



FD.io : The Universal Dataplane

Ray Kinsella & Hongjun Ni

March 2018



Legal Disclaimers

- Intel technologies may require enabled hardware, specific software, or services activation. Check with your system manufacturer or retailer.
- No computer system can be absolutely secure.
- Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.
- Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.
- All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps
- No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.
- Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.
- Intel, the Intel logo, and other Intel product and solution names in this presentation are trademarks of Intel . . .
- *Other names and brands may be claimed as the property of others.
- © 2018 Intel Corporation.

AGENDA

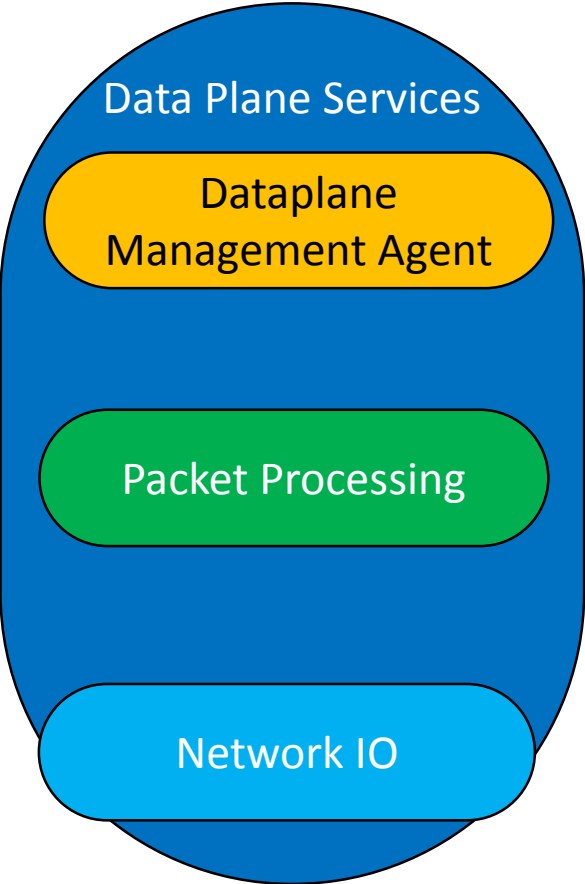
- The FD.io Networking Ecosystem
- Introducing FD.io VPP
- What's new in FD.io?
 - Asymmetric Pipelines
 - Containers
 - TCP Acceleration
 - Accelerators
- Summary



The FD.io Networking Ecosystem

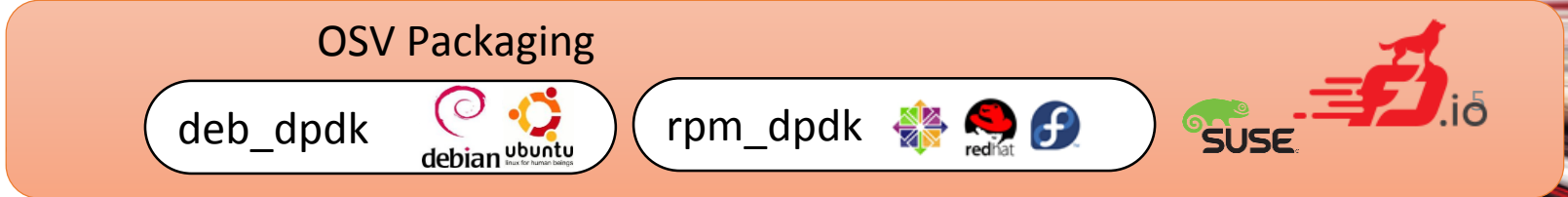
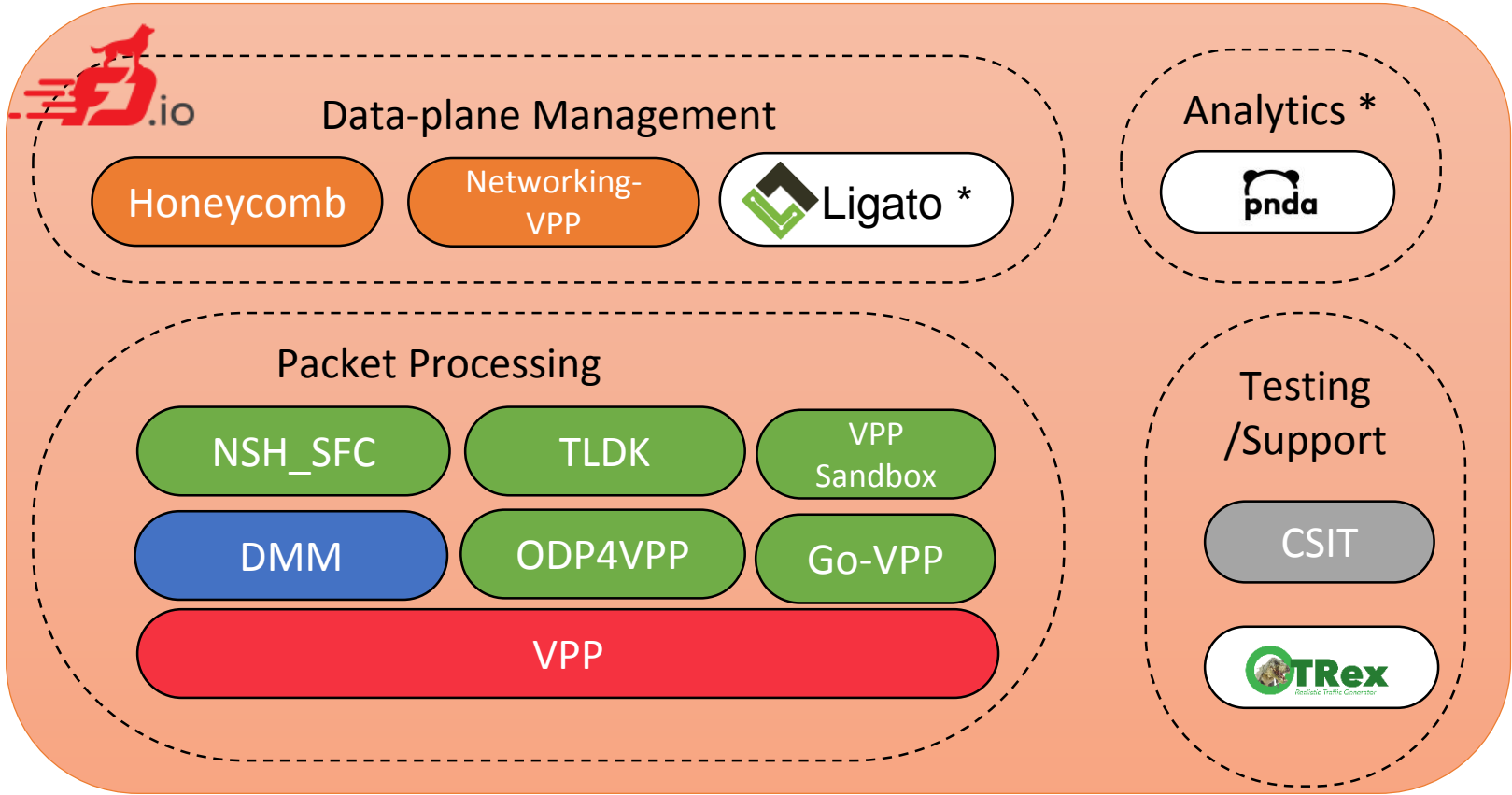


Network Controller



Operating Systems

* Supporting project





IT大咖说
知识共享平台

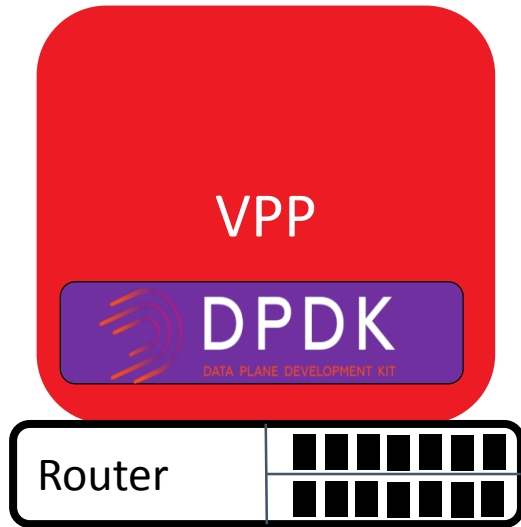
Introducing VPP



FD.io VPP, the new networking multi-uni!

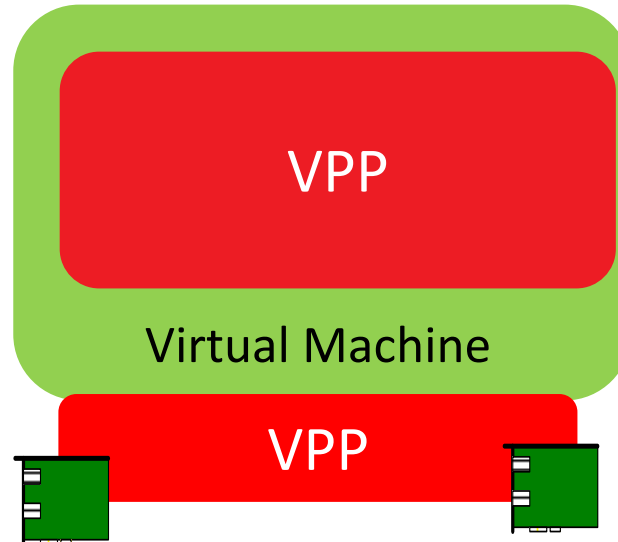


Discrete Appliance



Control Plane for routers and switches. Has been used in Router products since 2002.

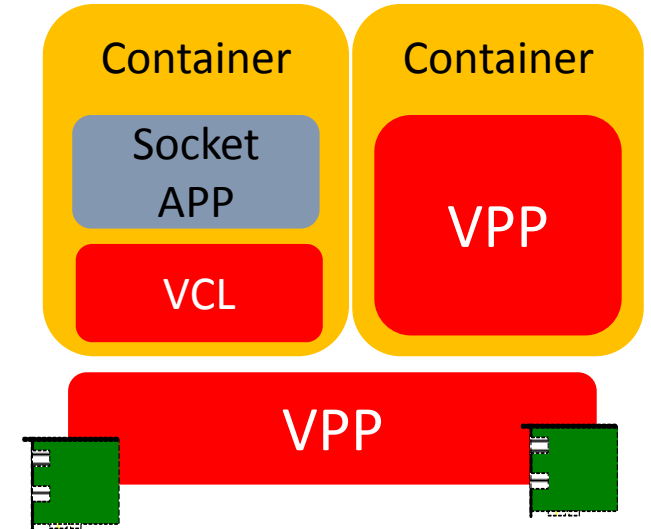
NFVi & VNF



NFVi : vSwitch and vRouter in OpenStack & KVM deployments.

VNF: Load Balancer, CG-NAT, Firewall, Forwarder, IPSEC Gateway, BRAs

Cloud Native



Infra: vRouter in Kubernetes & Docker deployments. TCP, UDP etc Host Stack for Socket apps.

VNF: Load Balancer, CG-NAT, Firewall, Forwarder, IPSEC Gateway, BRAs

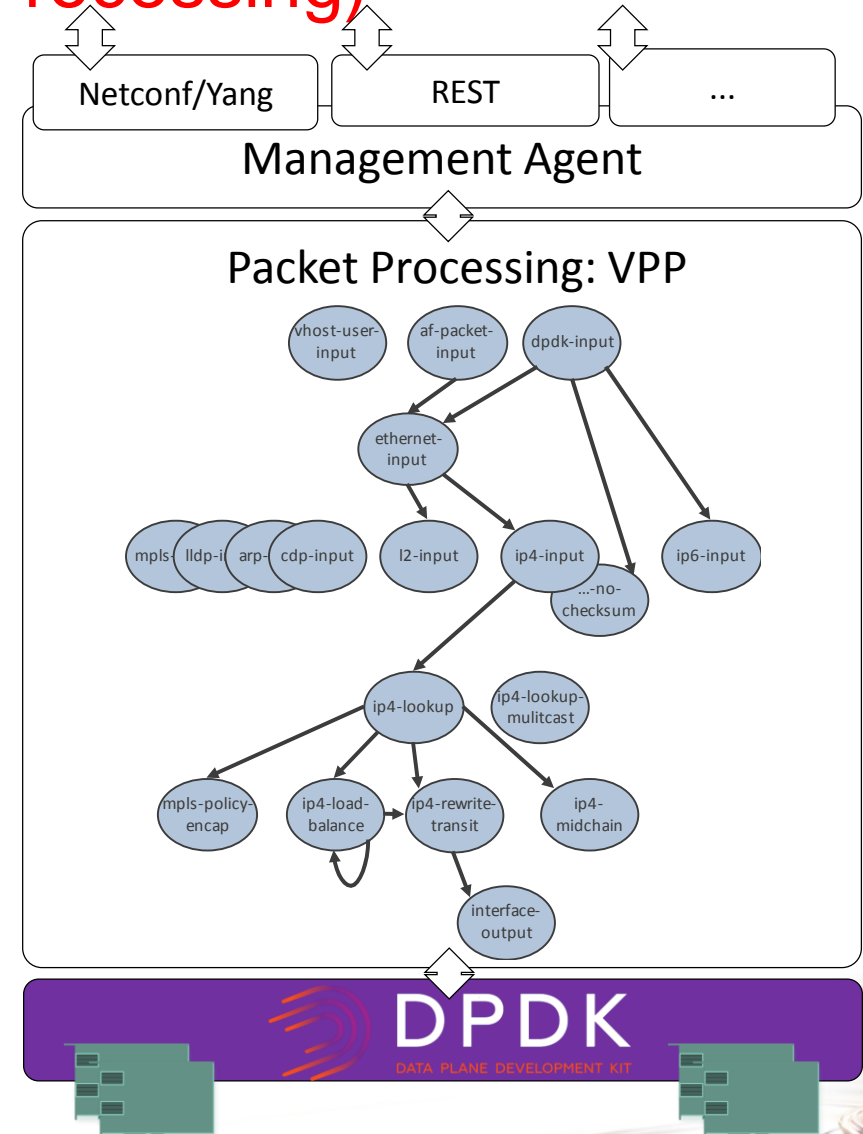
Introducing VPP (Vector Packet Processing)

Fast, Scalable and Deterministic

- 15+ Mpps per core
- Tested to 1TB
- Scalable routing/forwarding tables, supporting millions of concurrent entries.
- 0 packet drops, ~15µs latency

Optimized

- **Optimized for x86 and ARM Architectures.**
- **DPDK** for fast I/O
- **ISA:** SSE, AVX, AVX2, NEON ..
- **IPC:** Batching, no mode switching, no context switches, non-blocking
- **Multi-core:** Cache and memory efficient



VPP is fast!

Introducing VPP



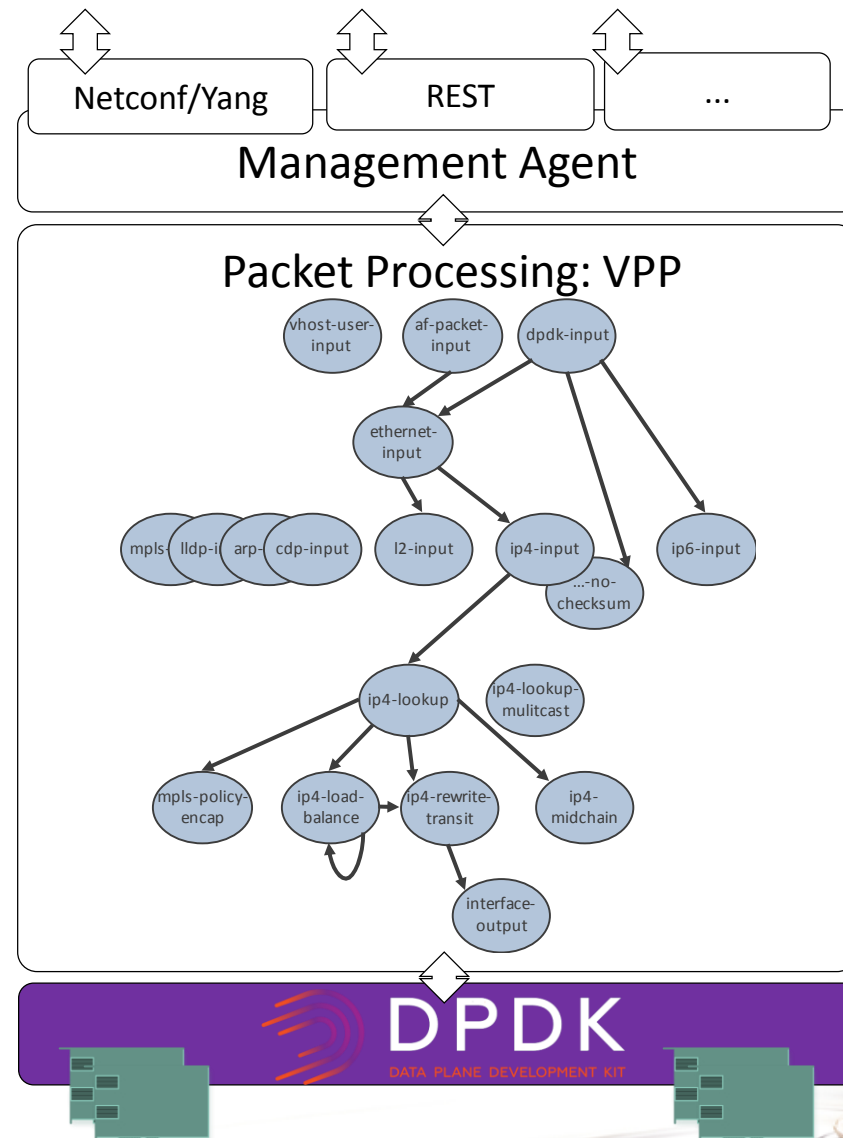
Extensible and Flexible modular design

- Implement as a directed graph of nodes
- Extensible with plugins, plugins are equal citizens.
- Configurable via CP and CLI

Developer friendly

- Deep introspection with counters and tracing facilities.
- Runtime counters with IPC and errors information.
- Pipeline tracing facilities, life-of-a-packet.
- Developed using standard toolchains.

VPP is extensible!



Introducing VPP



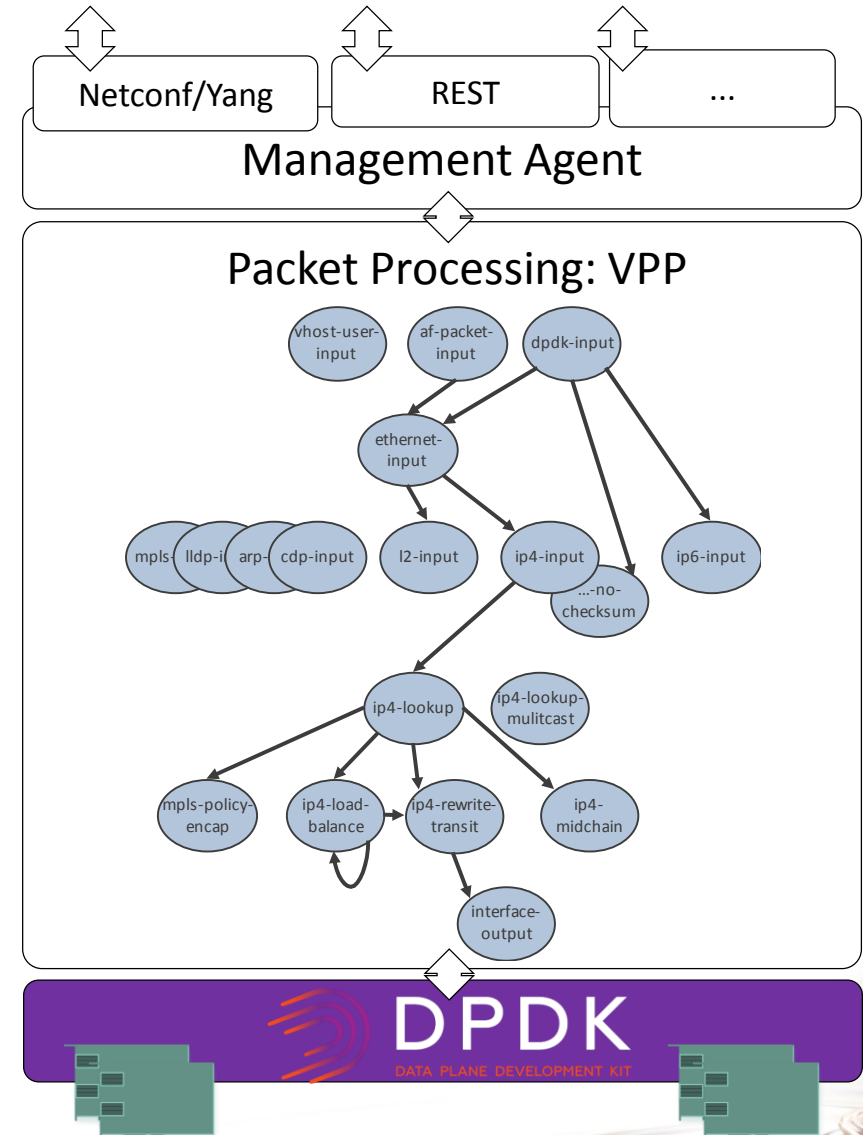
Fully featured

- **L2:** Vlan, Q-in-Q, Bridge Domains, LLDP ...
- **L3:** IPv4, DHCP, IPSEC ...
- **L3:** IPv6, Discovery, Segment Routing ...
- **L4:** SCTP, TCP, UDP ...
- **CP:** API, CLI, IKEv2 ...
- **Overlays:** GRE, VXLAN, Geneve ...

Usability

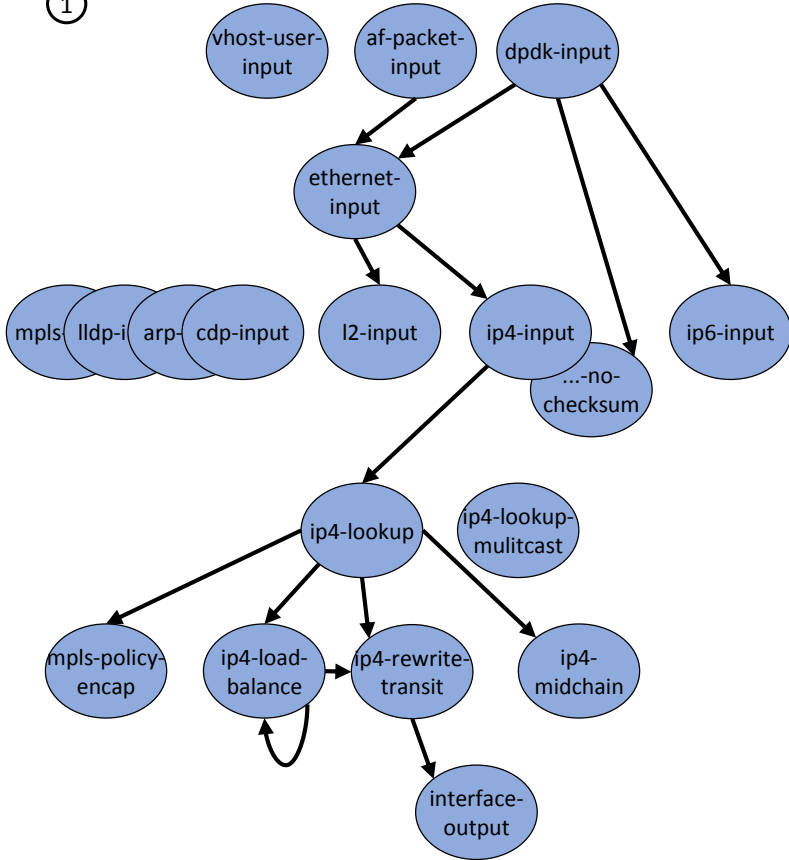
- Language bindings
- Open Stack/ODL (Netconf/Yang)
- Kubernetes/Contiv-VPP (Python API)
- OSV Packaging

VPP bootstraps VNF development!

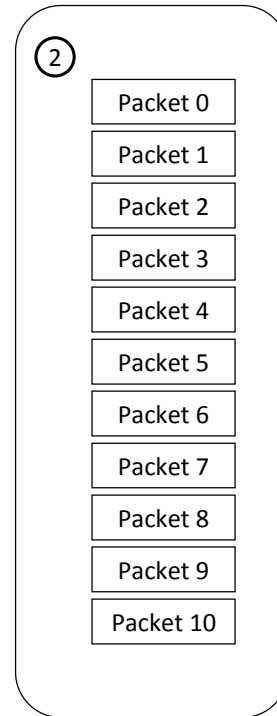


VPP: How does it work?

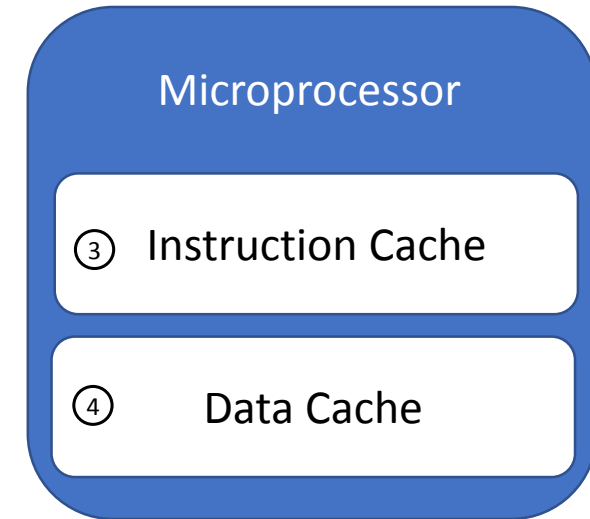
①



②



... graph nodes are optimized to fit inside the instruction cache ...



... packets moved through graph nodes in vector ...

... packets are pre-fetched, into the data cache ...

Packet processing is decomposed into a directed graph node ...

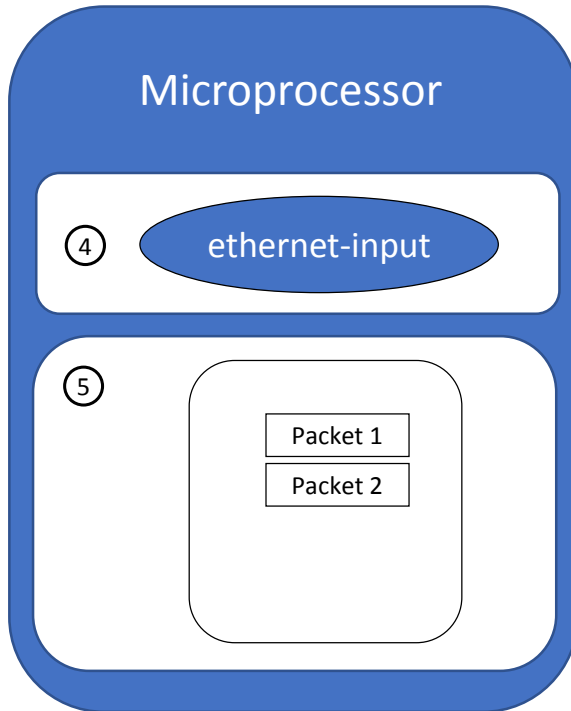
* approx. 173 nodes in default deployment

VPP: How does it work?

⑥

dispatch fn()

... instruction cache is warm with the instructions from a single graph node ...



... data cache is warm with a small number of packets ..

while packets in vector

Get pointer to vector

while 4 or more packets

PREFETCH #3 and #4

PROCESS #1 and #2

ASSUME next_node same as last packet

Update counters, advance buffers

Enqueue the packet to next_node

while any packets

<as above but single packet>

... packets are processed in groups of four, any remaining packets are processed on by one ...

Related Projects

- OPNFV Fast Data Stacks (FDS)
- CSIT: Continuous System Integration and Testing
- Cloud Native : Kubernetes, Contiv-VPP & Ligato.

OPNFV FastDataStacks (FDS)

“The 20th century was about invention, the 21st is about mashups and integration” – Toby Ford, AT&T

- Integrate VPP into existing OPNFV test scenarios
- Scenarios
 - OpenStack – ODL (Layer2) – VPP
 - OpenStack – ODL (Layer3) – VPP
 - OpenStack – Networking-VPP – VPP
 - etc

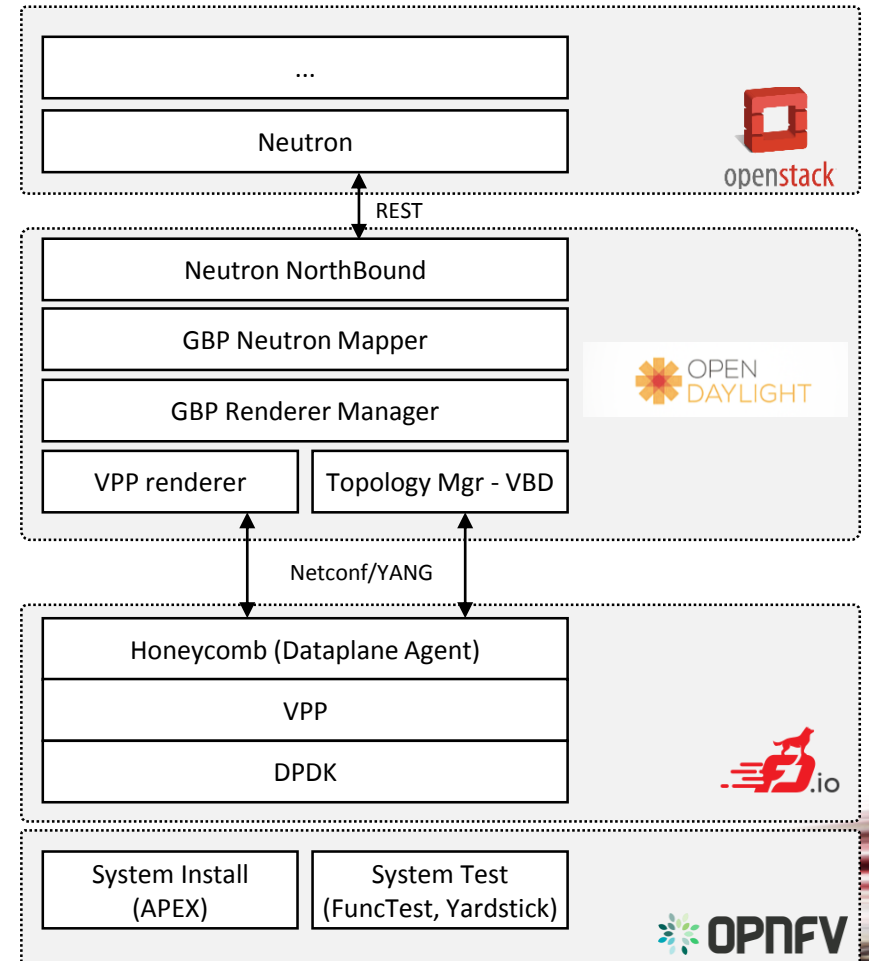
- Diverse set of contributors:



- <https://wiki.opnfv.org/display/fds>

See also:

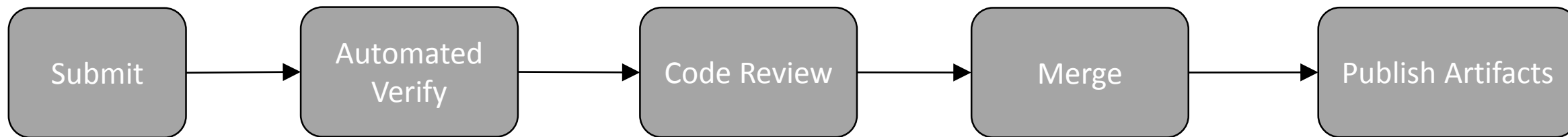
FDS Architecture: <https://wiki.opnfv.org/display/fds/OpenStack-ODL-VPP+integration+design+and+architecture>



CSIT: Continuous System Integration and



Built into the development process – patch by patch



Build/Unit Testing
120 Tests/Patch

Build binary packaging for
Ubuntu 14.04
Ubuntu 16.04
Centos 7

Automated Style Checking

Unit test :

IPFIX	IPv6
BFD	IP Multicast
Classifier	L2 FIB
DHCP	L2 Bridge Domain
FIB	MPLS
GRE	SNAT
IPv4	SPAN
IPv4 IRB	VXLAN
IPv4 multi-VRF	

System Functional Testing
252 Tests/Patch

DHCP – Client and Proxy
GRE Overlay Tunnels
L2BD Ethernet Switching
L2 Cross Connect Ethernet Switching
LISP Overlay Tunnels
IPv4-in-IPv6 Software Tunnels
Cop Address Security
IPSec
IPv6 Routing – NS/ND, RA, ICMPv6
uRPF Security
Tap Interface
Telemetry – IPFIX and Span
VRF Routed Forwarding
iACL Security – Ingress – IPv6/IPv6/Mac
IPv4 Routing
QoS Policer Metering
VLAN Tag Translation
VXLAN Overlay Tunnels

Performance Testing
144 Tests/Patch, 841 Tests

L2 Cross Connect
L2 Bridging
IPv4 Routing
IPv6 Routing
IPv4 Scale – 20k,200k,2M FIB Entries
IPv4 Scale - 20k,200k,2M FIB Entries
VM with vhost-user
PHYS-VPP-VM-VPP-PHYS
L2 Cross Connect/Bridge
VXLAN w/L2 Bridge Domain
IPv4 Routing
COP – IPv4/IPv6 whiteless
iACL – ingress IPv4/IPv6 ACLs
LISP – IPv4-o-IPv6/IPv6-o-IPv4
VXLAN
QoS Policer
L2 Cross over
L2 Bridging

Usability

Merge-by-merge:
apt installable deb packaging
yum installable rpm packaging
autogenerated code documentation
autogenerated cli documentation

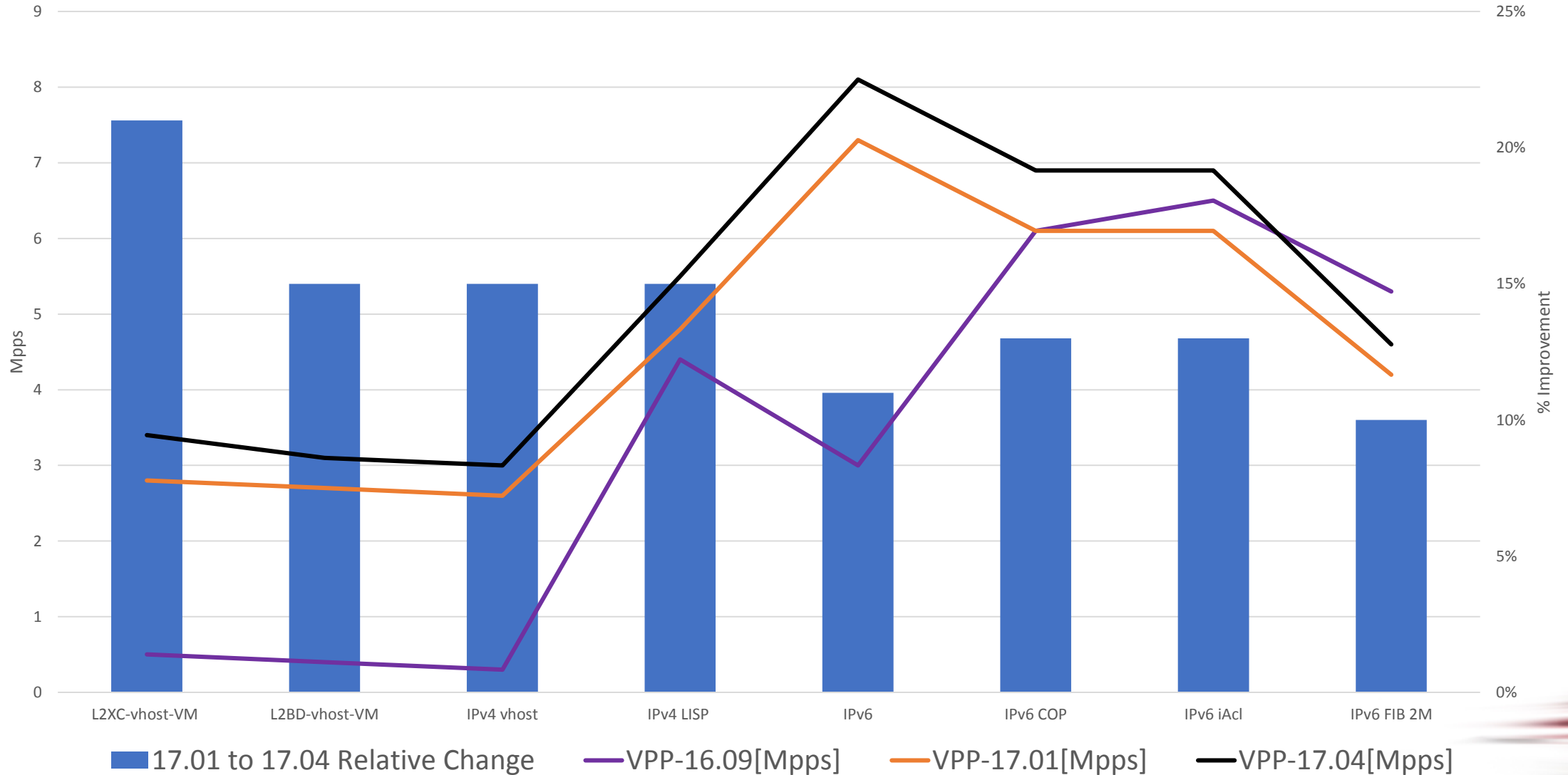
Per release:
autogenerated testing reports
report perf improvements

Puppet modules
Training/Tutorial videos
Hands-on-usecase documentation

Merge-by-merge packaging feeds
Downstream consumer CI pipelines

Run on real hardware in fd.io Performance Lab

CSIT NDR Throughput Improvement



Cloud Native : Kubernetes, Contiv-VPP & L

(a.k.a. Cloud Native Network Microservices)



Production-Grade
Container Orchestration



Performance-Centric
Container Networking



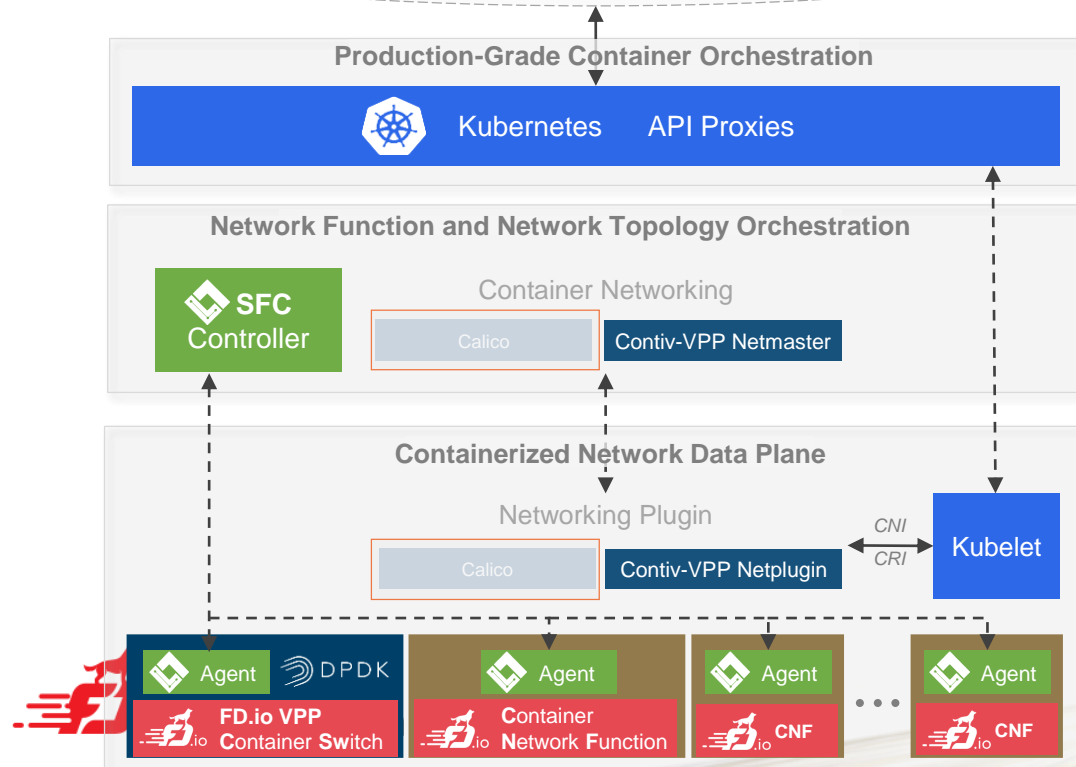
Cloud-native NF Orchestration
Cloud-native NF Agent platform



Containerized Fast
Data Input/ Output

Enabling Production-Grade Native Cloud Network Services at Scale

Service Policy Service Topology Lifecycle





What's new in FD.io?



FD.io Release Notes



17-01
Release:
VPP, Honeycomb,
NSH_SFC, ONE

17-04
Release:
VPP, Honeycomb,
NSH_SFC, ONE...

17-07
Release:
VPP, Honeycomb,
NSH_SFC, ONE...

17-10
Release:
VPP, Honeycomb,
NSH_SFC, ONE...

18-01
Release:
VPP, Honeycomb,
NSH_SFC, ONE...

17-01 New Features

DPDK 16.11 integration
Performance
IPSEC Performance
Network features
Hierarchical FIB
HQoS support
Simple Port Analyzer
BFD, ACL, IPFIX, SNAT
L2 GRE over IPsec tunnels
LLDP
LISP Enhancements
Flow Per Packet

17-04 New Features

DPDK 17.02 integration
Network features
Host Stack
DHCP & ND Relay/Proxy
SNAT NAT64, LW46
SRv6 support
iOAM improvements
IPFIX IPv6 Support

17-07 New Features

DPDK 17.05 integration
Interfaces
MemIF
Virtio-user
Network features
TCP RFC Compatibility
TCP Loss Recovery
MPLS Multicast & SRv6
Bidirectional Fwd Detection
GRE over IPV6
iOAM for SRv6
GTP-U support
LISP NSH Support

17-10 New Features

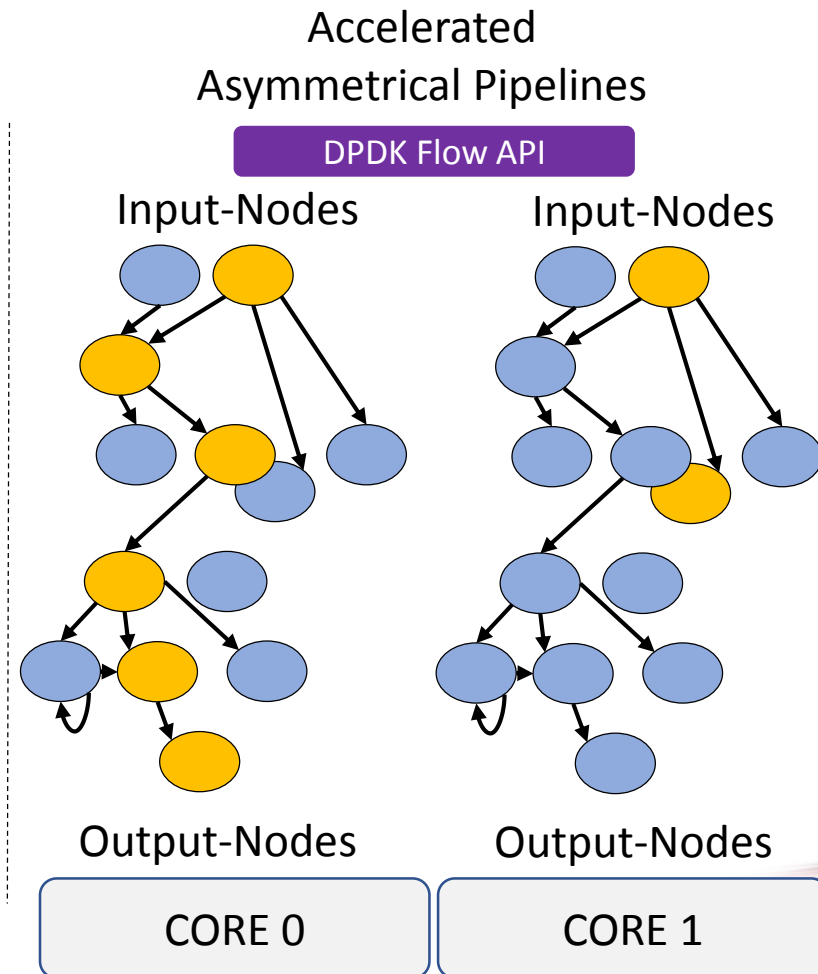
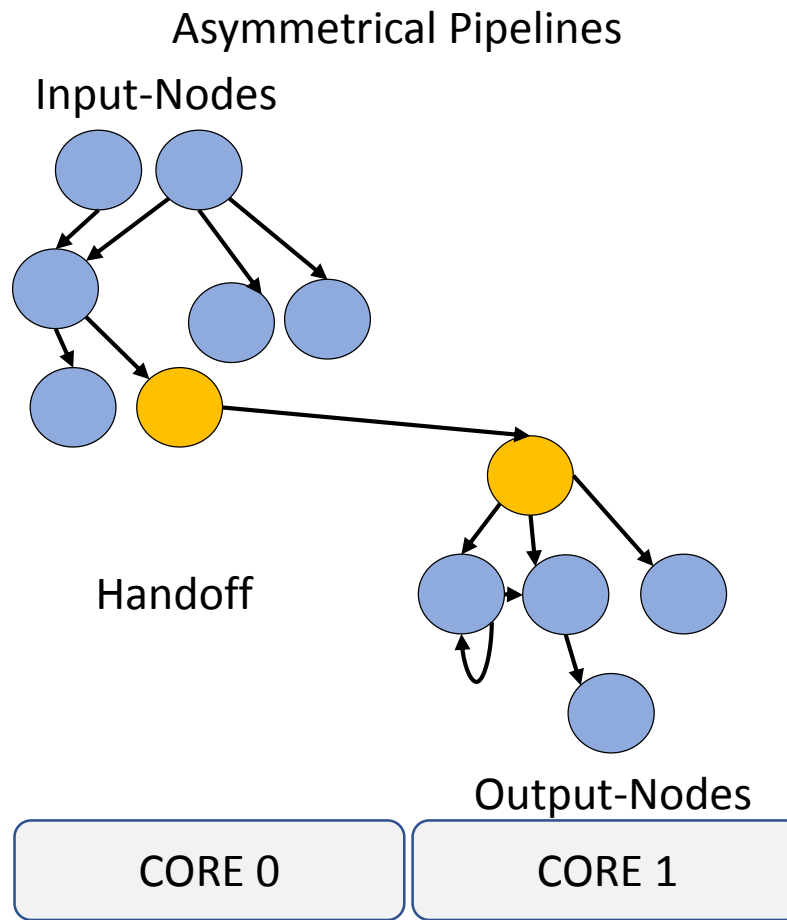
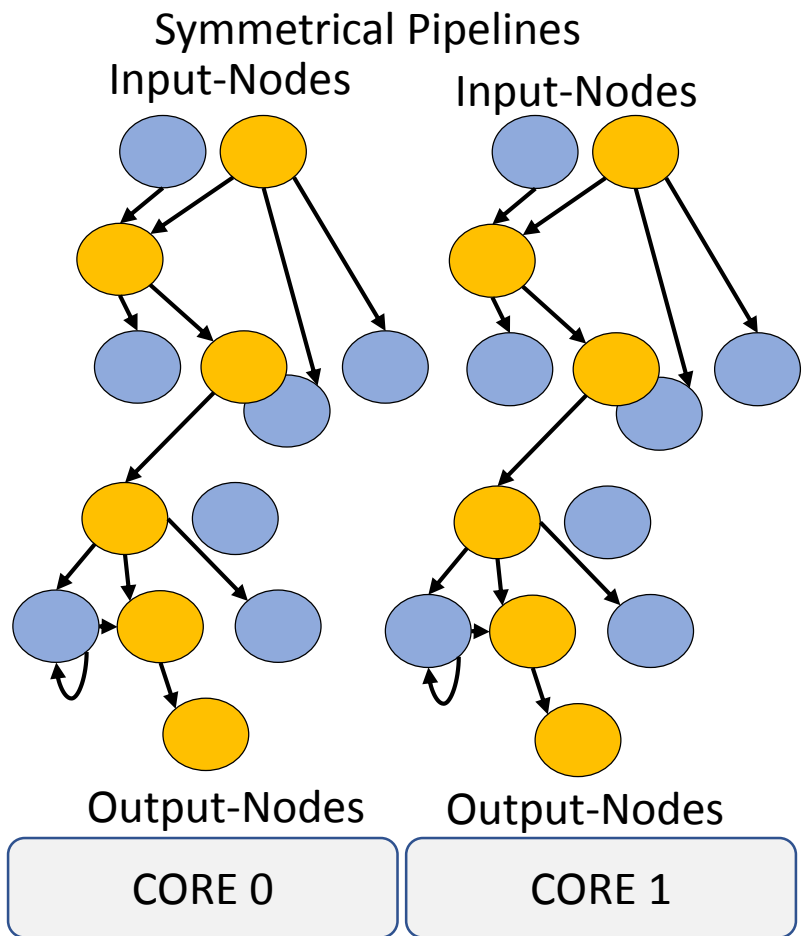
DPDK 17.08 integration
Network features
IPSec support hFIB
VPLS and VPWS support
SNAT / NAT
NAT64 Fragmentation
One-armed NAT
TCP stack scaling
VPP Comms Library (VLC)
PPPoE Control Plane
PPPoE Encap & Decap

18-01 New Features

DPDK 17.11 integration
Arm64/Arm-v8 support
Interfaces
FastTap
MemIF
Network Features
SNAT / NAT
GENEVE
Host Stack
ACL stats
Plugins
Kube-proxy
L2 Emulation
Memif

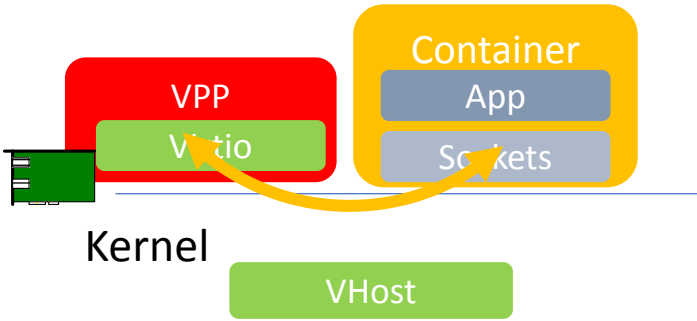
VPP is rapidly evolving!

New Directions: Asymmetrical Pipelines



New Directions: Containers

Container Connectivity



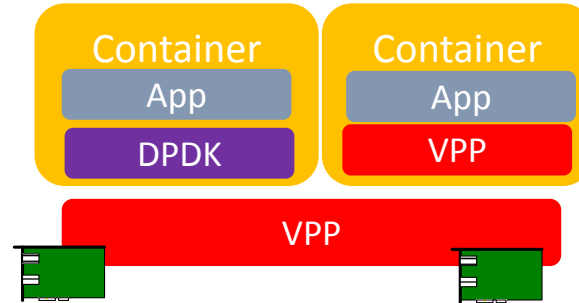
vEth TAPv2 (FastTap)

What is it?

Classic Linux Kernel Container interface, widely-used.

Throughput: **< 1 mpps**

Scaling: **> 1000 Containers**



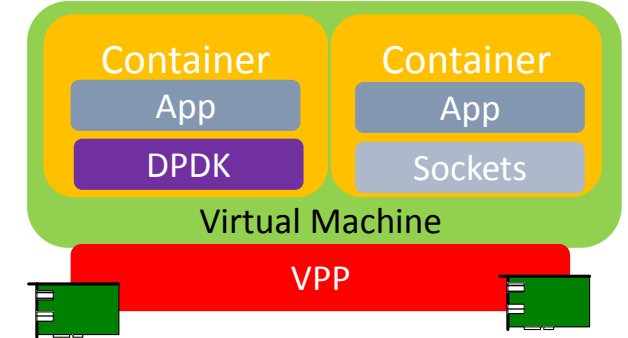
MemIF & Virtio-User:

What is it?

Container Virtual Interface for **Bare-metal** deployments

Throughput: **< 10 mpps**

Scaling: **< 1000 Containers**



Master-VM:

What is it?

Container Virtual Interface for **Virtualized** deployments

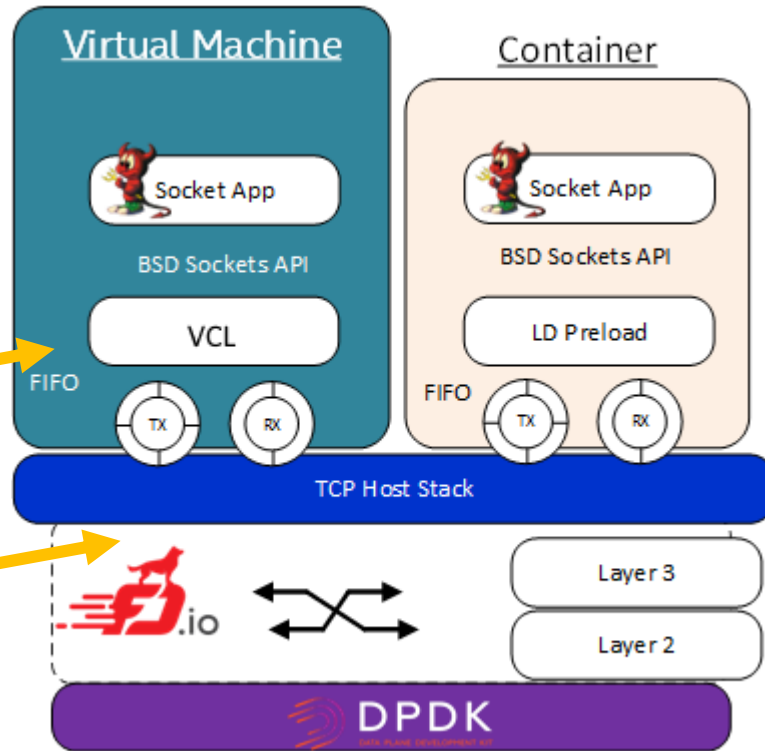
Throughput: **< 10 mpps**

Scaling: **< 1000 Containers**

New Directions: TCP Acceleration

VPP & DMM

VPP TCP Host Stack

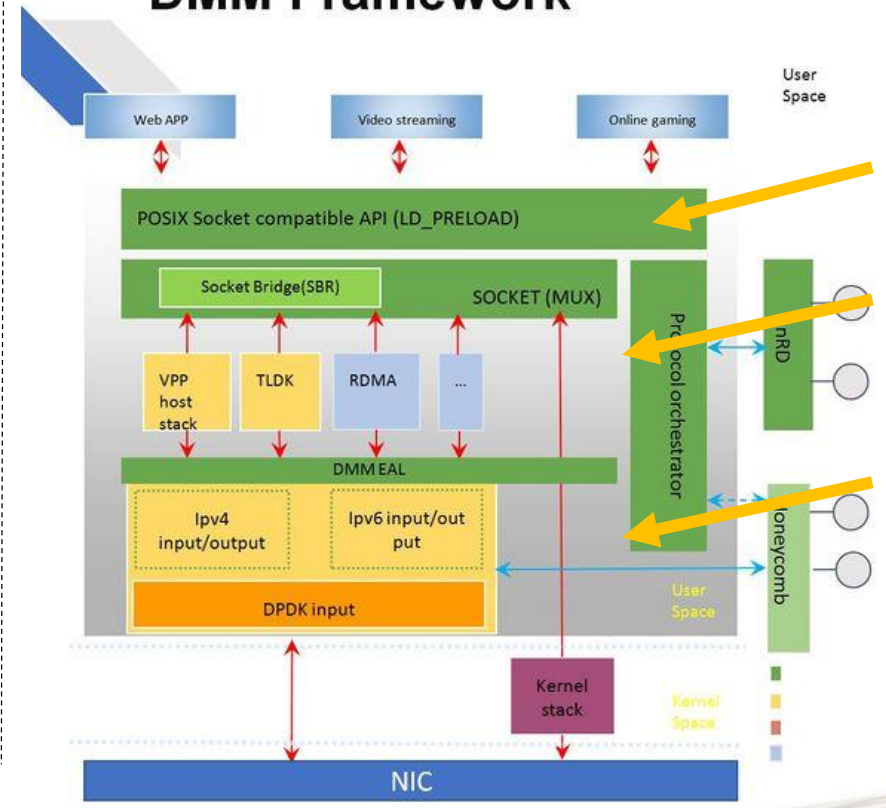


VCL Library, directs the TCP bit stream to FD.io VPP.

TCP Streams are aggregated in FD.io VPP



DMM Framework



Socket API Library

Network intelligence to route the packet

Stack Abstraction Layer

New novel approaches to accelerating TCP!

New Directions: Accelerators

Accelerating IPSEC with FD.io VPP and DPDK



VPP

Cryptodev API



DPDK



Vector & Crypto
Instructions



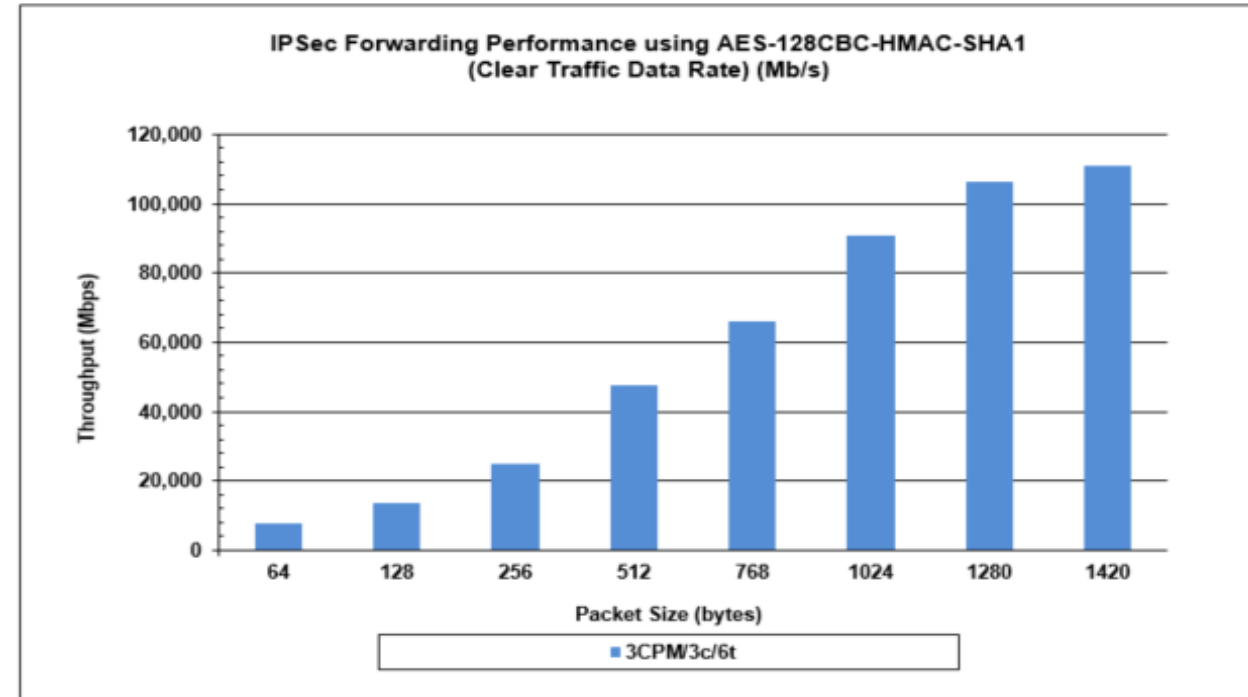
Quick Assist Crypto
Accelerators

Performance numbers are subject to change with later versions.

Disclaimer: 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. Source: Intel internal testing as of July 18, 2016.

Intel® Xeon® Scalable Processors @2.1GHz

VPP 17.01 IPsec Performance 1:1 Nodes (bi-directional flows)



100Gb of IPSEC with 3 cores!



FD.io Foundation

* Other names and brands may be claimed as the property of others.

Summary

- FD.io VPP is a **multi-vendor** packet processing technologies that is leading the **network transformation**.
- FD.io VPP is a fast and easy-to-use tools, so please try them today!
- Please join us in FD.io Today!

Email : ray.kinsella [at] intel.com
IRC: mortderire

Email : hongjun.ni [at] intel.com