ADVANCED NETWORK SCHEDULING AND ISOLATION IN KUBERNETES

KURALAMUDHAN RAMAKRISHNAN
KURALAMUDHAN.RAMAKRISHNAN@INTEL.COM

SENIOR SOFTWARE ENGINEER
DATA CENTER NETWORK SOLUTION GROUP
INTEL
LEGAL NOTICES AND DISCLAIMERS

• Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

• No computer system can be absolutely secure.

• Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

• Intel, the Intel logo, Xeon, and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.

• © 2017 Intel Corporation.
WHAT WILL YOU LEARN TODAY?

• Containers in the NFV ecosystem
• Introduction to Kubernetes and CNI
• Kubernetes Networking and Scheduling Challenges
• Advanced Networking Solution for Kubernetes
• Summary and Next Steps
• High Performance Networking Demo
VIRTUALIZATION OF NETWORK FOR SDN/NFV – MULTIPLE DEPLOYMENT MODELS
CONTAINERS IN NFV ECOSYSTEM

VNF Deployment model

VM

Containers

VM

Coexistence and Unified Orchestration

VNFs

vEPC

vNAT

vIMS

vRouter

vGGSN

vFirewall

vCPE

vRNC

vHLR

vSGSN

vMME

vIDS

NFV Orchestration

ONAP

MANO

kubernetes

openstack

NFVi- Network

OvS

DPDK

SR-IOV
**Kubernetes Introduction**

Open-source Platform for containerised applications
- Automates Deployments
- Manages application lifecycle and Scaling

Originated from Google and contributed to Cloud Native Computing Foundation (CNCF)

**Building Blocks of Kubernetes**

- Pod Spec
- K8s Master
  - etcd
  - API-SERVER
  - SCHEDULER
  - CONTROLLER MANAGER

- NODE 1
  - KUBELET
  - KUBE-PROXY
- NODE 2
  - POD
KUBERNETES NETWORKING VIA CNI

- Container Orchestration Engines
- Container Runtime
- Container Network Interface
- Container Network Interface Plugins

Containers
Node - Environment
Management
KUBERNETES NETWORKING VIA CNI

Management

Containers Node - Environment

Open Source Projects

Container Network Interface

- docker
- rkt
- hyper_

Frakti

- MULTUS
- Calico
- canal
- flannel
- ROMANA
- weave
- SR-IOV
- Contiv
- MACVLAN
- DHCP
- cilium
- IPVLAN
INDUSTRY CHALLENGES

- Multiple network interfaces for VNFs
- Support for resource isolation
- Support for Data Plane Networking
- Ability to request/allocate platform capabilities
- Support for CPU Core pinning for Kuryr-K8s pods
MULTIPLE NETWORK INTERFACES

PROBLEM
Lack of multi-network support in k8s
No network traffic separation for management, control and data planes.
No ability to implement different network SLAs

SOLUTION
Introducing MULTUS as a CNI plugin to support multi-homed pods in k8s
Working on Multiple Network proposal in Kubernetes Network SIG

REFERENCE
https://github.com/Intel-Corp/multus-cni
Multus CNI referenced in the following:
Containers in NFV, March 2017 Peter Willis, BT
Enter Multus CNI, Feb. 2017 Doug Smith, RedHat
A Hacker’s Guide to Kubernetes Networking, Feb 2017, Yaron Haviv, Iguazio
**SRIOV CNI PLUGIN**

**PROBLEM**
Lack of support for physical platform resource isolation
No guaranteed network IO performance
No support for Data Plane Networking

**SOLUTION**
Allows SRIOV support in Kubernetes via a CNI plugin
Intel contributor and maintainer of SR-IOV CNI plugin
Supports two modes of operation:
- **SR-IOV**: SR-IOV VFs are allocated to pod network namespace
- **DPDK**: SR-IOV VFs are bounded to DPDK drivers in the userspace

**REFERENCE**
https://github.com/Intel-Corp/sriov-cni
MULTI HOMED POD WITH MULTUS AND DPDK-SRIOV CNI PLUGIN
Multus CNI with DPDK-SRIOV CNI

```json
{
    "name": "dpdk-demo-network",
    "type": "multus",
    "delegates": [
        {
            "type": "sriov",
            "if0": "net2",
            "if0name": "south0",
            "ipam": {
                "type": "host-local",
                "subnet": "10.56.217.0/24",
            }
        },
        {
            "type": "sriov",
            "if0": "net3",
            "if0name": "north0",
            "dpdk": {
                "kernel_driver": "ixgbevf",
                "dpdk_driver": "igb_uio",
                "dpdk_tool": ".../dpdk/tools/dpdk-devbind.py"
            }
        },
        {
            "type": "flannel",
            "masterplugin": true,
            "delegate": {
                "isDefaultGateway": true
            }
        }
    ]
}
```

In the server backend

*Source: https://docs.oracle.com/cd/E19076-01/sparc.t2k/819-7988-10/rack_install.html*
PHYSICAL TOPOLOGY FROM MULTUS CNI CONF

K8s Master + Minion

POD #1
DPDK
Testpmd

MULTUS
SR-IOV
flannel

docker
ubuntu

Port 1 Port 2 Port 3 Port 4
PF 1 PF 2

Fortville 4 x 10

Switch - 1 GbE

SW Port SW Port SW Port SW Port

POD #2
DPDK
Testpmd

MULTUS
SR-IOV
flannel

docker
ubuntu

Port 1 Port 2 Port 3 Port 4
PF 1 PF 2

Fortville 4 x 10

Switch - 10 GbE

SW Port SW Port SW Port SW Port

Internet
NODE FEATURE DISCOVERY
FEATURE LABELS

K8s MASTER

ETCD

API-SERVER
SCHEDULER
CONTROLLER MANAGER

NODE 1
KUBELET
KUBE-PROXY
DOCKER
DISCOVERY POD
MUTUS-CNI
SRIOV-CNI
SRIOV-NIC

NODE 2
KUBELET
KUBE-PROXY
DOCKER
DISCOVERY POD

NODE 3
KUBELET
KUBE-PROXY
DOCKER

MUTUS-CNI
SRIOV-CNI
SRIOV-NIC

DATACENTER NETWORK SOLUTIONS GROUP
```yaml
apiVersion: v1
kind: Pod
metadata:
  name: nginx
spec:
  containers:
  - name: nginx
    image: nginx
  nodeSelector:
    SRIOV-NIC: true
    SRIOV-CNI: true
    MULTUS-CNI: true
```

![Node Selection Diagram](image)
apiVersion: v1
kind: Pod
metadata:
  name: nginx
spec:
  containers:
  - name: nginx
    image: nginx
nodeSelector:
  - name: nginx
    image: nginx
  - nodeSelector:
      SRIOV-NIC: true
      SRIOV-CNI: true
      MULTUS-CNI: true
INTEL NFV EXPERIENCE KITS ADDRESSING NFV USES IN KUBERNETES

EXISTING SOLUTIONS
MULTUS
- Multi-network Support
SR-IOV CNI PLUGIN
- Resource Isolation
DPDK CNI PLUGIN
- Data Plane Support
Node Feature Discovery
- Data Center Heterogeneity

NEW SOLUTIONS: WIP
Resource Management
- Core pinning
- Hugepages
- RDT

SW Contributions
Demo Integration
Open Source Projects

VNF
- Kubernetes v1.5.0
- Docker v1.12
- flannel v0.6.0
- Multus V1.11
- SR-IOV v6.3-49p1
- DPDK v16.11
- CentOS Linux 7 (Core)
  - Linux 3.10.0-327.36.3.e17.x86_64
- SRIOV-IPMI-Device
  - X715c-1000GFP

Best Practice Guidelines
Intel Experience Kits

Identified Gaps
Integrate
Build

Kubernetes Networking
Integrated
Kubernetes Compute
CALL TO ACTION

• Talk to us about your NFV use cases in K8s
• Join K8s Network SIG and collaborate with us on Multi-network proposal
• Join K8s Resource Management SIG for performance sensitive NFV use cases
• Contribute and provide feedback for Intel Multus & SR-IOV CNI plugin
TALK TO US FOR YOUR CONTAINERS USE CASE?

Email: kuralamudhan.ramakrishnan@intel.com
Slack: https://intel-corp.herokuapp.com
Github: rkamudhan
Twitter: @kuralamudhan