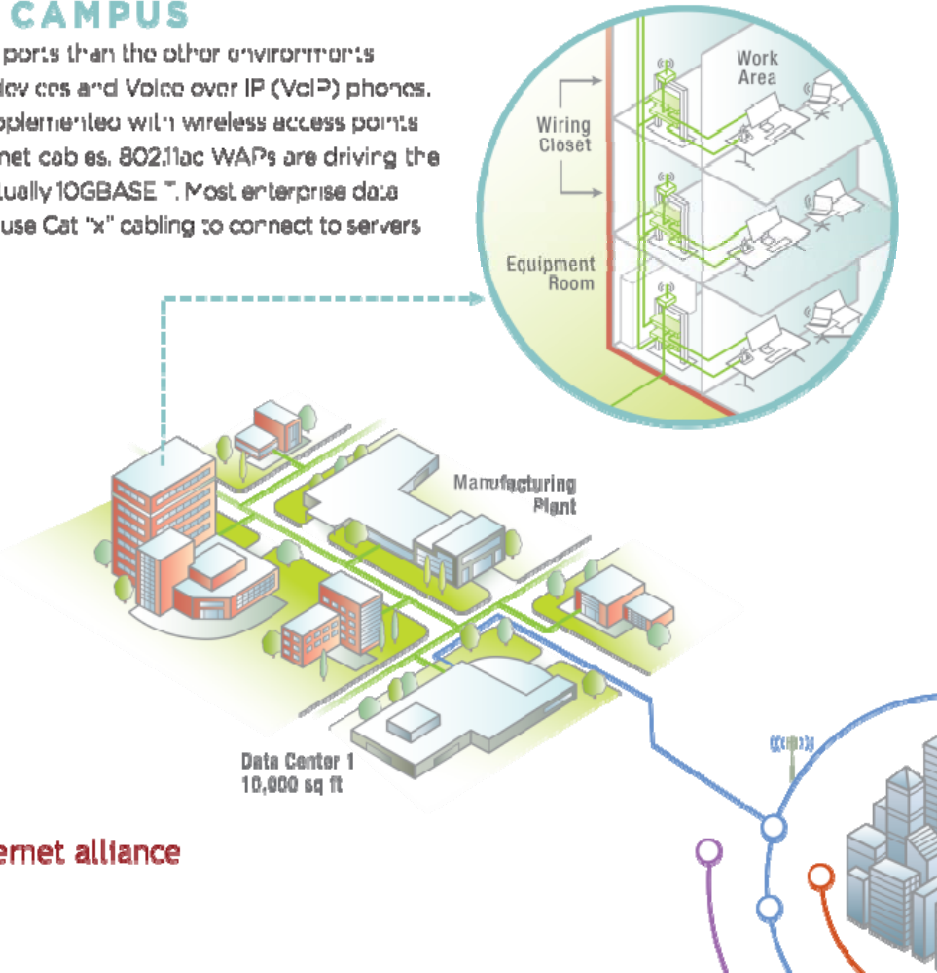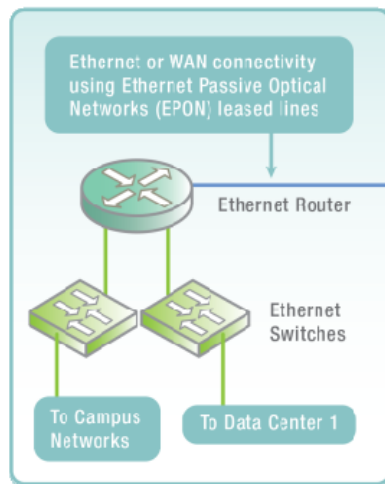
Low energy

Source: Ethernet alliance
http://www.ethernetalliance.org/roadmap/

Market trends of networking infrastructure for servers
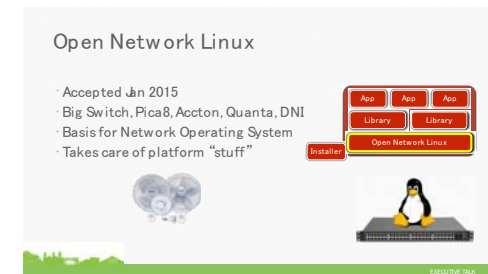
- CPU: Increasing # of CPU cores
  - Intel Haswell E5 Xeon 18 cores / CPU socket
  - Cavium ThunderX 48 cores / CPU socket
- Memory: Increasing size and speed
  - Size: up to 2 TB
  - Speed: 400Gbps class
- NIC: Increasing interface speed
  - From 10GbE to 40GbE or multiple 25GbE or 50GbE on single node configuration
  - 100GbE on multiple-node configuration
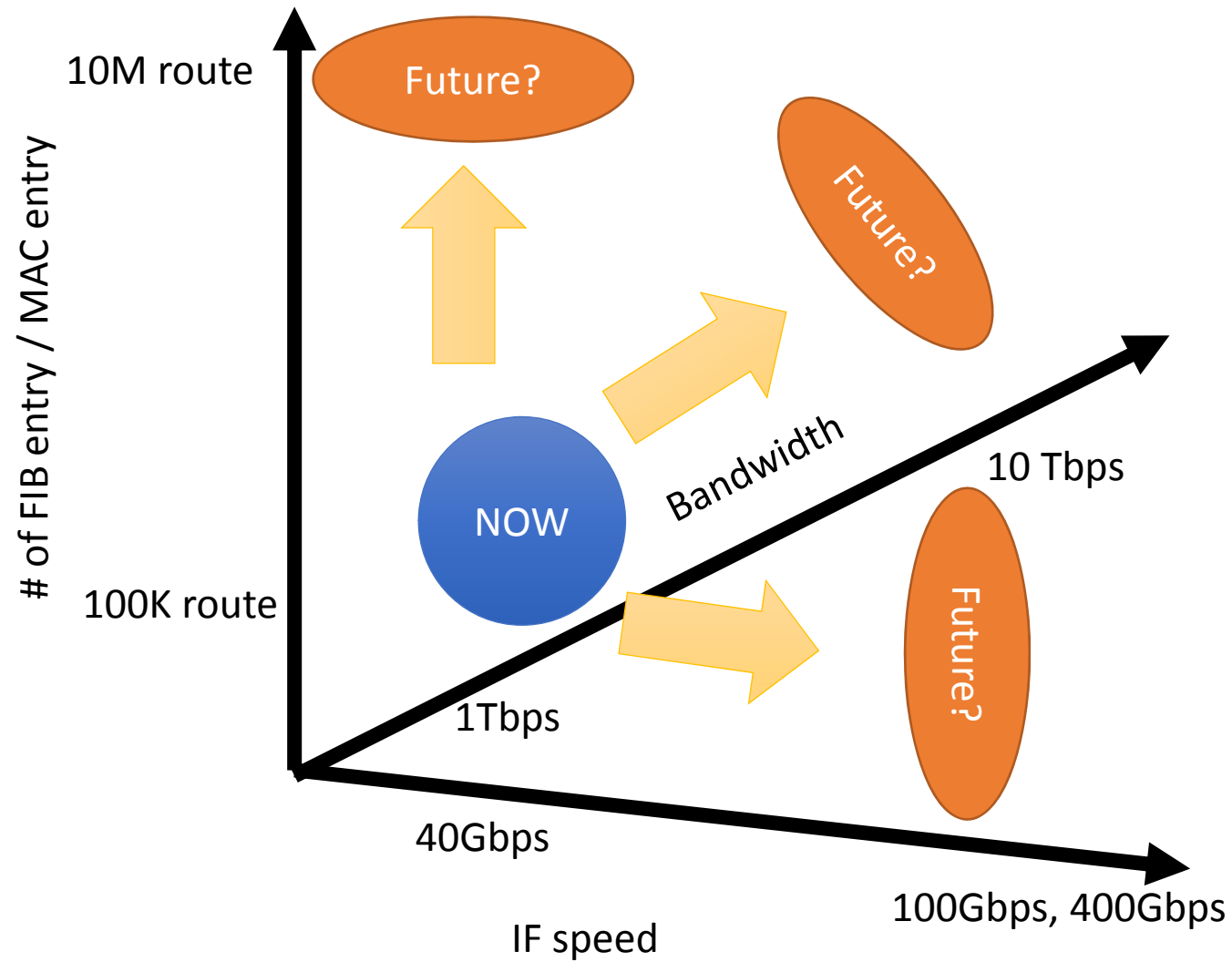
white box switch and open source network operating system

- Many ODM venders provide whitebox switch hardware
- Network operating system and
  tools for whitebox switch
  - Cumulus
  - OcNOS
  - OpenSwitch
  - Open Network Linux
  - ONIE
  - FBOSS

Open Network Linux
· Accepted Jan 2015
· Big Switch, Pica8, Accton, Quanta, DNI
· Basis for Network Operating System
· Takes care of platform "stuff"

Facebook – FBOSS Agent & OpenBMC
In production in FB data centers
· FBOSS Agent
  · Core library to switch ASIC
  · Thrift interfaces
· OpenBMC
  · Low-level system
    management of board

Switch abstraction for NOS and network control management system

- Open API and SDK for network operating system
  - OpenFlow
  - OCP SAI
  - Open Ethernet
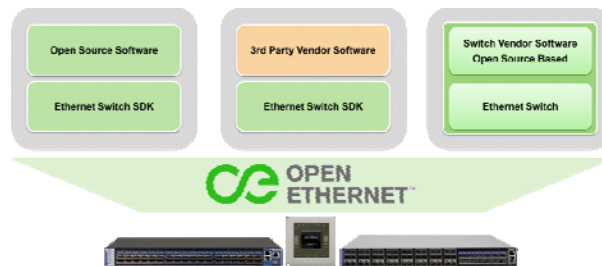 - Broadcom OpenNSL, OF-DPA
  - Linux kernel netlink
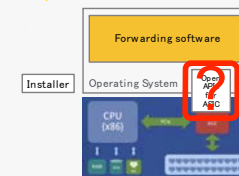
Switch ASIC company introduction

Switch ASIC company introduction (5-10 min by each company)

- Presentation on vision and direction of switch ASIC for next generation networking
  - high performance
  - improving functions and programmability
  - Products
  - APIs and SDK

- Broadcom (Eli)

- Mellanox (Benny)

- Barefoot (Prem)

- Cavium (Sachin)

Discussion

1. New requirements for next generation networking

- What is most significant aspects for ASIC/silicon chip set on networking switches in next 5 years?
    - High-performance (Bandwidth, scalability, latency)
    - functionality and protocol supports
    - Data plane programmability
    - Open API and SDK

2. SDN aspects

• Evolution of cloud computing and NFV are pushing networking trends to SDN strongly. What features are you planning to add to switch ASIC in the next 3 years?

• The switch ASIC design and implementation may be affected by new applications, such Internet of Things (IoT) and innovated big data analysis.
What kind of changes happen in the next 3 years in switch ASIC design?

3. Network virtualization

- Server virtualization technologies recently leverages overlay-typed networking virtualizations. Lots of changes in tunnel protocol formats have happened in these last 5 years, but it is so difficult to acquire deep knowledge of so many protocol formats which are rapidly developed.

- How do you catch up them immediately?

- How do you realize high-performance and scalability with tunnel protocol processing such as VxLAN, NSH, GRE, and MPLS on switch chip?

- What strategy do you have for emerging technologies such as SD-WAN, SDN applications, WAN optimization and service function chaining?

4. White box switch and Switch APIs

- What level of switch abstraction (API, SDK) is better for network OS developer and engineers from the network operating system point of view?

- What is interoperability in era of whitebox switch and open-API?

- What's next target or potential area of whitebox switch?