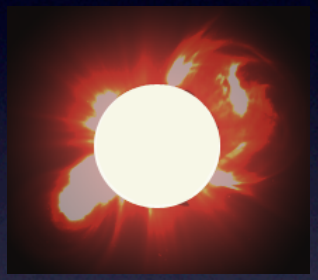
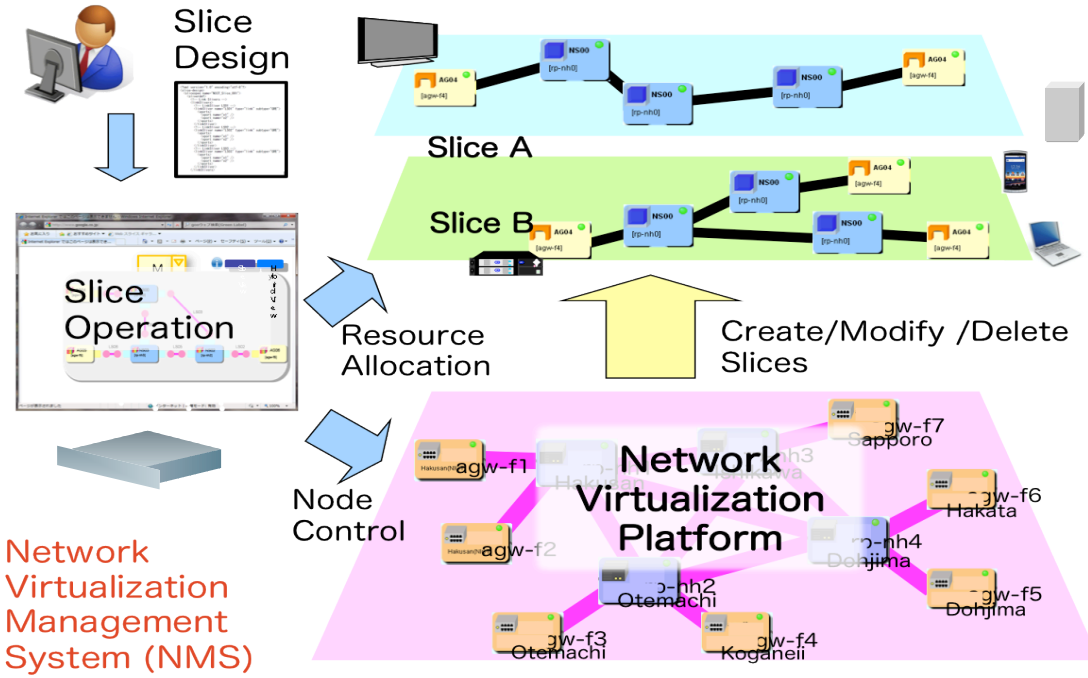
Revisiting "Clean-Slate" Approach to Re-designing the Internet



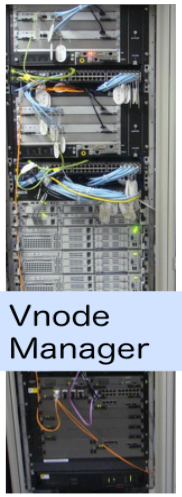
Aki Nakao
University of Tokyo
nakao@iii.u-tokyo.ac.jp
2012/7/10
GEC14

VNode Project

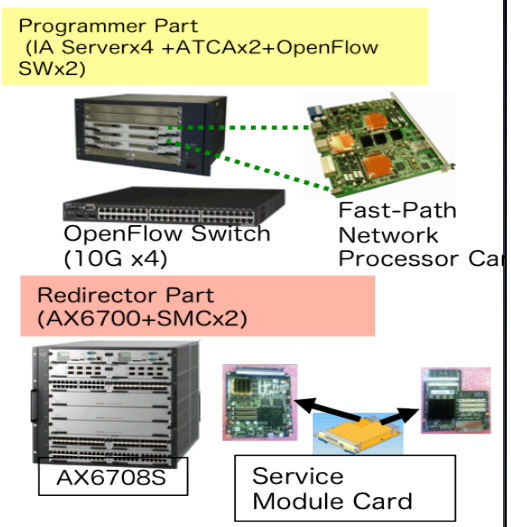
- VNode Infrastructure Enables Deeply Programmable Network
  (Project Leader: Aki Nakao)
- 2008-2010 Collaborative Research (NICT/Utokyo/NTT/NEC/Hitachi/Fujitsu)
- 2011-2014 Collaborative Research (Utokyo/NTT/NEC/Hitachi/Fujitsu/KDDI)



Network Virtualization Node (VNode)

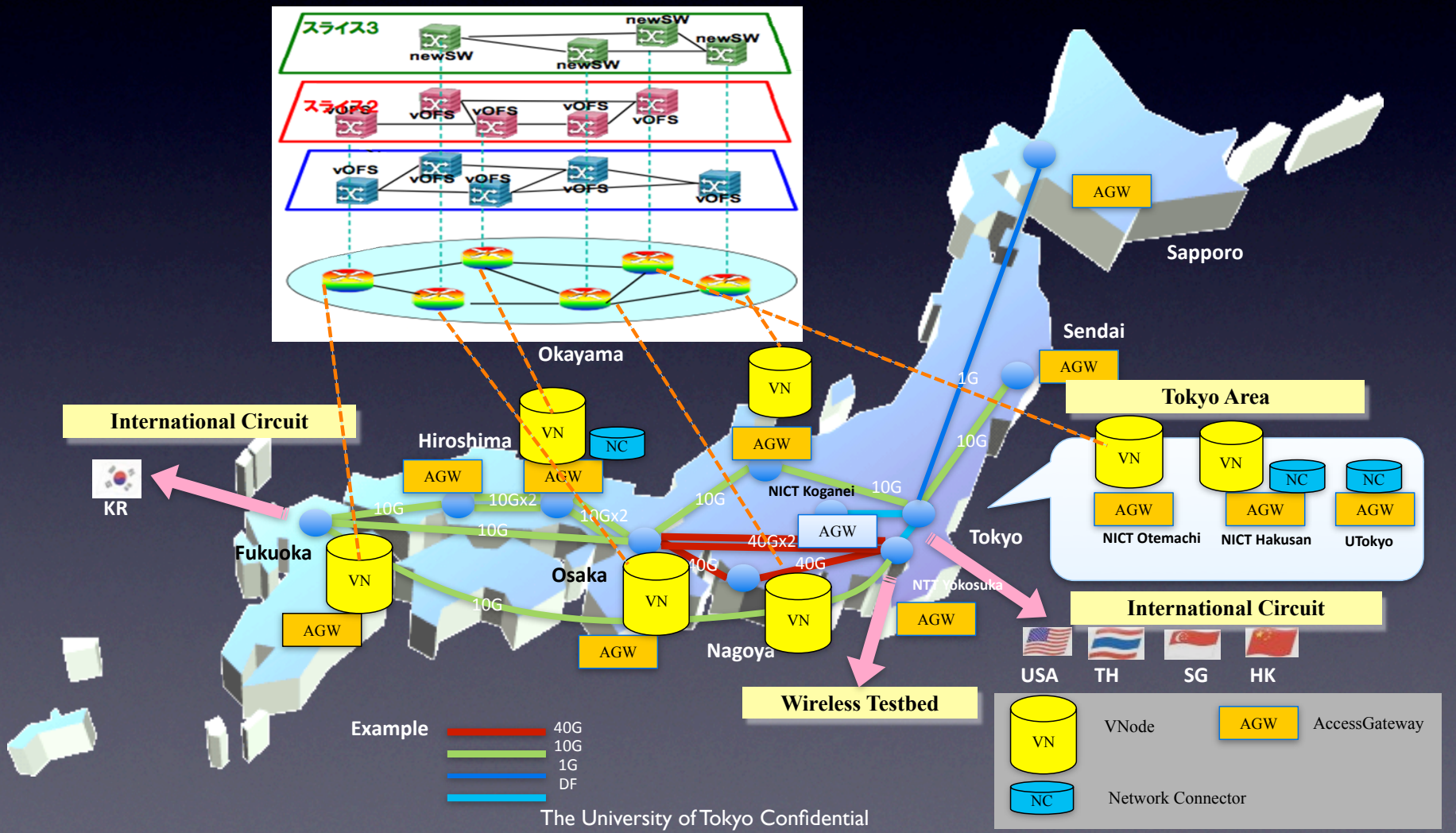


Vnode Manager



VNode Infrastructure (extended to US!)

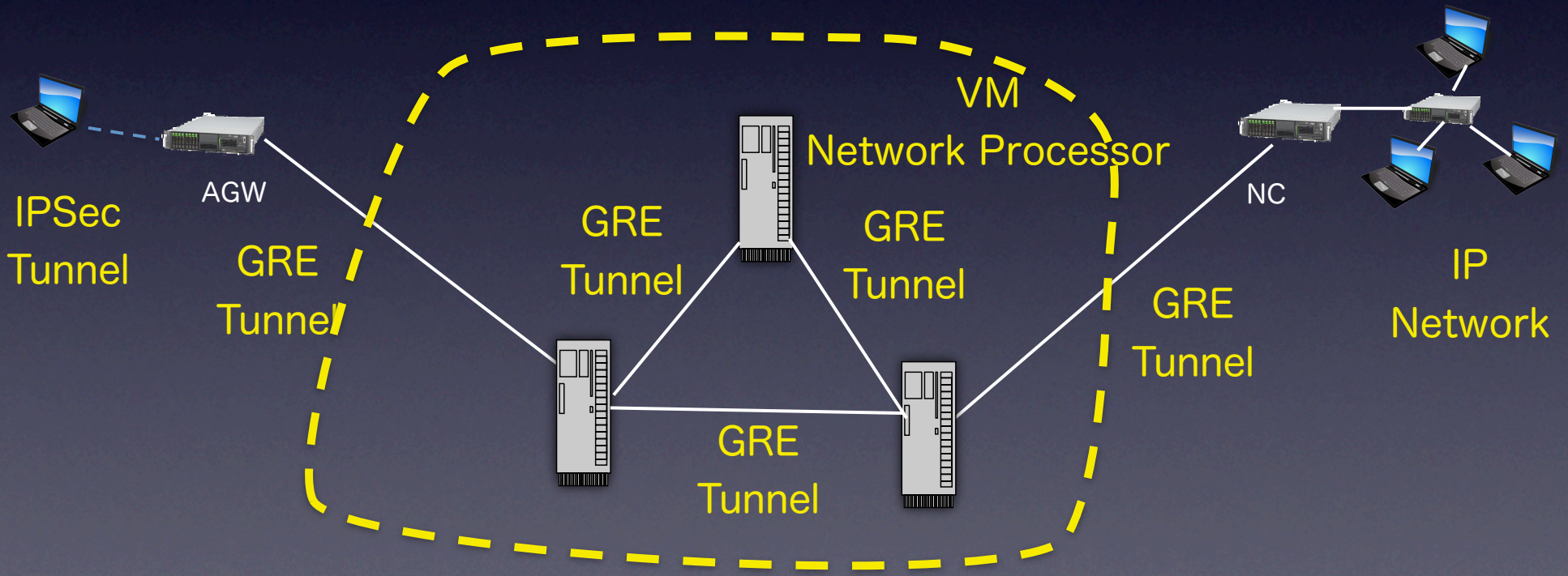
- 7 VNodes, 2 Network Connectors, 11 Access Gateways
- Deep Programmability for Experimenting with Arbitrary Protocols (Non-IP)
- Slice-Around-The-World Project (A VNode in U of Utah on ION/StarLight!)



Deep Programmability for Backbone Network

- Create slices for running arbitrary protocols
- Programmability for both control plane and data plane
- Mainly focused on wired / fixed network virtualization

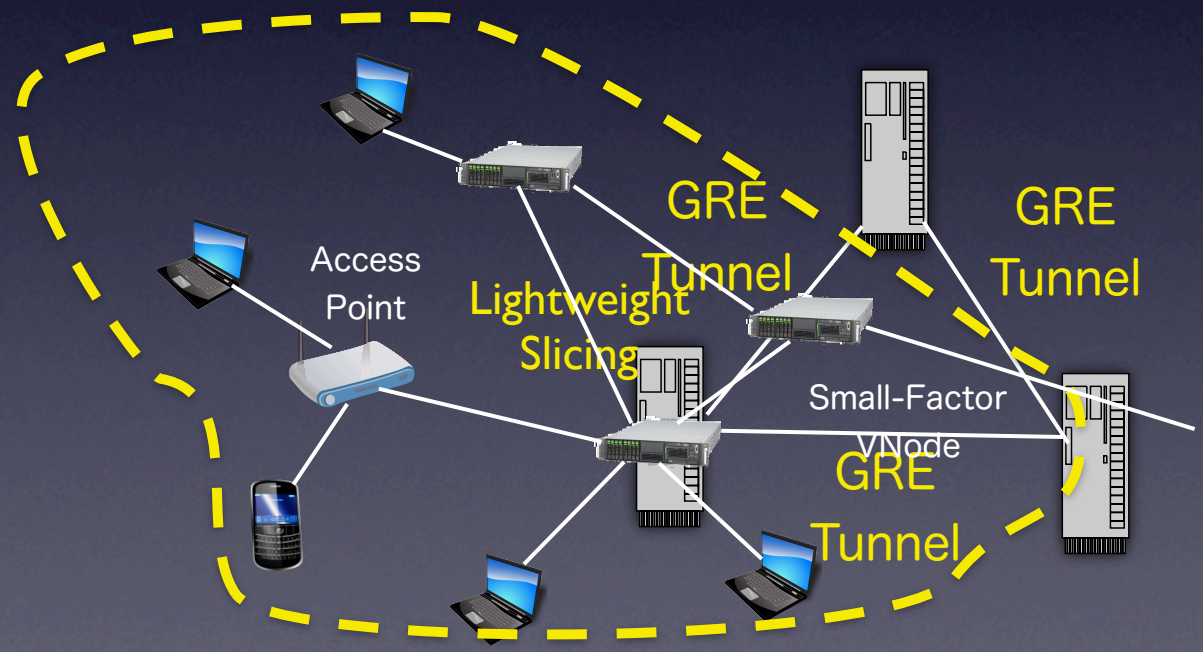
VNode Infrastructure



Deep Programmability for Network Edge

- "Tangible" small-form-factor (1U) VNode
- Deeply programmable, even at L2, yet high performance
- Fixed-mobile converged slicing

Network Edge Slicing

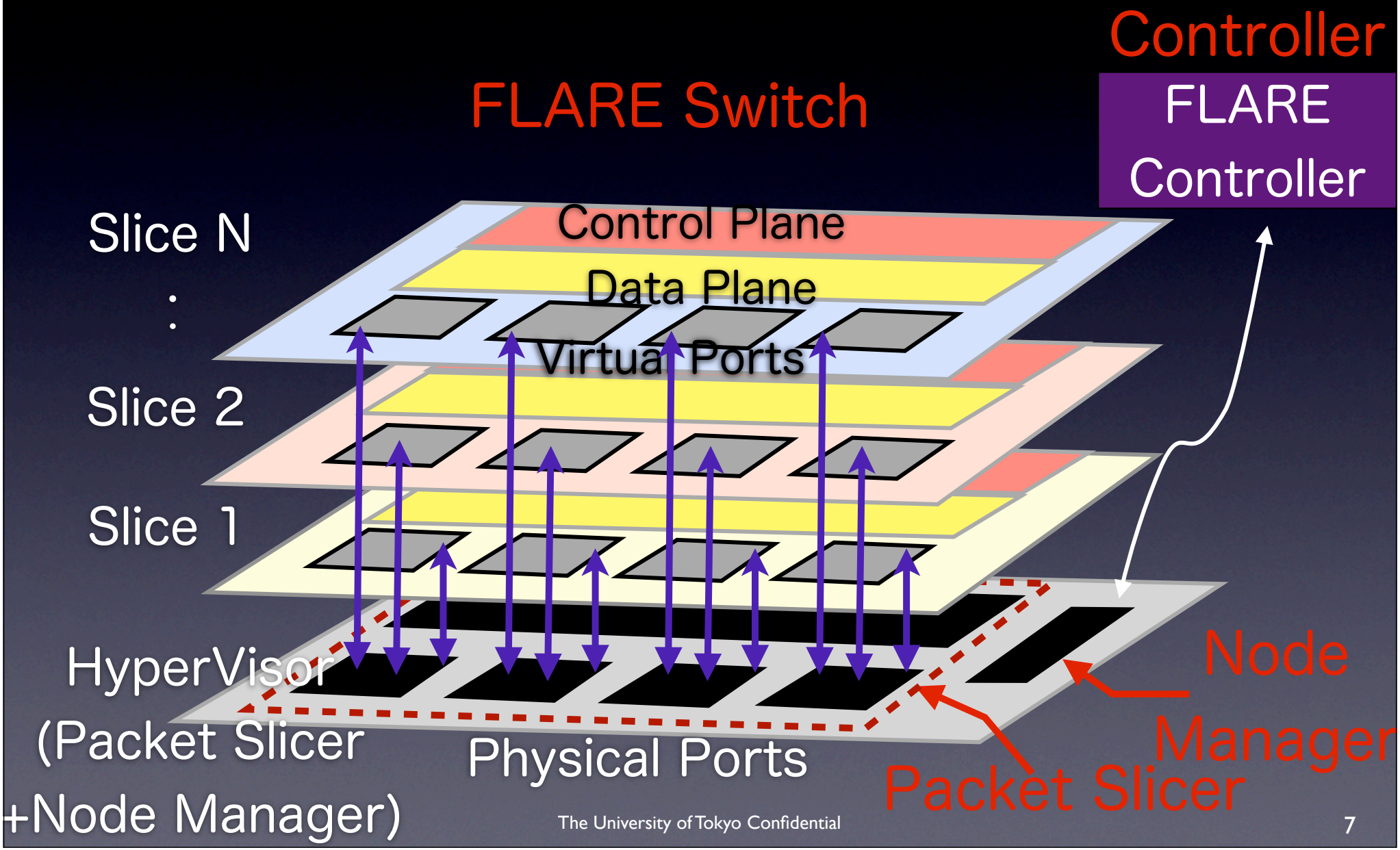


FLARE

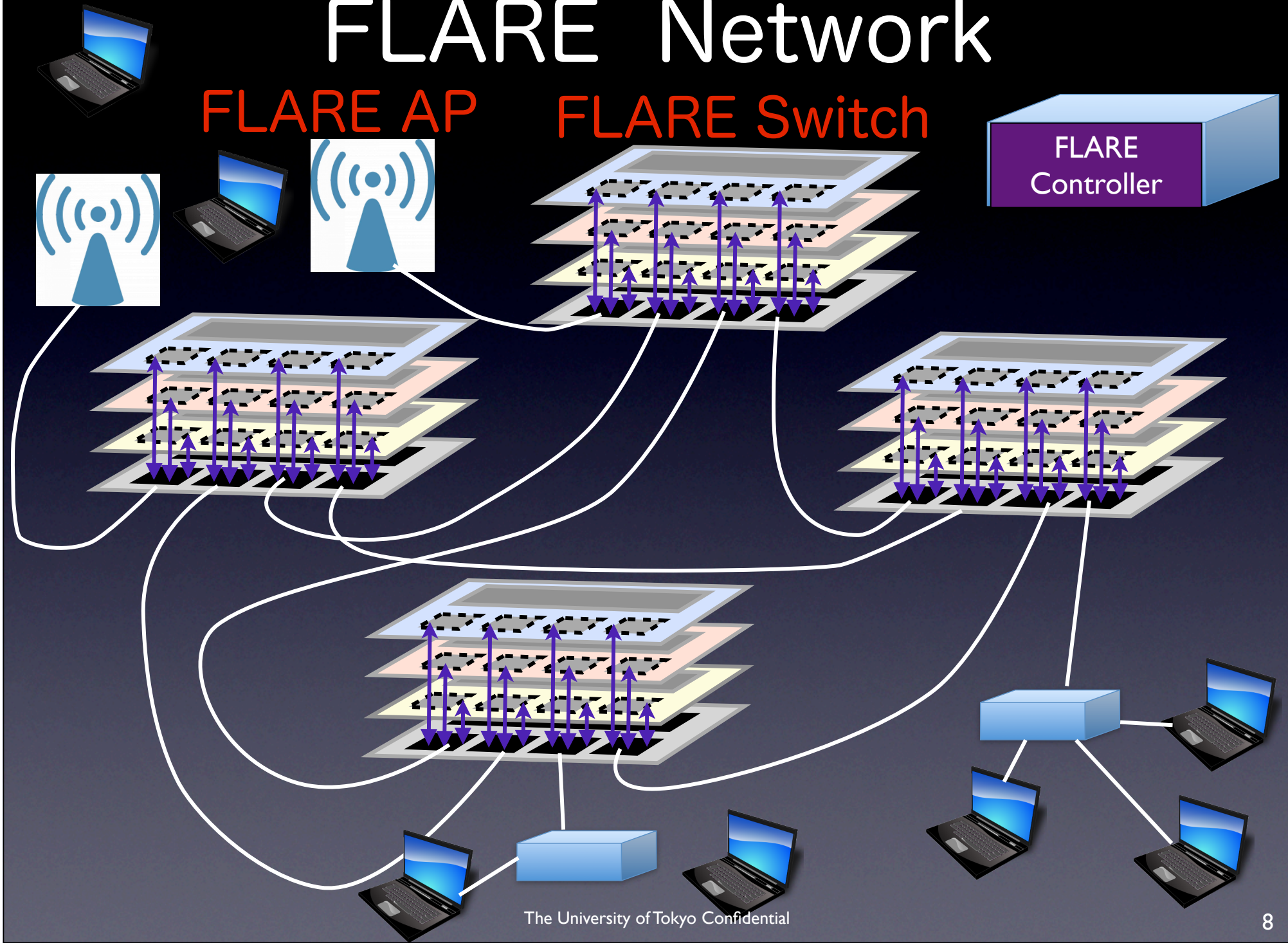
Open Deeply Programmable Switch

Challenges:

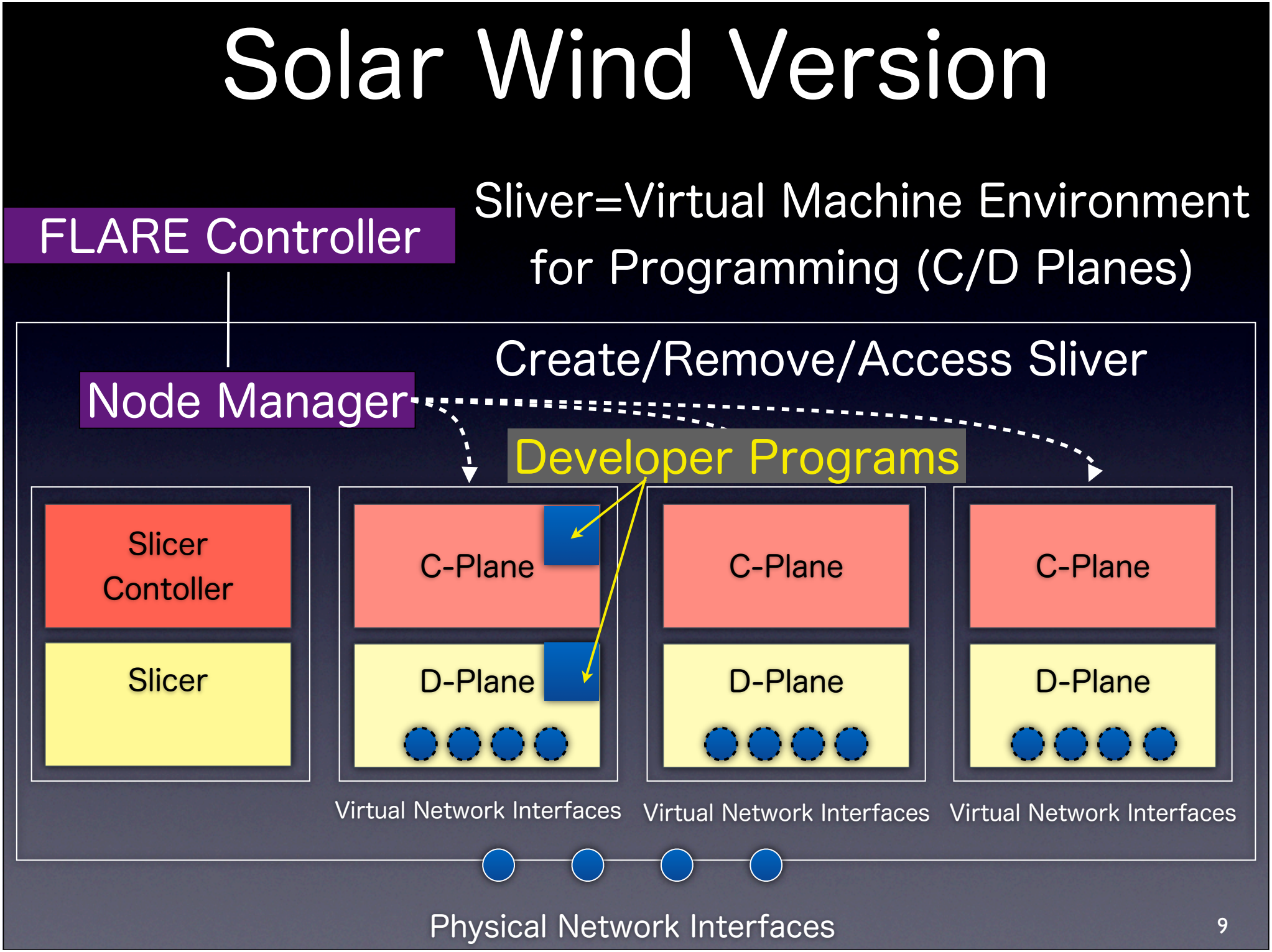
FLARE Architecture



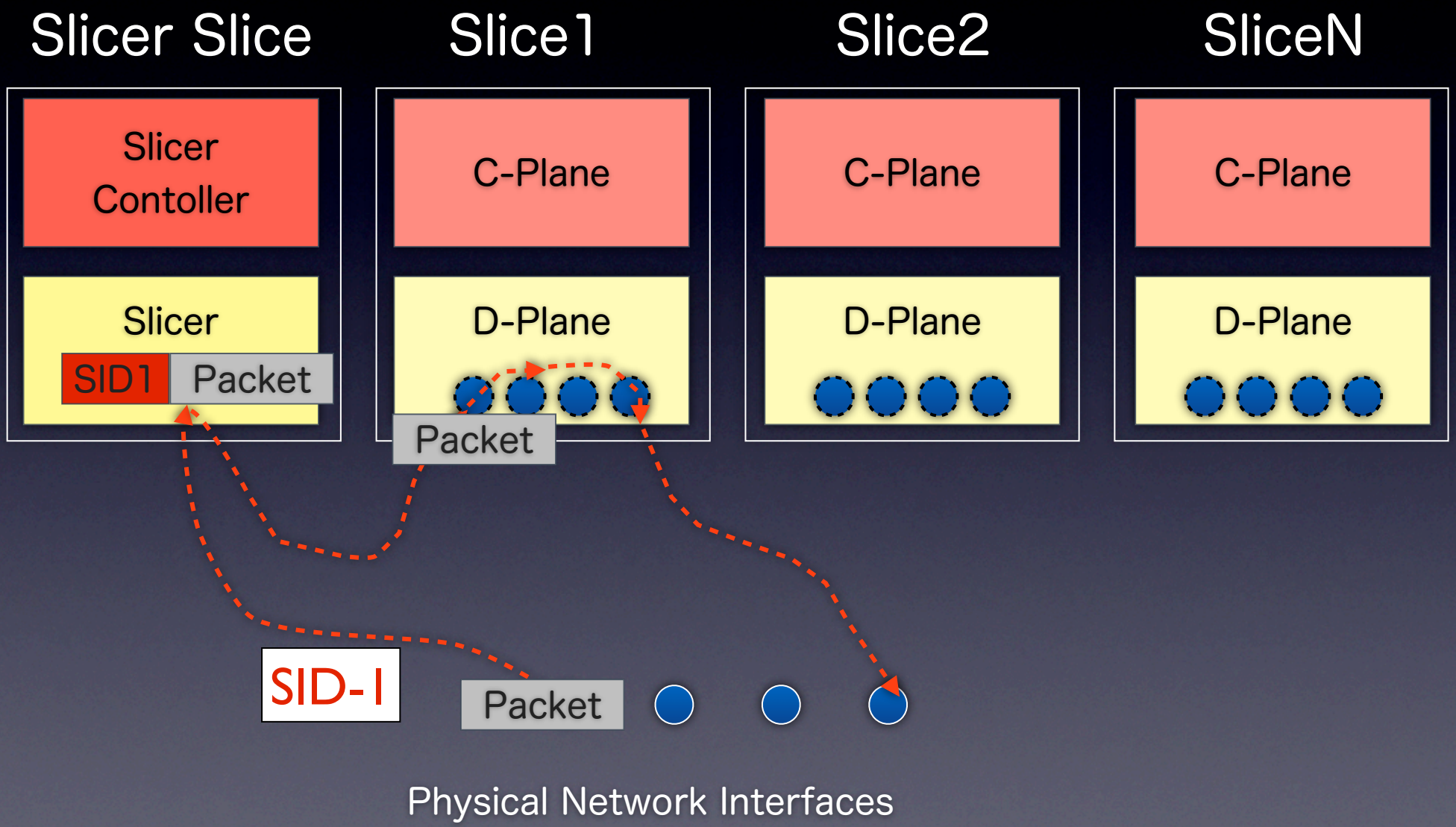
FLARE Network



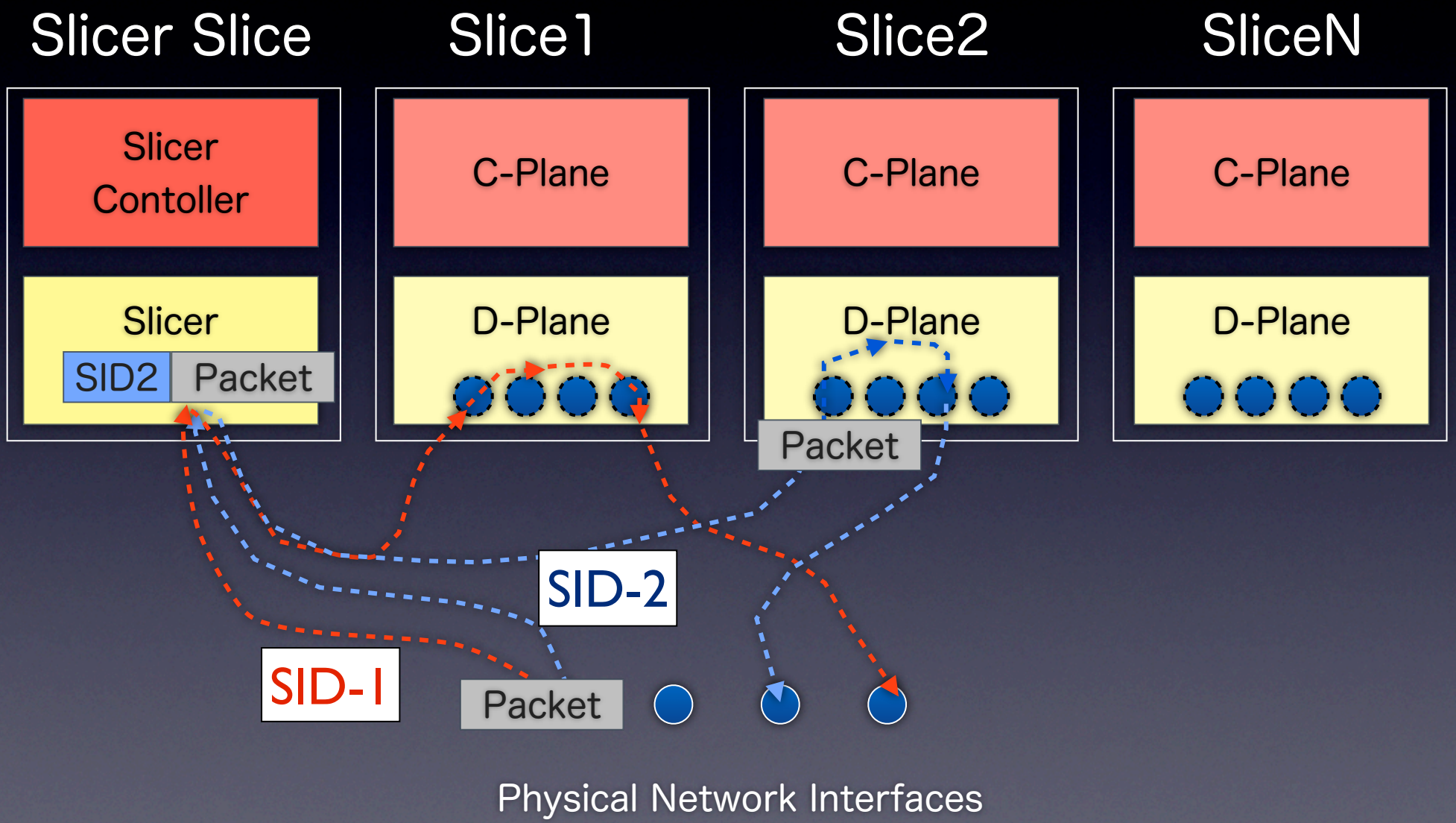
Solar Wind Version



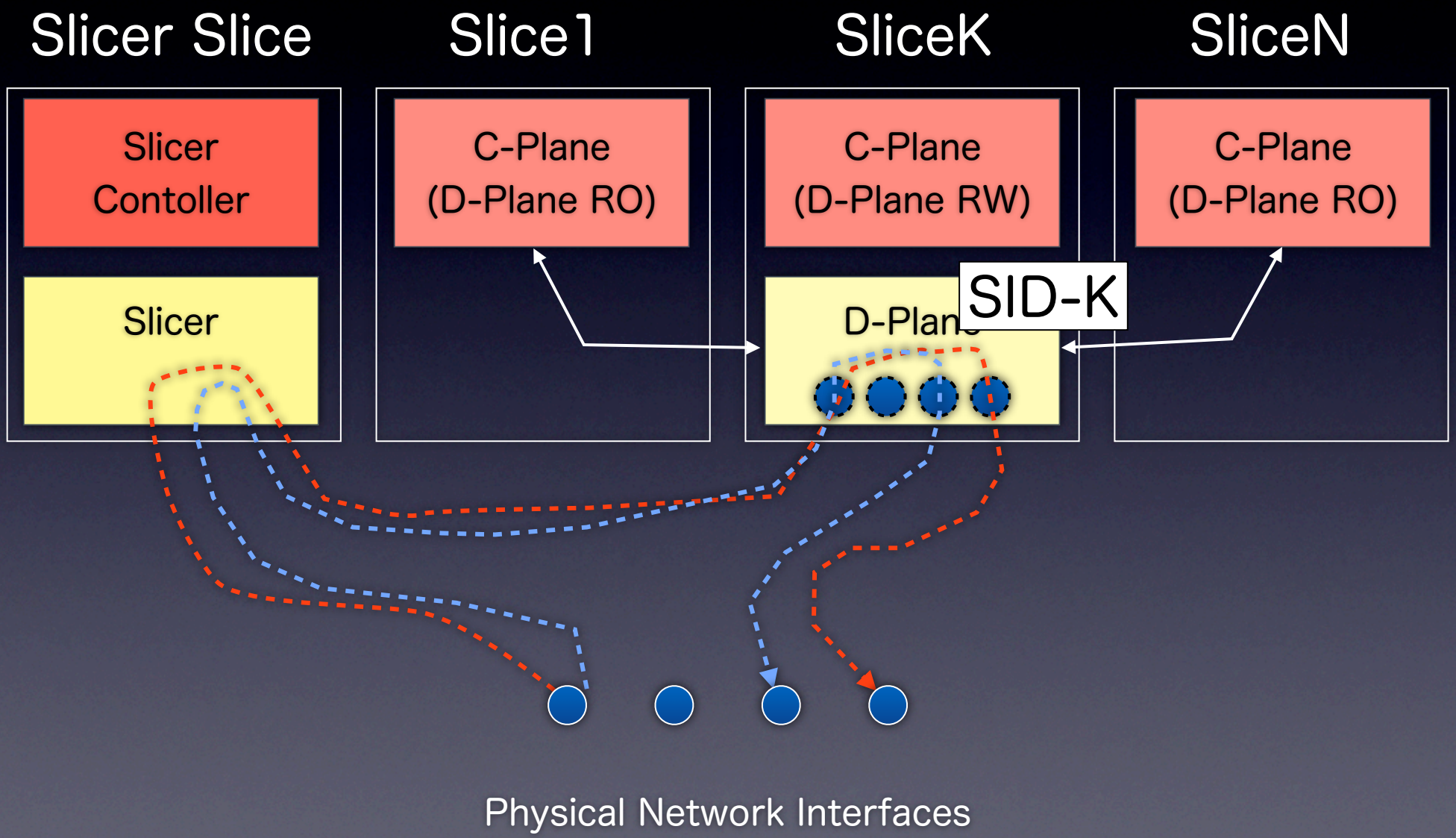
Per-Slice Switching



Per-Slice Switching

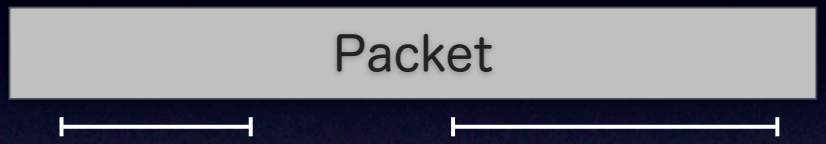


D-Plane Shared



Extracting Slice ID

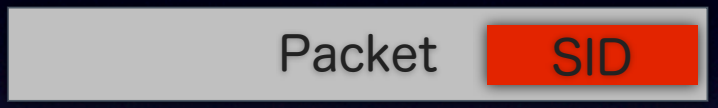
General Slicing
(Supported in DarkFlow Version)



Arbitrary Offset, Arbitrary Length

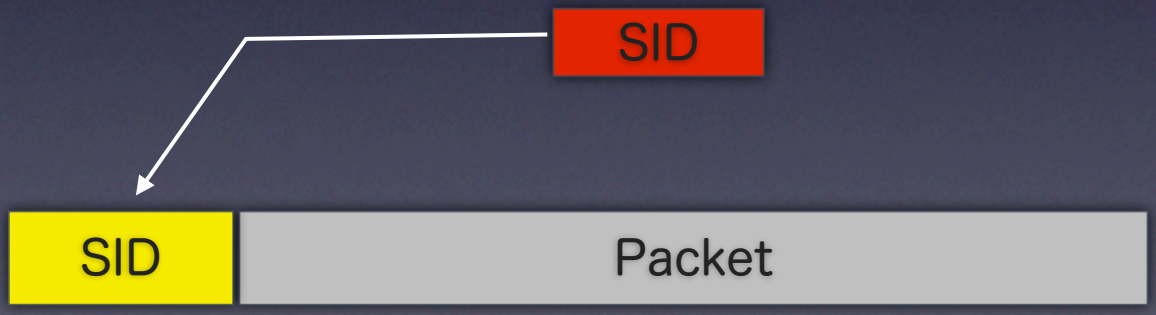
Multiple Parts

Trailer Slicing
(SolarWind Version)



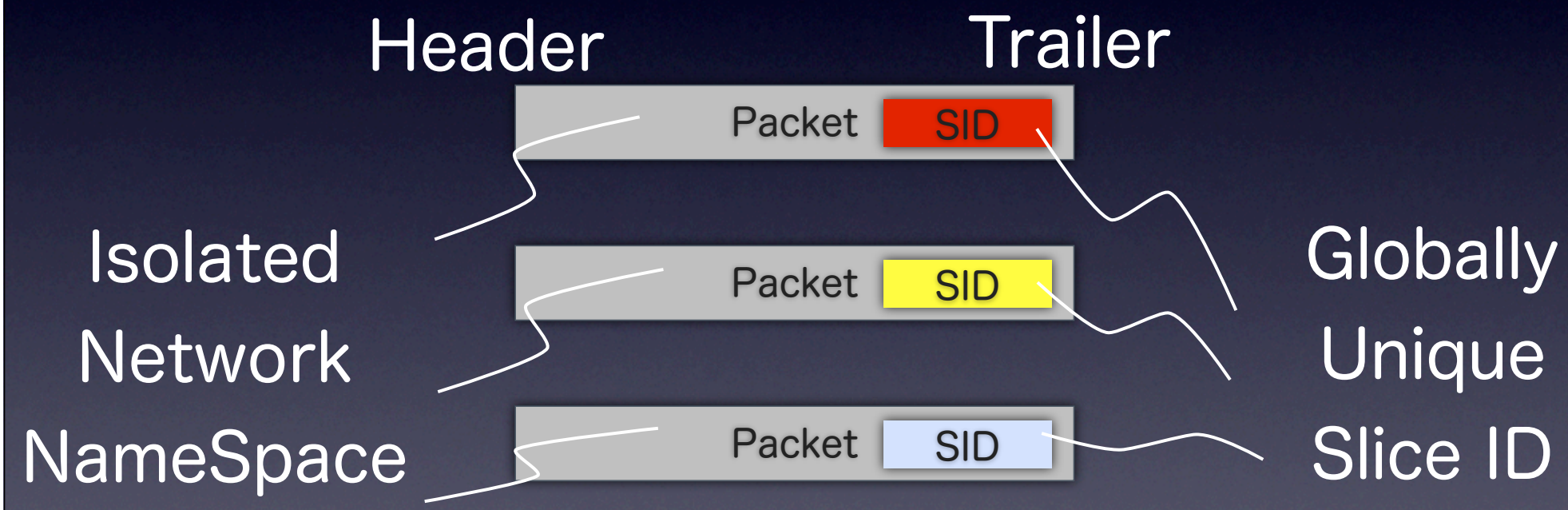
SID Embedded in Trailer

Slicer-slice identifies Slice ID(SID)



Internal Packet Format

Network Virtualization
Through Slice ID Embedded in Packet Trailers



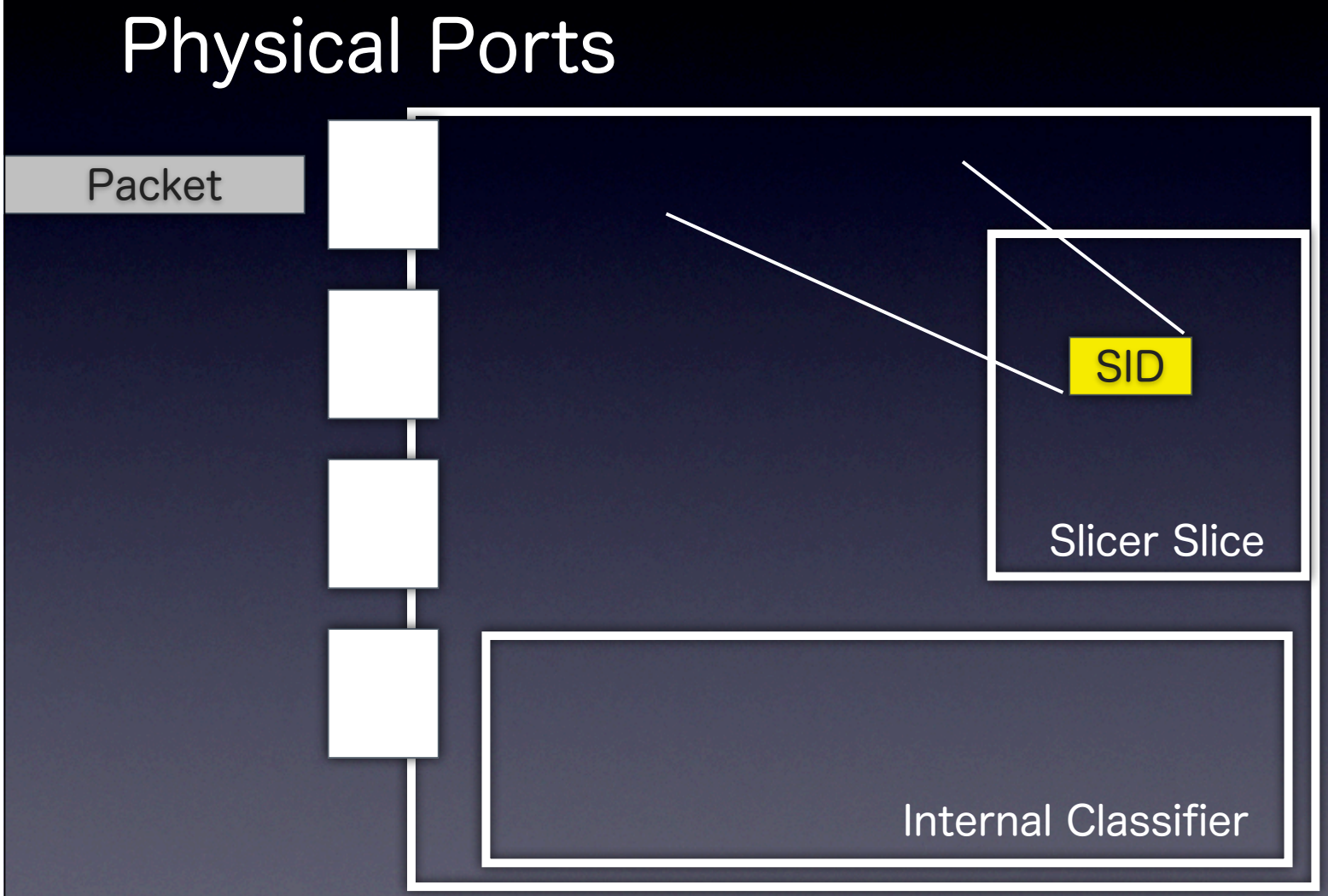
DarkFlow Version Slicer API

- Header Matching
- Trailer Matching
- Flow Hash
- Regular Expression Matching

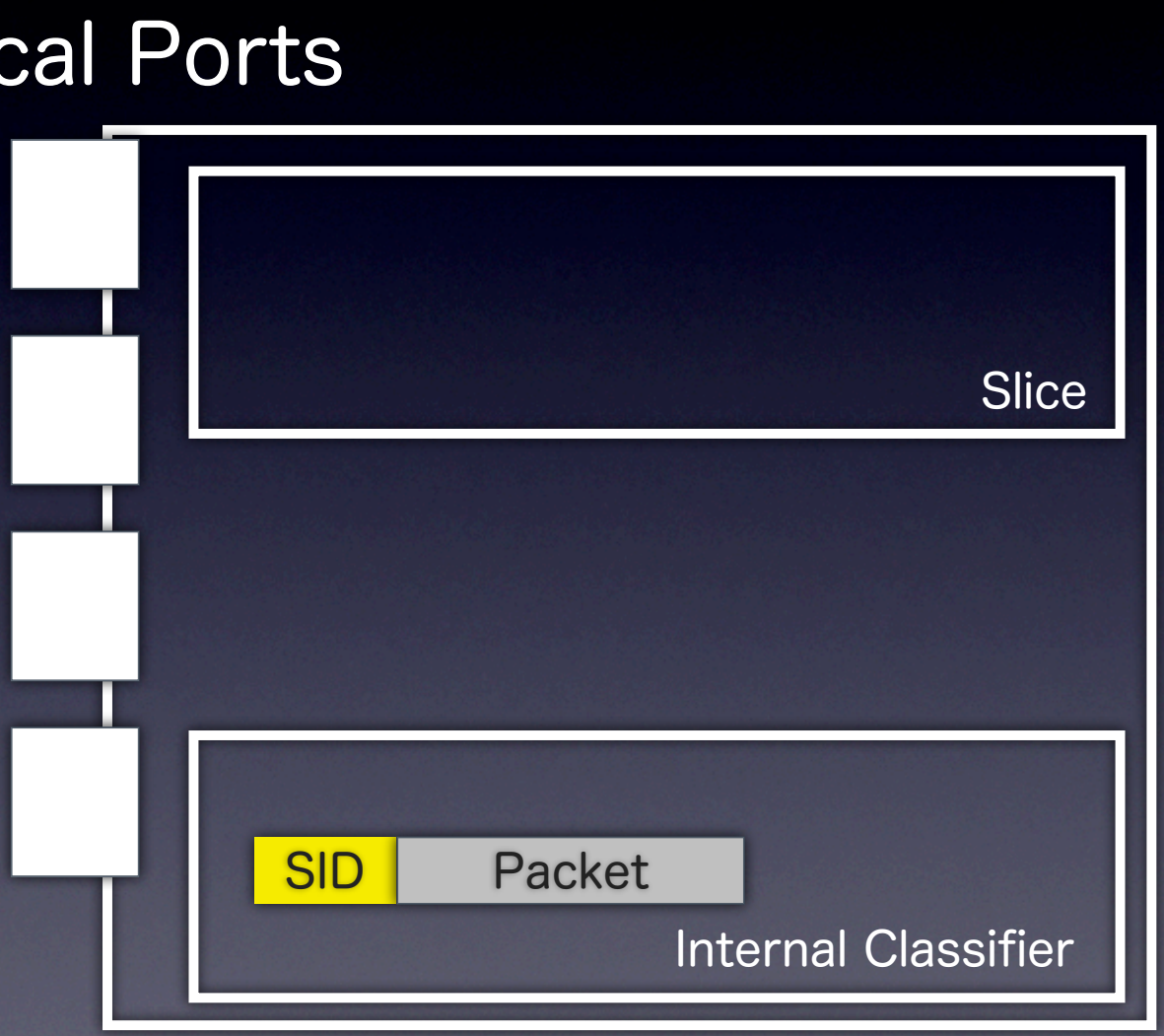
Three Primitives in FLARE

- Extract SliceID (Slicer)
- Demux/Mux to/from Slices (Redirector)
- Execute Programs on Slices (Programmer)

Slicer: Extract Slice ID

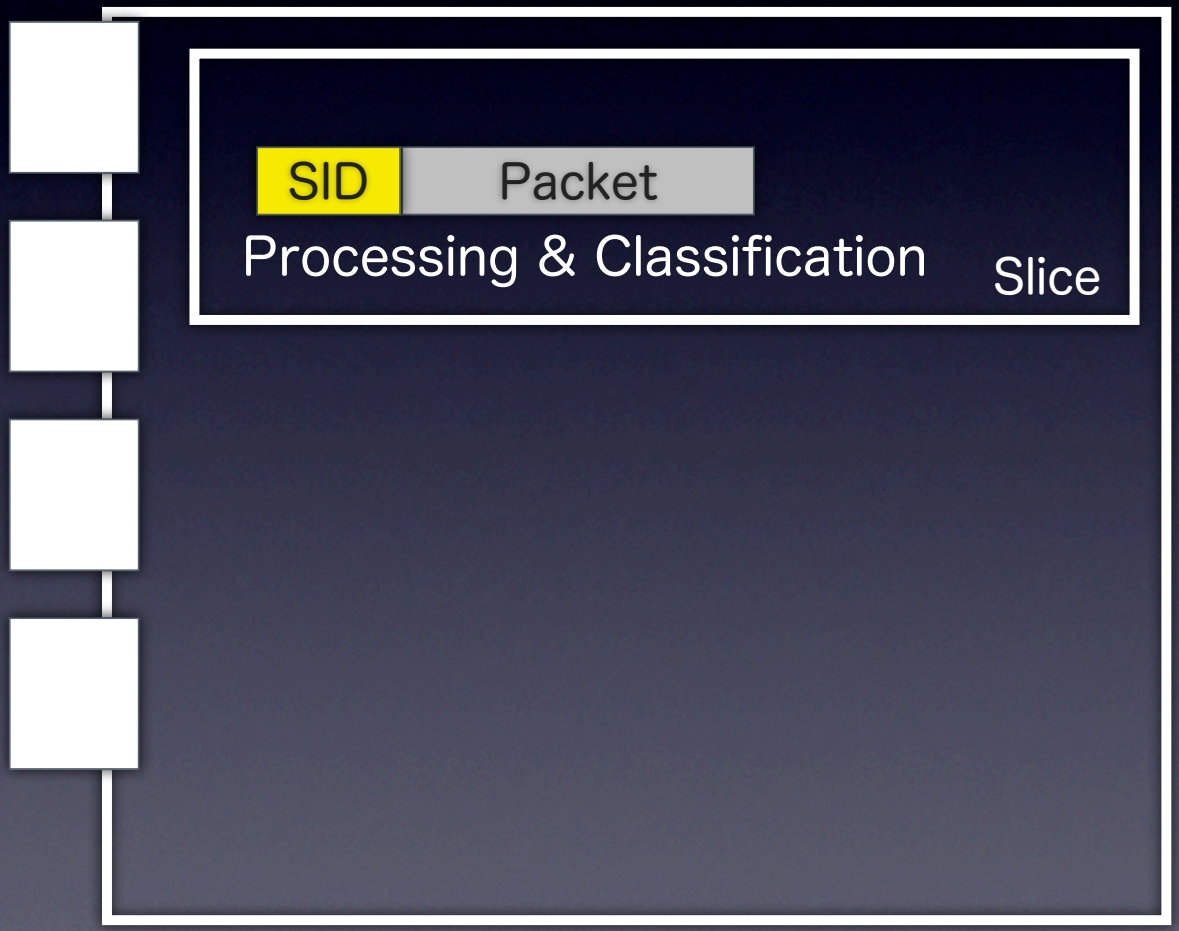


Redirector: Demux on Slice ID

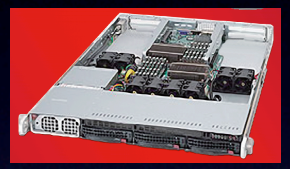


Programmer: Packet Processing

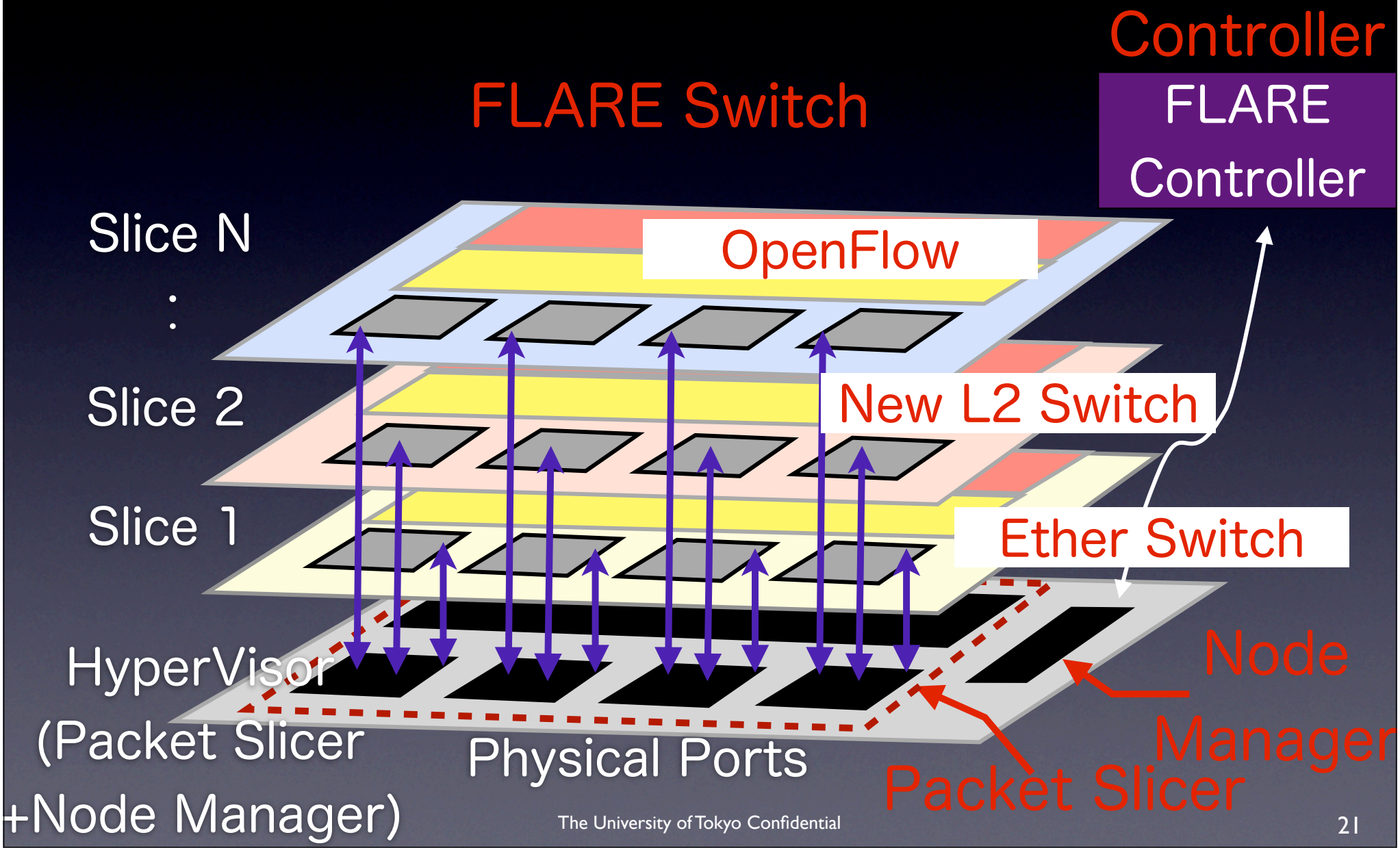
Physical Ports



SolarWind Version Implementation

- Hybrid of many-core processors + x86 processors
- 1U Form Factor 
- Trailer Slicing (Embedding Slice ID in Packet Trailers)
- 4 x 10Gbps Nonblocking (up to 2 Data Plane Slices)
- Up to 5 Data Plane Slices
- Control Plane & Data Plane Linux Programmability (Multi-Threaded Click)
- OpenFlow In A Slice

Multi-Protocol/Control Coexistence



OpenFlow vs. FLARE Differences

	OpenFlow	FLARE
Arbitrary Frame	X (Internet Protocol)	O (Non-Internet Protocol)
Control Plane Programmability	O	Multiple APIs (OpenFlow Compatible)
Data Plane Programmability (Packet Process)	X	O
Data Plane Programmability (New Protocol)	X	O

OpenFlow vs. FLARE Differences

	OpenFlow	FLARE
Programming Model	Flow Pattern Match Action	Linux Container Multi-Threaded Click
Processing Environment	Not Integrated	Integrated
Slicing Programmability	X (Fixed to Flow Matching)	O (fully supported in DarkFlow version)

Disclaimer: Not competing with OpenFlow technologies
We may be able to help improving OpenFlow

How deep programmability do we want?

Several questions to ask:

- Control plane programmability only?
- Data plane too (cache, transcode, DPI)?
- Aren't we pursuing "Clean Slate" design?
  e.g., Can we change L2 protocol?

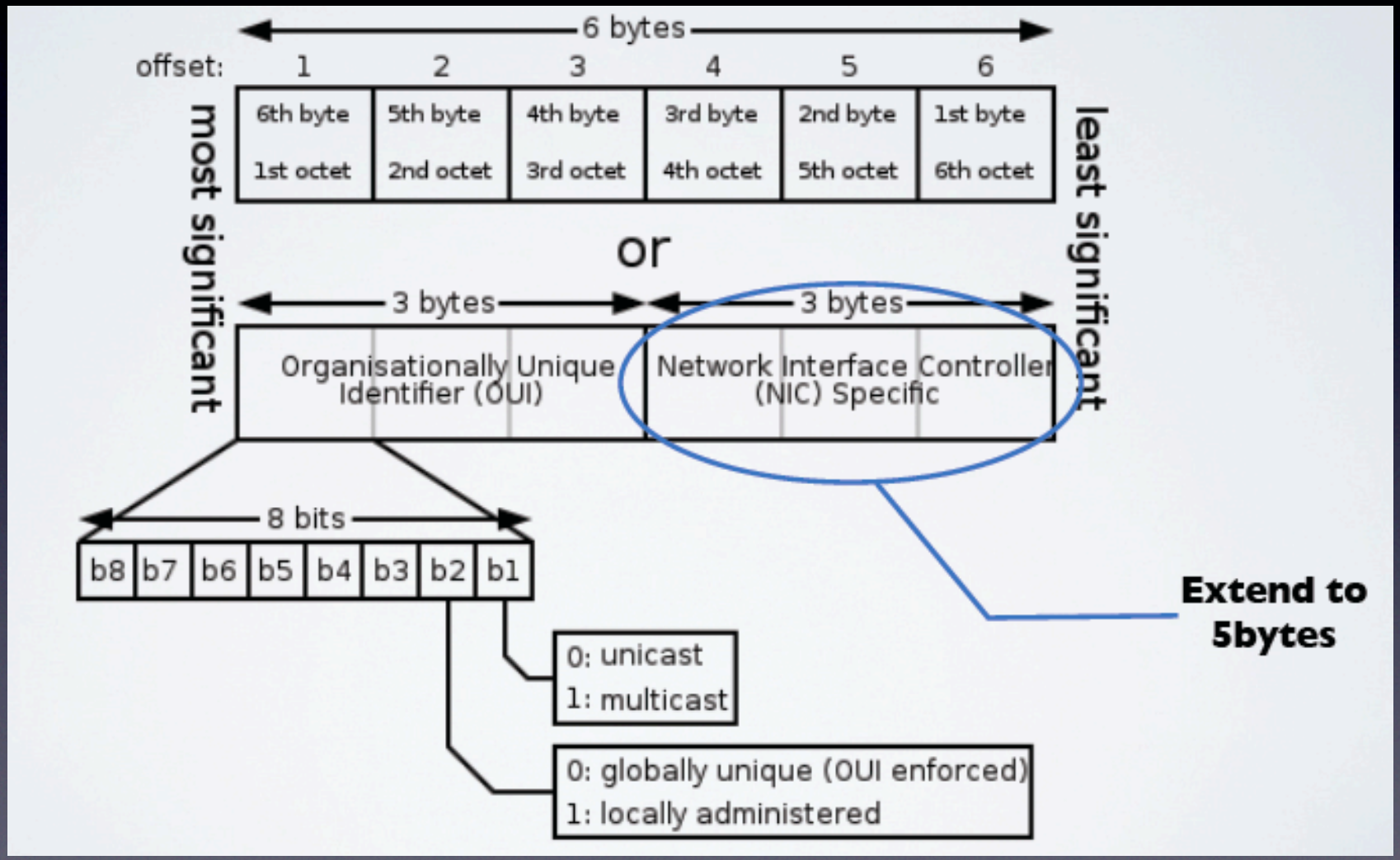
A Case in Data Center Network

- Limitation in MAC address space
  - Conflict of MAC addresses in VM migration
- Limitation in VID (802.1Q) space
  - The number of tenants increases in IaaS

Data Center Network depends heavily on L2 leading to solutions such as EUI-64 and VXLAN

Mac Address Extension

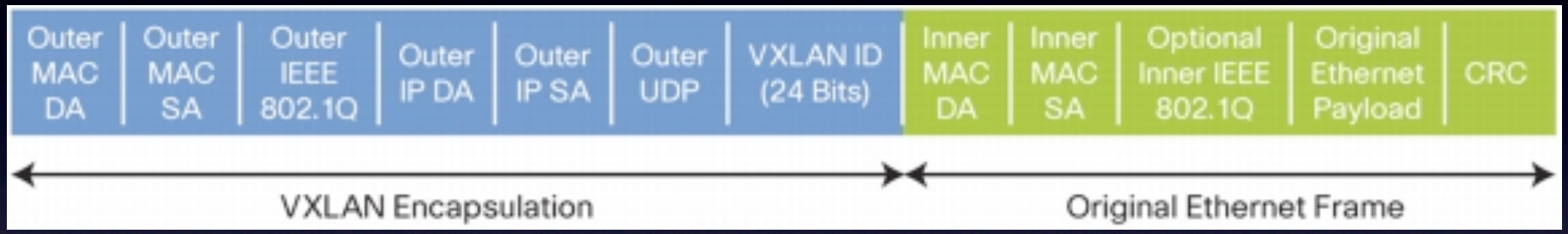
EUI-64 (64bit Extended Unique Identifier)



http://en.wikipedia.org/wiki/MAC_address

VXLAN

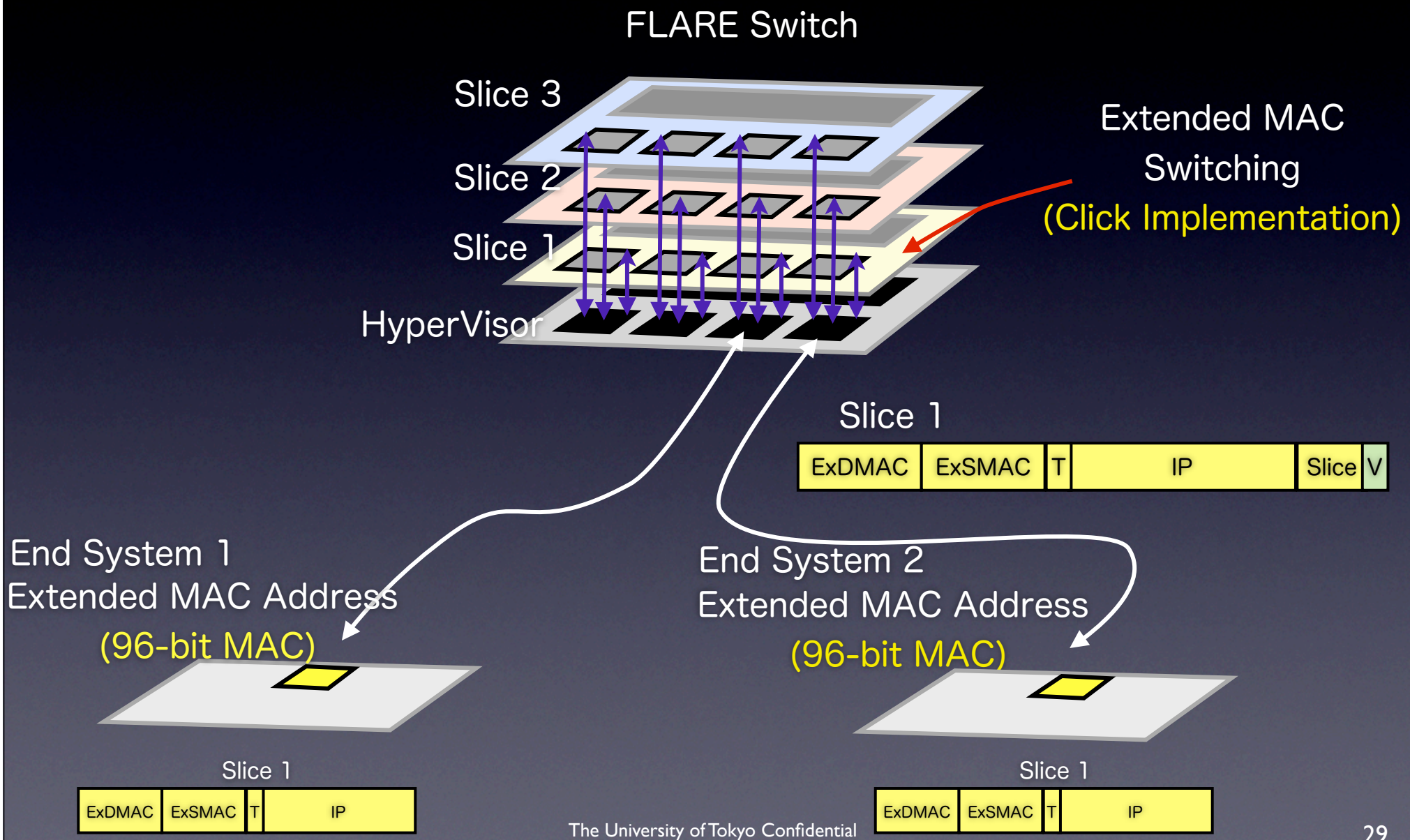
VXLAN(Virtual eXtensible Local Area Network)



http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9902/white_paper_c11-685115.html

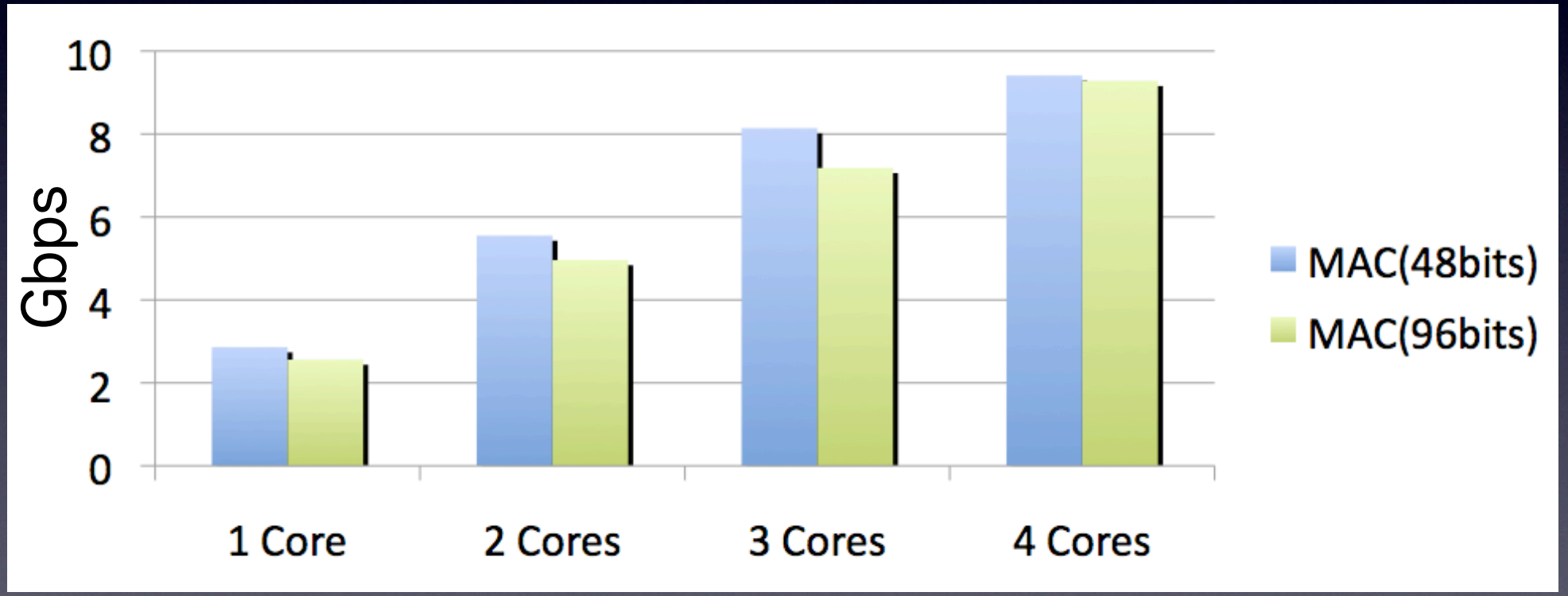
Why not just extend MAC ID space through deep programmability enabled by FLARE Switch?

Extended MAC (96bit) Switching

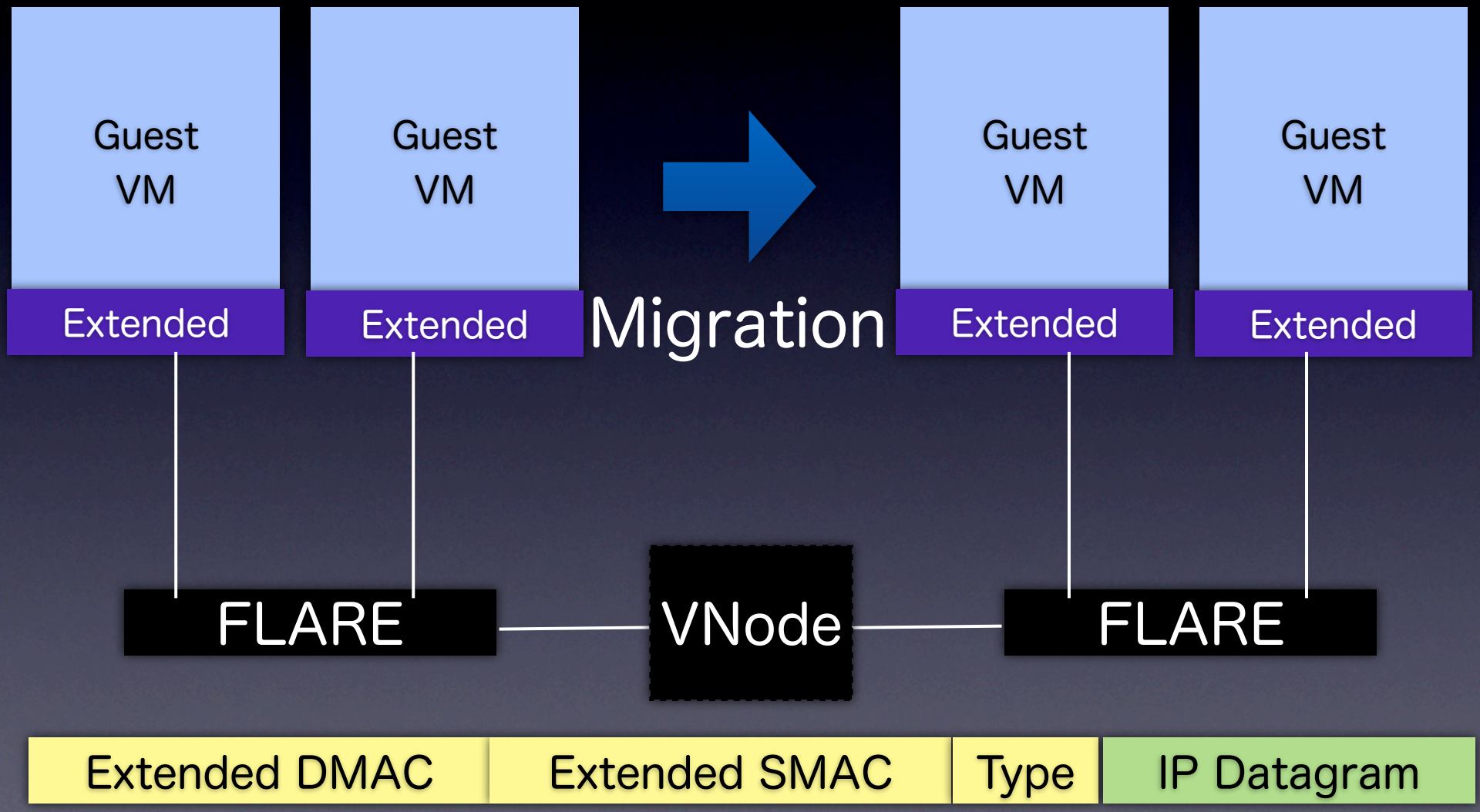


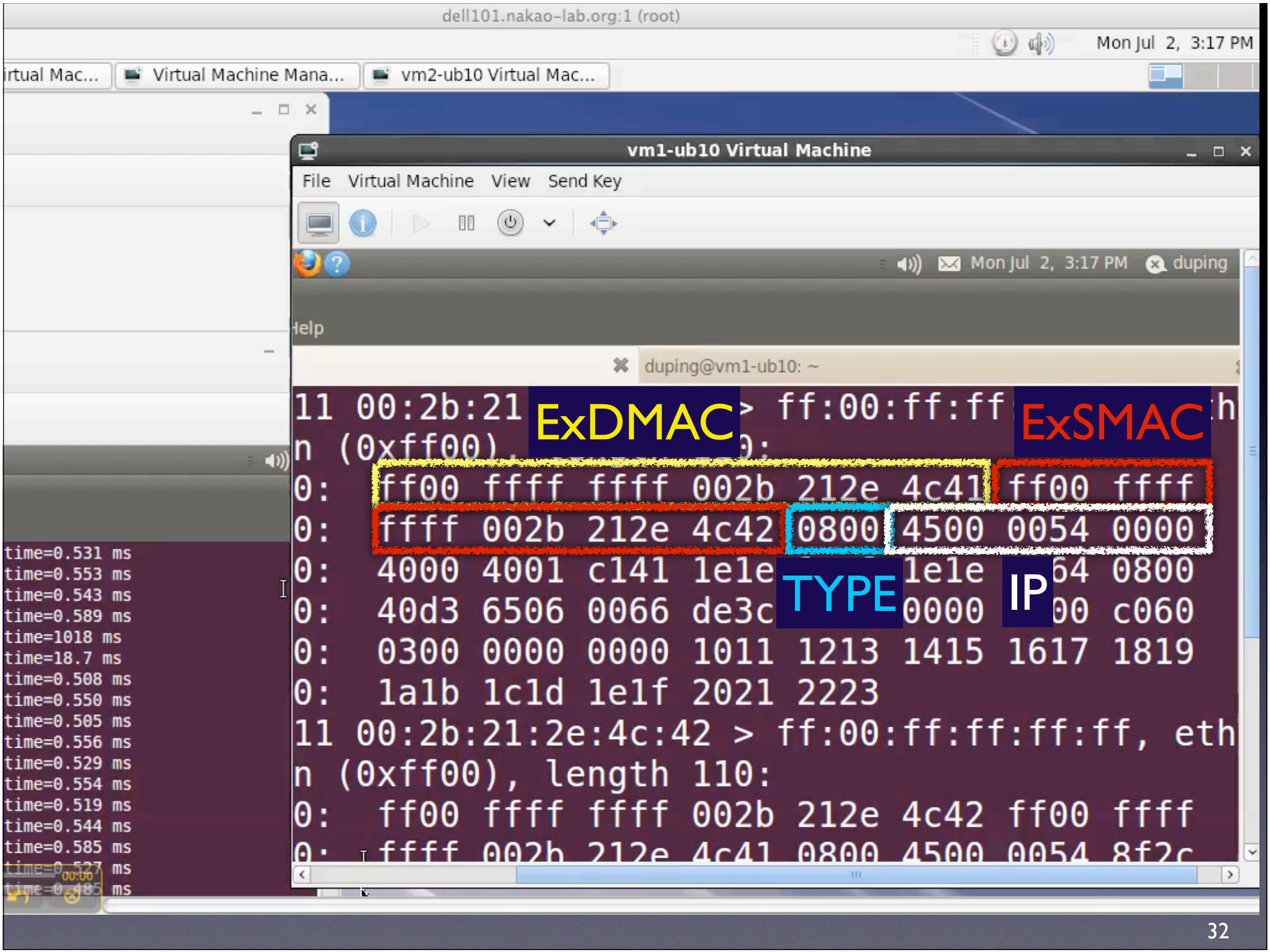
Performance Check

2 Port 10Gbps Switching Performance (Gbps)
MTU=1500 bytes

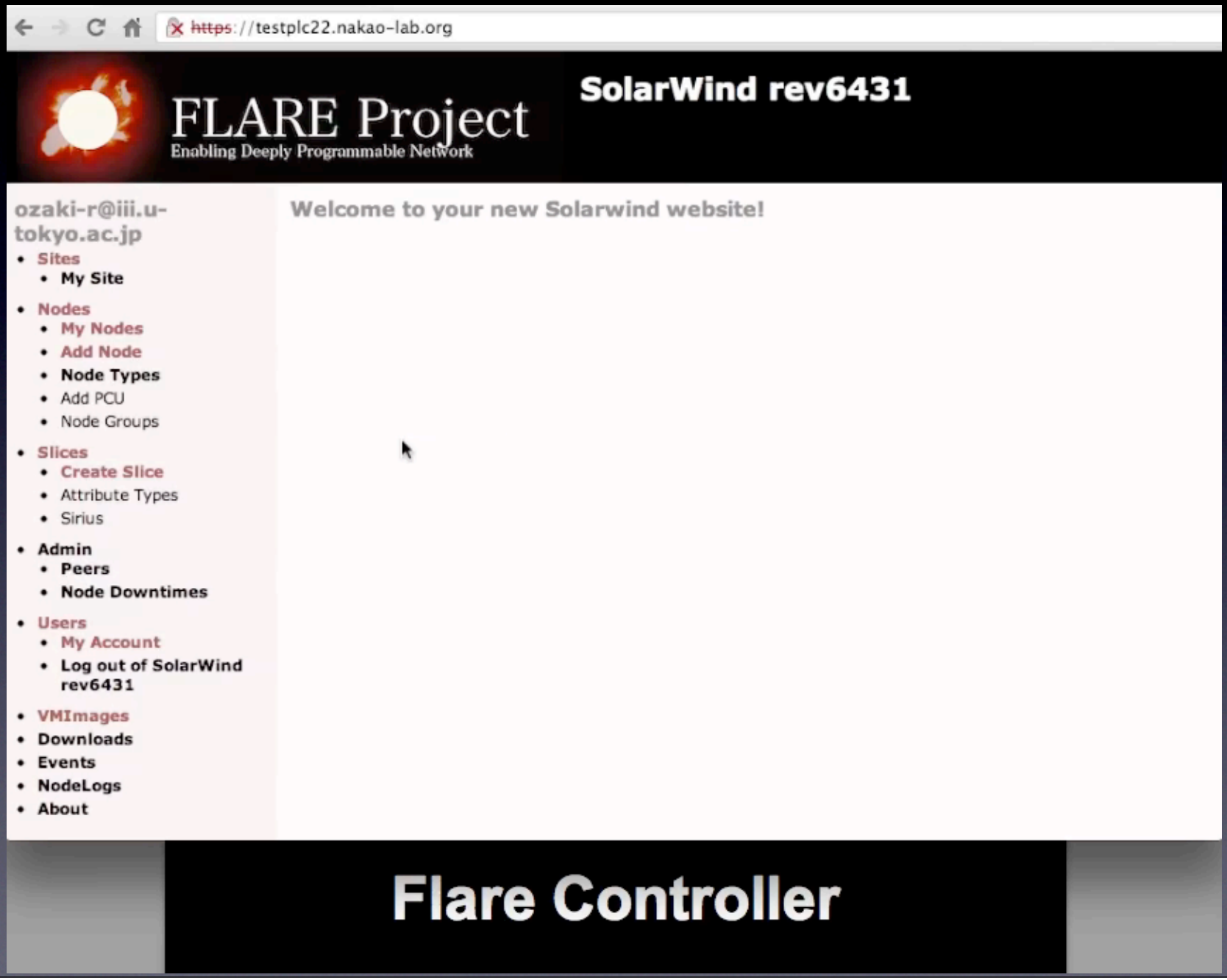


VM Migration With Extended MAC

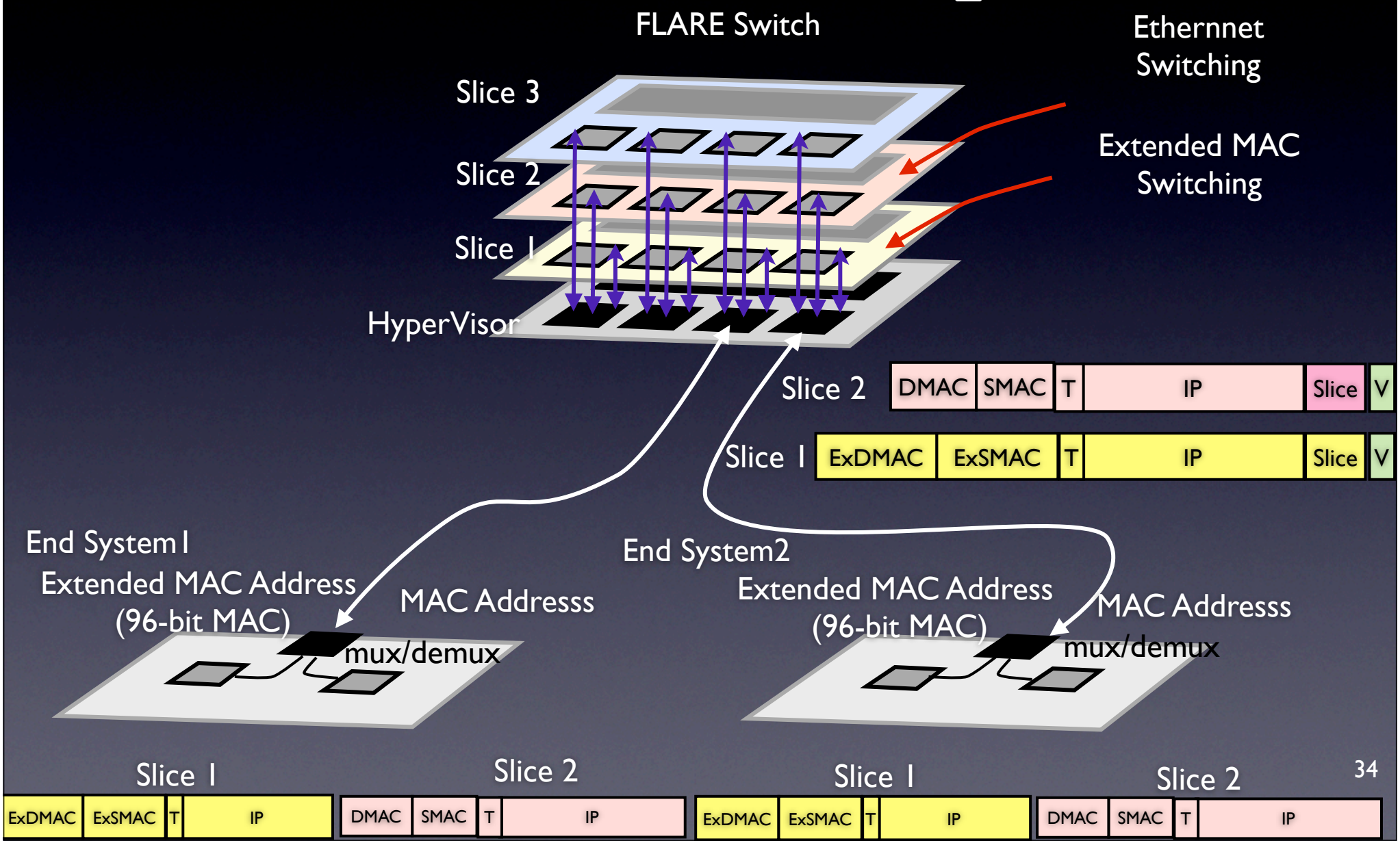



dell101.nakao-lab.org:1 (root)
Mon Jul 2, 3:17 PM
Virtual Machine Mana...
vm2-ub10 Virtual Mac...
vm1-ub10 Virtual Machine
File Virtual Machine View Send Key
Mon Jul 2, 3:17 PM duping
duping@vm1-ub10: ~
ExDMAC
ExSMAC
TYPE
IP
0: ff00 ffff ffff 002b 212e 4c41 ff00 ffff
0: ffff 002b 212e 4c42 0800 4500 0054 0000
0: 1a1b 1c1d 1e1f 2021 2223
11 00:2b:21:2e:4c:42 > ff:00:ff:ff:ff:ff, eth
n (0xff00), length 110:
0: ff00 ffff ffff 002b 212e 4c42 ff00 ffff
time=0.531 ms
time=0.553 ms
time=0.543 ms
time=0.589 ms
time=1018 ms
time=18.7 ms
time=0.508 ms
time=0.550 ms
time=0.505 ms
time=0.556 ms
time=0.529 ms
time=0.554 ms
time=0.519 ms
time=0.544 ms
time=0.585 ms

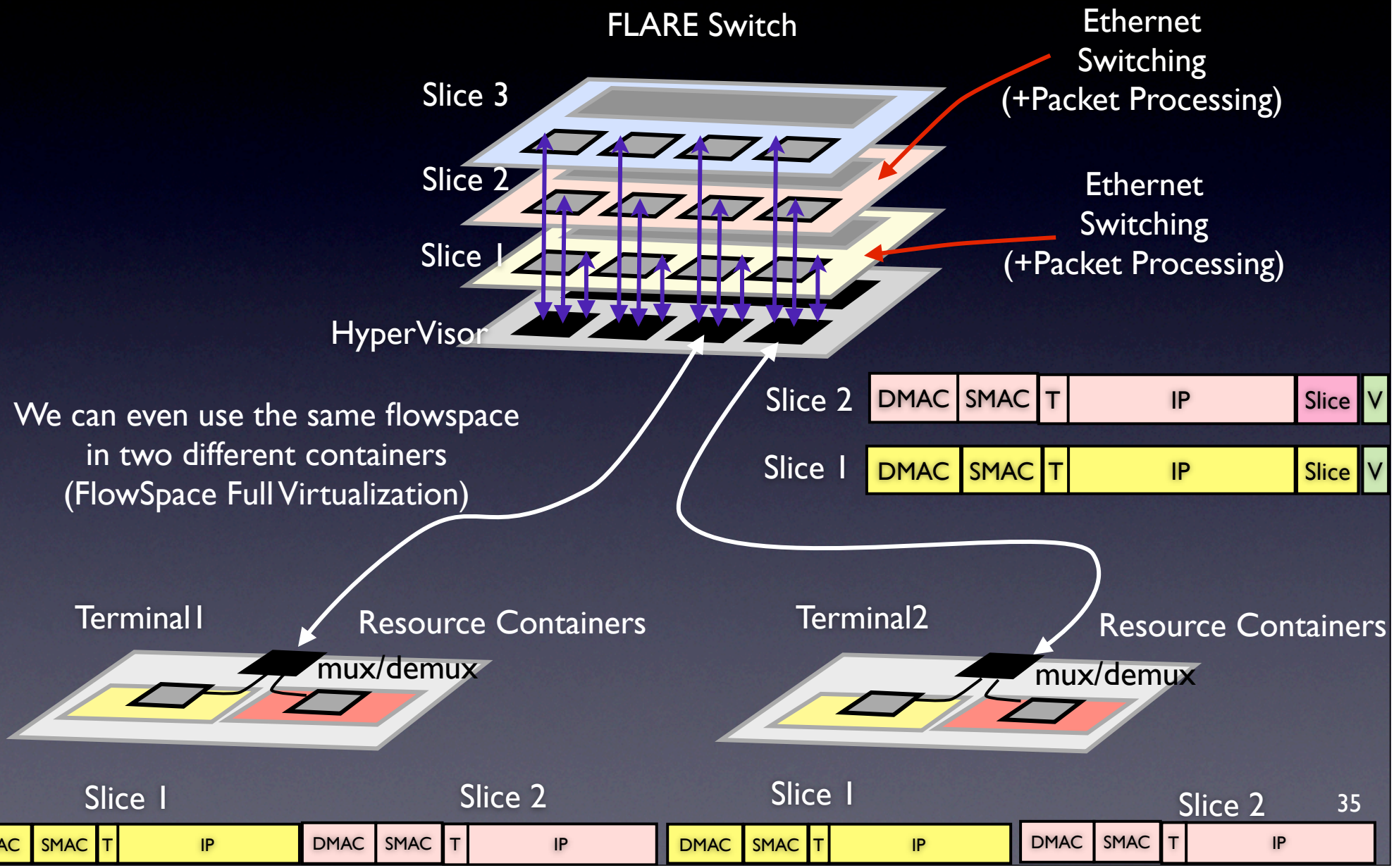
FLARE Controller



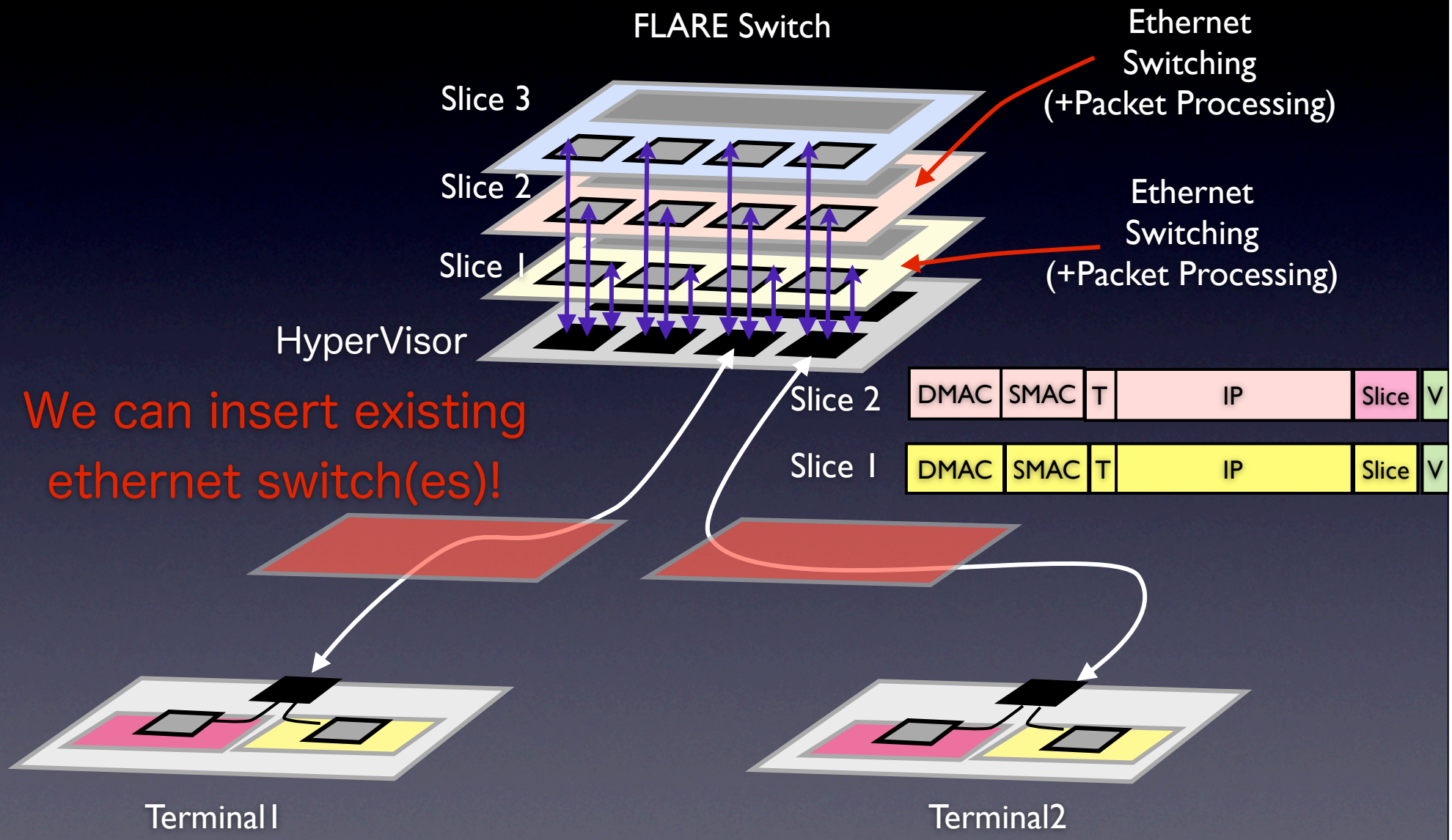
Two Different Kinds of Switching via Tailer Slicing



Network NameSpace Full Virtualization

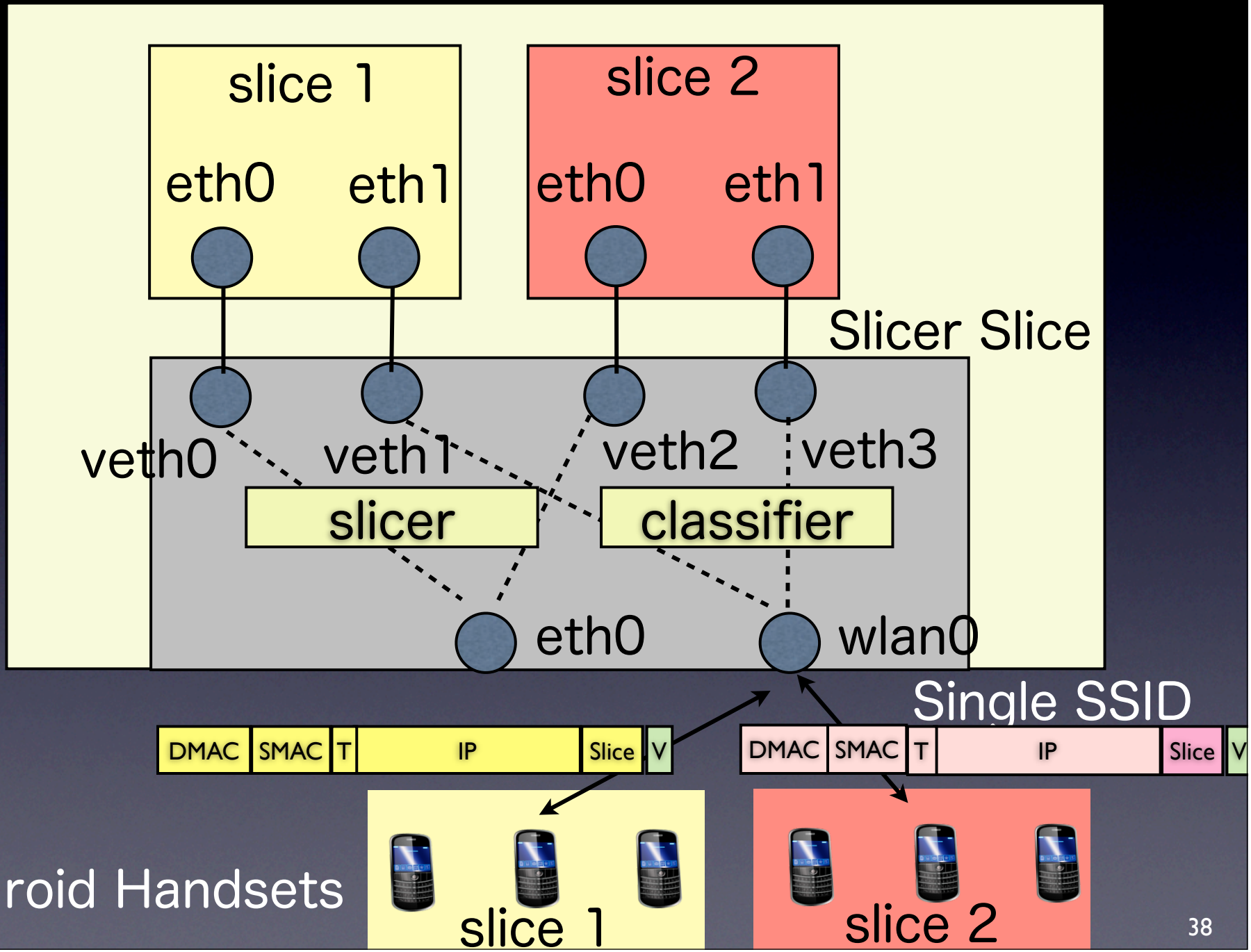


Beauty of Trailer Slicing

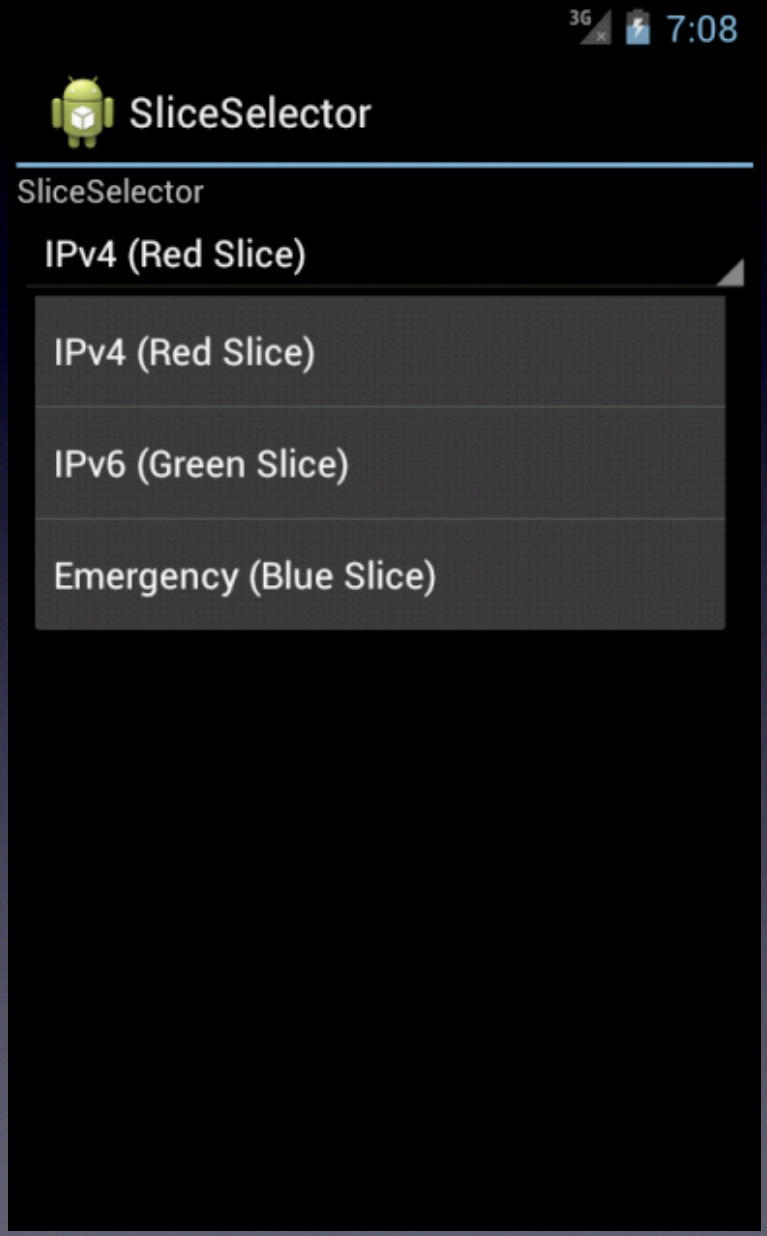


One more thing...

FLARE Wireless LAN AP



HandSet Slicing



- Slice Per Device
- Slice per OS
  (Android OS Virtualization)
- Slice per Application

Slice Selector (IPv4/IPv6)

Conclusion

- We might recall what we initially aimed at, namely, by "Clean Slate" re-design of the Internet
- VNode/FLARE projects now put forth edge area network virtualization and deep programmability

Credits

- FLARE Project Team @ UTokyo (Aki Nakao, Shu Yamamoto, Ryota Ozaki, Ping Du, Eiji Miyagaki, Hamid Farhady)
- VNode Project Team @ UTokyo (UTokyo, NTT, KDDI, Fujitsu, Hitachi, NEC)
- NICT & MIC for Funding the Projects
- GPO, especially Chip Elliot for discussion and giving us opportunity for this presentation