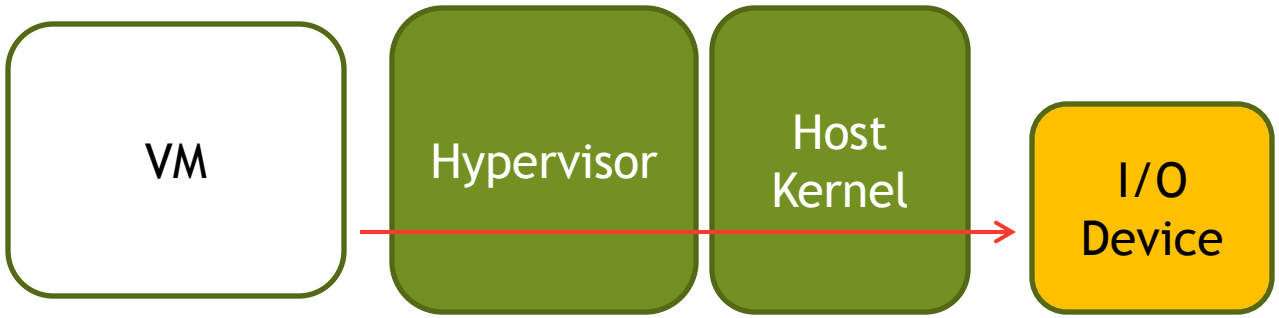

# PCI SR-IOV on FreeBSD

Ryan Stone

rstone@FreeBSD.org

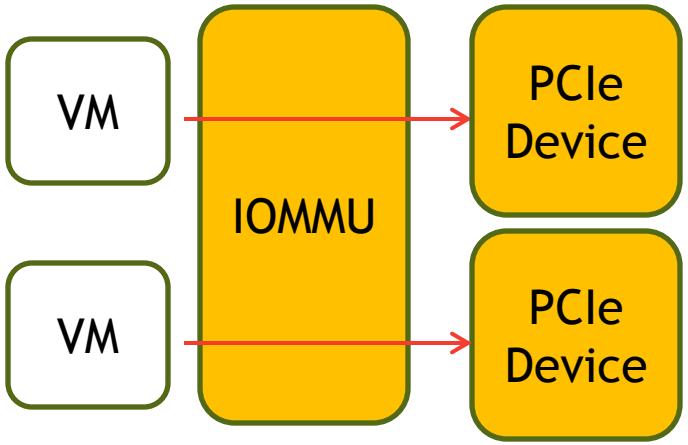
# Performing I/O from a VM

- I/O is through (para-)virtualized devices implemented in the hypervisor
  - Additional load on host; less CPU available to guests
- Accessing advanced NIC offloads can be tricky
- Not all I/O devices have paravirtualized equivalents
  - e.g. crypto/compression offload engines



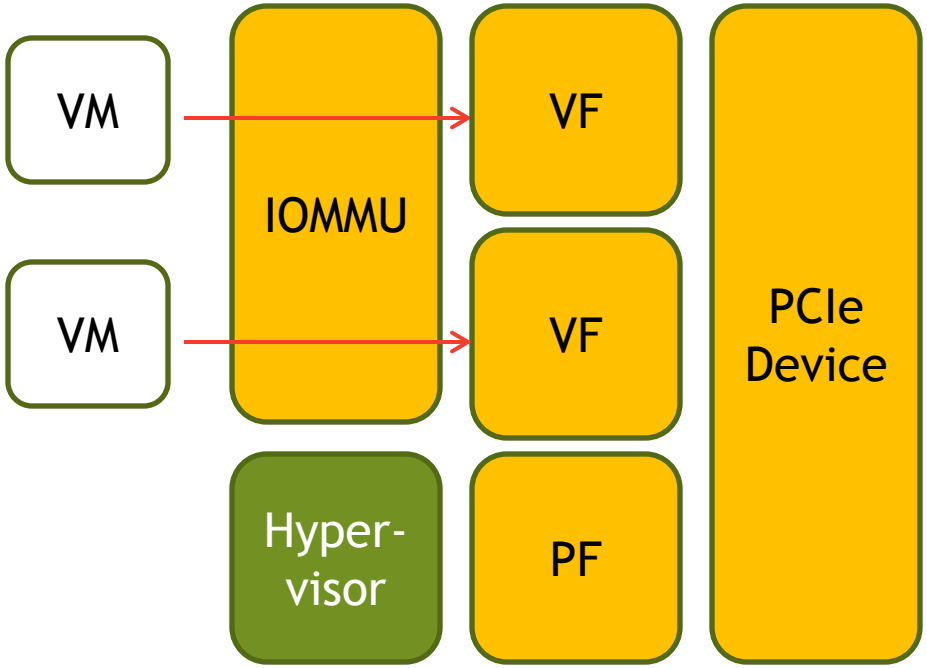
# PCI Passthrough

- Use HW offload features to give VM direct access to a PCIe device
- Impossible to share a device between multiple VMs



# SR-IOV

- Allow creation of virtual PCIe device (VF) sharing resources of a physical device
- Hypervisor can allocate 1 (or more) VF per VM
- VFs are accessed through PCI Passthrough



# SR-IOV Specification

- SR-IOV spec defines how to create VFs, enumerate them and assign resources
- The rest of the details are left to the implementation
- This gives HW makers lots of flexibility
  - But also means that a significant amount of driver code is needed to implement PF side of SR-IOV

# Configuring SR-IOV

- SR-IOV configuration needs to be flexible and extensible due to varying hw capabilities
- New PF drivers shouldn't to have to extend infrastructure to expose new capabilities
  - Critical that new drivers don't require ABI changes
- Need one unified tool for configuring all PF drivers
- Solution: config file is a hierarchy of K/V pairs
  - Drivers advertise which keys they accept

# Configuration Schemas

- PF drivers will advertise their capabilities via a configuration schema
- Config schemas define:
  - Name-value pairs accepted as configuration
  - The type of value accepted
  - Whether the param is required or optional
  - If optional, whether a default value is applied
- Parameters apply to the whole PF or one VF
  - Different VFs can have different config
- View config schema:
  - `iovctl -S -d <device>`

# Configuration Schema Example

```
# iovctl -S -d ixl0
The following configuration parameters may be
   configured on the PF:
         num_vfs : uint16_t (required)
         device : string (required)

The following configuration parameters may be
   configured on a VF:
         passthrough : bool (default = false)
         mac-addr : unicast-mac (optional)
         mac-anti-spoof : bool (default = true)
         allow-set-mac : bool (default = false)
         allow-promisc : bool (default = false)
```

# iovctl

- `iovctl(8)` is used to configure SR-IOV
- Configuration flow is:
  - Fetch config schema from kernel
  - Validate iovctl.conf against schema
  - Pass configuration up to kernel
  - PCI subsystem creates VFs
  - PF driver is informed that VFs have been created along with their configuration

# iovctl.conf

- File is in UCL format — same as pkg.conf
- Three types of sections:
  - PF section — configuration for whole device
  - Default section — Default config for all VFs
  - VF sections — configuration for single VF
    - » Can override values set in default section
- Sections with no parameters can be omitted
- One iovctl.conf file for each PF device
- To run iovctl at boot, set rc.conf var iovctl_files to list of iovctl.conf files
  - `iovctl_files="/etc/iov/ixl0.conf /etc/iov/ixl1.conf"`

# SR-IOV Infrastructure Parameters

- All PFs have the following required params:
  - device — Name of PF device
  - num_vfs — Number of VFs to create
- VFs accept the following optional param:
  - passthrough — reserve VF for bhyve PCI passthrough
    - » Defaults to false

# Example iovctl.conf

```
PF {
        device : ixl0;
        num_vfs : 3;
}
DEFAULT {
        passthrough : true;
}

# VF for use by host
VF-0 {
        mac-addr : "02:01:02:03:04:00";
        passthrough : false;
}
VF-1 {
        mac-addr : "02:01:02:03:04:01";
}
# VF-2 section is omitted: accept all default values for VF-2
```

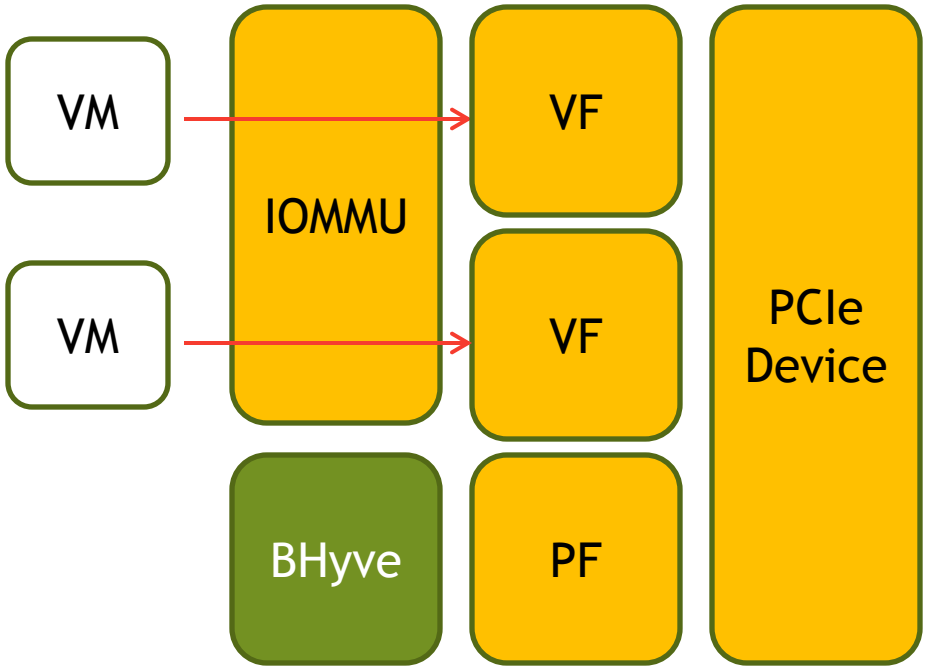
# Config Schema Stability

- PF device driver authors are required to treat their config schemas like an ABI
  - On stable branches, this means that existing iovctl.conf files MUST continue to work exactly the same
    - » No new required parameters
    - » No changes in default values or behavior
  - On head:
    - » New required parameters are discouraged
    - » Changes in default values/behavior is *strongly* discouraged
- PF drivers should default to most secure config

# Hardware/Driver Support

- Intel ixl driver has full SR-IOV support
- Intel ixgbe driver has support available as a Technology Preview
  - Only Intel 82599 and newer cards support SR-IOV

# Use Case: BHyve host

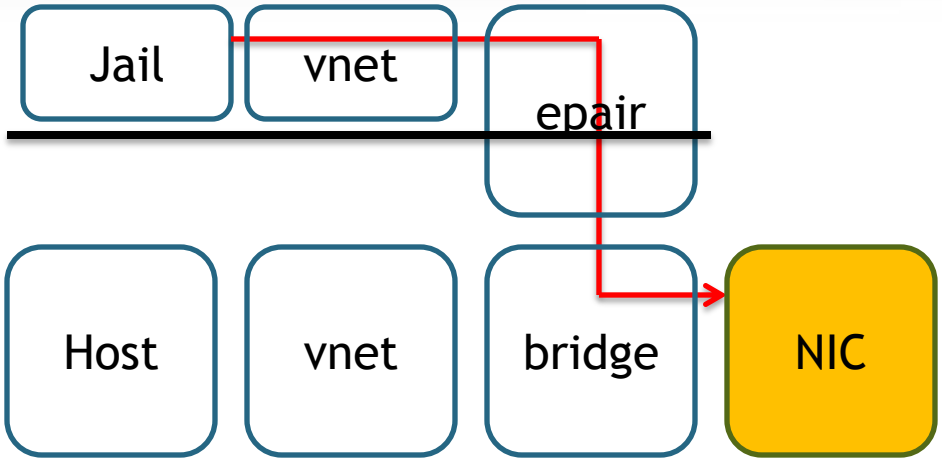


# Configuration for BHyve host

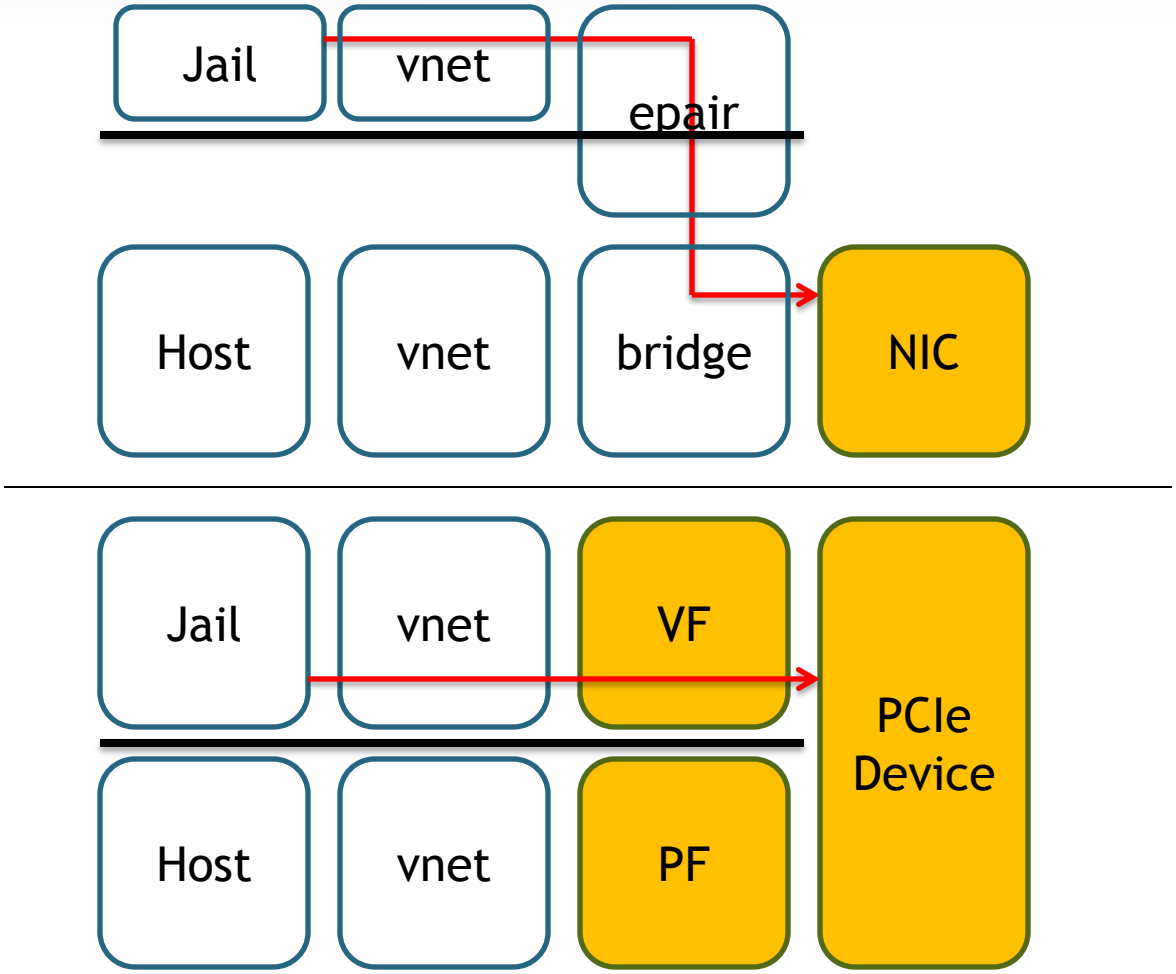
- iovctl.conf

```
PF {
 device : "ixl0";
 num_vfs : 4;
}
DEFAULT {
 passthrough : true;
}
```

# Use Case: VIMAGE jails



# Use Case: VIMAGE jails



# Configuration for VIMAGE jail

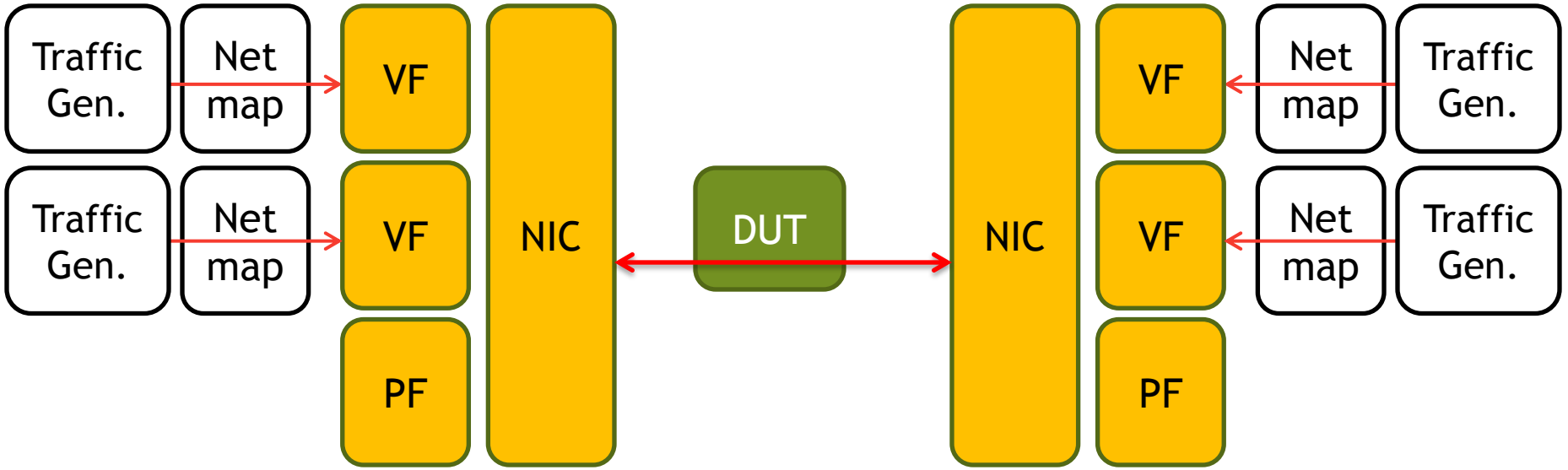
- iovctl.conf

```
PF {
 device : "ixl0";
 num_vfs : 4;
}
```

- jail.conf

```
testjail {
 vnet;
 vnet.interface = "ixlv1";
}
```

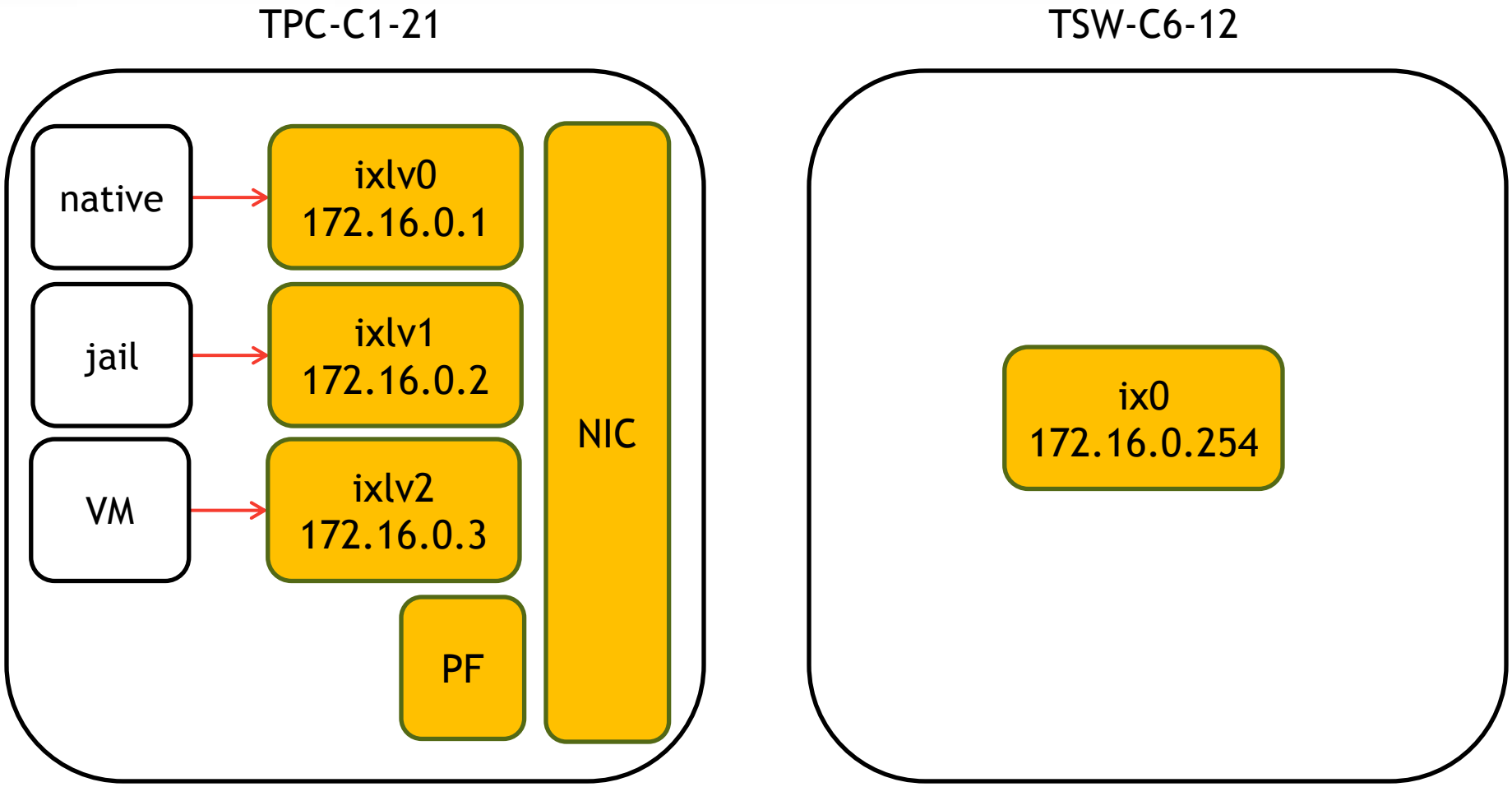
# Use Case: Netmap



# Special Thanks

- Reviewers
  - Mark Johnston (markj@) and Sean Mahood
  - John Baldwin (jhb@) and Jack Vogel (jfv@)
  - FreeBSD Documentation team
  - Many others
- Sandvine

# Demo


TPC-C1-21
native
jail
VM
ixlv0
172.16.0.1
ixlv1
172.16.0.2
ixlv2
172.16.0.3
PF
NIC
TSW-C6-12
ix0
172.16.0.254