Entender as redes virtuais NFVIS: OVS, DPDK e SR-IOV

Contents

Introdução

Este documento descreve o esquema de rede virtual que a plataforma NFVIS fornece para comunicação de VNFs em redes corporativas e de serviço.

Componentes Utilizados

As informações neste documento são baseadas nestes componentes de hardware e software:

- ENCS5412 executando NFVIS 4.7.1-FC4
- c8300 uCPE 1N20 executando NFVIS 4.12.1-FC2

As informações neste documento foram criadas a partir de dispositivos em um ambiente de laboratório específico. Todos os dispositivos utilizados neste documento foram iniciados com uma configuração (padrão) inicial. Se a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando.

Visão geral de rede no NFVIS

Uma rede de gerenciamento interno (int-mgmt-net) e uma bridge (int-mgmt-br) são usadas internamente para monitoramento de VNF, atribuindo endereços IP de gerenciamento da sub-rede 10.20.0.0/24.

Plataforma ENCS54XX

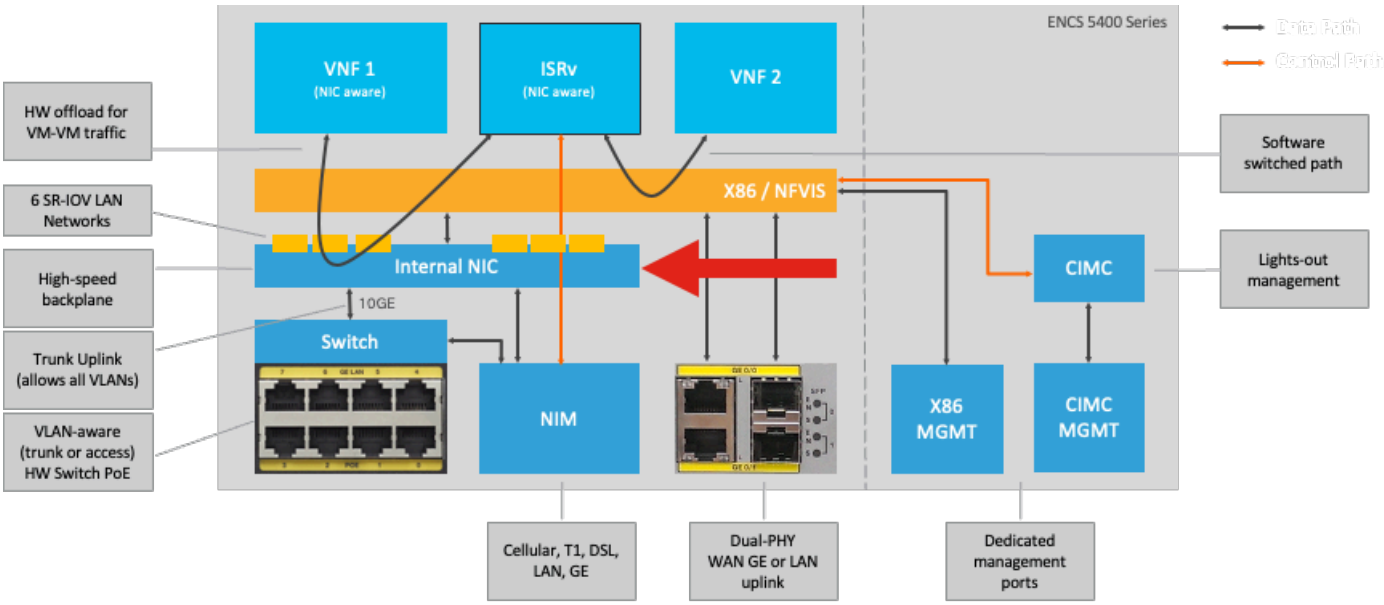


Figura 1. Conexões internas entre switch de hardware e placas de rede de uplink WAN/LAN

UCPE Catalyst 8200

- O NFVIS pode ser acessado por padrão através da porta WAN ou da porta LAN GE0/2 para gerenciamento.
- A rede WAN (wan-net e wan2-net) e a ponte WAN (wan-br e wan2-br) estão configuradas para ativar o DHCP por padrão. GE0-0 é associado à ponte WAN e GE0-1 à ponte WAN2 por padrão.
- O endereço IP de gerenciamento 192.168.1.1 no Catalyst 8200 UCPE é acessível através de GE0-2.
- GE0-2 está associado à bridge LAN.
- Uma rede de gerenciamento interno (int-mgmt-net) e uma ponte (int-mgmt-br) são criadas e usadas internamente para o monitoramento do sistema.

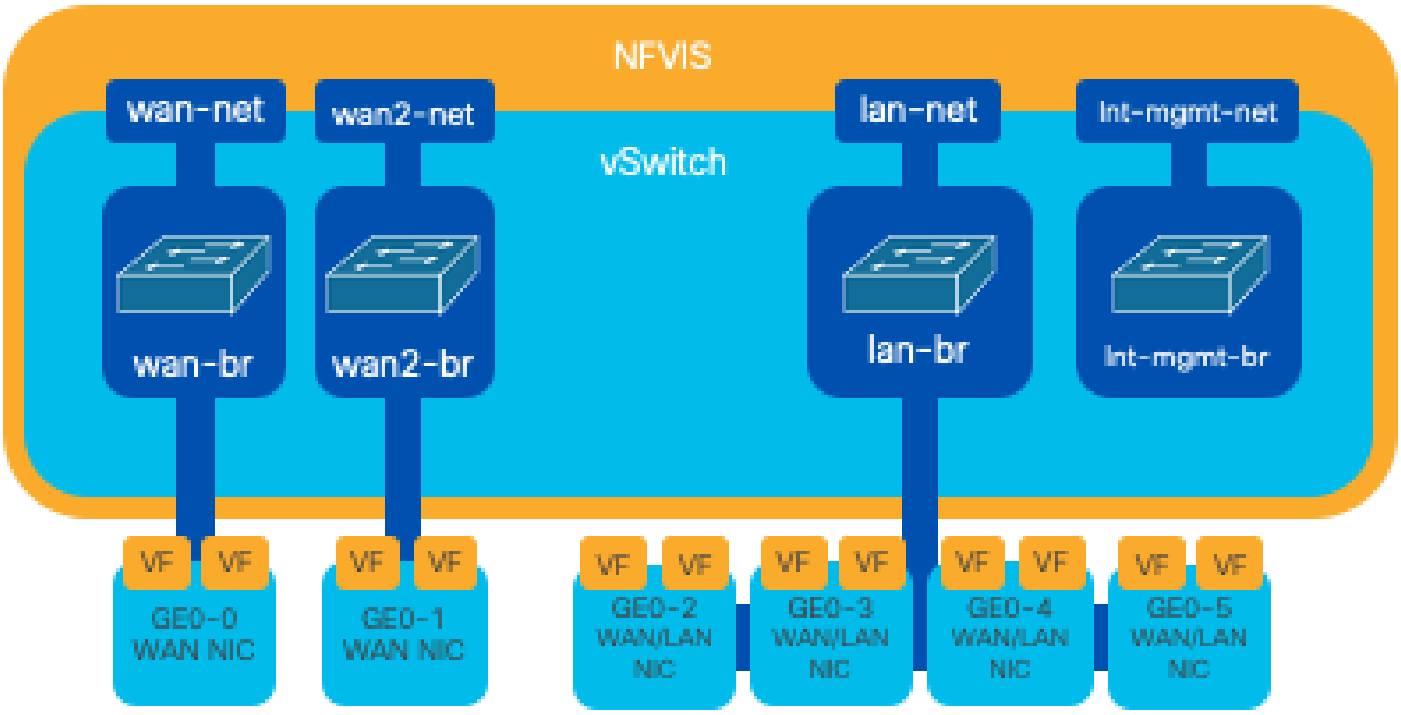


Figura 2. Bridging interno e switches virtuais atribuídos às placas de rede 8200

Catalyst 8300 uCPE 1N20

1. NFVIS pode ser acessado por padrão através das portas FPGE (Gigabit Ethernet do Painel Frontal)WAN ou através da porta LAN GE0-2 paraGerenciamento

2. A rede WAN (wan-net) e uma ponte WAN (wan-br) são definidas por padrão para ativar o DHCP. GE0-0 é associado por padrão à ponte WAN

3. A rede WAN (wan2-net) e uma ponte WAN (wan2-br) são criadas por padrão, mas não estão associadas a nenhuma porta física

4. GE0-2 está associado à bridge LAN, todas as outras portas não estão associadas ao OVS

5. O IP de gerenciamento 192.168.1.1 no C8300-uCPE é acessível via GE0-2

6. Uma rede de gerenciamento interno (int-mgmt-net) e uma ponte (int-mgmt-br) são criadas e usadas internamente para o monitoramento do sistema.

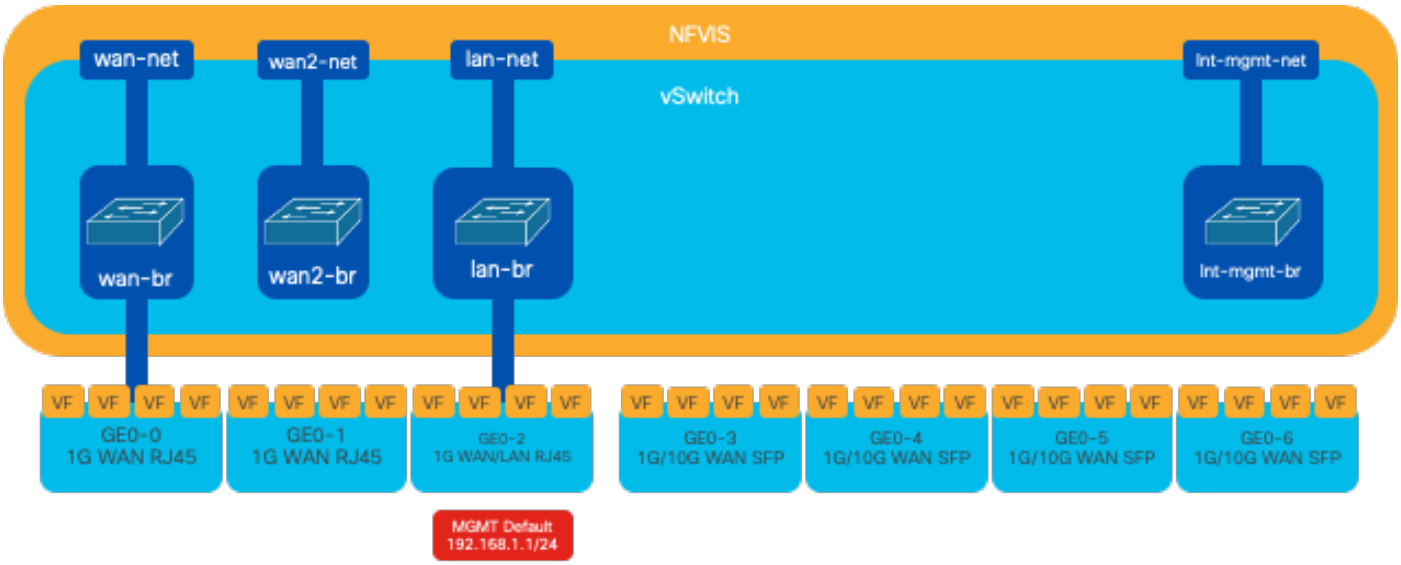


Figura 3. Bridging interno e switches virtuais atribuídos às placas de rede 8300

Tecnologias de virtualização de rede

Open vSwitch (OVS)

O Open vSwitch (OVS) é um switch virtual multicamada de código aberto projetado para permitir a automação da rede através de extensões programáticas, ao mesmo tempo em que fornece suporte para interfaces e protocolos de gerenciamento padrão, como NetFlow, sFlow, IPFIX, RSPAN, CLI, LACP e 802.1ag. É amplamente usado em grandes ambientes virtualizados, principalmente com hipervisores para gerenciar o tráfego de rede entre máquinas virtuais (VMs). Ele permite a criação de topologias de rede sofisticadas e políticas gerenciadas diretamente através da interface NFVIS, fornecendo um ambiente versátil para a virtualização da função de rede.

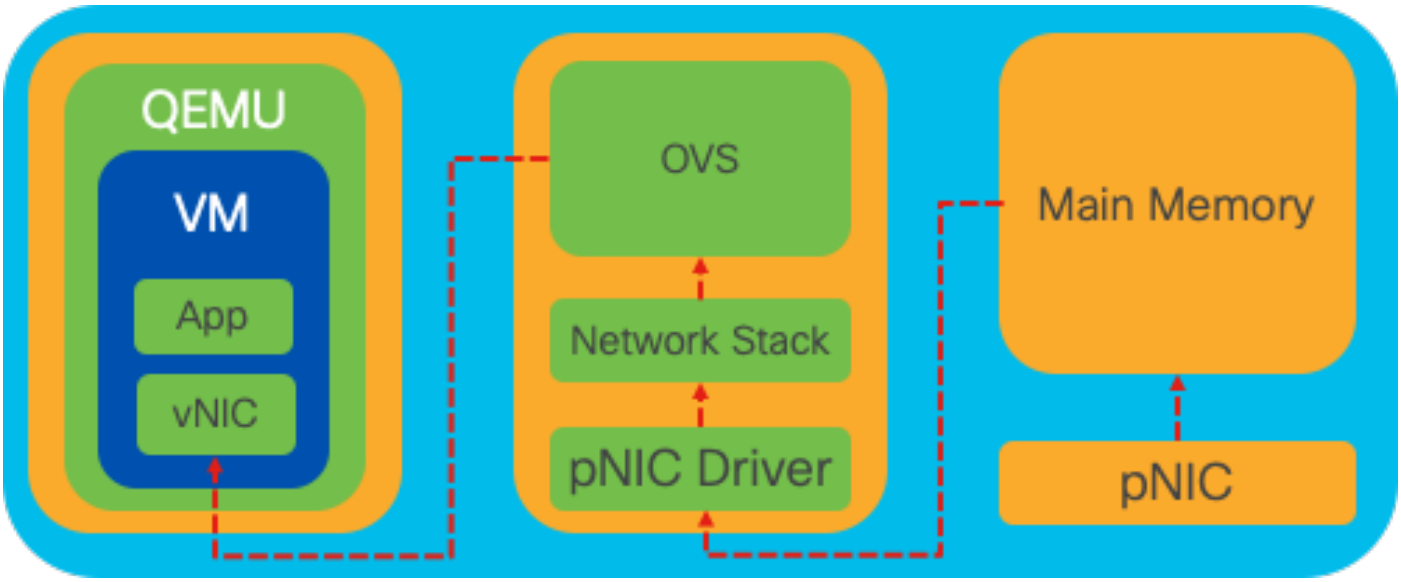


Figura 4. Configuração OVS dentro do kernel Linux

Bridges OVS

Ele usa bridges de rede virtual e regras de fluxo para encaminhar pacotes entre hosts. Ele se comporta como um switch físico, apenas virtualizado.

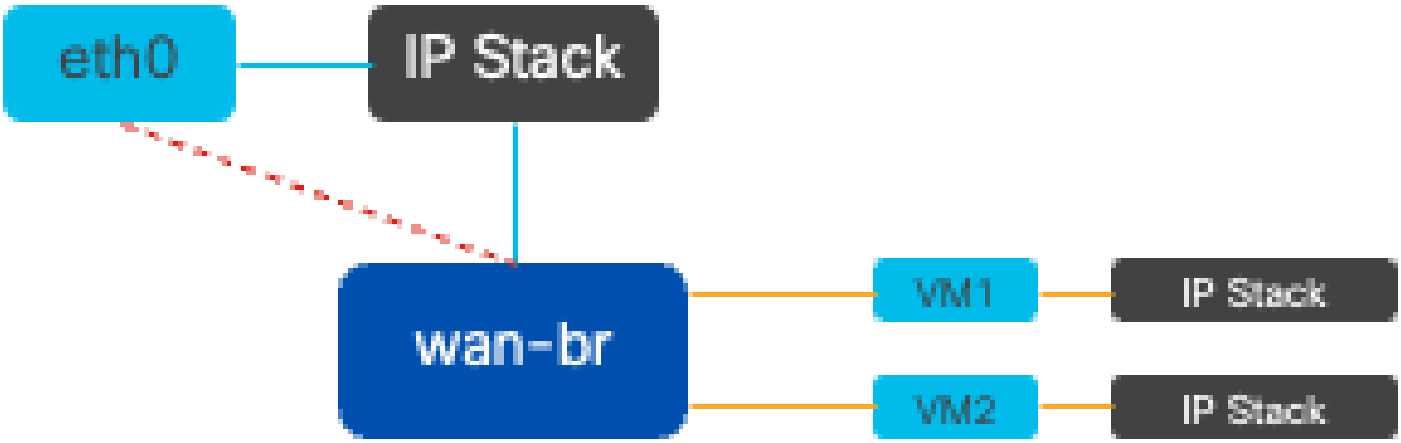


Figura 5. Exemplo de implementação de 2 VMs ou VNFs conectadas à bridge wan-br

Déficits de comutação de contexto

Quando um pacote de rede chega a uma placa de rede (NIC), ele dispara uma interrupção, um sinal para o processador indicando que precisa de atenção imediata. A CPU pausa suas tarefas atuais para lidar com a interrupção, um processo conhecido como processamento de interrupção. Durante essa fase, a CPU, sob o controle do kernel do sistema operacional, lê o pacote da placa de rede na memória e decide as próximas etapas com base no destino e na finalidade do pacote. O objetivo é processar ou rotear rapidamente o pacote para a aplicação pretendida, minimizando a latência e maximizando o throughput.

Comutação de contexto é o processo pelo qual uma CPU alterna a execução de tarefas em um ambiente (contexto) para outro. Isso é particularmente relevante quando se move entre o modo usuário e o modo kernel:

- Modo de usuário: este é um modo de processamento restrito no qual a maioria dos aplicativos é executada. Os aplicativos no modo usuário não têm acesso direto ao hardware ou à memória de referência e devem se comunicar com o kernel do sistema operacional para executar essas operações.

- Modo Kernel: este modo concede ao sistema operacional acesso total ao hardware e a toda a memória. O kernel pode executar qualquer instrução da CPU e fazer referência a qualquer endereço de memória. O modo kernel é necessário para a execução de tarefas como gerenciamento de dispositivos de hardware, memória e execução de chamadas do sistema.

Quando um aplicativo precisa executar uma operação que requer privilégios no nível do kernel (como a leitura de um pacote de rede), ocorre uma alternância de contexto. A CPU faz a transição do modo usuário para o modo kernel para executar a operação. Depois de concluído, outro switch de contexto retorna a CPU ao modo de usuário para continuar executando o aplicativo. Esse

processo de switching é essencial para manter a estabilidade e a segurança do sistema, mas introduz uma sobrecarga que pode afetar o desempenho.

O OVS é executado principalmente no espaço do usuário do sistema operacional, que pode se tornar um gargalo à medida que o throughput de dados aumenta. Isso ocorre porque são necessários mais switches de contexto para que a CPU mude para o modo kernel para processar pacotes, reduzindo o desempenho. Essa limitação é particularmente perceptível em ambientes com altas taxas de pacotes ou quando o tempo preciso é crucial. Para lidar com essas limitações de desempenho e atender às demandas de redes modernas e de alta velocidade, foram desenvolvidas tecnologias como DPDK (Data Plane Development Kit) e SR-IOV (Single Root I/O Virtualization).

Kit de desenvolvimento de plano de dados (DPDK)

O DPDK é um conjunto de bibliotecas e drivers projetados para acelerar as cargas de trabalho de processamento de pacotes em uma ampla variedade de arquiteturas de CPU. Ao ignorar a pilha de rede de kernel tradicional (evitando a comutação de contexto), o DPDK pode aumentar significativamente a taxa de transferência do plano de dados e reduzir a latência. Isso é particularmente benéfico para VNFs de alto throughput que exigem comunicação de baixa latência, tornando o NFVIS uma plataforma ideal para funções de rede sensíveis ao desempenho.

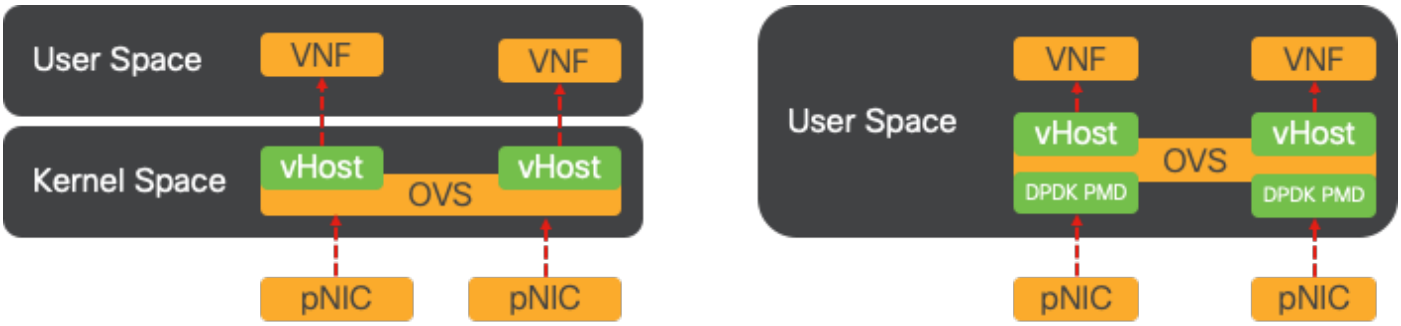


Figura 6. Otimizações tradicionais de comutação de contexto OVS (lado esquerdo) e DPDK OVS (lado direito)

O suporte para DPDK para OVS começou no NFVIS 3.10.1 para ENCS e 3.12.2 para outras plataformas.

- Taxa de transferência da cadeia de serviços perto do SRIOV, melhor que OVS não-DPDK.
- Driver de virtualização necessário para VNF.
- Plataformas suportadas:
- ENCS 3.10.1 em diante.
- UCSE, UCS-C, CSP5K 3.12.1 em diante.
- DPDK para canais de porta suportados desde 4.12.1.
- Captura de pacote/tráfego : sem suporte no DPDK.
- Tráfego de abrangência em PNIC: sem suporte em DPDK.
- Depois que o OVS-DPDK for habilitado, ele não poderá ser desabilitado como um recurso individual. A única maneira de desativar o DPDK seria uma redefinição de fábrica.

Cópia de dados

As abordagens de rede tradicionais frequentemente exigem que os dados sejam copiados várias vezes antes de chegar ao seu destino na memória da VM. Por exemplo, um pacote deve ser copiado da placa de rede para o espaço do kernel, depois para o espaço do usuário para processamento por um switch virtual (como OVS) e, finalmente, para a memória da VM. Cada operação de cópia incorre em um atraso e aumenta a utilização da CPU, apesar das melhorias de desempenho que o DPDK oferece, ignorando a pilha de rede de kernels.

Essas sobrecargas incluem cópias de memória e o tempo de processamento necessário para lidar com pacotes no espaço do usuário antes que eles possam ser encaminhados para a VM. A passagem PCIe e SR-IOV abordam esses gargalos, permitindo que um dispositivo de rede física (como uma placa de rede) seja compartilhado diretamente entre várias VMs sem envolver o sistema operacional do host na mesma extensão dos métodos de virtualização tradicionais.

Passagem PCIe

A estratégia envolve ignorar o hipervisor para permitir que as Virtual Network Functions (VNFs) acessem diretamente uma placa de rede (NIC), alcançando um throughput quase máximo. Essa abordagem é conhecida como passagem de PCI, que permite que uma placa de rede completa seja dedicada a um sistema operacional convidado sem a intervenção de um hipervisor. Nessa configuração, a máquina virtual opera como se estivesse diretamente conectada à placa de rede. Por exemplo, com duas placas de rede disponíveis, cada uma pode ser atribuída exclusivamente a diferentes VNFs, fornecendo-lhes acesso direto.

No entanto, esse método tem uma desvantagem: se apenas duas placas de rede estiverem disponíveis e forem usadas exclusivamente por duas VNFs separadas, qualquer VNF adicional, como uma terceira, ficaria sem acesso à placa de rede devido à falta de uma placa de rede dedicada disponível para ela. Uma solução alternativa envolve o uso de SR-IOV (Single Root I/O Virtualization, virtualização de E/S de raiz única).

Virtualização de E/S de Raiz Única (SR-IOV)

É uma especificação que permite que um único dispositivo PCI físico, como uma placa de rede, apareça como vários dispositivos virtuais separados. Essa tecnologia fornece acesso direto de VM a dispositivos de rede físicos, reduzindo a sobrecarga e melhorando o desempenho de E/S. Ele funciona dividindo um único dispositivo PCIe em várias fatias virtuais, cada uma atribuível a diferentes VMs ou VNFs, resolvendo efetivamente a limitação causada por um número finito de NICs. Essas fatias virtuais, conhecidas como Funções Virtuais (VFs), permitem recursos de NIC compartilhados entre várias VNFs. A função física (PF) refere-se ao componente físico real que facilita os recursos de SR-IOV.

Aproveitando o SR-IOV, o NFVIS pode alocar recursos de NIC dedicados para VNFs específicos, garantindo alto desempenho e baixa latência, facilitando o acesso direto à memória (DMA) dos pacotes de rede diretamente na respectiva memória da VM. Essa abordagem minimiza o envolvimento da CPU para apenas processar os pacotes, reduzindo assim o uso da CPU. Isso é especialmente útil para aplicativos que exigem largura de banda garantida ou têm requisitos de

desempenho rigorosos.

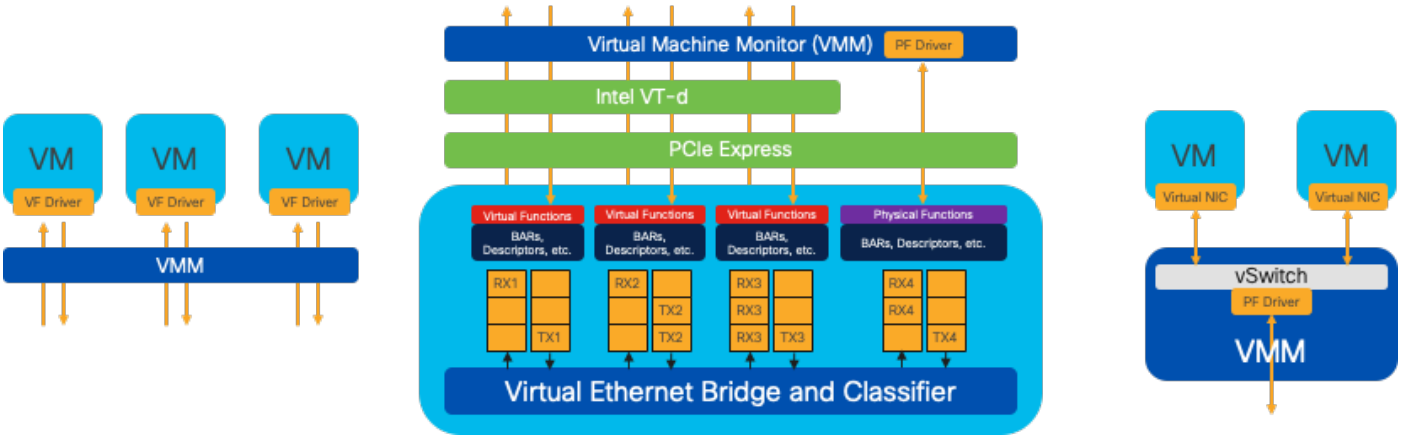


Figura 7. Separação de recursos de PCIe NFVIS SR-IOV por meio de funções de hardware

Funções físicas (PFs)

Elas são funções de PCIe completas e se referem a uma caixa de hardware desenvolvida especificamente que fornece uma função de rede específica; essas são funções de PCIe completas que podem ser descobertas, gerenciadas e manipuladas como qualquer outro dispositivo PCIe. As funções físicas incluem os recursos SR-IOV que podem ser usados para configurar e controlar um dispositivo PCIe.

Funções virtuais (VFs)

São funções otimizadas com recursos mínimos de configuração (lightweight), concentrando-se exclusivamente no processamento de E/S como funções PCIe simples. Cada função virtual se origina de uma função física. O hardware do dispositivo limita o número possível de funções virtuais. Uma porta Ethernet, o Dispositivo físico, pode corresponder a várias Funções virtuais, que podem ser alocadas para diferentes máquinas virtuais.

Drivers recomendados para aceleração de SR-IOV em hardware compatível com NFVIS

Platform	NIC(s)	Driver NIC
ENCS 54XX	Switch de backplane	i40e
ENCS 54XX	GE0-0 e GE0-1	IGB
UCPE Catalyst 8200	GE0-0 e GE0-1	ixgbe
UCPE Catalyst 8200	GE0-2 e GE0-5	IGB

Casos de uso para DPDK e SR-IOV

Preferência de DPDK

Particularmente em cenários onde o tráfego de rede flui principalmente de leste para oeste (o que

significa que ele permanece dentro do mesmo servidor), o DPDK supera o SR-IOV. O raciocínio é simples: quando o tráfego é gerenciado internamente no servidor sem a necessidade de acessar a placa de rede, o SR-IOV não oferece nenhum benefício. Na verdade, o SR-IOV pode potencialmente levar a ineficiências, estendendo desnecessariamente o caminho do tráfego e consumindo os recursos da placa de rede. Portanto, para o gerenciamento de tráfego de servidor interno, aproveitar o DPDK é a escolha mais eficiente.

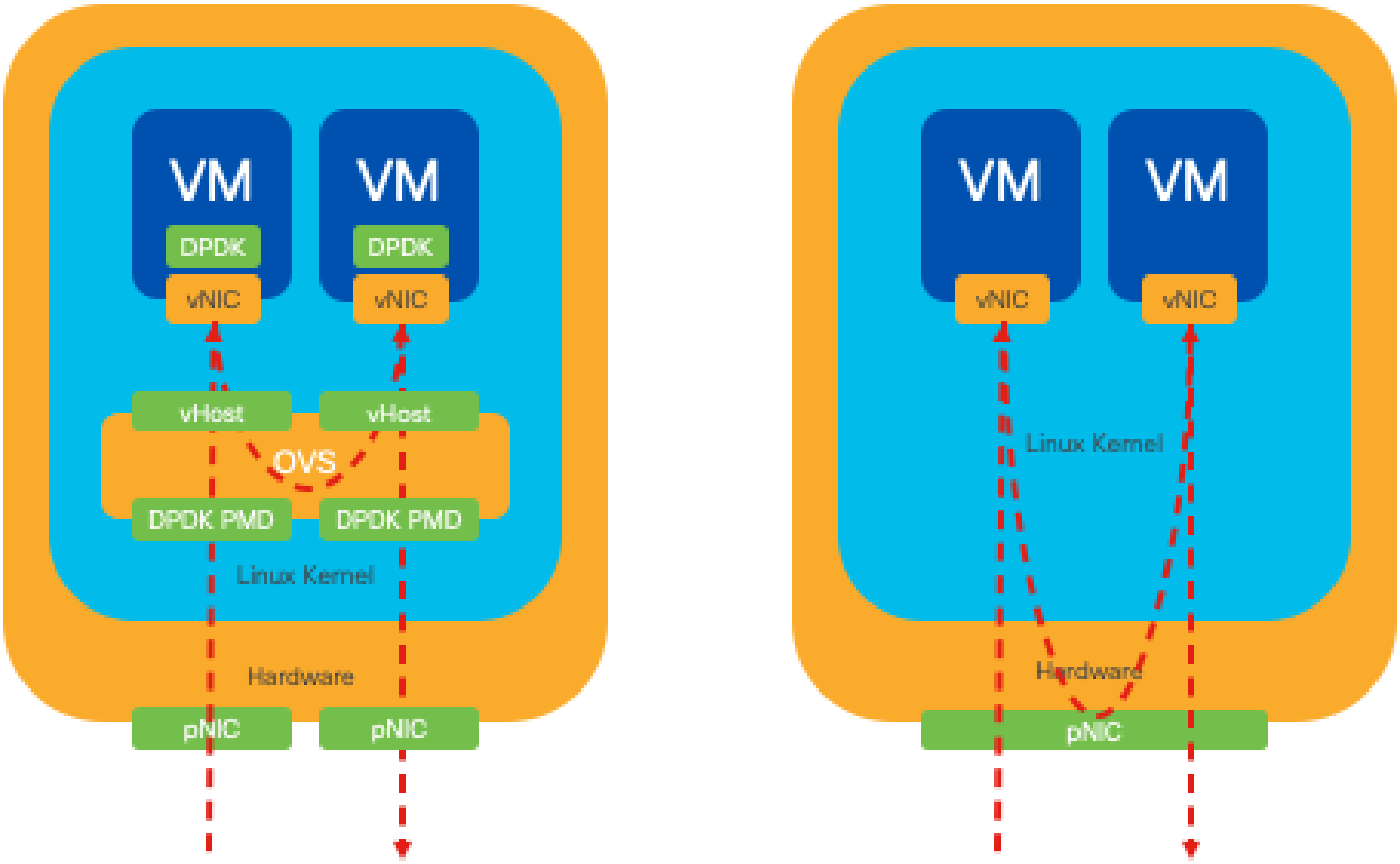


Figura 8. Pacote DPDK e SR-IOV transversal no tráfego leste-oeste

Preferência de SR-IOV

Em situações onde o tráfego de rede flui de norte a sul, ou até mesmo de leste a oeste, mas especificamente entre servidores, o uso de SR-IOV se mostra mais vantajoso que o DPDK. Isso é particularmente verdadeiro para a comunicação de servidor para servidor. Como esse tráfego inevitavelmente tem que atravessar a placa de rede, optar por OVS com DPDK avançado pode introduzir desnecessariamente complexidade adicional e possíveis restrições de desempenho. Portanto, o SR-IOV surge como a escolha preferível nessas circunstâncias, oferecendo um caminho direto e eficiente para lidar com o tráfego entre servidores.

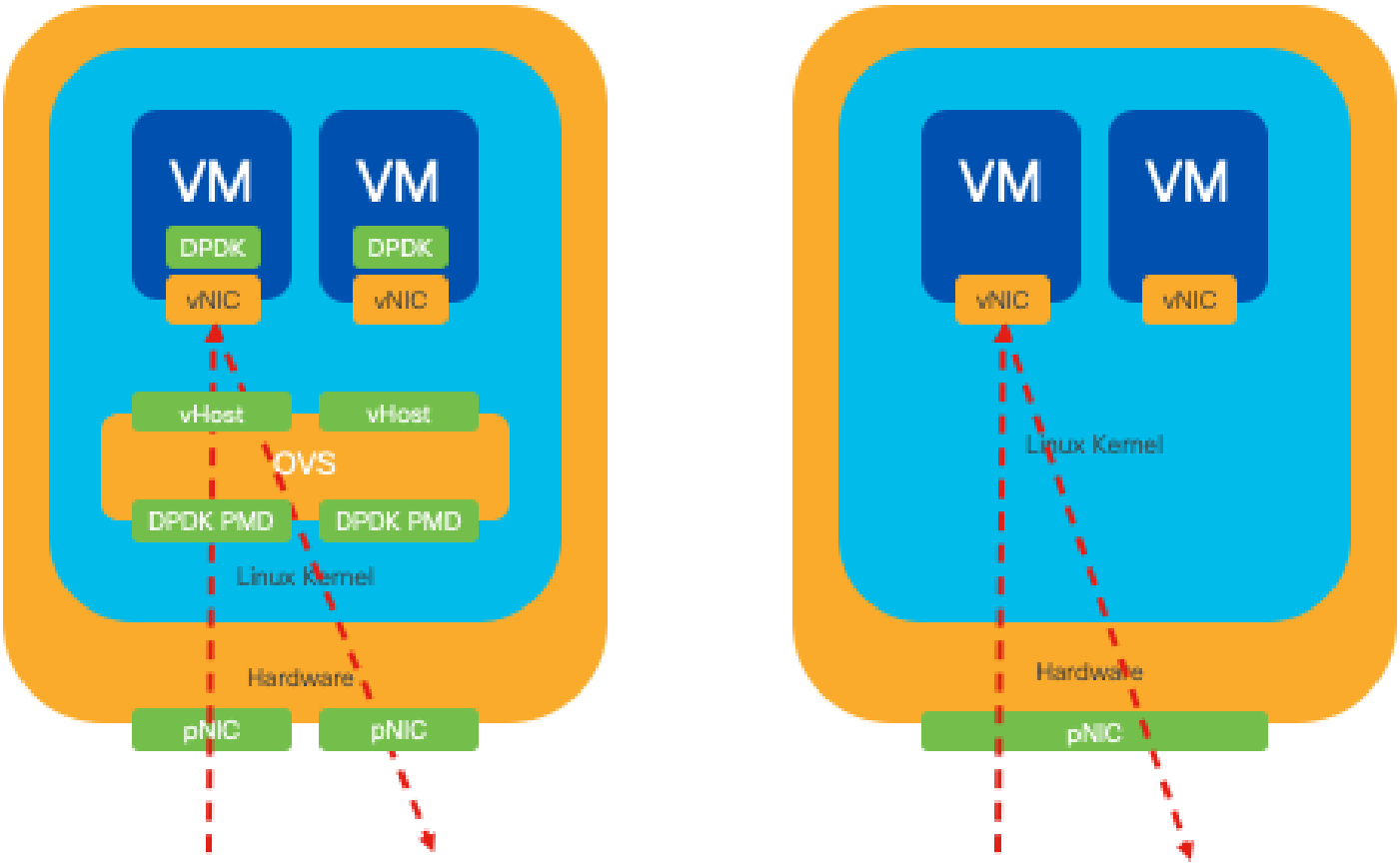
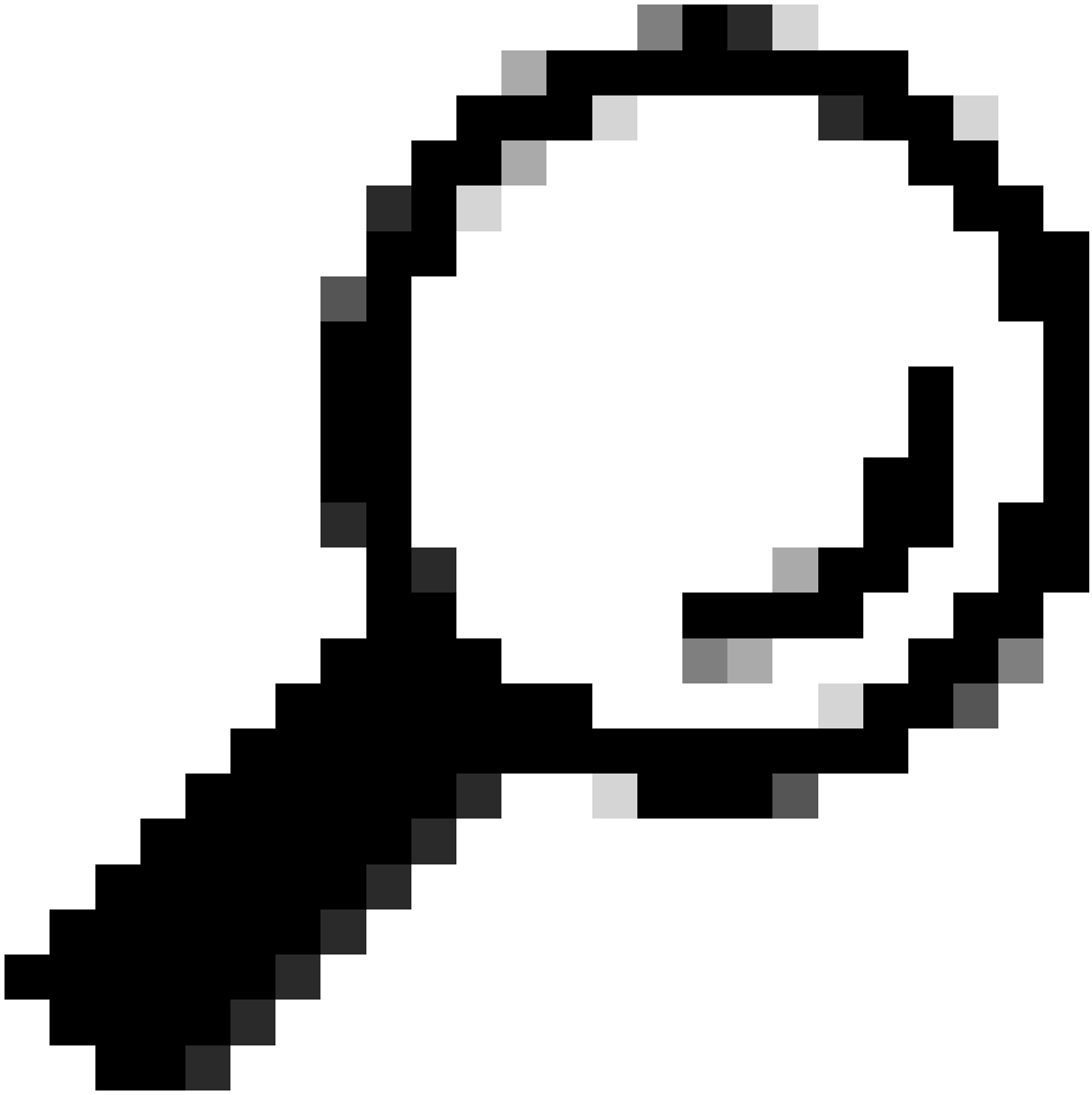


Figura 9. Tráfego de pacotes DPDK e SR-IOV no tráfego Norte-Sul



Dica: lembre-se de que é possível melhorar o desempenho de uma configuração baseada em SR-IOV integrando SR-IOV com DPDK em uma Virtual Network Function (VNF), excluindo o cenário em que o DPDK é usado em conjunto com OVS, conforme descrito anteriormente.

Configuração

Ativação do DPDK

Para ativar o DPDK na GUI, você deve navegar para Configuration > Virtual Machine > Networking > Networks. Quando estiver no menu, clique no interruptor para ativar o recurso

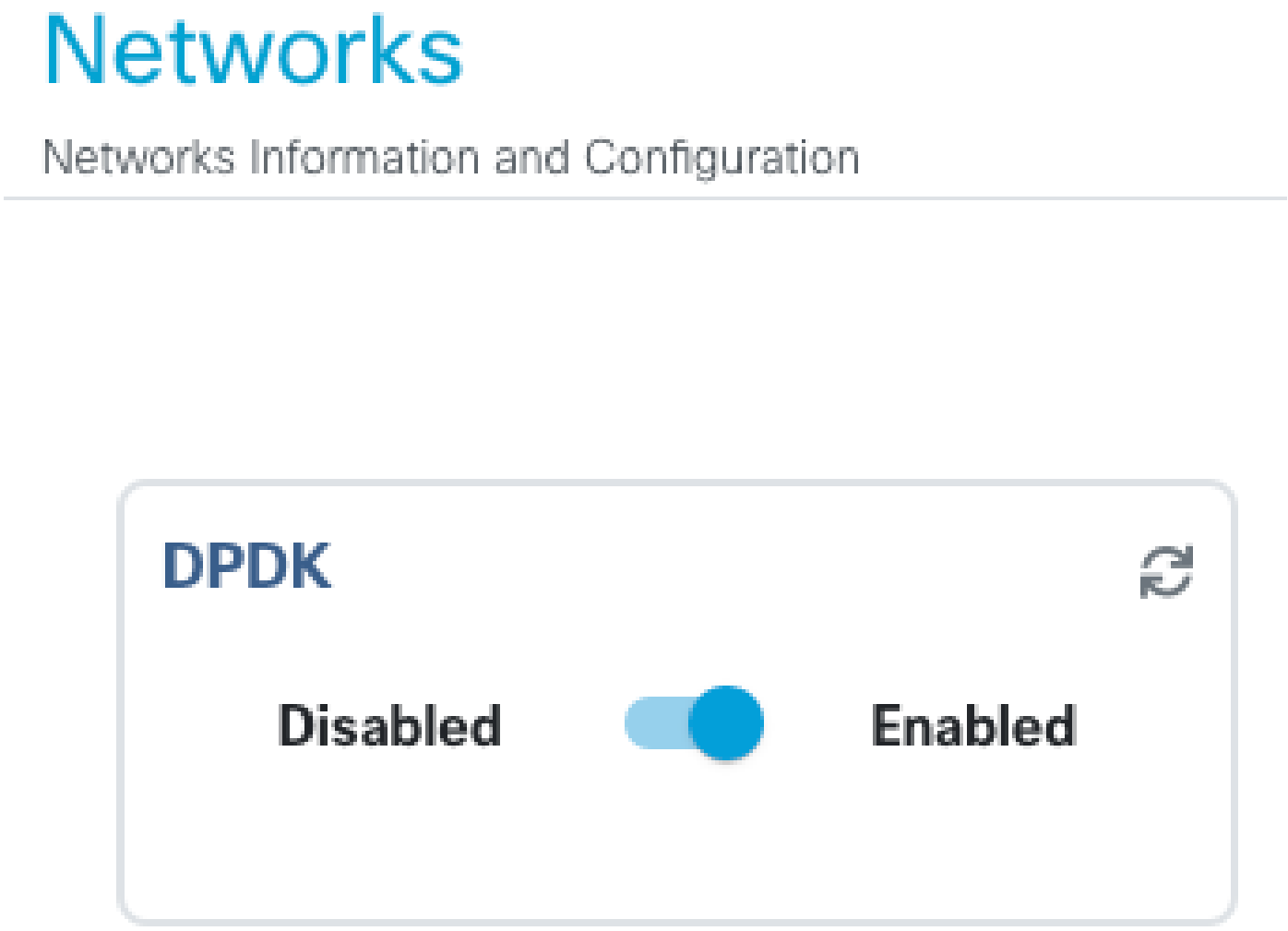


Figura 10. Botão deslizante disponível na GUI para ativação do DPDK

Para a CLI, você deve ativá-la nas configurações globais do sistema no modo de configuração.

```
nfvis(config)# system settings dpdk enable
```

Cuidado: o DPDK não pode ser desabilitado a menos que uma redefinição de fábrica seja executada a partir do NFVIS.

Criar uma nova rede e associá-la a uma nova ponte OVS

Navegue até Configuration > Virtual Machine > Networking > Networks. Quando estiver na página Redes, clique no sinal de mais à esquerda (+) da tabela Redes,

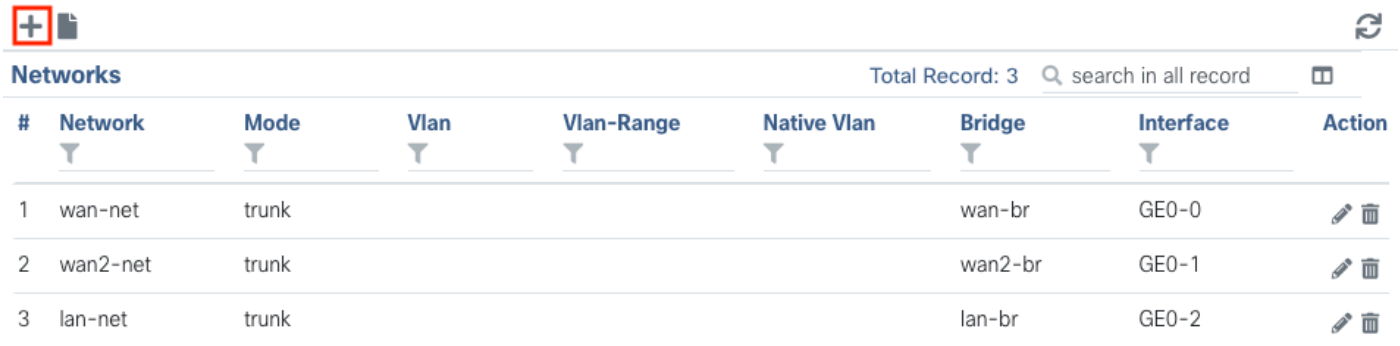


Figura 11. Visualização da tabela de redes na GUI do NFVIS

Nomear a rede e associá-la a uma nova bridge. As opções de vinculação de VLAN e interface podem depender das necessidades da infraestrutura de rede.

Add Network

Network *

inter-vnf-net

Mode *

trunk

Vlan

Vlan-Range

Native Vlan

1

Bridge *

○ Existing ◉ Create New

Bridge

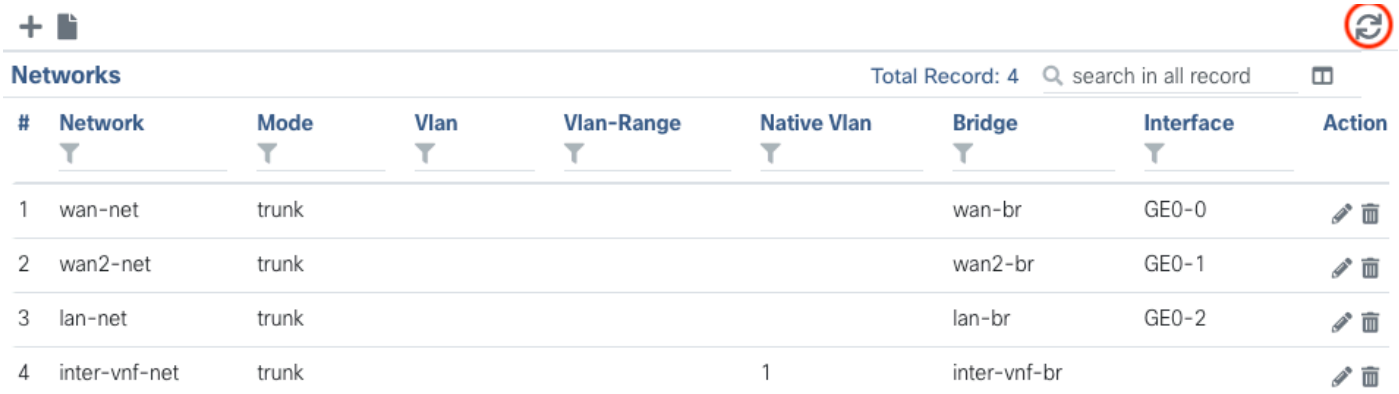
inter-vnf-br

Interface

Submit Cancel Reset

Figura 12. Modal "Add Network" (Adicionar rede) para criação de redes virtuais na GUI do NFVIS

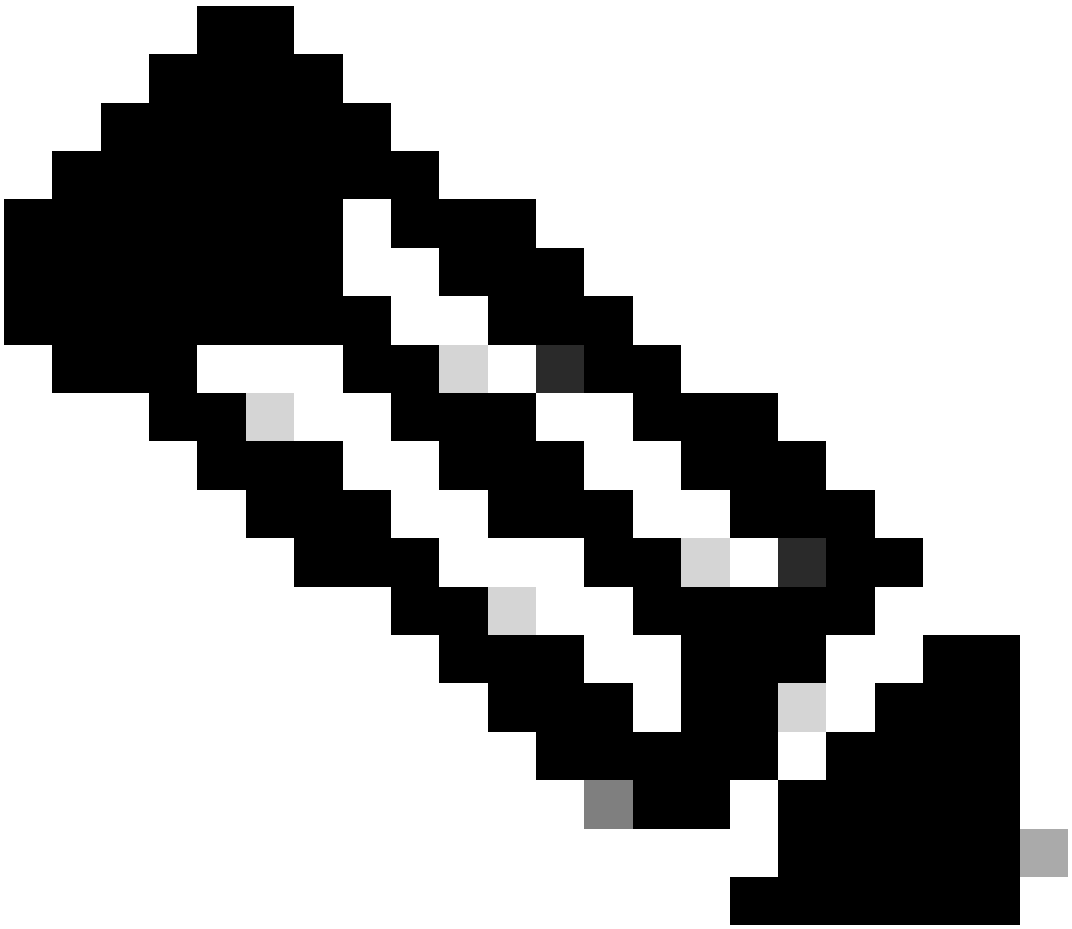
Depois de clicar no botão submit, você deverá ser capaz de revisar a rede recém-criada anexada à tabela Networks.



Networks Total Record: 4 search in all record

#	Network	Mode	Vlan	Vlan-Range	Native Vlan	Bridge	Interface	Action
1	wan-net	trunk				wan-br	GE0-0	
2	wan2-net	trunk				wan2-br	GE0-1	
3	lan-net	trunk				lan-br	GE0-2	
4	inter-vnf-net	trunk			1	inter-vnf-br		

Figura 13. Visualização da tabela de redes na GUI do NFVIS, onde o "ícone de atualização" está no canto superior direito (destacado em vermelho)



Observação: se a nova rede não for observada na tabela, clique no botão de atualização superior direito ou atualize a página inteira.

Se executado na CLI, toda rede e ponte são criadas a partir do modo de configuração, o fluxo de trabalho é o mesmo da versão da GUI.

1. Crie a nova ponte.

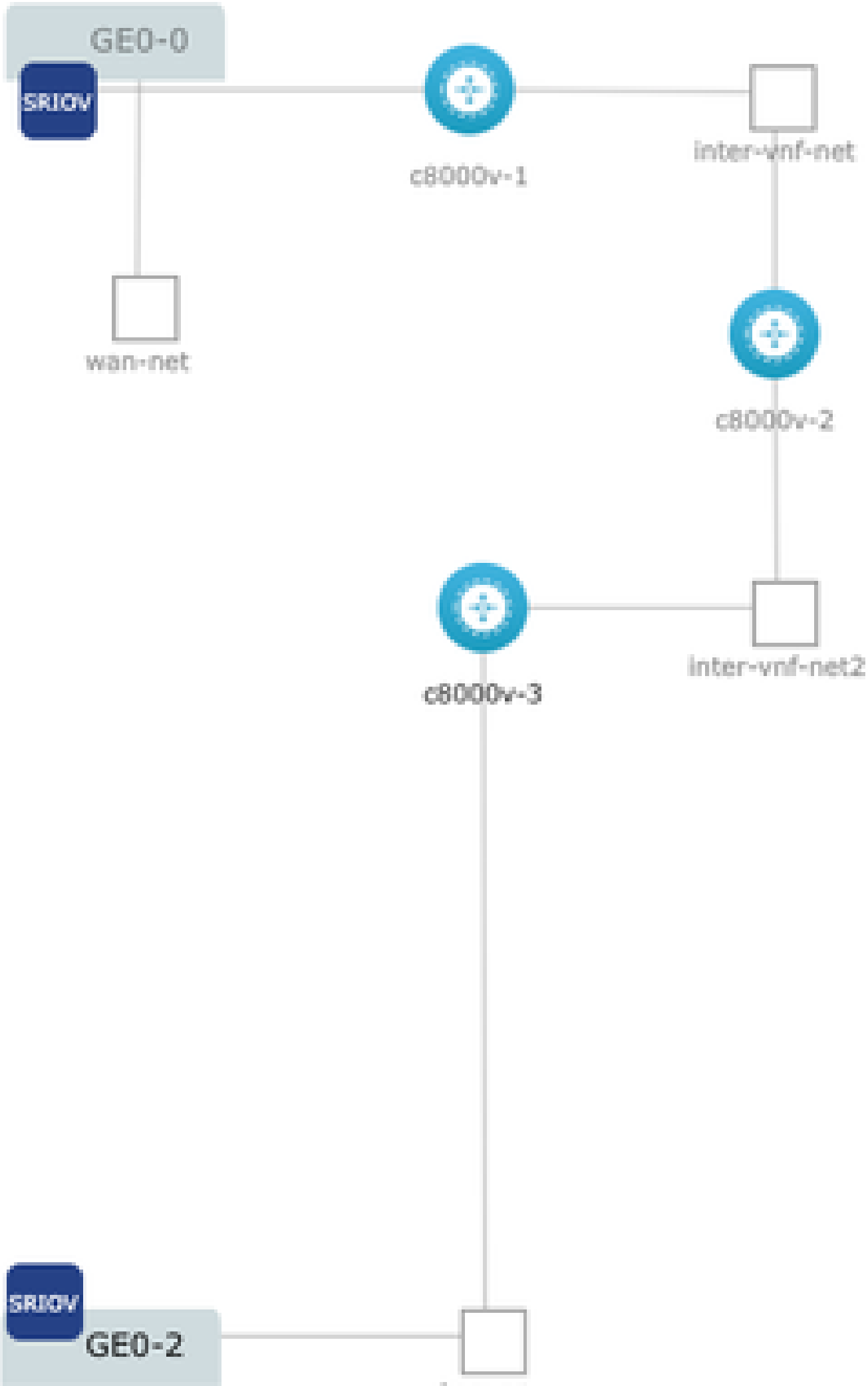
```
nfvis(config)# bridges bridge inter-vnf-br2
nfvis(config-bridge-inter-vnf-br2)# commit
```

2. Crie uma nova Rede e associe-a à bridge criada anteriormente

```
nfvis(config)# networks network inter-vnf-net2 bridge inter-vnf-br2 trunk true native-vlan 1
nfvis(config-network-inter-vnf-net2)# commit
```

Conexão de VNFs

Para começar com uma topologia de rede ou uma única implantação de VFN, você deve navegar para Configuration > Deploy. Você pode arrastar uma VM ou um contêiner da lista de seleção para a área de criação de topologia para começar a criar sua infraestrutura virtualizada.


GE0-0
SRIOV
c8000v-1
inter-vnf-net
wan-net
c8000v-2
c8000v-3
inter-vnf-net2
SRIOV
GE0-2

Sobre esta tradução
A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês (link fornecido) seja sempre consultado.