High-performance vSwitch of the user, by the user, for the user

Yoshihiro Nakajima, Wataru Ishida, Tomonori Fujita, Takahashi Hirokazu, Tomoya Hibi, Hitoshi Matsutahi, Katsuhiro Shimano

NTT Labs

Agenda

Innovative R&D by NTT

- **vSwitch of the user**
  - Background and motivation
- **vSwitch by the user**
  - High-performance software-switch
    - Concept
    - Design
    - Implementation & techniques
    - Performance Evaluation
- **vSwitch for the user**
  - PoC of carrier usecase
    - Segment Routing or Service chaining (Tentative)
  - Open source community

NTT

lagopus

Our mission

Innovative R&D by NTT

Open innovation with open source for networking

NTT

vSwitch of the user

- Background
- Motivation

Trend shift in networking

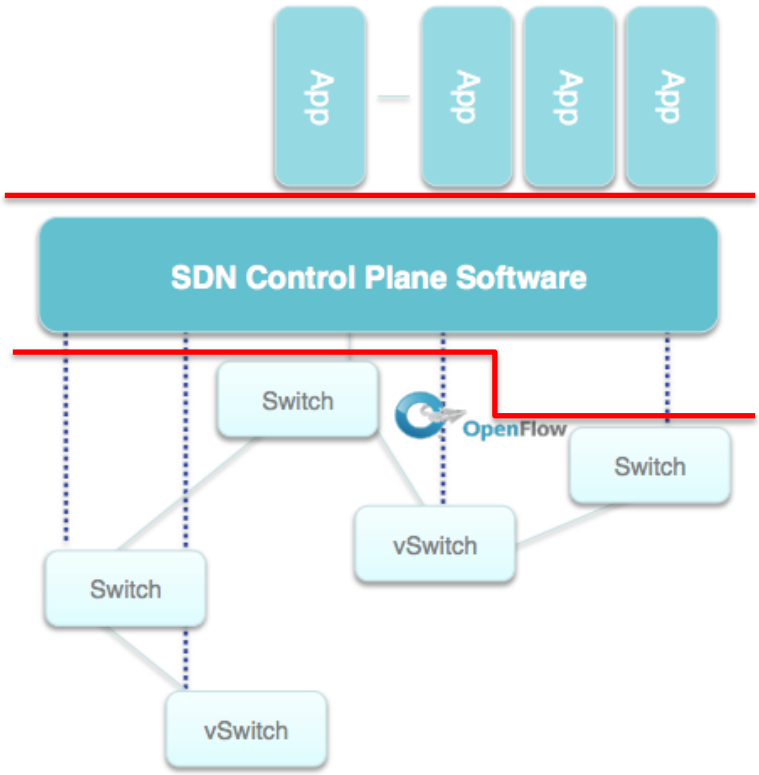
- ✓ Closed (Vender lock-in)
- ✓ Yearly dev cycle
- ✓ Waterfall dev
- ✓ Standardization
- ✓ Protocol
- ✓ Special purpose HW / appliance
- ✓ Distributed cntrl
- ✓ Custom ASIC / FPGA

- ✓ Open (lock-in free)
- ✓ Monthly dev cycle
- ✓ Agile dev
- ✓ DE fact standard
- ✓ API
- ✓ Commodity HW/ Server
- ✓ Logically centralized cntrl
- ✓ Merchant Chip

What is Software-Defined Networking?

Innovative R&D by NTT

SDN conceptual model



Programmability via abstraction layer
→Enables flexible and rapid service/application development

Logically centralized view
→Hide and abstract complexity of networks, provide entire view of the network

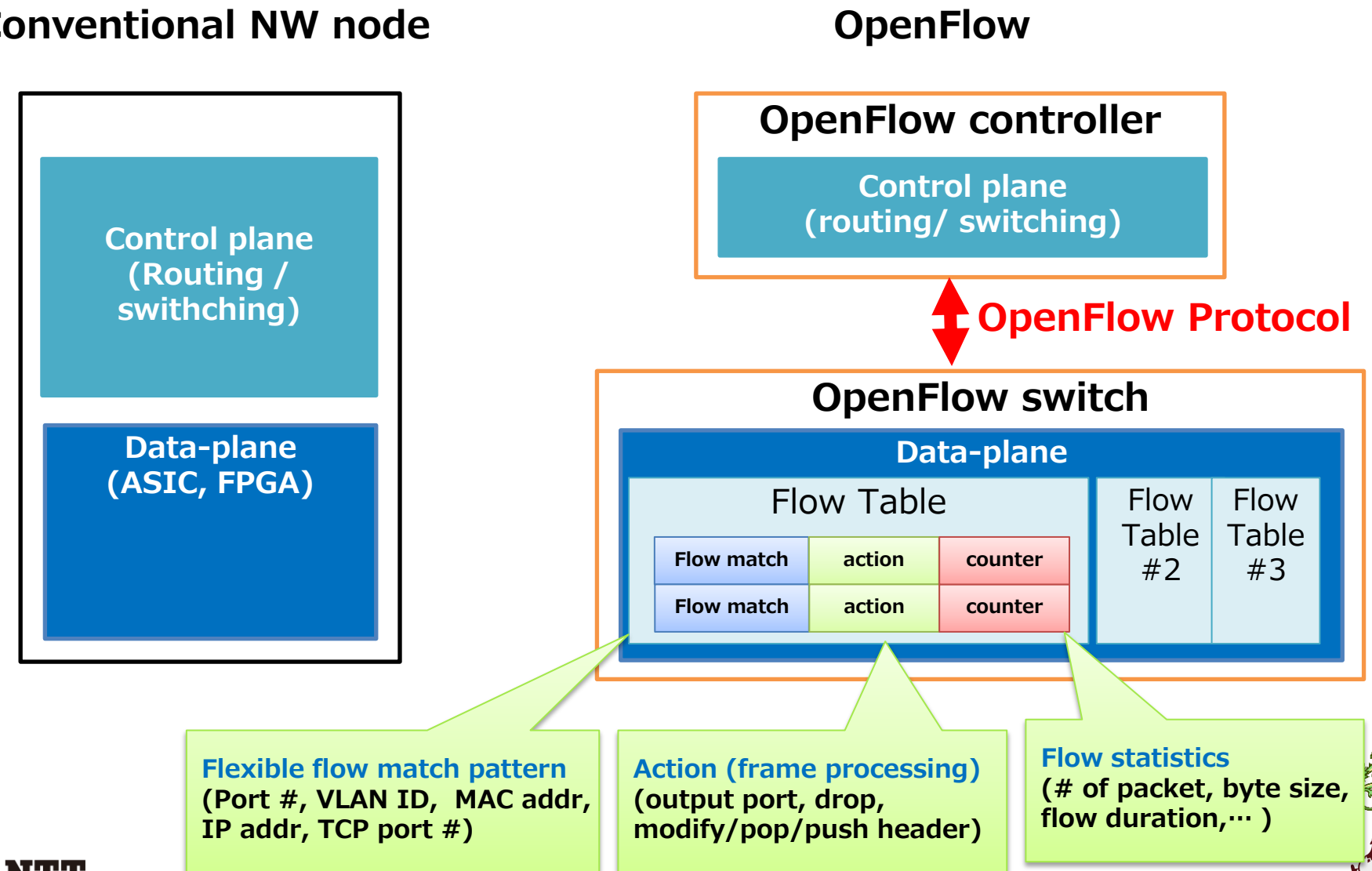
Decouple control plane and data plane
→Free control plane out of the box
(OpenFlow is one of APIs)

Reference: http://opennetsummit.org/talks/ONS2012/pitt-mon-ons.pdf

Innovate services and applications in software development speed.

NTT

OpenFlow vs conventional NW node



Evaluate the benefits of SDN by implementing our control plane and switch

7

Motivation of vSwitch

Innovative R&D by NTT

- **Agile and flexible networking**
  - Full automation in provisioning, operation and management
  - Seamless networking for customers
- **Server virtualization and NFV needs a high-performance software switch**
  - Short latency
  - Wire-rate in case of short packet (64B)
- **No High-performance OpenFlow 1.3 software switch for wide-area networks**
  - 1M flow rules
  - 10Gbps-wire-rate
  - Multiple table supports

NTT

vSwitch requirement from user side

Innovative R&D by NTT

1. **Run on the commodity PC server and NIC**
2. **Provide a gateway function to allow connect different various network domains**
   - Support of packet frame type in DC, IP-VPN, MPLS and access NW
3. **Achieve 10Gbps-wire rate with >= 1M flow rules**
   - low-latency packet processing
   - flexible flow lookup using multiple-tables
   - High performance flow rule setup/delete
4. **Run in userland and decrease tight-dependency to OS kernel**
   - easy software upgrade and deployment
5. **Support various management and configuration protocols.**

NTT

Why Intel x86 server for networking?

Innovative R&D by NTT

- **Strong processing power**
  - Many CPU cores
    - 8 cores/CPU (2012), 12 cores/CPU (2013),…
  - Rapid development cycle
- **Available everywhere**
  - Can be purchased in Akihabara!
  - 3 month lead time in case of special purpose HW
- **Reasonable price and flexible configuration**
  - 8 core CPU (ATOM) with 1,000 USD
  - Can chose configuration according to your budget
- **Many programmers and engineers**
  - Porting to a NPU needs special skill and knowledge

NTT

vSwitch by the user

- Concept
- Design
- Implementation & techniques
- Performance Evaluation

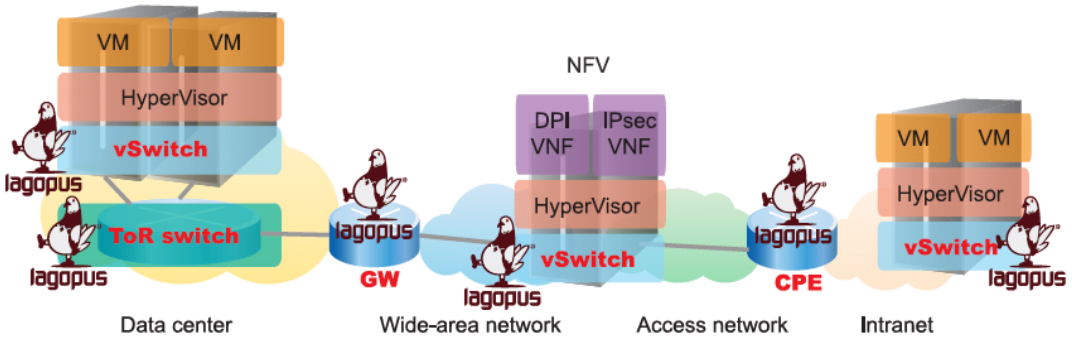
NTT

vSwitch by the user Concept

Target of Lagopus vSwitch

Innovative R&D by NTT

- **High performance soft-based OpenFlow switch**
  - 10Gbps wire-rate packet processing / port
  - 1M flow rules
- **Expands the SDN apps to wide-area networks**
  - Not only for data centers
  - WAN protocols, e.g. MPLS and PBB
  - Various management /configuration interfaces
- **Open Innovation**
  - Community-based development



NTT

Approach for vSwitch development

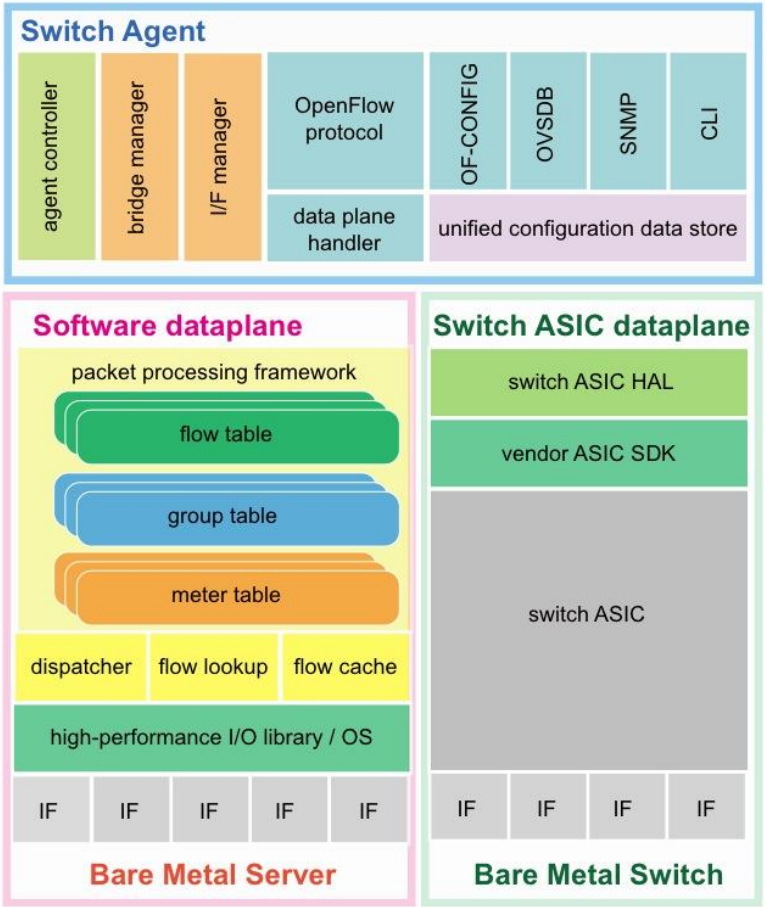
- **Simple is better for everything**
  - Packet processing
  - Protocol handling
- **Straight forward approach**
  - No OpenFlow 1.0 support
  - Full scratch (No use of existing vSwitch code)
- **User land packet processing as much as possible, keep kernel code small**
  - Kernel module update is hard for operation
- **Every component & algorithm can be replaced**
  - Flow lookup, library, protocol handling

NTT

vSwitch design

Innovative R&D by NTT

- **Simple modular-based design**
  - Switch control agent
  - Data-plane
- **Switch control agent**
  - Unified configuration data store
  - Multiple management IF
  - Data-plane control with HAL
- **Data-plane**
  - High performance I/O library (Intel DPDK)
  - Raw socket for test



NTT

Implementation strategy for vSwitch

- Massive RX interrupts handling for NIC device

=> Polling-based packet receiving

- Heavy overhead of task switch

=> Thread assignment
(one thread/one physical CPU)

- Lower performance of PCI-Express I/O and memory bandwidth compared with CPU

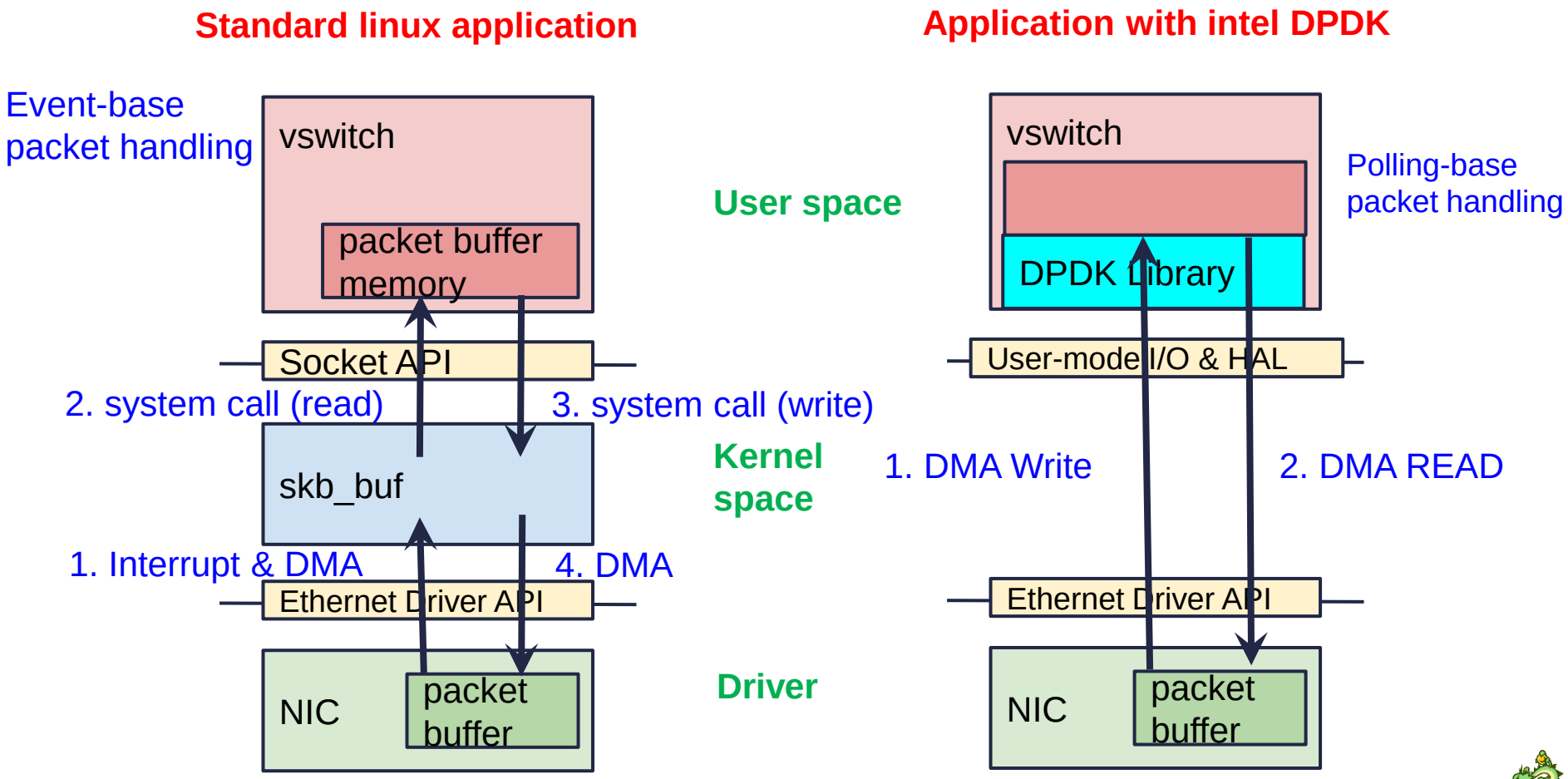
=> Reduction of # of access in I/O and memory

- Shared data access is bottleneck between threads

=> Lockless-queue, RCU, batch processing

NTT

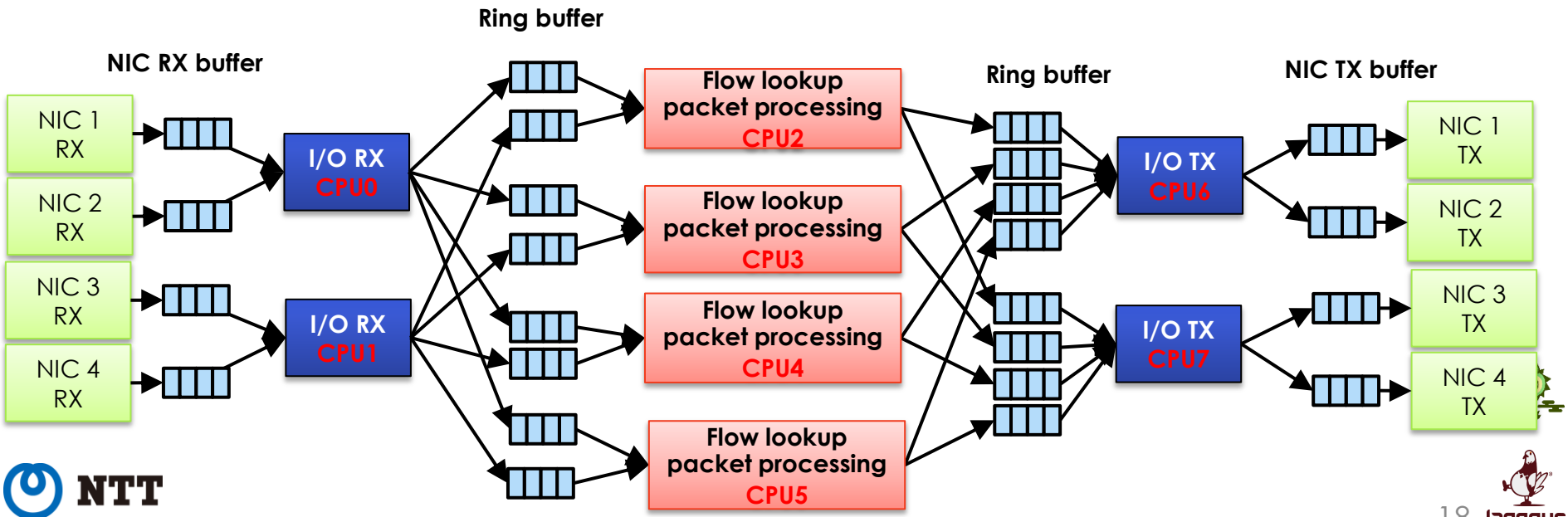
Processing bypass for speed



Packet processing using multi core CPUs

Innovative R&D by NTT

- **Exploit many core CPUs**
  - Reduce data copy & move (reference access)
  - Simple packet classifier for parallel processing in I/O RX
- **Decouple I/O processing and flow processing**
  - Improve D-cache efficiency
- **Explicit thread assign to CPU core**

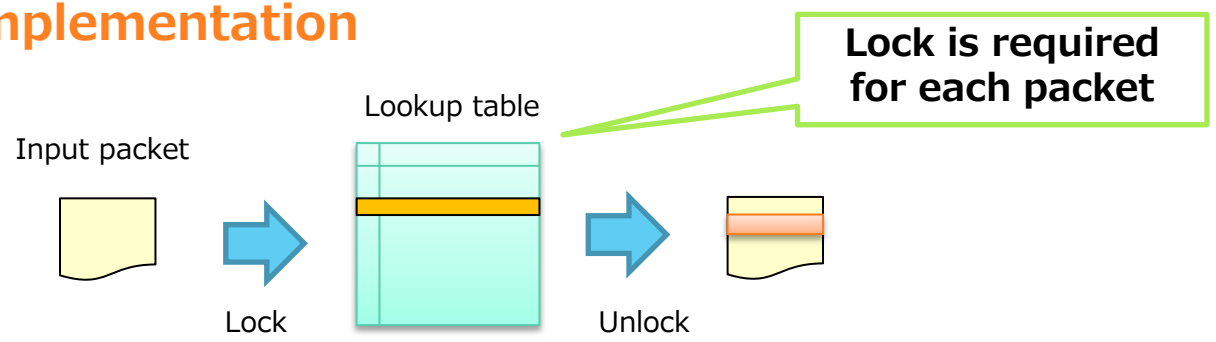


Packet batching

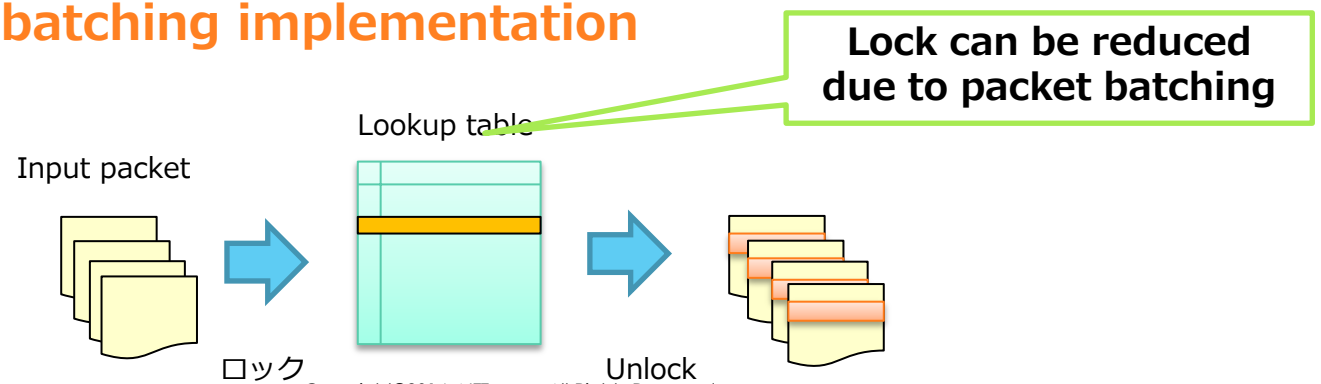
Reduce # of lock in flow lookup table

- Frequent locks are required
  - Switch OpenFlow agent and counter retrieve for SNMP
  - Packet processing

- **Naïve implementation**



- **Packet batching implementation**



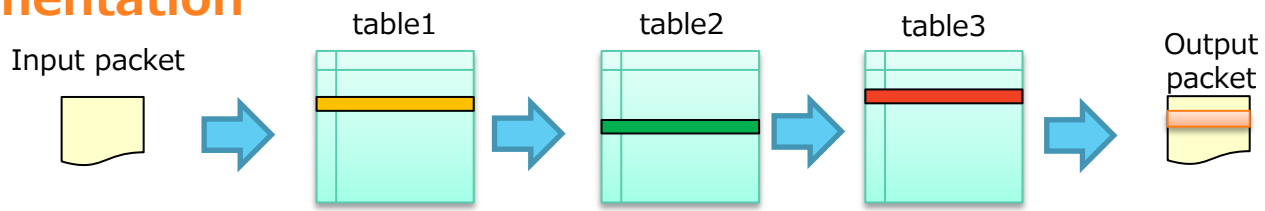
Bypass pipeline with flow $

Innovative R&D by NTT

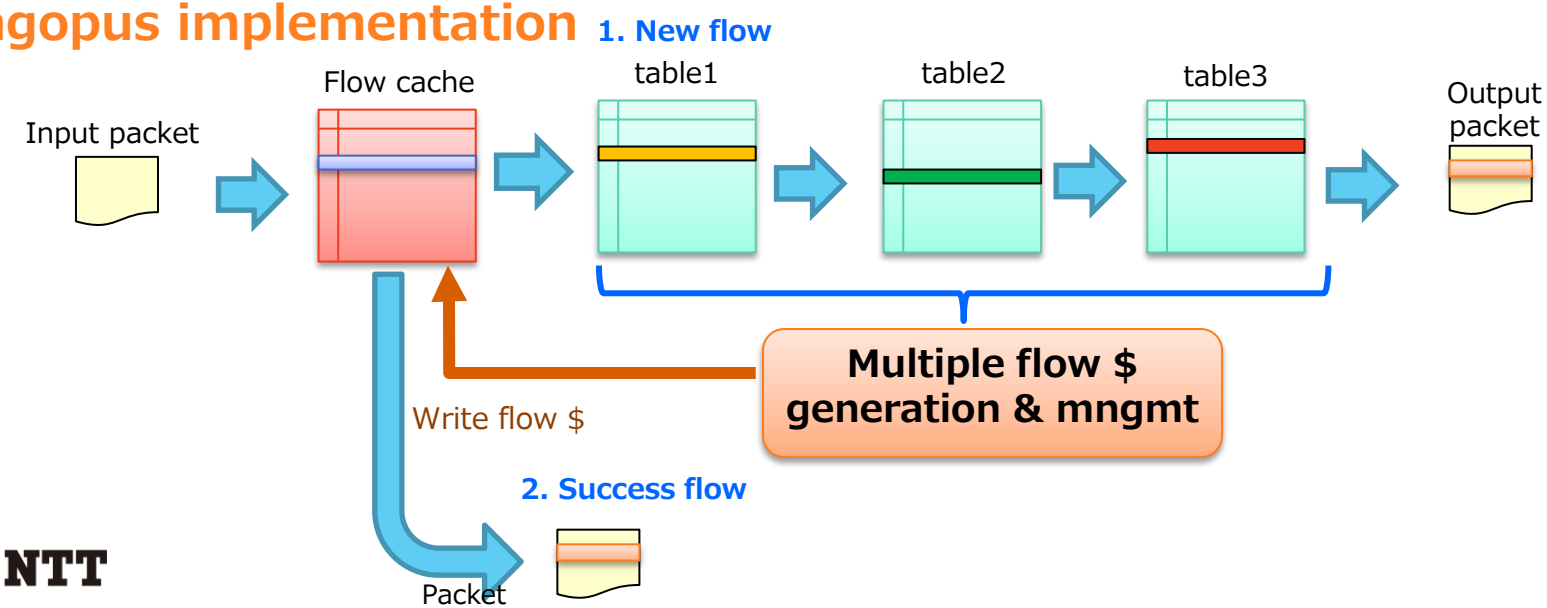
Reduce # of flow lookup in multiple table

- Exploit flow $ for

Naïve implementation



Lagopus implementation



NTT

lagopus

vSwitch by the user Performance Evaluation

NTT

Performance Evaluation

Innovative R&D by NTT

Summary

- Throughput: **10Gbps wire-rate**
- Flow rules: **1M flow rules**

Evaluation models

- WAN-DC gateway
  - MPLS-VLAN mapping
- L2 switch
  - Mac address switching

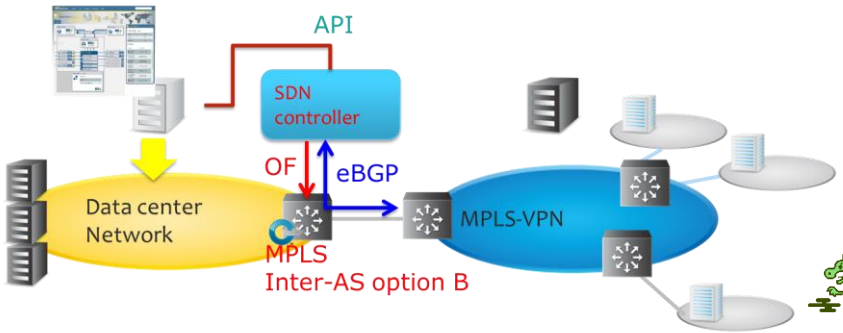
Evaluation scenario

Usecase : Cloud-VPN gateway

From ONS2014 NTT COM Ito-san presentation

Tomorrow :

- Automatic connection setup via North Bound APIs
- SDN controller maintain mapping between tenant logical network and VPN
- Routes are advertised via eBGP, no need to configure ASBRs on provider side

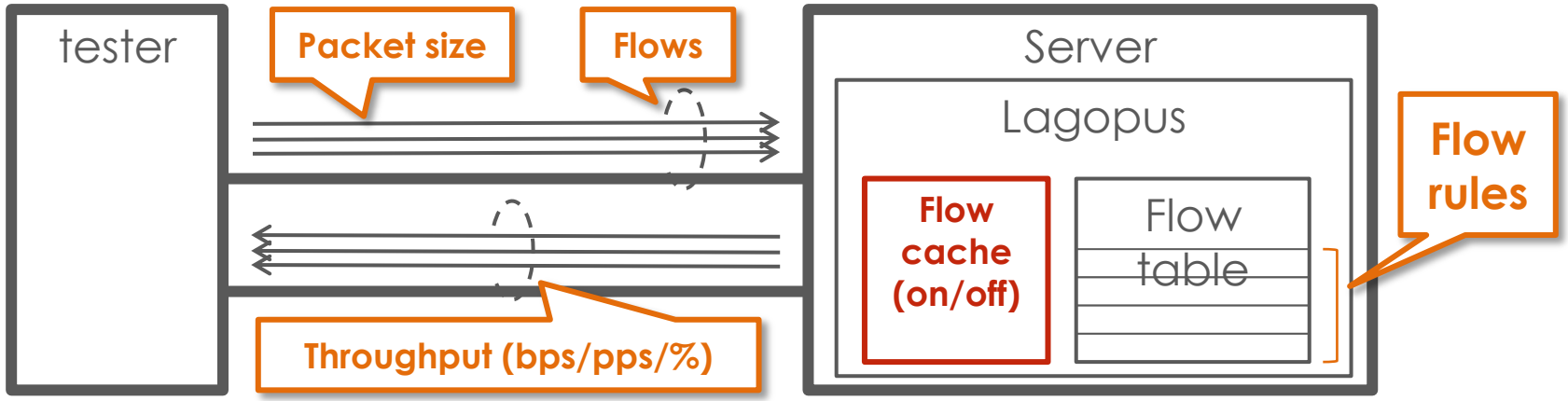


NTT

Performance Evaluation

Innovative R&D by NTT

Evaluation setup



Server spec.

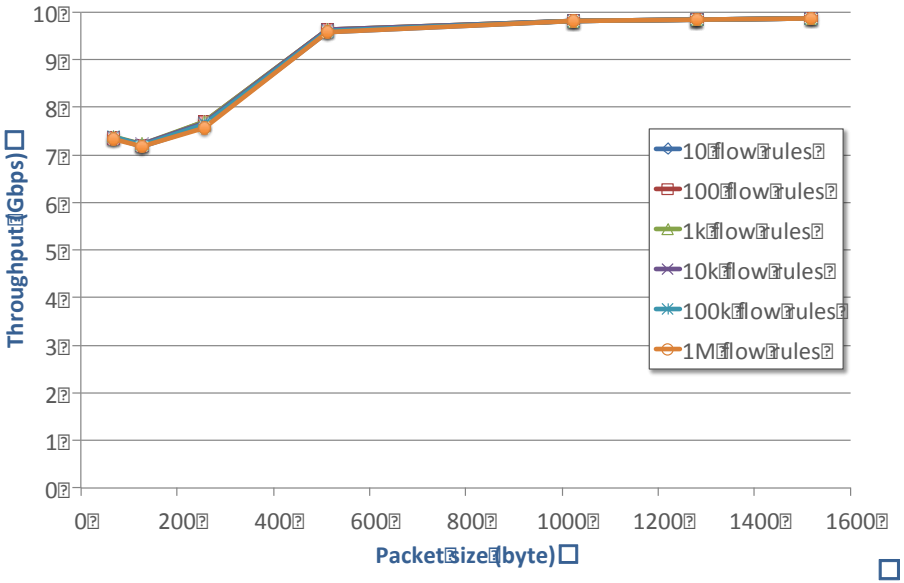
- CPU: Dual Intel Xeon E5-2660
  - 8 core(16 thread), 20M Cache, 2.2 GHz, 8.00GT/s QPI, Sandy bridge
- Memory: DDR3-1600 ECC 64GB
  - Quad-channel 8x8GB
- Chipset: Intel C602
- NIC: Intel Ethernet Converged Network Adapter X520-DA2
  - Intel 82599ES, PCIe v2.0

NTT

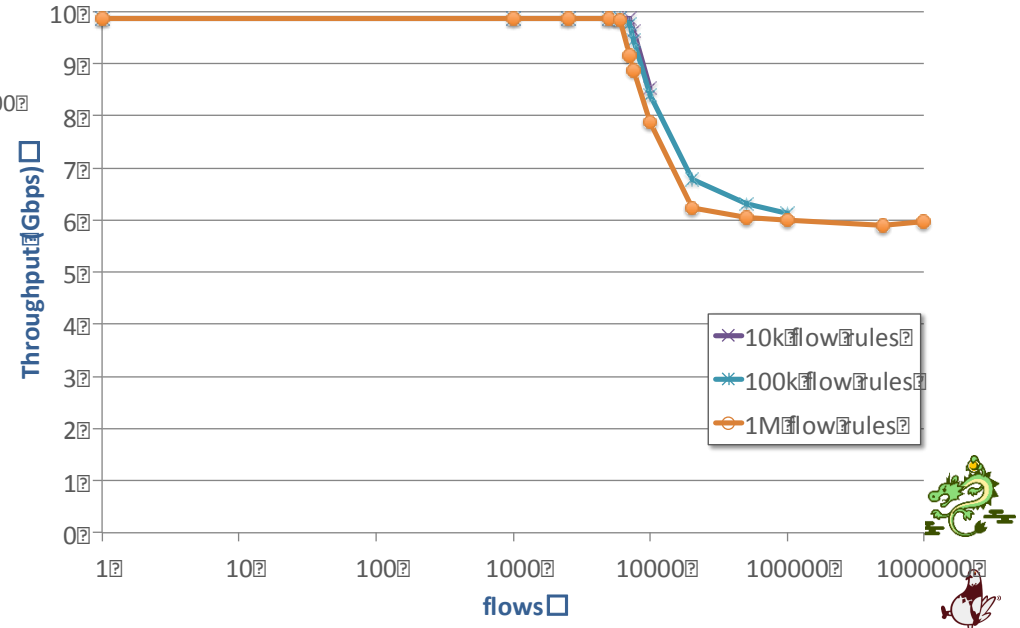
WAN-DC Gateway

Innovative R&D by NTT

Throughput vs packet size, 1 flow, flow-cache

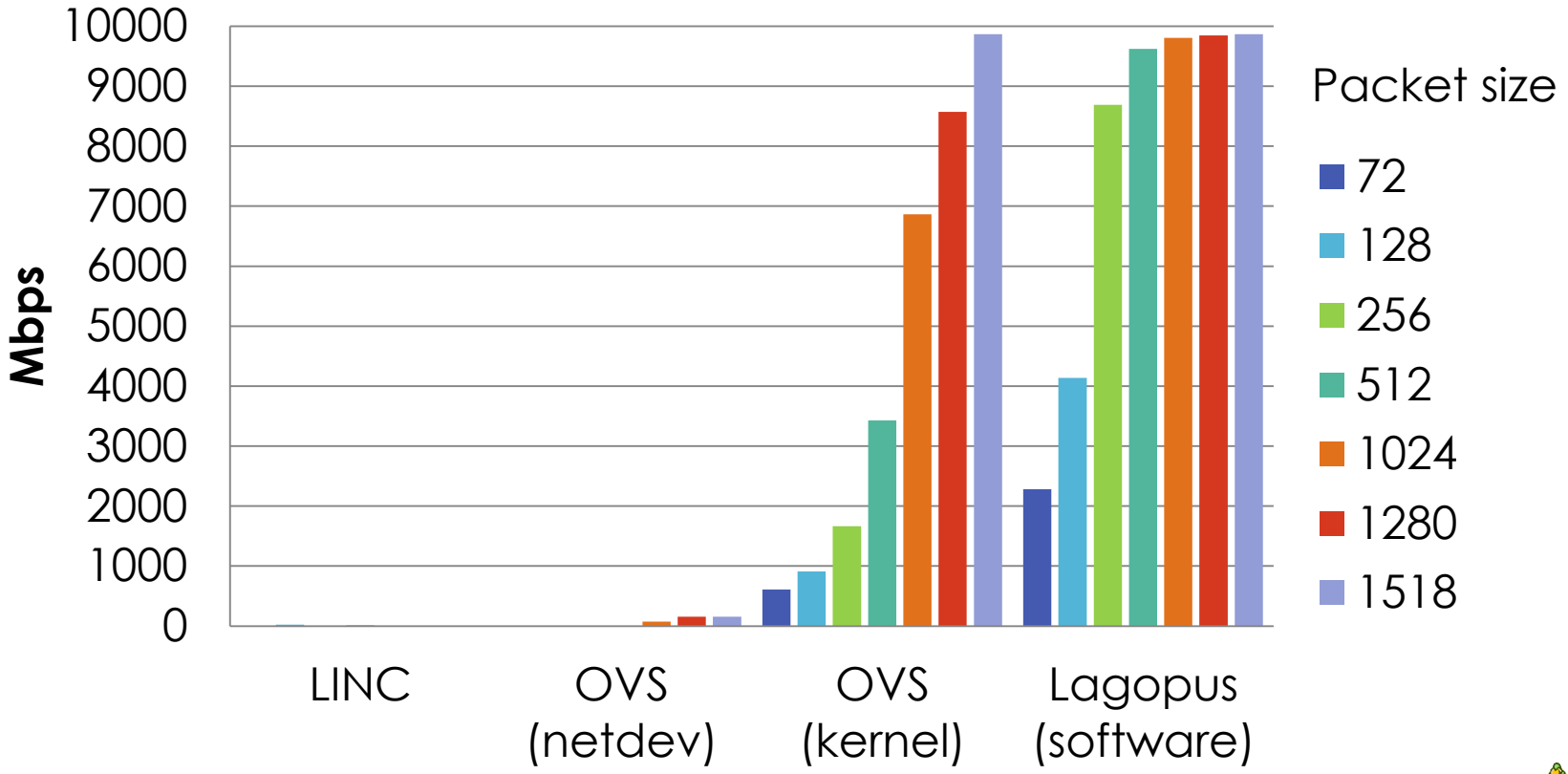


Throughput vs flows, 1518 bytes packet



L2 switch performance (Mbps)
10GbE x 2 (RFC2889 test)

Innovative R&D by NTT



Intel xeon E5-2660 (8cores, 16threads),
Intel X520-DA2, DDR3-1600 64GB

NTT

vSwitch for the user

PoC of carrier usecase

Open source community

NTT

SPRING

Innovative R&D by NTT

alternative way of IP forwarding

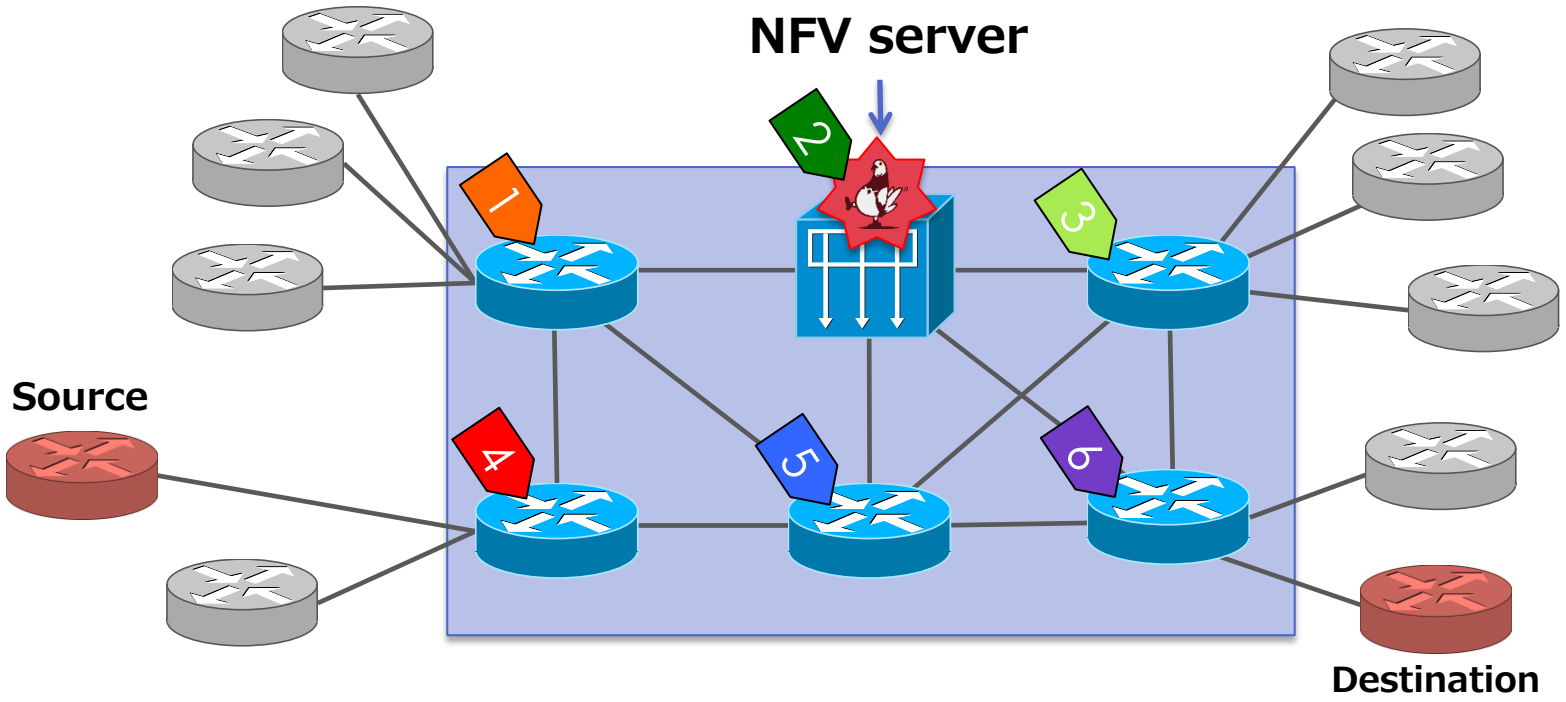
- use “labels” to forward packets
- much simpler architecture than MPLS network
- can utilize traditional MPLS routers
- good chemistry with NFV

For what ?

- policy-based routing
  - explicit routing instead of shortest path routing
- service-chaining (NFV)
  - apply services to specific traffic pattern

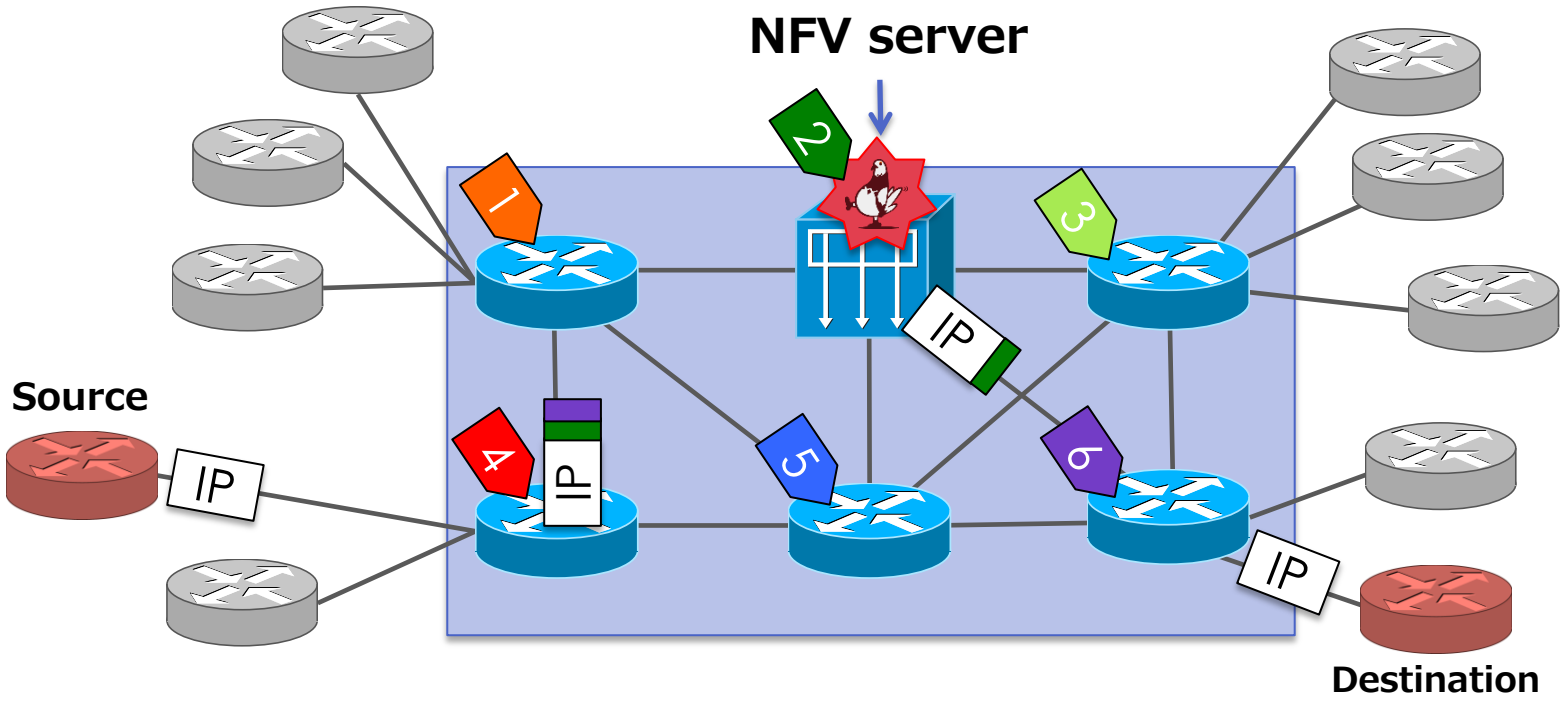
NTT

Career Network with SPRING



- use lagopus as NFV server with career grade routers
- all nodes include NFV server are labeled

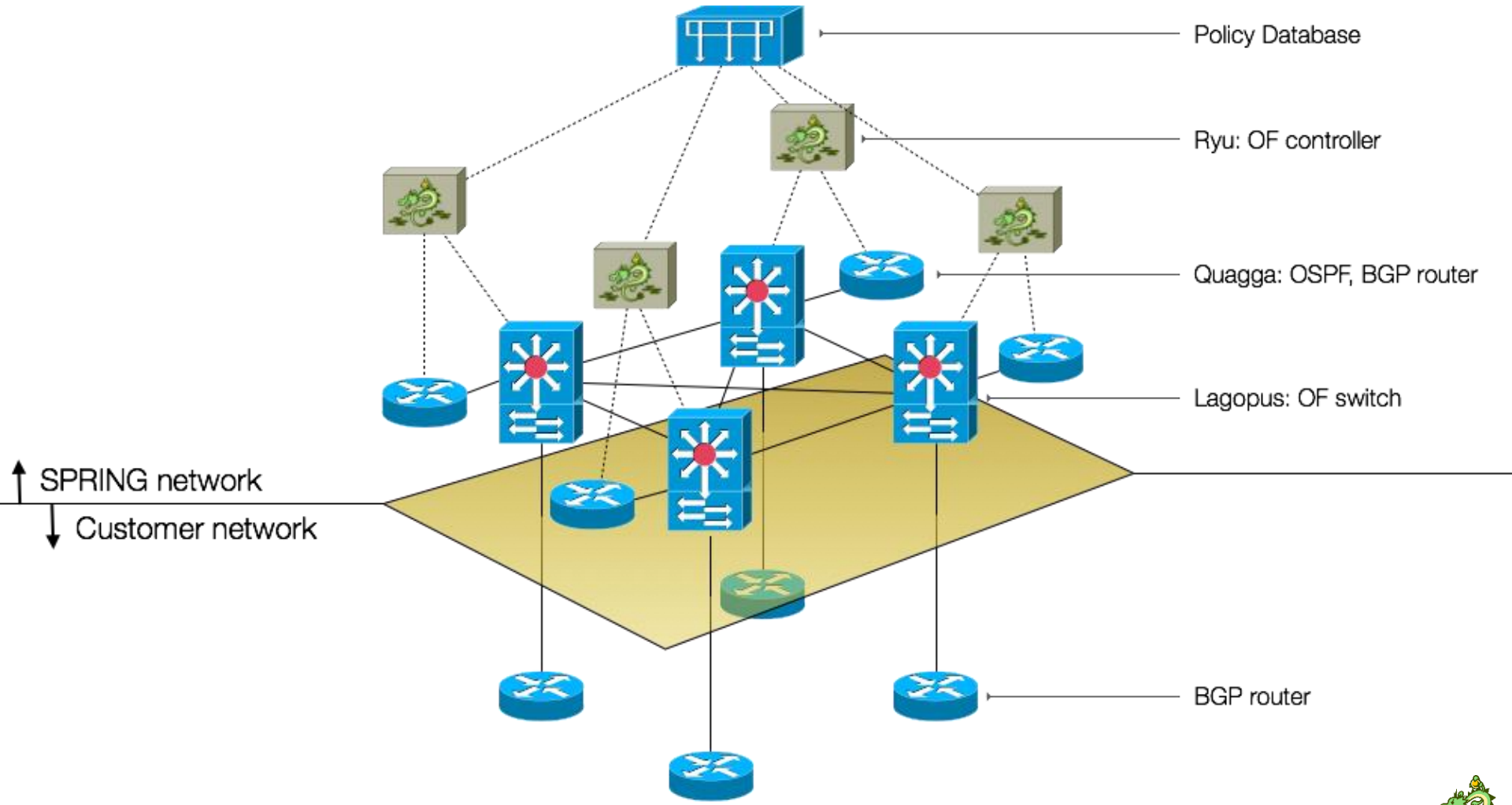
Career Network with SPRING



- at the ingress router, label stack is pushed
- label stack represents the policy

PoC Implementation

Innovative R&D by NTT



NTT

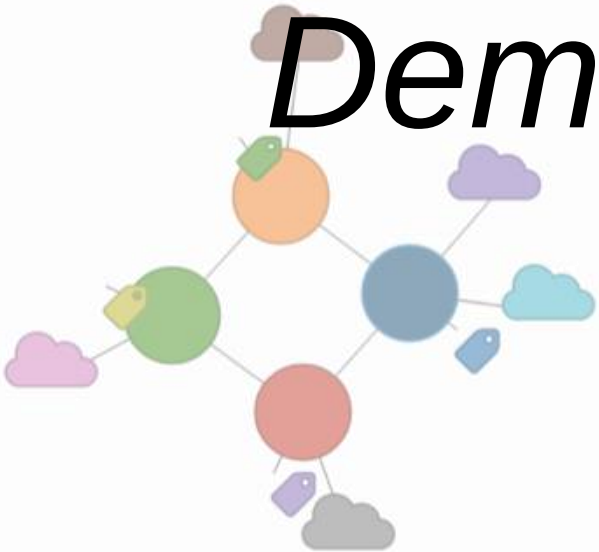
SPRING WEB-UI

localhost:4000

SPRING WEB-UI

Welcome to SPRING WEB-UI

Demo



Router | Network | Label

ID: 192.168.0.2

ID: 192.168.0.3

ID: 192.168.0.4

ID: 192.168.0.5

Router 192.168.0.2

Associated Labels

ID: 1

Label input | add label

Associated Networks

ID: 20.0.0.0/24

ID: 50.0.0.0/24

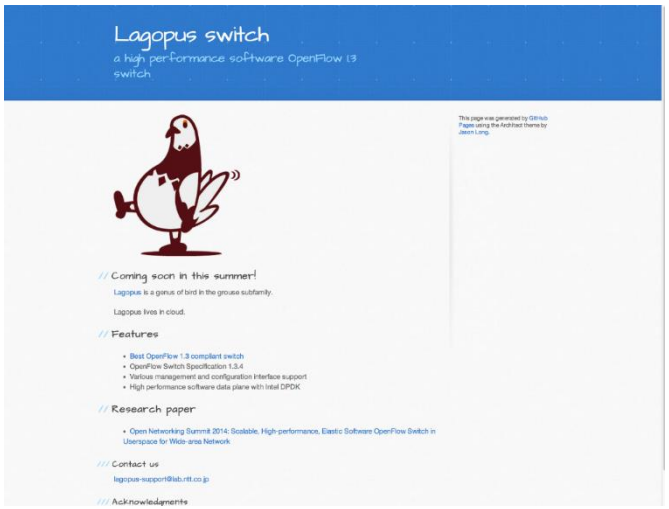
Associated Services

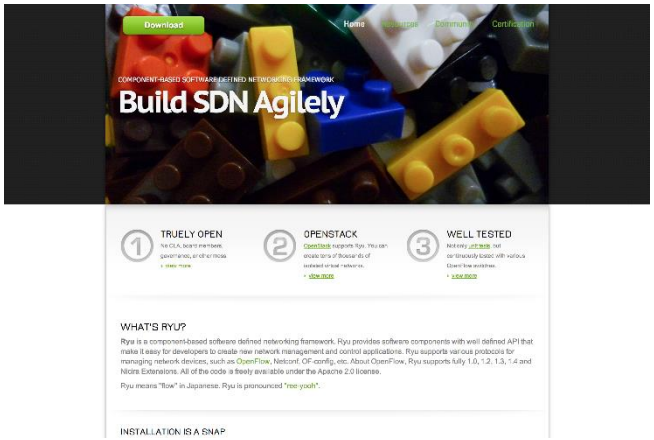
boot service

Open Source Activity for the user

Innovative R&D by NTT

- **Lagopus goes wild since this summer**
  - Your contribution are very welcome!
  - http://lagopus.github.io/
  - Twitter: lagopusvswitch
- **Ryu: component-based SDN framework**
  - OpenFlow 1.3, 1.4, BGP, BMP, etc..
  - http://osrg.github.io/ryu/





NTT

Summary

Innovative R&D by NTT

We contribute vSwitch for user

- Improve performance and scalability
- Enhance the functionality and capability

DPDK is great library for user

- Easy to develop & implement
- Easy to deploy like UNIX apps

NTT

Thank you!

This research is a part of the project for "Research and Development of Network Virtualization Technology" supported by the Ministry of Internal Affairs and Communications.

NTT