

NFVIS仮想ネットワークについて：OVS、DPDK、およびSR-IOV

内容

[はじめに](#)

[使用するコンポーネント](#)

[NFVISにおけるネットワーキングの概要](#)

[ENCS54XXプラットフォーム](#)

[Catalyst 8200 uCPE](#)

[Catalyst 8300 uCPE 1N20](#)

[ネットワーキング仮想化テクノロジー](#)

[オープンvSwitch\(OVS\)](#)

[OVSブリッジ](#)

[コンテキストスイッチングの障害](#)

[データプレーン開発キット\(DPDK\)](#)

[データのコピー](#)

[PCIeバススルー](#)

[シングルルートI/O仮想化\(SR-IOV\)](#)

[物理機能\(PF\)](#)

[仮想関数\(VF\)](#)

[NFVIS対応ハードウェアでのSR-IOV促進に推奨される促進要因](#)

[DPDKとSR-IOVの使用例](#)

[DPDKプリファレンス](#)

[SR-IOVプリファレンス](#)

[コンフィギュレーション](#)

[DPDKの有効化](#)

[新しいネットワークを作成して新しいOVSブリッジに関連付ける](#)

[VNFの接続](#)

[関連記事およびドキュメント](#)

はじめに

このドキュメントでは、NFVISプラットフォームがエンタープライズネットワークおよびサービスネットワークでVNFと通信するために提供する仮想ネットワーキング方式について説明します。

使用するコンポーネント

このドキュメントの情報は、次のハードウェアおよびソフトウェア コンポーネントに基づくものです。

- NFVIS 4.7.1-FC4を実行するENC5412
- nfvis 4.12.1-FC2を実行するc8300 uCPE 1N20

このドキュメントの情報は、特定のラボ環境にあるデバイスに基づいて作成されました。このドキュメントで使用するすべてのデバイスは、クリアな（デフォルト）設定で作業を開始していません。本稼働中のネットワークでは、各コマンドによって起こる可能性がある影響を十分確認してください。

NFVISにおけるネットワークングの概要

内部管理ネットワーク(int-mgmt-net)とブリッジ(int-mgmt-br)は、VNFモニタリングのために内部で使用され、10.20.0.0/24サブネットから管理IPアドレスを割り当てます。

ENC54XXプラットフォーム

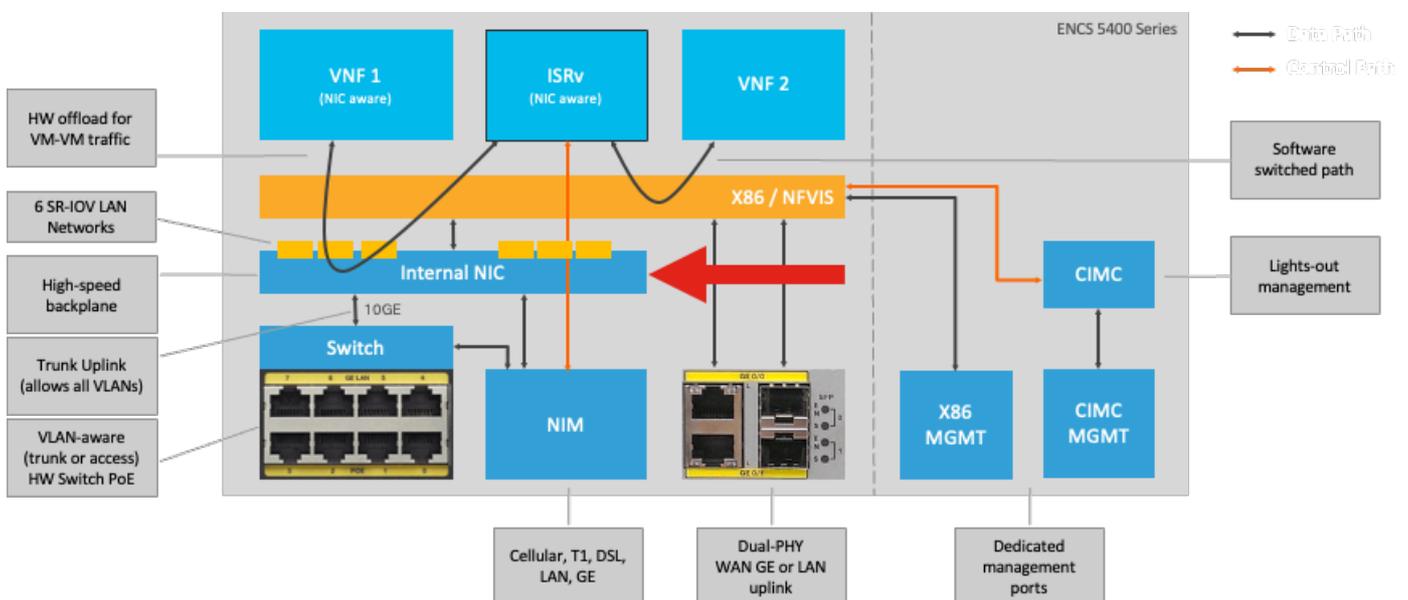


図 1.ハードウェアスイッチとWAN/LANアップリンクNICの内部接続

Catalyst 8200 uCPE

- NFVISには、デフォルトで、管理用のWANポートまたはGE0/2 LANポートを介してアクセスできます。
- WANネットワーク (wan-netおよびwan2-net) とWANブリッジ (wan-brおよびwan2-br) は、デフォルトでDHCPを有効にするように設定されています。デフォルトでは、GE0-0はWANブリッジに関連付けられ、GE0-1はWAN2ブリッジに関連付けられます。
- Catalyst 8200 UCPEの管理IPアドレス192.168.1.1は、GE0-2を介してアクセス可能です。
- GE0-2はLANブリッジに関連付けられています。
- 内部管理ネットワーク(int-mgmt-net)とブリッジ(int-mgmt-br)が作成され、システムモニタリング用に内部で使用されます。

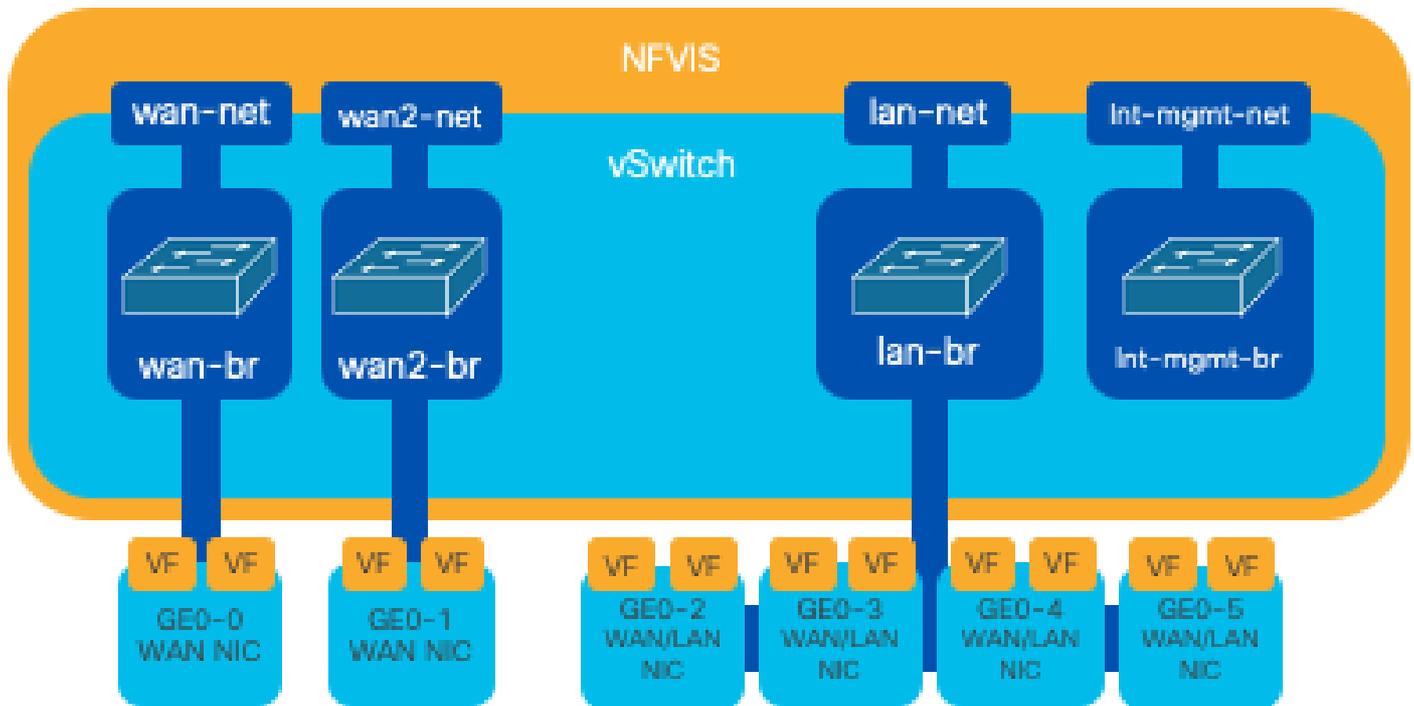


図 2 : 8200 NICに割り当てられた内部ブリッジングおよび仮想スイッチ

Catalyst 8300 uCPE 1N20

1. NFVISには、管理用にFPGE (前面パネルギガビットイーサネット) WANポートまたはGE0-2 LANポートを介してデフォルトでアクセスできます
2. WANネットワーク(wan-net)とWANブリッジ(wan-br)は、デフォルトでDHCPを有効にするように設定されています。GE0-0は、デフォルトでWANブリッジに関連付けられます
3. WANネットワーク(wan2-net)とWANブリッジ(wan2-br)はデフォルトで作成されますが、物理ポートには関連付けられません
4. GE0-2はLANブリッジに関連付けられており、他のすべてのポートはOVSに関連付けられていません。
5. C8300-uCPEの管理IP 192.168.1.1には、GE0-2を介してアクセスできます
6. 内部管理ネットワーク(int-mgmt-net)とブリッジ(int-mgmt-br)が作成され、システムモニタリング用に内部で使用されます。

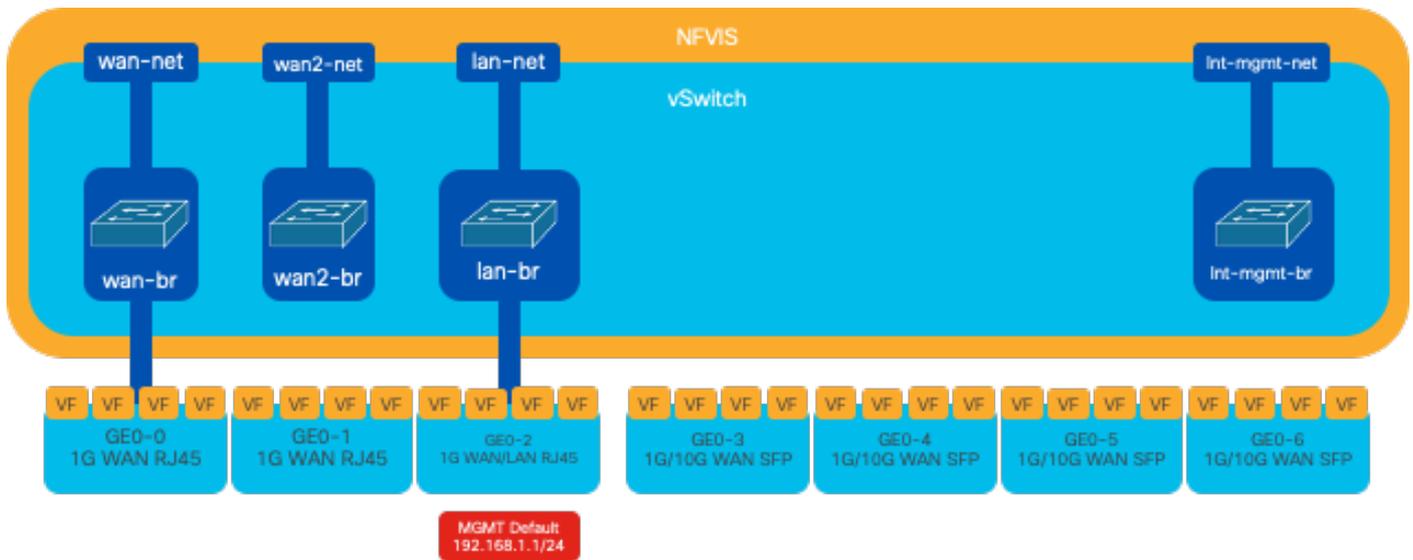


図 3： 8300 NICに割り当てられた内部ブリッジおよび仮想スイッチ

ネットワーキング仮想化テクノロジー

オープンvSwitch(OVS)

Open vSwitch(OVS)は、オープンソースのマルチレイヤ仮想スイッチであり、プログラムによる拡張を通じてネットワークの自動化を実現すると同時に、NetFlow、sFlow、IPFIX、RSPAN、CLI、LACP、802.1agなどの標準的な管理インターフェイスおよびプロトコルをサポートします。大規模な仮想化環境で広く使用されており、特にハイパーバイザで仮想マシン(VM)間のネットワークトラフィックを管理するために使用されます。NFVISインターフェイスを通じて直接管理される高度なネットワークトポロジとポリシーの作成が可能になり、ネットワーク機能の仮想化に適した汎用的な環境が提供されます。

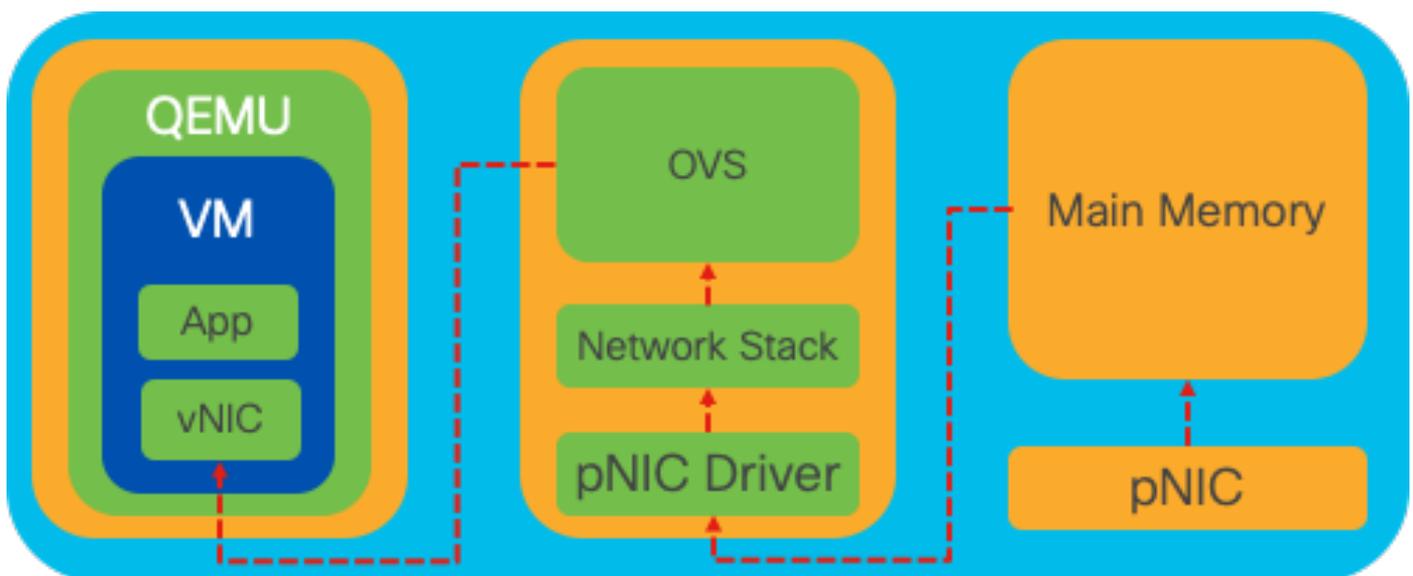


図 4： Linuxカーネル内でのOVS設定

OVSブリッジ

仮想ネットワークブリッジを使用し、ホスト間でパケットを転送するルールを流します。物理スイッチのように動作し、仮想化されるだけです。

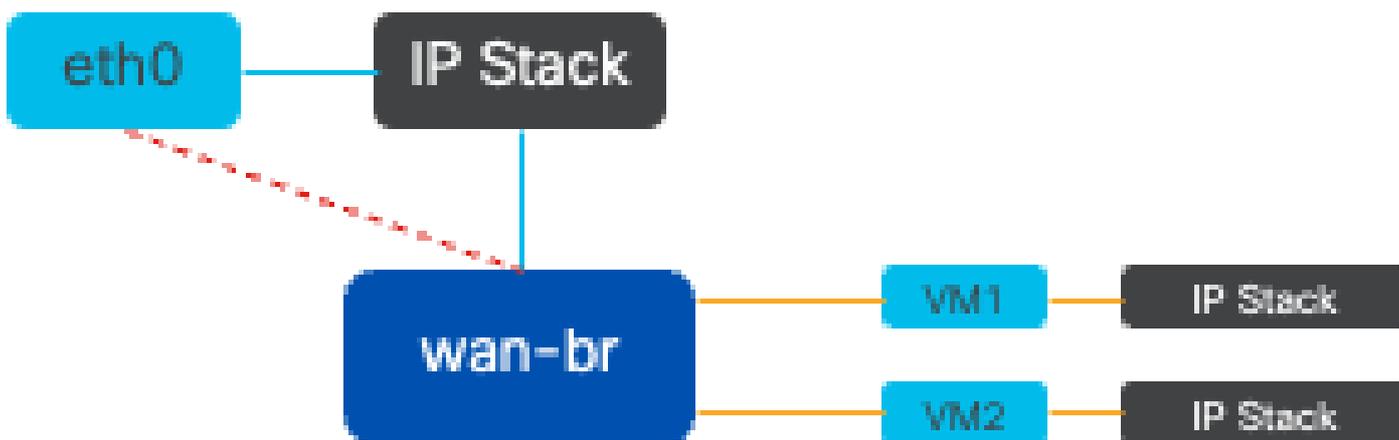


図 5.wan-brブリッジに接続された2つのVMまたはVNFの実装例

コンテキストスイッチングの障害

ネットワークパケットがネットワークインターフェイスカード(NIC)に到着すると、プロセッサに対して割り込みをトリガーし、プロセッサが即座に対応を必要としていることを示す信号を送信します。CPUは現在のタスクを一時停止して、割り込みを処理します。このプロセスは割り込み処理と呼ばれます。このフェーズでは、オペレーティングシステムカーネルの制御下にあるCPUが、NICからメモリにパケットを読み込み、パケットの宛先と目的に基づいて次のステップを決定します。目標は、目的のアプリケーションにパケットを迅速に処理またはルーティングし、遅延を最小限に抑えてスループットを最大化することです。

コンテキストスイッチングは、ある環境(コンテキスト)で実行されているタスクを別の環境(コンテキスト)に切り替えるプロセスです。これは、ユーザモードとカーネルモードの間を移動する場合に特に重要です。

- ユーザモード：これは、ほとんどのアプリケーションが実行される制限処理モードです。ユーザモードのアプリケーションは、ハードウェアや参照メモリに直接アクセスすることはできず、これらの操作を実行するにはオペレーティングシステムのカーネルと通信する必要があります。
- カーネルモード：オペレーティングシステムにハードウェアとすべてのメモリへのフルアクセスを許可します。カーネルは任意のCPU命令を実行し、任意のメモリアドレスを参照できます。カーネルモードは、ハードウェアデバイスの管理、メモリ、システムコールの実行などのタスクを実行するために必要です。

アプリケーションがカーネルレベルの特権を必要とする操作(ネットワークパケットの読み取りなど)を実行する必要がある場合、コンテキストスイッチが発生します。CPUはユーザモードからカーネルモードに移行して動作を実行します。完了すると、別のコンテキストスイッチがCPUをユーザモードに戻し、アプリケーションの実行を続行します。このスイッチングプロセスは、システムの安定性とセキュリティを維持するために重要ですが、パフォーマンスに影響を与える可能性のあるオーバーヘッドが発生します。

OVSは主にオペレーティングシステムのユーザ領域で実行され、データのスループットが増加するとボトルネックになる可能性があります。これは、パケットを処理するためにCPUをカーネルモードに移行するには、より多くのコンテキストスイッチが必要になり、パフォーマンスが低下するためです。この制限は、パケットレートが高い環境や正確なタイミングが重要な場合に特に顕著です。これらのパフォーマンスの制限に対処し、最新の高速ネットワークの要求を満たすために、DPDK (データプレーン開発キット) やSR-IOV (シングルルートI/O仮想化) などのテクノロジーが開発されました。

データプレーン開発キット(DPDK)

DPDKは、幅広いCPUアーキテクチャでパケット処理のワークロードを高速化するために設計されたライブラリとドライバのセットです。従来のカーネルネットワーキングスタックをバイパスすることで (コンテキストスイッチングを回避)、DPDKはデータプレーンのスループットを大幅に向上させ、遅延を削減できます。これは、低遅延通信を必要とする高スループットのVNFに特に有効で、パフォーマンスの影響を受けやすいネットワーク機能の理想的なプラットフォームとなります。

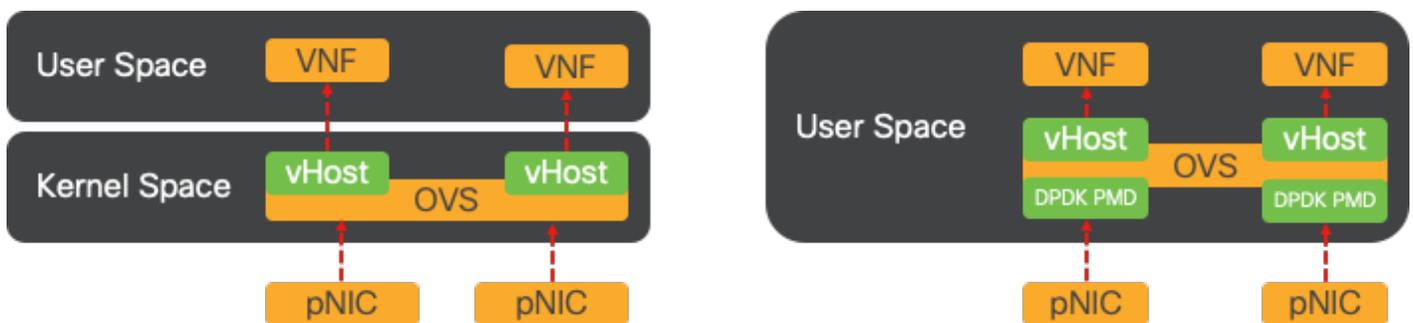


図 6.従来のOVS (左側) およびDPDK OVS (右側) コンテキストスイッチングの最適化

OVS用のDPDKのサポートは、ENCs用のNFVIS 3.10.1およびその他のプラットフォーム用の3.12.2で開始されました。

- SRIOVに近いサービスチェーンスループット、DPDK以外のOVSよりも優れている
- VNFに必要な仮想ドライバ。
- 対応プラットフォーム:
 - ENCS 3.10.1以降
 - UCSE、UCS-C、CSP5K 3.12.1以降
- ポートチャネルのDPDKは4.12.1からサポートされています。
- パケット/トラフィックキャプチャ：DPDKではサポートされていません。
- PNICでのトラフィックのスパン：DPDKではサポートされていません。
- OVS-DPDKを有効にした後は、個別の機能として無効にすることはできません。DPDKを無効にする唯一の方法は、工場出荷時の状態にリセットすることです。

データのコピー

従来のネットワーキングアプローチでは、VMメモリ内の宛先に到達する前に、データを何度もコピーする必要があります。たとえば、パケットはNICからカーネル空間にコピーされた後、仮想スイッチ (OVSなど) で処理するためにユーザ空間にコピーされ、最後にVMメモリにコピーさ

れます。各コピー操作は、カーネルネットワーキングスタックをバイパスすることでDPDKが提供するパフォーマンスの向上にもかかわらず、遅延が発生し、CPU使用率が増加します。

これらのオーバーヘッドには、メモリコピーや、ユーザ空間のパケットをVMに転送する前に処理するのに必要な処理時間が含まれます。PCIeパススルーおよびSR-IOVは、ホストのオペレーティングシステムを従来の仮想化方法と同じ程度に使用せずに、物理ネットワークデバイス（NICなど）を複数のVM間で直接共有できるようにすることで、これらのボトルネックに対処します。

PCIeパススルー

この戦略では、ハイパーバイザをバイパスして、仮想ネットワーク機能(VNF)がネットワークインターフェイスカード(NIC)に直接アクセスできるようにし、ほぼ最大のスループットを実現します。このアプローチはPCIパススルーと呼ばれ、ハイパーバイザの介入なしに完全なNICをゲストオペレーティングシステム専用にすることができます。この設定では、仮想マシンはNICに直接接続されているかのように動作します。たとえば、2枚のNICカードが使用可能な場合、それぞれが異なるVNFに排他的に割り当てられ、直接アクセスが提供されます。

ただし、この方法には欠点があります。2つのNICだけが使用可能で、2つの別個のVNFによって排他的に使用される場合、使用可能な専用NICがないため、追加のVNF（3番目のVNFなど）はNICアクセスなしで残されます。代替ソリューションとして、シングルルートI/O仮想化(SR-IOV)を使用します。

シングルルートI/O仮想化(SR-IOV)

ネットワークインターフェイスカード(NIC)などの単一の物理PCIデバイスを複数の個別の仮想デバイスとして認識できるようにする仕様です。このテクノロジーは、物理ネットワークデバイスへの仮想マシンの直接アクセスを提供し、オーバーヘッドを削減してI/Oパフォーマンスを向上させます。単一のPCIデバイスを実数の仮想スライスに分割し、各スライスを異なるVMまたはVNFに割り当てることで、限られた数のNICによる制限を効果的に解決します。これらの仮想スライスは仮想機能(VF)と呼ばれ、複数のVNF間でNICリソースを共有できます。物理機能(PF)とは、SR-IOV機能を促進する実際の物理コンポーネントを指します。

SR-IOVを活用することで、NFVISは専用のNICリソースを特定のVNFに割り当て、ネットワークパケットのダイレクトメモリアクセス(DMA)を各VMメモリに直接容易にすることで、高パフォーマンスと低遅延を実現できます。このアプローチでは、パケットを処理するだけでCPUの影響を最小限に抑えるため、CPU使用率が低下します。これは、帯域幅の保証が必要なアプリケーションや、パフォーマンス要件が厳しいアプリケーションで特に役立ちます。

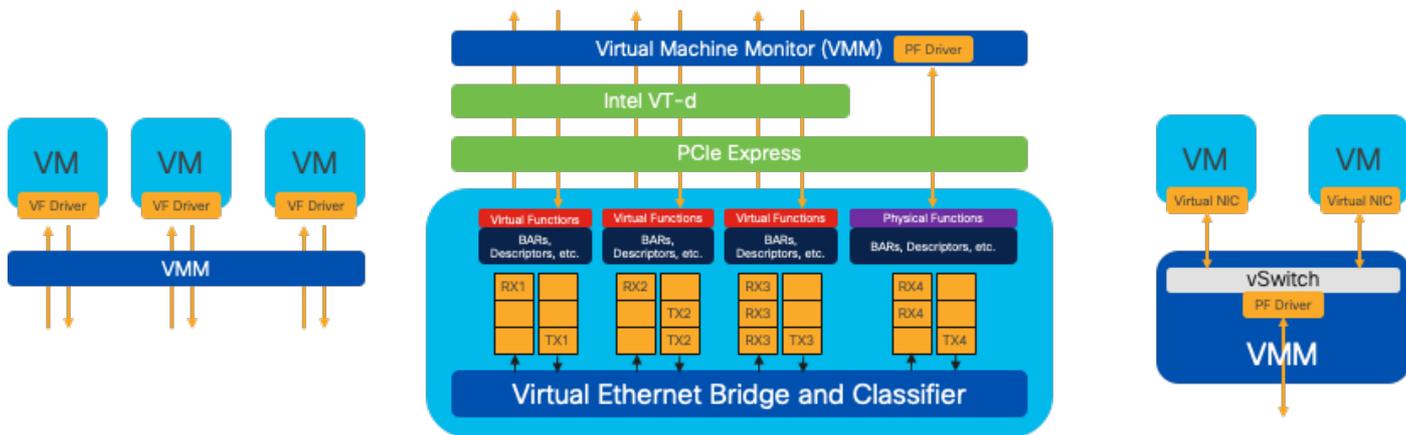


図 7NFVIS SR-IOV PCIeリソースのハードウェア機能による分離

物理機能(PF)

これらの機能はフル装備のPCIe機能であり、特定のネットワーク機能を提供する専用のハードウェアボックスを参照します。これらの機能はフル装備のPCIe機能であり、他のPCIeデバイスと同様に検出、管理、および操作できます。物理機能には、PCIeデバイスの設定と制御に使用できるSR-IOV機能が含まれます。

仮想関数(VF)

シンプルなPCIe機能としてのI/O処理のみに重点を置き、最小限の構成リソース（軽量）で機能を合理化します。すべての仮想関数は、物理関数から生成されます。デバイスのハードウェアによって、仮想機能の数が制限されます。1つのイーサネットポートである物理デバイスは、多数の仮想機能に対応できます。その後、それらの機能を異なる仮想マシンに割り当てることができます。

NFVIS対応ハードウェアでのSR-IOV促進に推奨される促進要因

Platform	NIC	NICドライバ
ENCS 54XX	バックプレーンスイッチ	i40e
ENCS 54XX	GE0-0およびGE0-1	igb
Catalyst 8200 uCPE	GE0-0およびGE0-1	ixgbe
Catalyst 8200 uCPE	GE0-2およびGE0-5	igb

DPDKとSR-IOVの使用例

DPDKプリファレンス

特に、ネットワークトラフィックが主にイースト/ウェスト（つまり、同じサーバ内に留まる）を流れるシナリオでは、DPDKがSR-IOVを上回ります。理論的根拠は単純です。NICにアクセスせずにサーバ内でトラフィックを管理しても、SR-IOVには利点がありません。実際、SR-IOVは、トラフィックパスを不必要に拡張し、NICリソースを消費するため、非効率性を招く可能性があります。したがって、内部サーバトラフィック管理では、DPDKを活用する方がより効率的な選

択です。

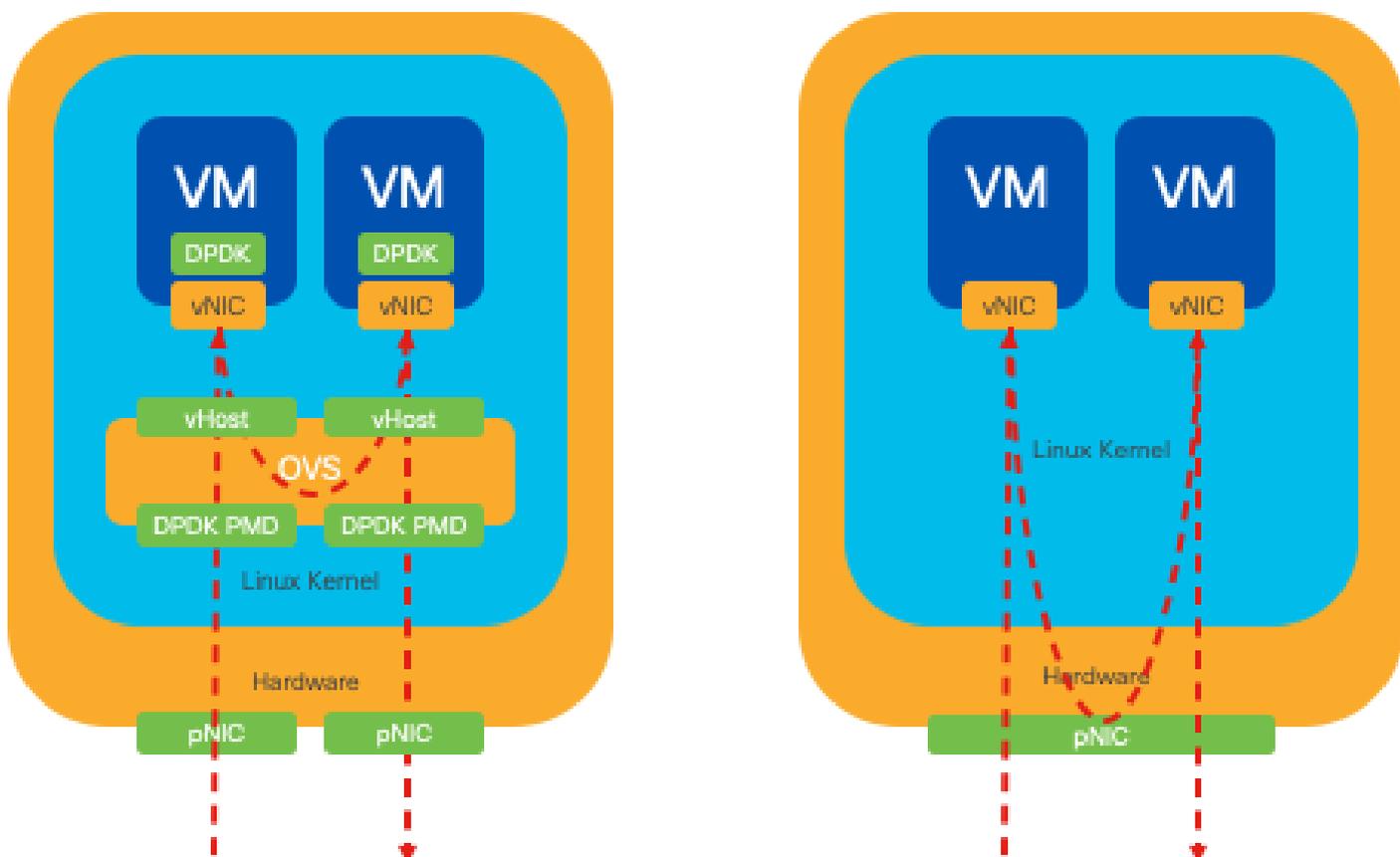


図 8.East-to-WestトラフィックでのDPDKおよびSR-IOVの packets トラバース

SR-IOVプリファレンス

ネットワークトラフィックが北から南へ、または東から西へ、特にサーバ間を流れる状況では、SR-IOVを使用する方がDPDKよりも有利です。これは特に、サーバ間通信に当てはまります。このようなトラフィックは必ずNICを通過する必要があるため、DPDK拡張OVSを選択すると、不必要に複雑さが増し、パフォーマンスの制約が発生する可能性があります。したがって、このような状況ではSR-IOVが望ましい選択肢となり、サーバ間トラフィックを処理するための簡単で効率的な経路が提供されます。

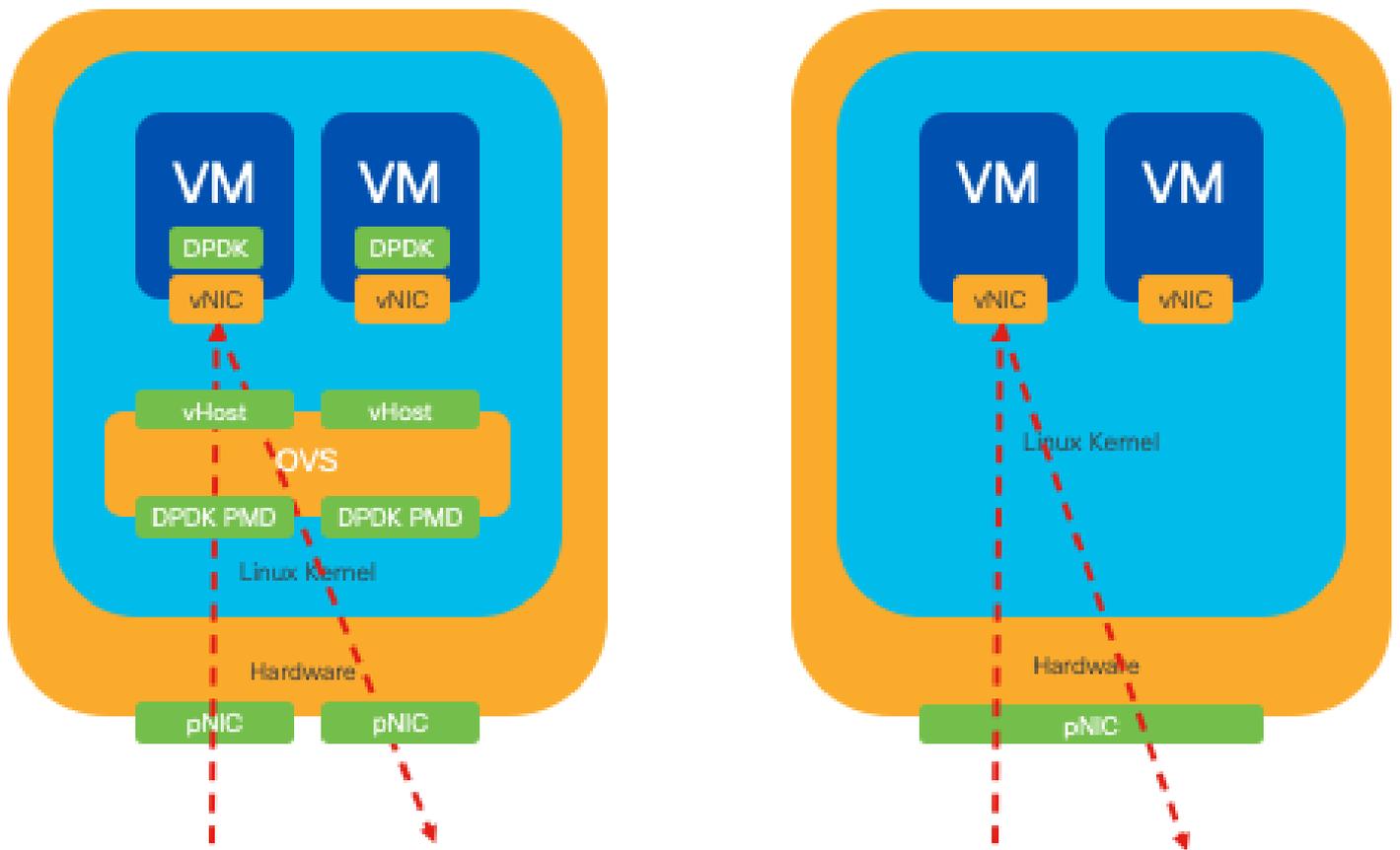
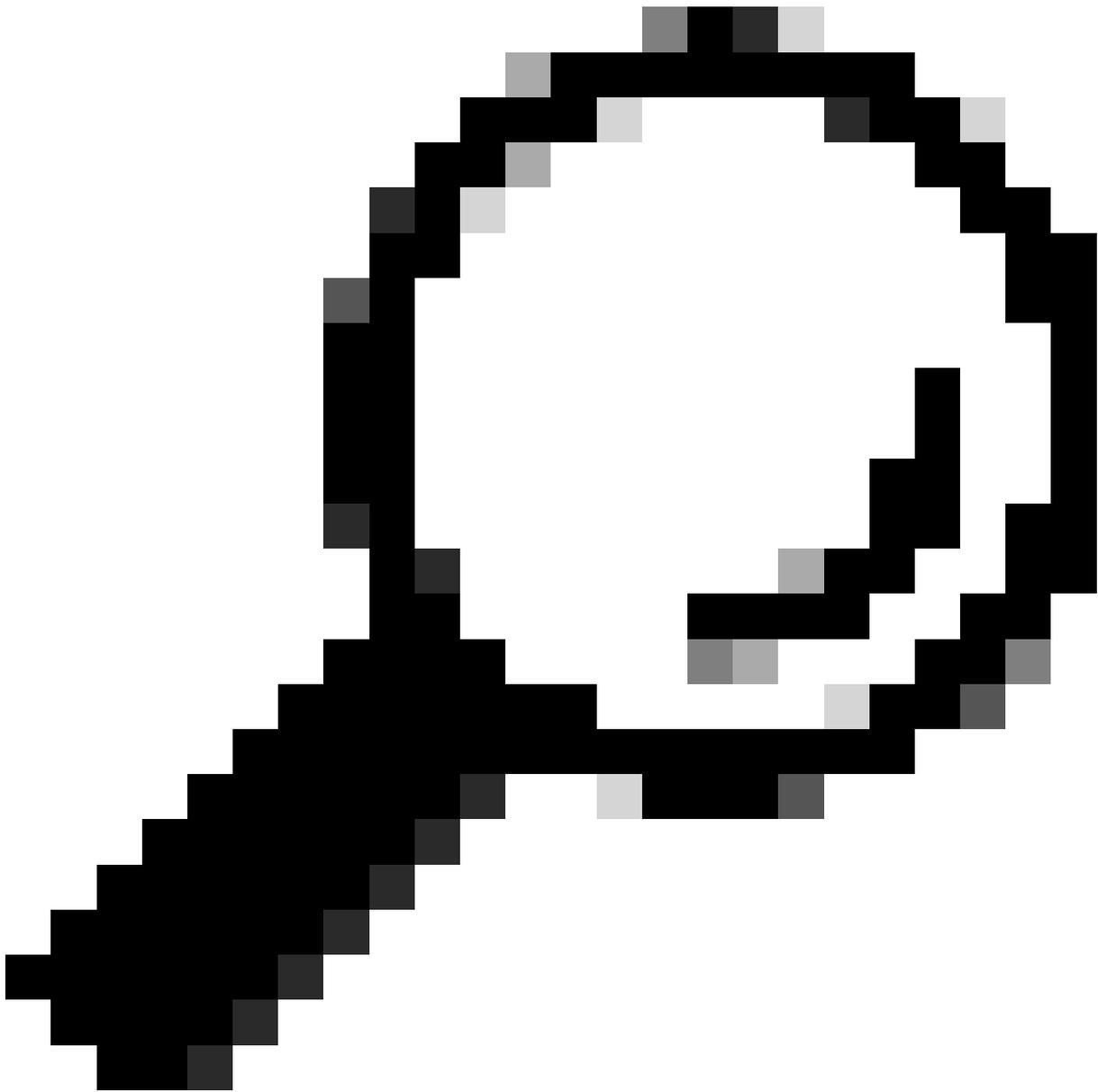


図 9. ノースツーサウスのトラフィックでのDPDKとSR-IOVのパケットトラバースル



ヒント：前述のように、DPDKがOVSとともに使用されるシナリオを除き、仮想ネットワーク機能(VNF)内でSR-IOVをDPDKと統合することによって、SR-IOVベースのセットアップのパフォーマンスを向上させることは可能です。

コンフィギュレーション

DPDKの有効化

GUIからDPDKを有効にするには、Configuration > Virtual Machine > Networking > Networksの順に移動する必要があります。メニューに移動したら、スイッチをクリックして機能をアクティブにします

Networks

Networks Information and Configuration

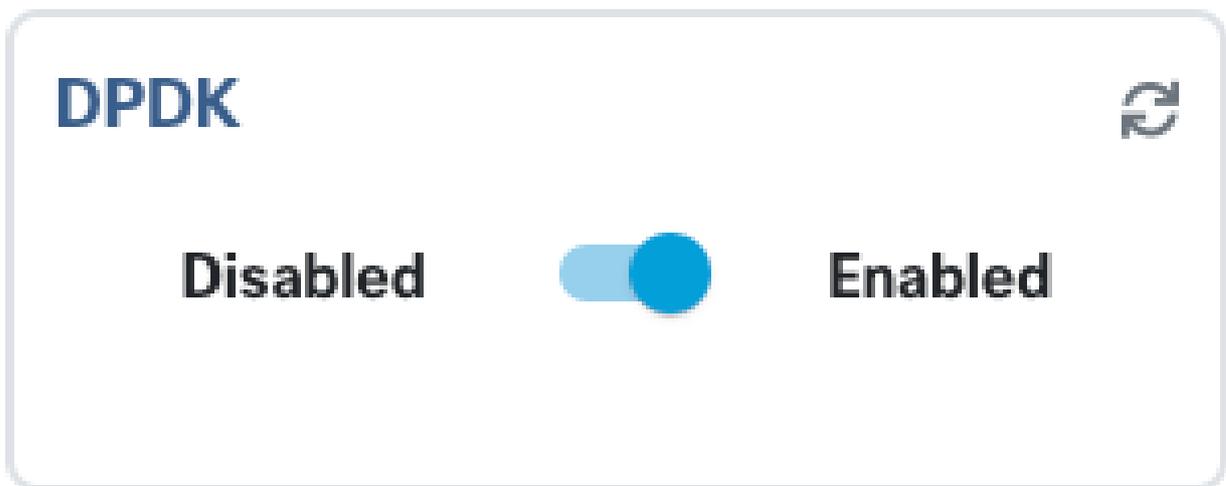
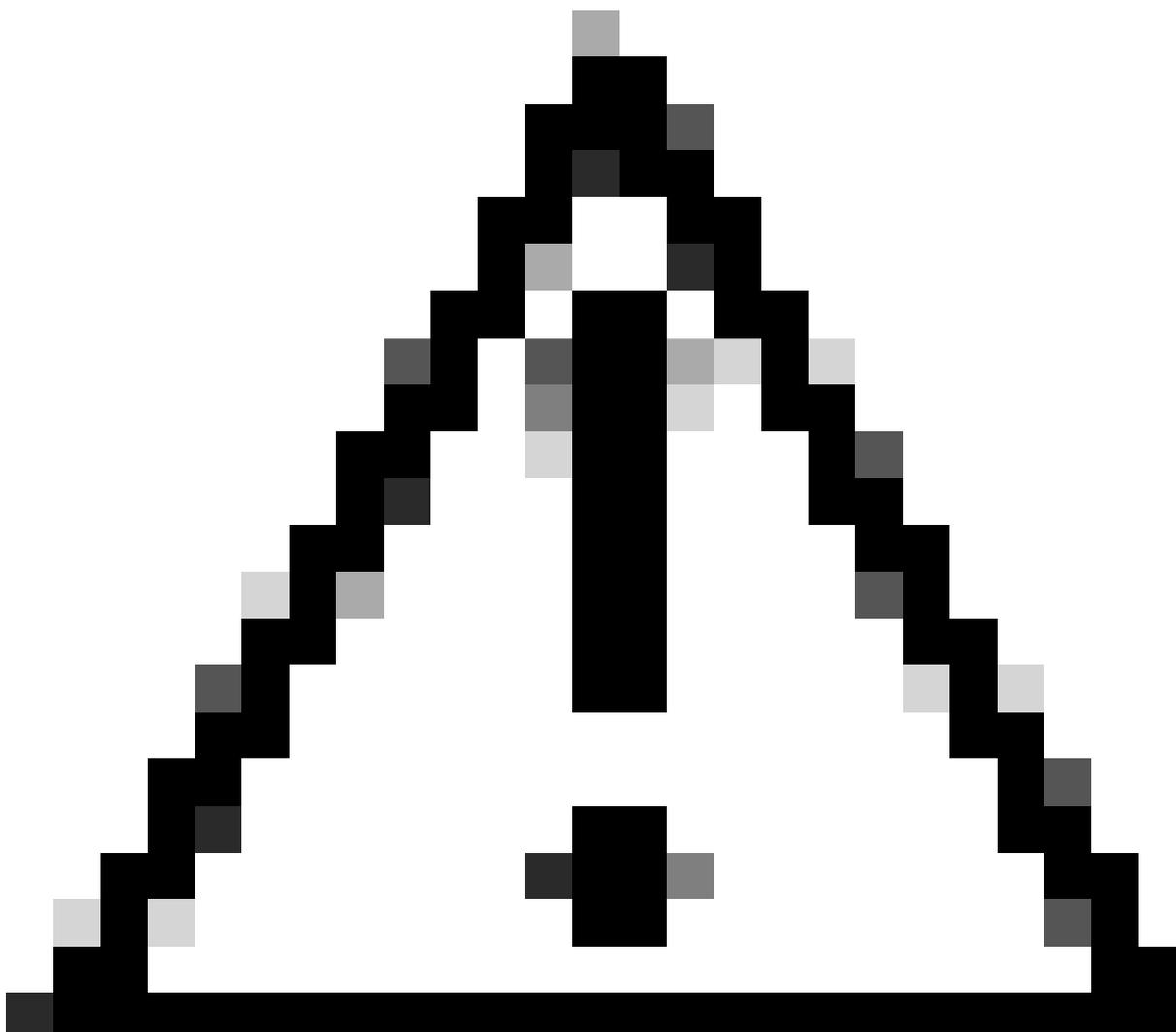


図 10 DPDK アクティベーションの GUI で使用可能なスライドボタン

CLI の場合、コンフィギュレーションモードでグローバルシステム設定からこのコマンドを有効にする必要があります。

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



注意:NFVISから工場出荷時の初期状態にリセットしない限り、DPDKを無効にすることはできません。

新しいネットワークを作成して新しいOVSブリッジに関連付ける

Configuration > Virtual Machine > Networking > Networksの順に移動します。Networksページで、Networksテーブルの左上のプラス記号(+)をクリックします。

#	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	 

図 11NFVIS GUIからのネットワークテーブルビュー

ネットワークに名前を付け、新しいブリッジに関連付けます。VLANとインターフェイスのバインドオプションは、ネットワークインフラストラクチャのニーズによって異なります。

Add Network

Network *

inter-vnf-net

Mode *

trunk

Vlan

Vlan-Range

Native Vlan

1

Bridge *

Existing Create New

Bridge

inter-vnf-br

Interface

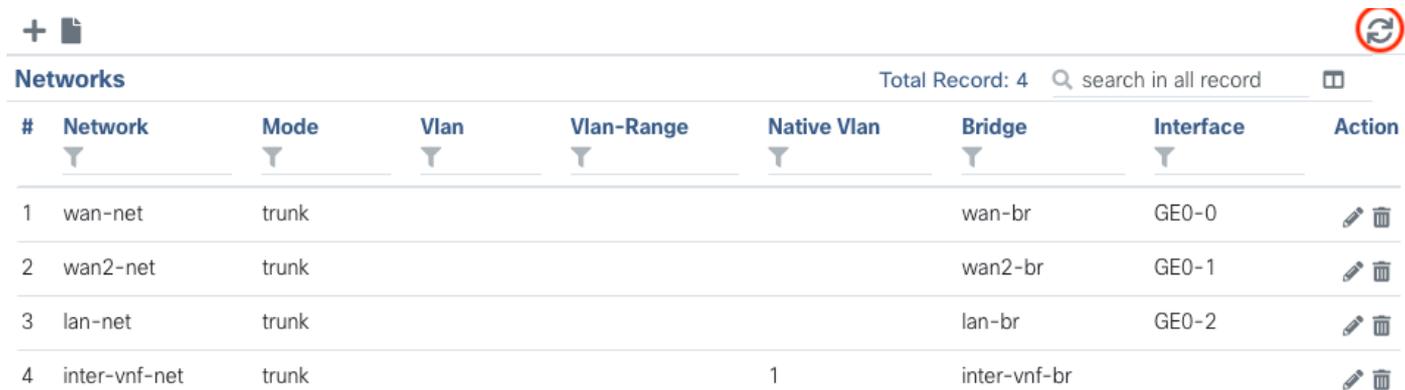
Submit

Cancel

Reset

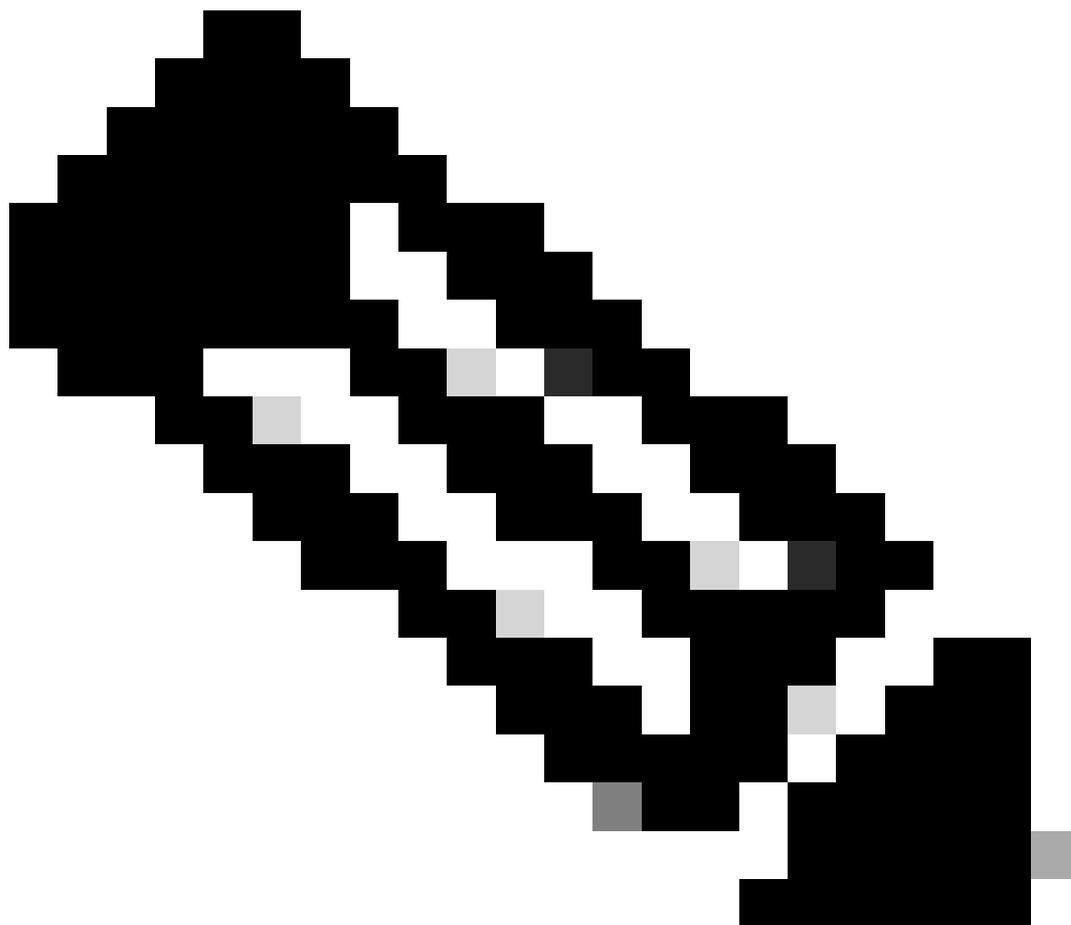
図12. NFVIS GUIで仮想ネットワークを作成する「ネットワークの追加」モーダル

submitボタンをクリックすると、Networksテーブルに追加された、新しく作成したネットワークを確認できる必要があります。



#	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		 

図 13.NFVIS GUIからのネットワークテーブルビュー(「Refresh Icon」が右上隅 (赤で強調表示))



注：新しいネットワークがテーブルに表示されない場合は、右上の更新ボタンをクリックするか、ページ全体を更新してください。

CLIからで実行した場合、すべてのネットワークとブリッジはコンフィギュレーションモードで作成され、ワークフローはGUIバージョンと同じです。

1. 新しいブリッジを作成します。

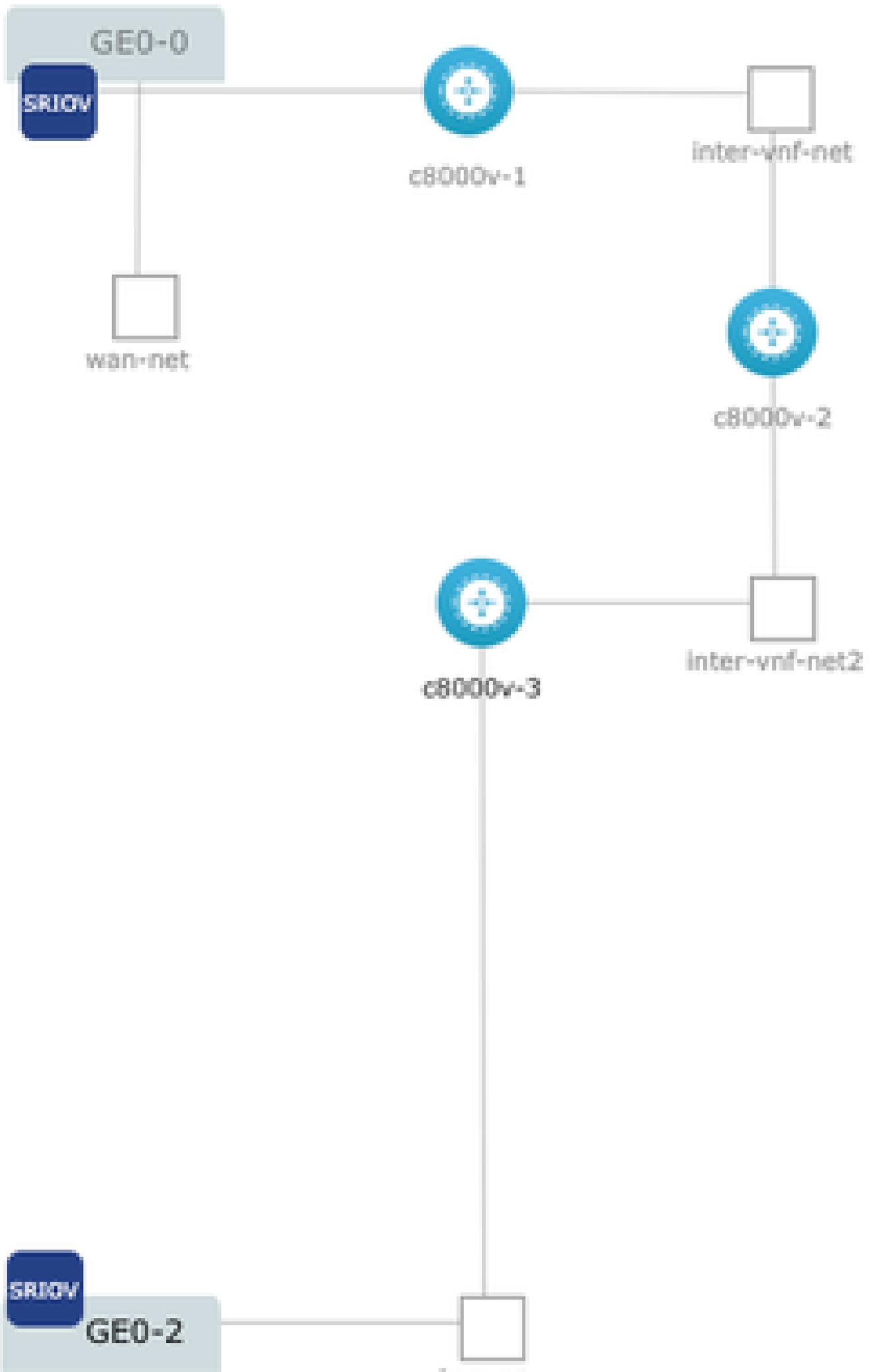
```
nfvis(config)# bridges bridge inter-vnf-br2
nfvis(config-bridge-inter-vnf-br2)# commit
```

2. 新しいネットワークを作成し、以前に作成したブリッジに関連付けます

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

VNFの接続

ネットワークトポロジまたは単一のVFNの導入から開始するには、Configuration > Deployの順に移動する必要があります。VMまたはコンテナを選択リストからトポロジ作成領域にドラッグして、仮想インフラストラクチャの作成を開始できます。



翻訳について

シスコは世界中のユーザにそれぞれの言語でサポート コンテンツを提供するために、機械と人による翻訳を組み合わせて、本ドキュメントを翻訳しています。ただし、最高度の機械翻訳であっても、専門家による翻訳のような正確性は確保されません。シスコは、これら翻訳の正確性について法的責任を負いません。原典である英語版（リンクからアクセス可能）もあわせて参照することを推奨します。