デジタルプラクティス　Vol.9 No.4 (Oct. 2018)

一般投稿論文

NFVを利用したサービスチェイニングの設計と運用の実践

堀場 勝広[1]　中村 遼[2]　鈴木 茂哉[1]　関谷 勇司[2]　村井 純[1]

[1]慶應義塾大学　[2]東京大学

Network Function Virtualization（NFV）を利用したサービスチェイニング（NFV-SC）は，ソフトウェアによる動的なネットワークの構成変更を可能とし，通信事業者の機器や運用のコストを低減することが期待されている．しかし，Interop Tokyo 2014 ShowNetにおいてNFV-SCを実装・運用した結果，Virtual Network Function（VNF）の連結によってパケット転送性能の低下が確認され，スケールアウトに課題が残った．そこで本研究では，スケールアウトとその前提となる相互接続性が実現可能なVNF 構成を検討し，その知見に基づき筆者らが提案しているNFV-SCの方式であるFlowFallを設計・実装するとともに，Interop Tokyo 2015 ShowNetにおいて，実際のネットワーク装置を利用してFlowFallを構築・運用し，商用ネットワークサービスとして20の出展者に対して3日間のNFV-SCを提供した．本稿は，これらの実践から得られたNFV-SCにおける相互接続性とスケールアウトの実現に必要な知見を述べる．

1. はじめに

サービスチェイニング[1]は，ネットワーク機能（FirewallやDeep Packet Inspectionなど特定のアプリケーションやプロトコルに特化したパケット処理）を連結し，利用者の要求に合致したサービスを提供するネットワークである．従来のサービスチェイニングは，特定のネットワーク機能に特化した機器を物理的に連結して構成されてきた．そのため，トラフィック量の増加やトラフィック傾向の変化に応じて，柔軟なネットワークの構成変更が困難であった[2]．

これらの問題を解決するために，NFV（Network Function Virtualization）[3]を利用したサービスチェイニング（本稿ではNFV-SCと呼ぶ）が提案された[4]．**図1**にEuropean Telecommunications Standards Institute（ETSI）のNFV-ISGで提唱されているNFV-SCのネットワーク構成を示す[3]．NFV-SCでは，ネットワーク機能をVirtual Network Function（VNF）と呼び，汎用サーバ上のハイパーバイザ（HV）で動作する仮想マシン（VM）として実装する．そして，OpenStack[5]に代表されるVirtual Infrastructure Manager（VIM）によるVNFの管理と，Software Dened Networking（SDN）によるVNF間のトラフィック操作をソフトウェアで制御することによって，動的なネットワークの構成変更を可能にする．その結果，通信事業者の機器や運用のコストを低減することが期待されている．

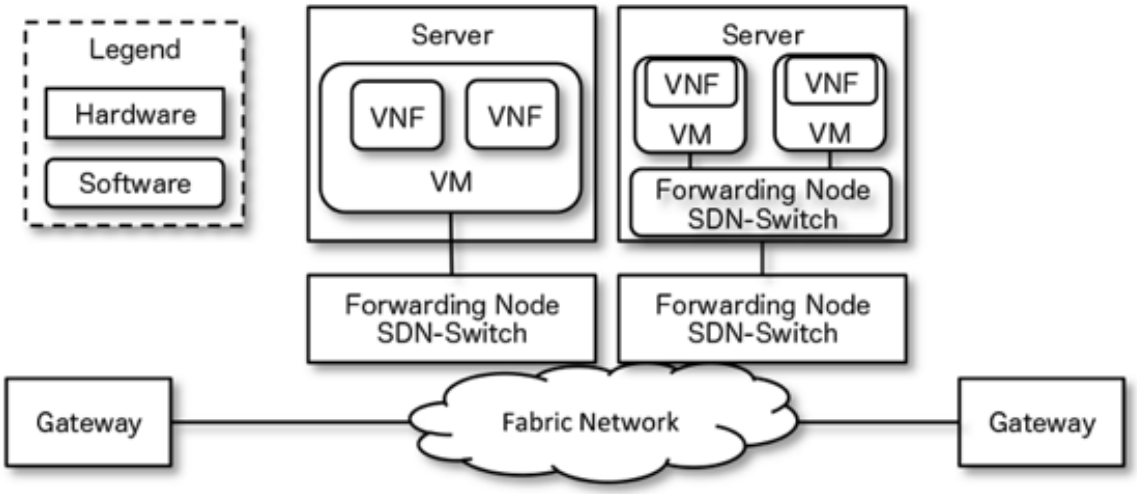


図1　ETSI NFV-ISGが提唱しているNFV-SCの構成

筆者らはInterop Tokyo☆1[6]のShowNet☆2[7]において，2013年から出展者を収容するネットワーク（以下出展者収容ネットワーク）のNFV化を試み[8]，2014年に開催されたInterop TokyoのShowNet（以下ShowNet 2014 と略記）では，NFV-SC の適用領域の1つであるvirtual Customer Premises Equipment（vCPE）サービス[9] に挑戦した[10]．その結果，ポータルサイトを介して利用者の要求に応じて動的に構成変更が可能なネットワークを提供できた．しかし，VNFの連結によるパケット転送性能の低下を確認した．対策としてVNFに割り当てるCPUやメモリなどの資源を追加したが，十分な性能の向上が得られなかった．

そこで本研究では，VNFに資源を追加することによって性能が向上するスケールアウトが可能なVNF構成について検討し，その知見に基づくNFV-SCの方式としてFlowFall[11]を設計・実装するとともに，ShowNet 2015においてFlowFallを構築・運用した．その結果，スケールアウトを実現するVNF構成には，VMのネットワークI/OにSR-IOV[12]を利用し，HV外部のハードウェアスイッチを利用したVMの連結方式と，複数のVM に分散した資源の割り当て方式が適していることが分かった．また，このVNF構成を利用したFlowFallは，VMとHVの追加によって線形にパケット転送性能が向上することを確認し，得られた知見の実用性を示した．Interop Tokyo 2015 ShowNetでは，FlowFallを実際のネットワーク装置を利用して構築し，20の出展者に対して3日間の商用インターネット接続性を提供し，vCPEサービスの実現性を示した．

本稿の構成は以下の通りである．第2章では，vCPEサービスの概要と要件およびShowNetへの適用について述べる．第3章では，ShowNet 2014におけるNFVの課題分析とスケールアウト可能なVNF構成の検討について述べる．第4章では，第3章での検討結果に基づいてvCPEサービスを実現するFlowFallの設計・実装について述べる．第5章では，ShowNet 2015におけるFlow Fallの構築・運用について述べる．第6章でFlowFallの実装・構築・運用から得られたNFV-SCの実現に必要なプラクティスと，それに伴うトレードオフについて述べる．最後に7章で結論を述べる．

2. ShowNet におけるNFV-SC の適用領域

本章では，vCPEサービスの概要を説明し，それらがShowNetの出展者収容ネットワークのネットワーク要求と構成に合致しており，vCPEサービスの概念実証に適していることを示す．

2.1　vCPEサービスの概要

vCPEサービスは，従来のCPEで動作していたさまざまなネットワーク機能を通信事業者のHV上でVNFとして動作させ，各利用者が要求するサービスに合致したネットワークをサービスチェイニングによって提供する．**図2**にvCPEサービスのネットワーク構成を示す．vCPEサービスでは，VNFの動作するHVが利用者のアクセス回線を集約する場所に設置される．サービスチェイニングは，利用者が要求するサービスに基づいて，トラフィックが通過するVNFを動的に変更することで実現される．

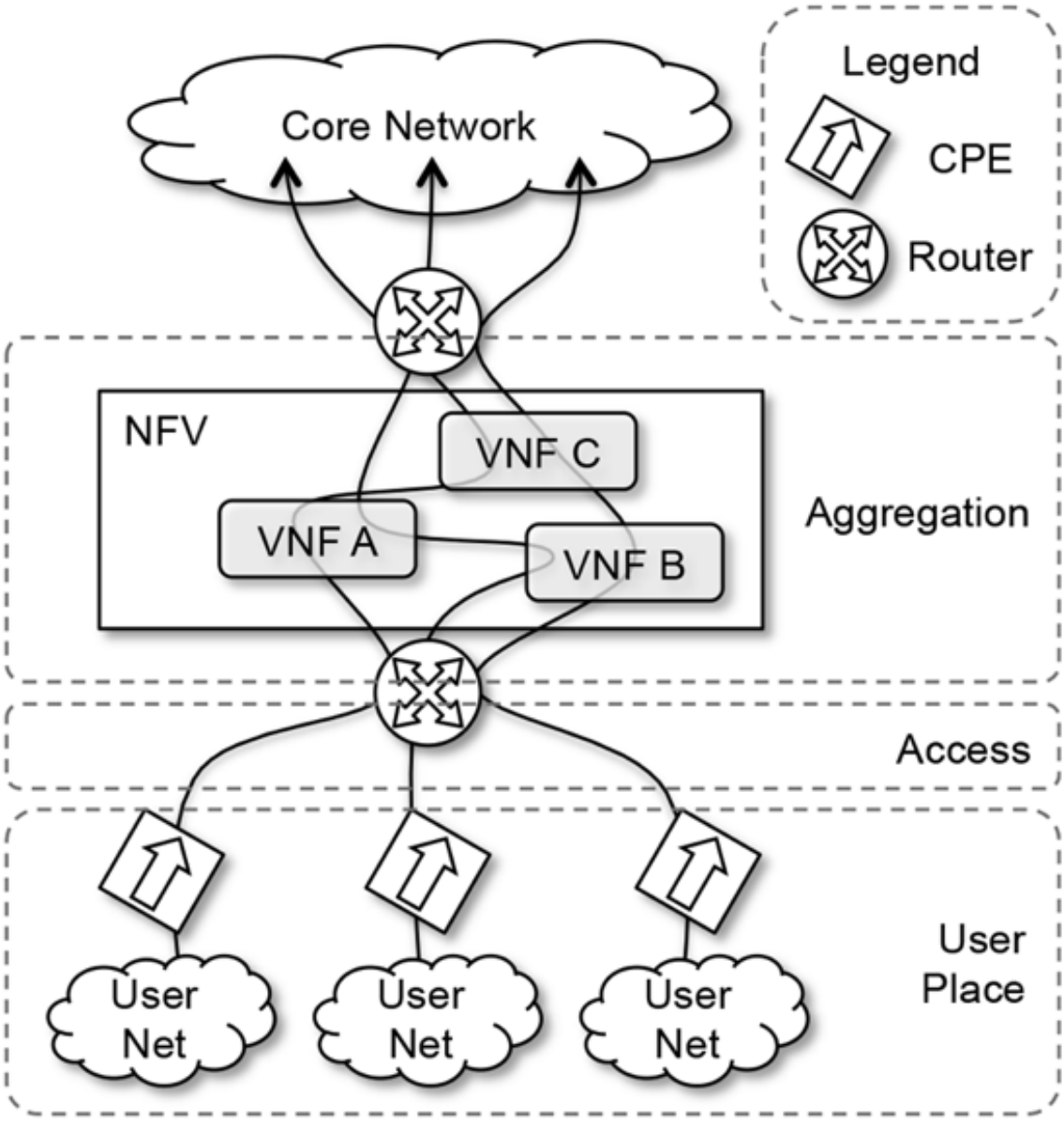


図2　EvCPEサービスのネットワーク構成

vCPEの実現には，一般的なNFVの要求事項[13]であるセキュリティ，サービス保証，耐障害性，既存ネットワークとの共存，電力効率などを考慮する必要はあるが，これまでの継続的なShowNetの構築・運用実績から，相互接続性を前提としたスケールアウトの確保が基本的な重要課題となっている．以下に本稿における相互接続性とスケールアウトの定義を示す．

（1）相互接続性：

ネットワーク機器のコントロールプレーンとデータプレーンに標準化された技術を利用し，さまざまなベンダのネットワーク機器を組み合わせてネットワークを設計できること．NFV-SCの利点は，さまざまなVNFを組み合わせたネットワークの動的な構成変更であり，それを制限する特定ネットワークベンダに限定された技術や標準化が進行中の技術の利用は避ける必要がある．

（2）スケールアウト：

VNFのパケット転送性能が，VNFに割り当てる資源の追加によって段階的に増強可能なこと．vCPEサービスでは，VNFは集約されたアクセス回線分のパケット転送性能を実現しなければならない．そのため，単一のVMでVNFのパケット転送性能を実現するのは困難であり，複数のVMを用いてトラフィックを分散処理する必要がある．

2.2 ShowNetとvCPEサービスの類似点

ShowNetの出展者収容ネットワークの要求と構成は，vCPEサービスに求められるものと類似しており，筆者らはNFV-SCの実践的な概念実証に適していると考えている．ShowNetにおけるネットワークの利用者とは出展者であり，出展者がShowNetに対して要求するサービスとは，出展者のデモに応じたトラフィック処理方法である．例えば，セキュリティ製品のデモを行う出展者は，外部からの攻撃トラフィックを自社の出展者収容ネットワークに流入させるため，ShowNetに対してFirewallによるトラフィックフィルタの解除を要求する．こうした要求に耐え得る柔軟なネットワークの構成変更を実現するために，ShowNetはvCPEサービスと同様に利用者ごとに適用するVNFの組合せを動的に変更できる必要がある．

ShowNetの出展者収容ネットワークの構成は，出展者ごとのアクセス回線をバックボーンの手前で集約する．ShowNetは出展者収容ネットワークに対して1Gbpsのアクセス回線と，プライベートまたはグローバルIPアドレスのスタブネットワークを提供する．ShowNet2015における138の出展者収容ネットワークのうち，最終的には20の出展者収容ネットワークをvCPEサービスで収容することとなった．そのため集約したアクセス回線分のパケット転送性能（20Gbps）を実現する必要がある．また，ShowNetの役割の1つは，Interop参加企業から提供されるさまざまな機器を組み合わせて構築し，異なるネットワークベンダの製品を組み合わせて複雑なネットワークが構成可能なことを示すことである．そのため，出展者収容ネットワークもShowNetの一部として，vCPEと同様にさまざまなVNF製品を組み合わせた構成が要求されており，相互接続性が必要になる．

3. スケールアウトが可能なVNF構成の検討

本章ではShowNet 2014の課題がスケールアウトである点について述べ，スケールアウトが可能なVNF構成を実現するために検討したVMのネットワークI/O技術とVMの連結方式およびVMの資源割り当て方式の組合せに関する評価結果について述べる．

3.1 ShowNet 2014のNFV-SCにおける課題分析

ShowNet 2014のNFV-SCではVNFの連結によるパケット転送性能の低下が確認され，スケールアウトが課題として残った[10]．ShowNet 2014のNFV-SCでは，VNFを構成するVMのネットワークI/O技術に仮想NICを利用し，HV内の仮想スイッチを利用してVMを連結した．その結果，VNFの連結数が増えるに従って，パケット転送性能が低下することを確認した．その対策として，VNFを構成するVMにCPUコアやメモリなどの資源を追加して性能の向上を試みたが，

パケット転送性能の十分な向上が得られなかった．この課題を分析した結果，NFV-SCの構成要素間による資源の競合が原因だと分かった．HVでの資源利用状況を計測した結果，仮想NIC，仮想スイッチ，VMの仮想CPUのスレッドが同一のCPUコアに集中し，大量のコンテキストスイッチが発生していた．この資源競合は，VMのネットワークI/O技術，VMの連結方式，VMの資源割り当て方式の組合せに起因する問題である[14]．この問題は，汎用サーバにおける高性能なパケット転送技術[15][16]によって解決できるという報告があるが，VMに対して特別なネットワークI/O技術を要求するため相互接続性の点で問題があり，ShowNet 2015の用途には適合しなかった．

3.2　検討内容と評価環境

そこで本研究では，相互接続性を維持したままスケールアウトが可能なVNF構成を検討するため，多くのVNF製品が対応しているVMのネットワークI/O技術と，VMの連結方式およびVMの資源割り当て方式の組合せを評価し，NFV-SCの構成要素間による資源競合が発生しにくい方式を検討した

（1）VMの連結方式：

VMの連結方式では，**表1**に示すVMのネットワークI/O技術と，VMの連結方式を組み合わせたVNF構成を比較し，資源競合によりパケット転送性能が低下しない方式を検討する．**図3**に，表1に示したVNF構成におけるトラフィックの通過パスを示す．利用したVMのネットワークI/O技術は，ShowNet 2014で利用したソフトウェアで用いた仮想スイッチであるOpenVswitch（OVS）[17]と，準仮想NIC（Virtio）[18]を組み合わせた構成に加えて，PCIカードのハードウェア補助による資源競合の回避が期待できるSR-IOVを検討対象とした．SR-IOVは，VF（Virtual Function）と呼ばれるPCIカード内の仮想NICを持ち，VMはVFをPCIPassThroughによって直接利用する．VMの連結方式には，ShowNet 2014で採用したHV内で動作する仮想スイッチ（OVS），HV外部のハードウェアスイッチ，SR-IOVをサポートするPCIカード内のスイッチを利用したVMの連結を対象にした．

表1　評価したVNF構成で利用した技術

VNF 構成	VM のネットワーク I/O 技術	VM の連結方式
(1) Virtual-Switch_OVS	OVS[17] + Virtio-Net[18]	HV 内部の仮想スイッチ (OVS)
(2) Exter-Switch_OVS	OVS[17] + Virtio-Net[18]	HV 外部のハードウェアスイッチ
(3) PCI-Builtin-Switch_SR-IOV	SR-IOV[12]	PCI カード内のスイッチ
(4) PCI-External-Switch_SR-IOV	SR-IOV[12]	HV 外部のハードウェアスイッチ

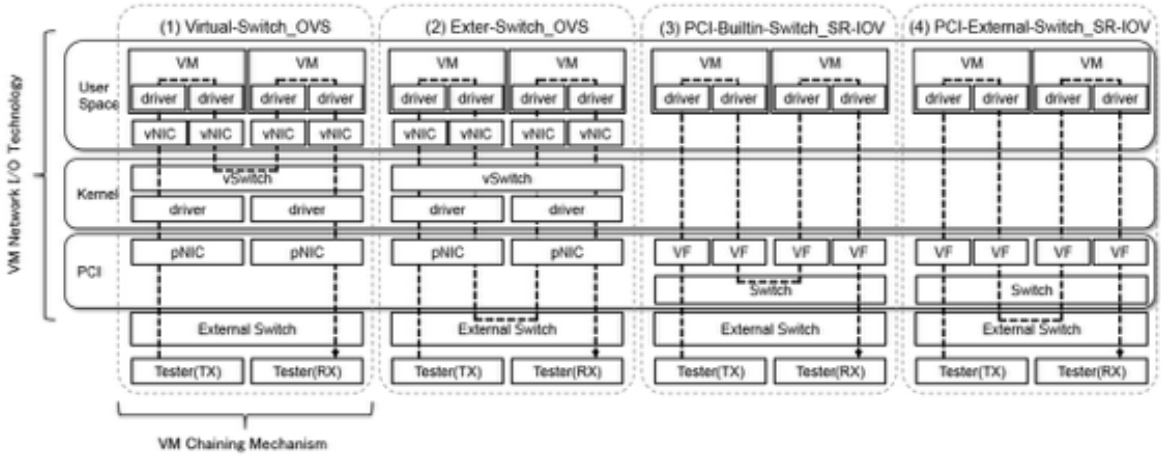


図3　VNF構成ごとのトラフィックの通過パス

（2）VMの資源割り当て方式：

VMの資源割り当て方式では，表1に示したVMのネットワークI/O技術とVMの資源割り当て方式を組み合わせたVNF構成を比較し，資源の追加に対してパケット転送性能の向上率が高い方法を検討する．VMの資源割り当て方式には，ShowNet 2014で採用した単一のVMに資源を集約してVNFを構成する方法（以下，資源集約と記載する）と，複数のVMに資源を分散してVNFを構成する方法（以下，資源分散と記載する）を対象とした．

表2に検討に使用した評価環境を示す．評価対象のVNFは，VMのOS上で動作するカーネルに付属するパケット 転送機能とiptablesを利用して構成した．パケット転送性能の評価は，負荷試験機器としてnetmap[16]に付属するトラフィック生成アプリケーションを利用して，フレームサイズ64バイトで10Gigabit Ethernetの理論値でトラフィックを送信し，通過できたパケット数を10秒間計測した．

表2　評価環境

プロセッサ	Intel(R) Xeon(R) E5-2650 2.00GHz
CPU コア数	8 コア
メモリ	メモリ 64G バイト
ホスト OS	Linux kernel 4.4.0-62
ハイパーバイザ	QEMU KVM 2.5.0
VNF OS	Linux kernel 4.4.0-62
トラフィック生成	netmap
計測時間	10 秒

3.3　評価結果

（1）VM の連結方式：

図4に同一HVにおけるVMの連結数と性能低下率の関係を示す．評価対象は，表1に記載したVMのネットワークI/O技術とVMの連結方式の組み合わせたVNF構成である．縦軸はVMが1個の場合のパケット転送性能の低下率を示し，横軸はVMの連結数を示す．

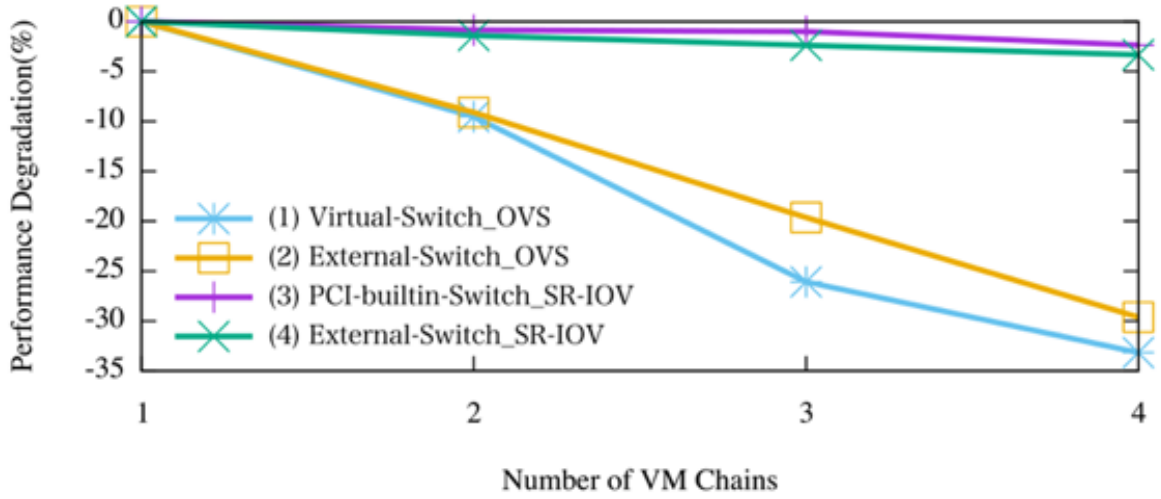


図4　VM連結方式の評価

仮想スイッチと仮想NICを利用した場合は，HV内部の仮想スイッチ，HV外部のスイッチのどちらを利用した場合（図中（1）および（2））も，連結するVM数が増加するとパケット転送性能が大きく低下する．一方，SR-IOVを利用した場合は，PCIカード内部のハードウェアスイッチ，HV外部のハードウェアスイッチのどちらを利用した場合（図中（3）および（4））でも，パケット転送性能は低下しない．

仮想スイッチと仮想NICを利用した場合に性能が低下する理由は，仮想スイッチからVMの仮想NICにパケットのデータを転送する際に発生するカーネルとユーザスペースでのデータコピーや，HVからのコンテキストスイッチのオーバーヘッドがあるためである[16]．

検討の結果，SR-IOVを利用してHV外部のハードウェアスイッチもしくはPCIカード内部のハードウェアスイッチを利用したVMの連結は，ShowNet2014におけるVNF構成（同一HV内で仮想スイッチを利用したVMの連結）と比較して，VNFの連結による性能低下の抑制に適した接続方式であることが分かった．

（2）VMの資源割り当て方式：

図5にVMに割り当てる資源の量とパケット転送性能の関係を示す．評価対象は，図3の（1）および（3）のVNF構成かつVMの連結数が1の場合である．縦軸は1つのVMに対して，1つのCPUコアを割り当てた場合を100としたパケット転送性能を示し，横軸はVMに割り当てたCPUコア数を示す．

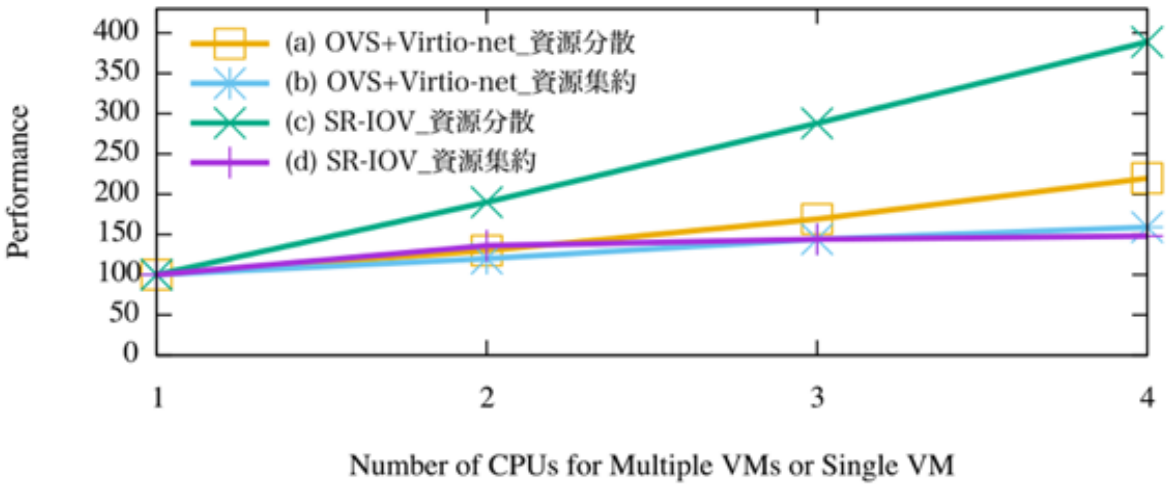


図5　VM資源割り当て方式の評価，IP Forwarding

仮想スイッチと仮想NICを利用した資源分散を行った場合（図中（a）），パケット転送性能が線形に向上した．しかし，資源集約（図中（b））を行った場合はパケット転送性能が線形に向上しなかった．一方，SR-IOVを利用した資源分散を行った場合（図中（c）），パケット転送性能が線形に向上した．しかし，資源集約を行った場合（図中（d））はCPUコア3個以上になった時点からパケット転送性能が向上しなかった．SR-IOVを利用した資源分散を行った方が，仮想スイッチと仮想NICを用いた資源分散と比較して，性能の向上率が高かった．

SR-IOVを利用した資源集約を行った場合に性能が線形に向上しなかった理由は，実験に利用したPCIカードのVFに対して割り当てるキューが2個であり，3個目以降のCPUコアがReceiver Side Scaling（RSS）[19]による負荷分散の対象にならないためである．なお，このキュー数はデバイスドライバの実装に依存する．SR-IOVを利用した資源分散の方が，仮想スイッチと仮想NICを利用した資源分散と比較して性能の向上率が高かった理由は，前述のSR-IOVには仮想スイッチと仮想NICを利用した際に発生するオーバーヘッドがないためである．

図6に，特定のアプリケーションやプロトコルに特化したVNF利用を想定し，VMでFirewall機能を有効にし，100個のルールを設定した場合のVNF割り当て資源量とパケット転送性能の関係を示す．これらのルールは，すべてのトラフィックで評価対象となるが，マッチしない条件設定としている．図5と同様に，資源分散を行った場合（図中（a）（c））は性能が線形に向上し，資源集約を行った場合（図中（b）（d））はスケールアウトに適した結果にはならなかった．

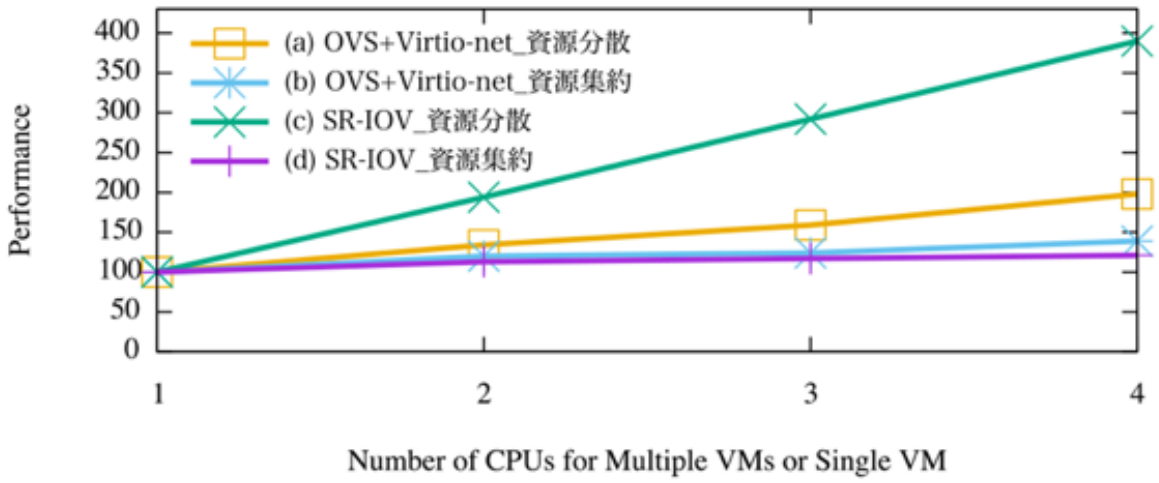


図6　VM資源割り当て方式の評価，Firewall

検討の結果，SR-IOVを利用した資源分散は，特定のアプリケーションやプロトコルに特化したVNFを利用した場合であっても，ShowNet2014におけるVNF構成（仮想スイッチと仮想NICを利用した資源集約）と比較して，スケールアウトに適したVMの資源割り当て方式であることが分かった．

4. FlowFallの設計・実装

本章では，第3章で得られたスケールアウト可能なVNF構成の知見に基づいて検討したFlowFallの設計・実装について述べる．FlowFallはvCPEサービスを実現するNFV-SCである．FlowFallがvCPEの要件を満たしていることを示すため，アーキテクチャとトラフィック制御について説明するとともに，プロトタイプ実装による相互接続性とスケールアウトの検証結果を示す．

4.1　アーキテクチャ

FlowFallは，相互接続性の要件を満たすため，OpenFlowを利用してネットワークを構成する．**図7**にFlowFallのネットワーク構成を示す．FlowFallのネットワークは，複数のVNFレイヤ，バイパスリンク，CPE，アグリゲーションルータによって構成される．VNFレイヤは，複数のVM，HV，OpenFlowスイッチによって構成され，複数のVNFレイヤを重ねることで複数のネットワーク機能の連結したNFV-SCを構成する．バイパスリンクは，サービスチェイニングにおいて特定のVNFレイヤを通過する必要がないトラフィックの転送に利用するリンクである．アグリゲーションルータは，利用者ネットワークのアクセス回線を集約し，NFV-SCにトラフィックを転送する．CPEは利用者ネットワークのデフォルトゲートウェイとして動作し，利用者の要求するサービスに基づいて後述するIPヘッダのType-of-Service（ToS）フィールドを用いたサービス識別子を記述する．NFV-SCの内部では，このサービス識別子に基づいてトラフィックが通過するVNFレイヤを決定する．

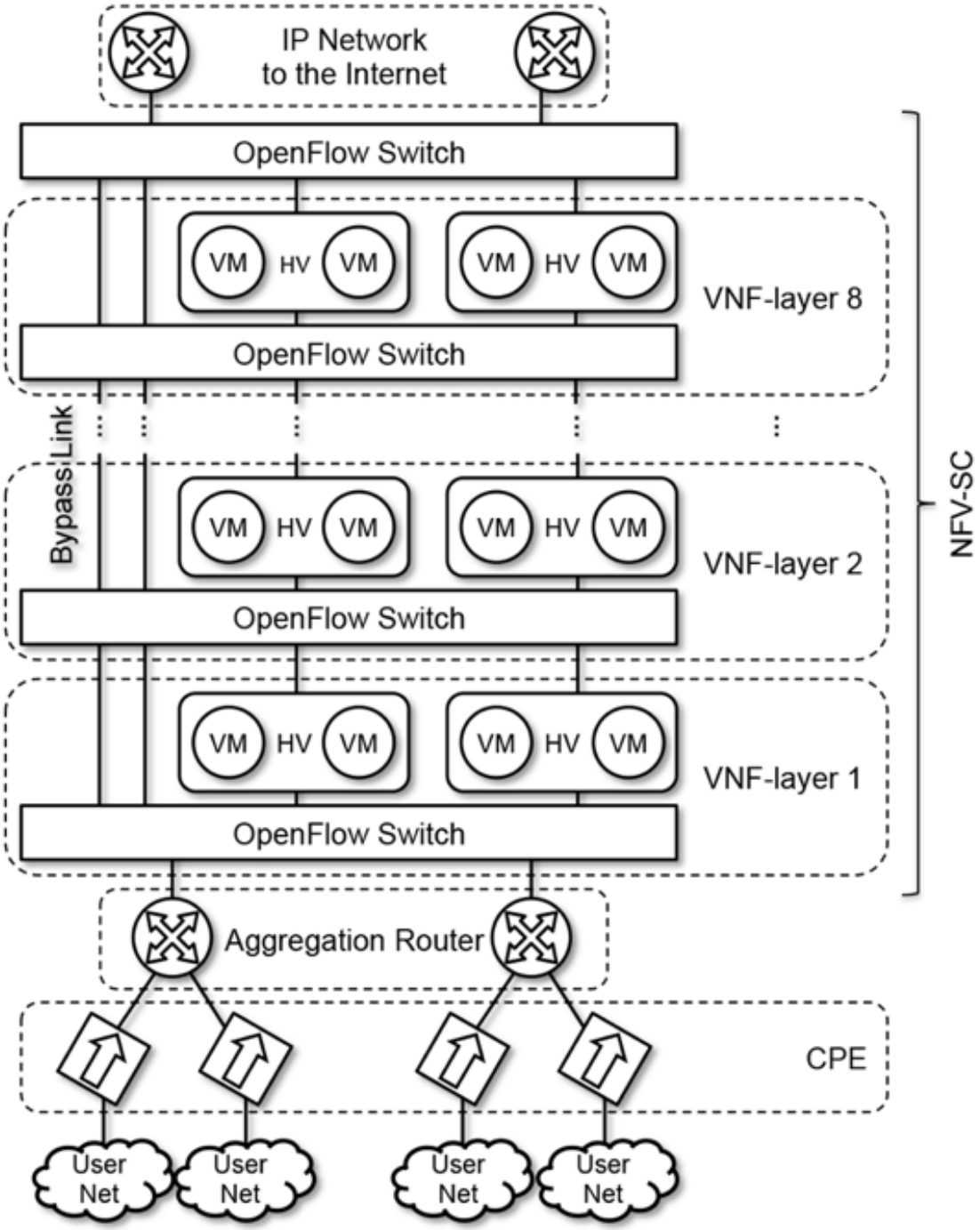


図7　FlowFall のアーキテクチャ

VNFレイヤは，スケールアウトの要件を満たすため，．第3章の検討で得られた知見に基づいた構成となっている．**図8**に，HVにおけるVMの構成を示す．各HVはアップリンクとダウンリンクの物理NICを搭載する．VNFを構成するVMのネットワークI/Oは，SR-IOVで分離された仮想NIC（VF）を利用し，それぞれ異なる物理NICに所属するVFを割り当てる．VMの連結とトラフィックの分散は，物理NICが接続されているOpenFlowスイッチで行う．各VMのCPUコアおよびメモリは，ほかのVMと重複しないように割り当てる．このようなVMの構成によってVM間の資源競合を排除し，複数のVMを用いたトラフィックの分散処理が可能になるため，VNFレイヤのスケールアウトが実現できる．この分散処理はHVの追加にも適用可能なため，1台のHVが保持するCPUコア数やメモリ量に依存せず，VNFレイヤのパケット転送性能を向上できる．

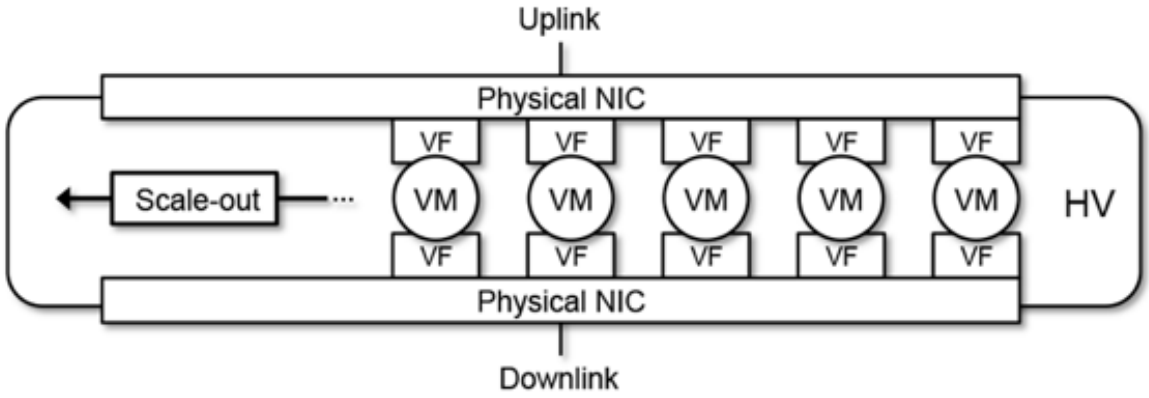


図8　VNFレイヤにおけるVMのネットワーク構成

4.2　トラフィック制御

FlowFall におけるトラフィック制御は，利用者ごとのサービスチェイニングを実現するため，IPヘッダのToSフィールドをサービスの識別子として用いる．OpenFlowスイッチは，ToSフィールドによってトラフィックの転送先をVMにするかバイパスリンクにするか，つまり各VNFレイヤでネットワーク機能を適用するか否かを決定する．FlowFallにおけるサービス識別子にToSフィールドを選択した理由は，トラフィックごとに適用すべきサービスを識別でき，相互接続性を確保できるためである．たとえば，ToS フィールドの書き換えは多くのCPE で可能であり，OpenFlowのマッチフィールドとしても利用可能である．

図9に，FlowFallにおけるToSフィールドを利用してサービスの識別子を記述する方法を示す．ToS フィールドはIPヘッダに含まれる8 ビット長のフィールドである．FlowFallにおけるToSフィールドの各ビットは，各VNFレイヤの適用の有無を示す．たとえば，DPIとFirewallサービスを通過するパケットでは，ToSフィールドの対応する3ビット目と4ビット目を1にする．利用者ネットワークを収容するCPEは，通過するパケットのToSフィールドに利用者の選択したサービスに対応したビット列を設定する．

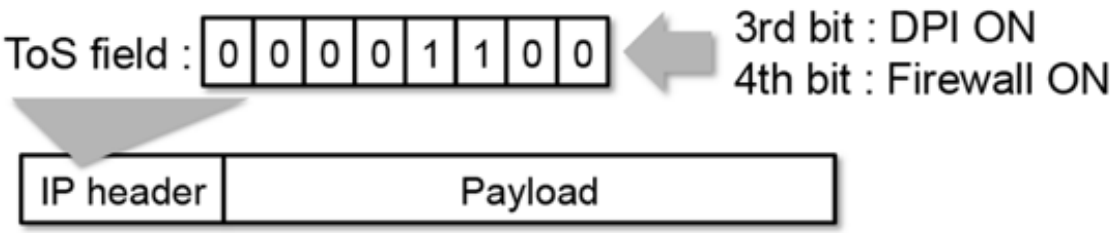


図9　ToSフィールドをビットマップとしたサービス識別子

このToSフィールドを用いたサービス適用の可否判断とVNFを構成する複数VMへのトラフィックの分散のために，FlowFallはOpenFlowスイッチのポートをトラフィック制御の観点から4種類に分類して管理する．**図10**にFlowFallにおけるOpenFlowのスイッチポートの分類を示す．BypassUp（BU）ポートは上位VNFレイヤのOpenFlowスイッチに，BypassDown（BD）ポートは下位VNFレイヤのOpenFlowスイッチに接続する．VNFUp（VU）ポートは上位VNFレイヤのHVに，VNFDown（VD）ポートは下位VNFレイヤのHVに接続する．同じ分類のスイッチポートが複数ある場合，OpenFlowスイッチはトラフィ

ックの分散処理を行う．OpenFlowコントローラは，各VNFレイヤに属するVMの接続情報を，OpenFlowスイッチのポートと宛先MACアドレスのペアで表現する．たとえば，VUポートではHVと接続されたポートと，そのHV上に存在するVMのMACアドレスによって表現する．これらの情報は，VMの起動と終了に応じて動的に設定する．

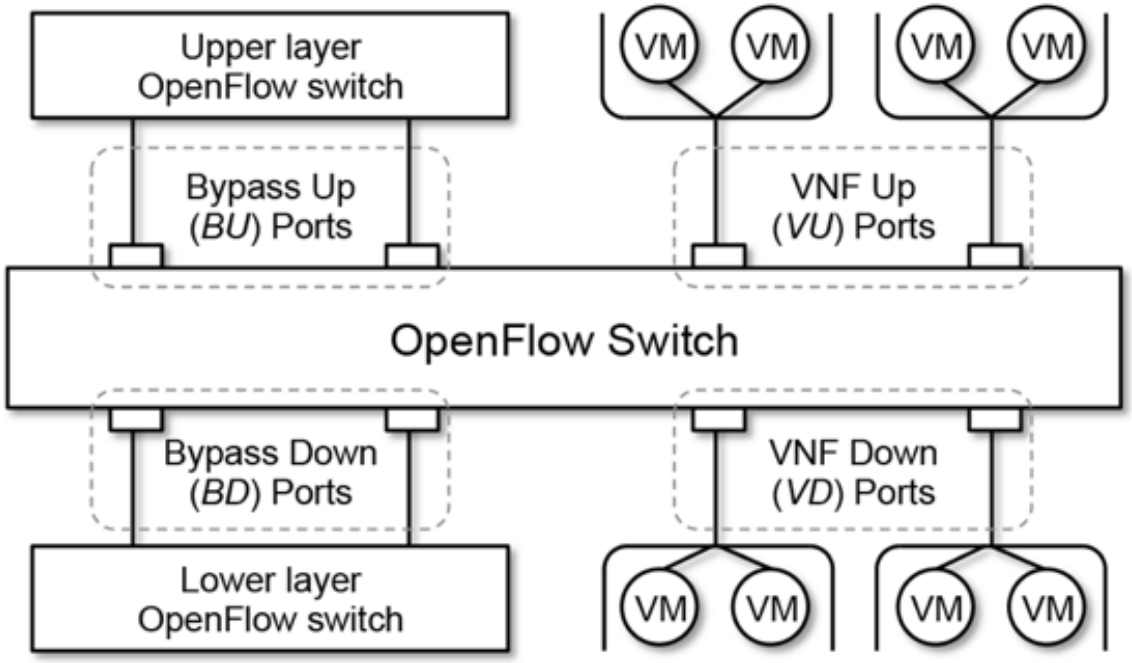


図10　FlowFallにおけるOpenFlowスイッチポート

出展者収容ネットワークからインターネットに向けたトラフィックの転送は，OpenFlowコントローラがあらかじめ設定された出展者収容ネットワークのIPアドレスとToSフィールドの値を用いて，OpenFlowスイッチにフローエントリを設定することで実現する．OpenFlowコントローラは，パケットの受信ポートがVDもしくはBDポートかつ，受信パケットのToSフィールド上で該当するVNFレイヤのビットが1の場合，VMの接続情報からIPアドレスのMD5ハッシュ値に基づいて出力先VUポートと宛先MACアドレスを選出する．その後，受信ポート，ToSフィールド，送信元IPアドレスをマッチフィールドとし，宛先MACアドレスを選出されたMACアドレスに書き換え（set-dl-dst），選出されたVUポートに出力するアクションを持つフローエントリを設定する．受信パケットのToSフィールド上で当該VNFレイヤのビットが0の場合は，バイパスリンクに対して同様の処理を行う．

インターネットから出展者収容ネットワークに向けたトラフィックの転送は，出展者収容ネットワークからインターネットに向けたトラフィックの転送に利用したフローエントリの送信元IPアドレスおよびMACアドレスを宛先とし，VDまたはBDポートを出力先としたフローエントリを設定することで実現する．このようにVNFレイヤごとに上下対称にフローを設定することで，あるフローが通過するVMの列が上下対称になる．これによって，セッション単位の処理が必要なFirewallやDPIといったネットワーク機能が利用可能となる．

4.3　プロトタイプ実装による検証

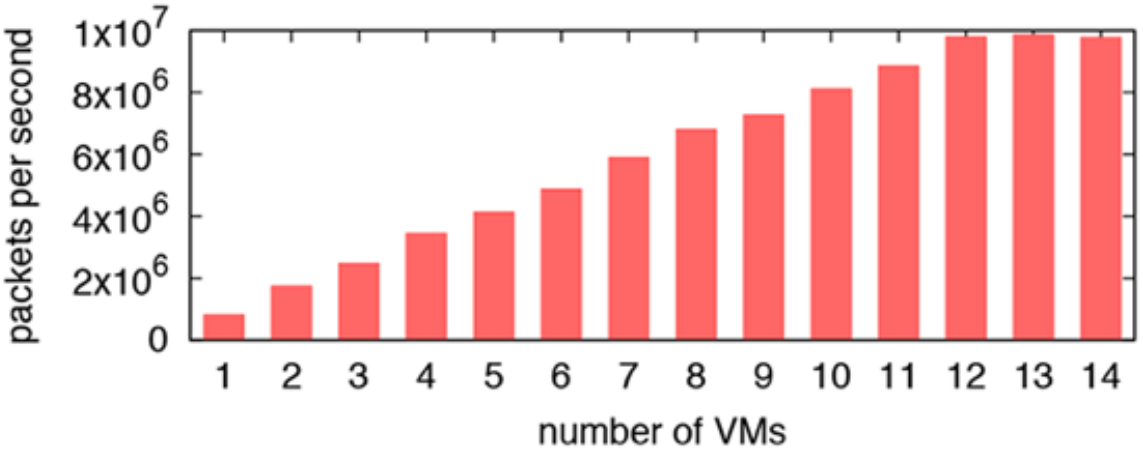
FlowFallの動作確認のためにプロトタイプを実装し，相互接続性とスケールアウトについて検証した．その結果，相互接続性については定性的，スケールアウトについては定量的な評価によって要件を満たしていることを確認した．

（1）相互接続性：

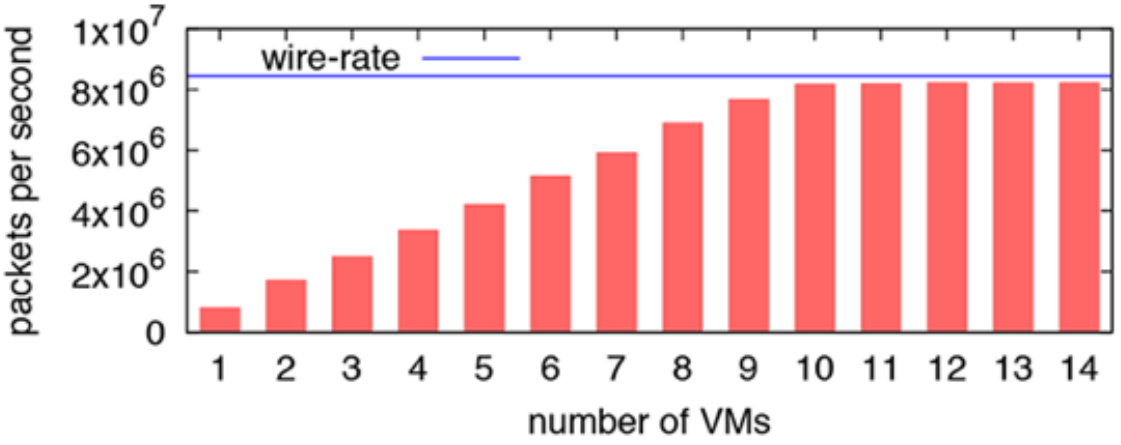
FlowFallは，汎用的なハードウェアと標準化された技術によって構成されており，プロトタイプ実装においても相互接続性を実現している．FlowFallのネットワークは，x86アーキテクチャの汎用サーバとOpenFlow 1.3に準拠した汎用OpenFlowスイッチによって構成され，サービスチェイニングのトラフィック制御に用いた識別子は，利用者のIPアドレスとToSフィールドである．また，VNFレイヤは，VMのネットワークI/O技術にSR-IOVを利用し，連結方式にHV外部のハードウェアスイッチを利用しており，一般的に入手可能なVNF製品で動作可能な構成である．つまり，FlowFallが利用するデータプレーンおよびコントロールプレーンは標準化された技術であり，さまざまな製品を組み合わせてネットワークを構成することができる．

（2）スケールアウト：

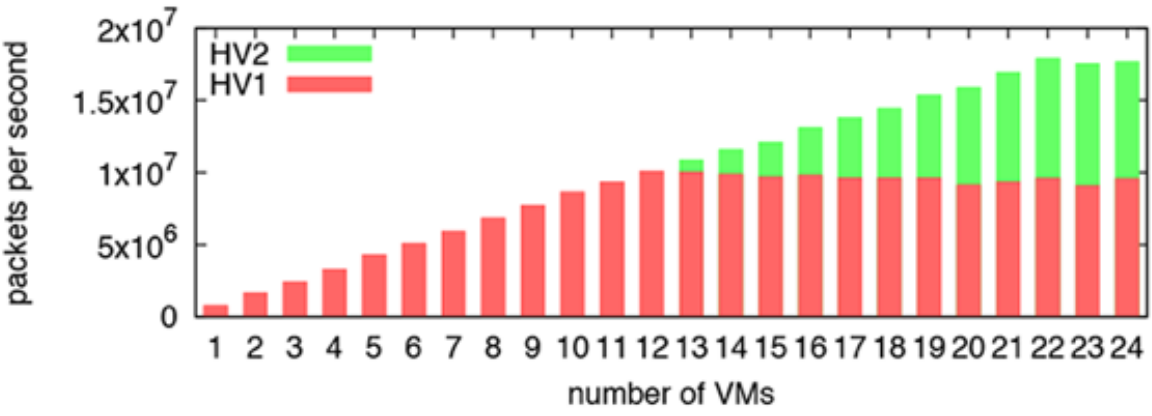
論文[11]において，VNFレイヤへのVMおよびHVの追加によるスケールアウトの実現を確認している．以下にこの論文における結論を引用し，スケールアウトの検証結果について述べる．**図11**に，1台のHVにおいて起動するVMを増加させていった場合のパケット転送性能（フレームサイズ64バイト，128バイト）と，HVを追加した場合のパケット転送性能（フレームサイズ64バイト）を示す．1台のHVにおいて起動するVMを増加させていった結果，フレームサイズ64バイトでは，12台のVMまで性能が線形に向上し，約7Gbpsのパケット転送性能を達成した．フレームサイズ128バイトでは，10台のVMまで性能が線形に向上し，10 Gigabit Ethernetの理論値性能を達成した．また，HVを追加することによって，パケット転送性能が線形に向上することを確認した．



(a) HV1 台，フレームサイズ 64 バイト



(b) HV1 台，フレームサイズ 128 バイト



(c) HV2 台，フレームサイズ 64 バイト

図11　単一VNFレイヤにおけるVNF数とHV数を増加させた場合のパケット転送性能

5. ShowNetにおけるFlowFallの構築・運用

本章では，ShowNet 2015で構築したFlowFallのネットワーク構成，出展者収容ネットワークとして運用した結果から見えてきた課題について述べる．

5.1 ネットワーク構成

ShowNet 2015の出展者収容ネットワークは，すべて市販されている汎用サーバ，VNF製品，ネットワーク装置によって構成された．**表3**にShowNet 2015の出展者収容ネットワークに利用したネットワーク機器の一覧を示し，**図12**にそれらを用いたネットワーク構成を示す．利用したハードウェアは，OpenFlowスイッチとx86アーキテクチャの汎用サーバである．OpenFlowスイッチにはNEC PF5248とPF5459を利用し，HVとして用いる汎用サーバにはDell PowerEdge R630，Huawei FusionServer X6800，Dell PowerEdge C6220を利用した．これらのハードウェアを用いて，3つのVNFレイヤから成るFlowFallを構築した．VNFレイヤは3つの機能によって構成した．1つ目のVNF レイヤはPalo Alto NetworksのPA-VMを利用したアプリケーションレベルでのトラフィック解析機能，2つ目のVNFレイヤはCisco SystemsのCSR1000vを利用したFirewall機能，3つ目のVNFレイヤはA10NetworksのThunder 6536 TPSを利用したDDoS対策機能である．Thunder 6536 TPSは専用ハードウェアを利用したネットワーク機器であるが，FlowFallのアーキテクチャではVNFがハードウェアかVMかを区別する必要はないため，動作確認のために導入した．

表3　ShowNet 2015におけるFlowFallで利用したネットワーク機器

装置名	タイプ	役割	HV
A10 Thunder 6435 TPS	Hardware	DDos Mitigation	N/A
Cisco csr1000v	VNF	Firewall	Dell PowerEdge C6220
PaloAlto PA-VM	VNF	DPI	Huawei Fusion Server X6800
Juniper EX4600	Hardware	Aggregation Router	N/A
Juniper vSRX	CPE	NAT	Dell PowerEdge R630

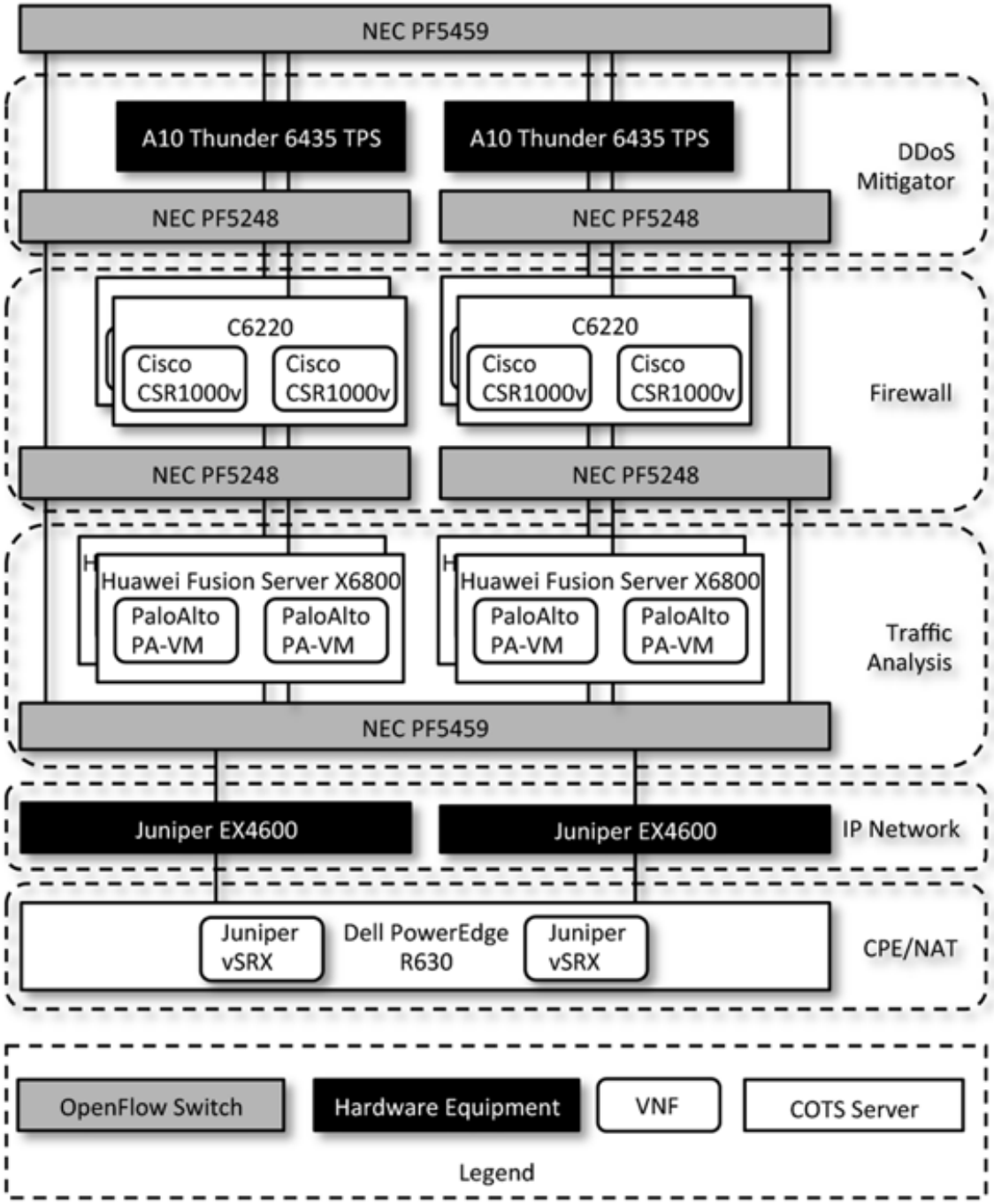


図12　ShowNet 2015におけるFlowFallのネットワーク構成

CPEにはJuniper NetworksのvSRXを用いて，NAT機能と，出展者の適用するサービスをToSフィールドに設定する機能を実装した．NAT機能をVNFレイヤではなくCPEレイヤに実装した理由は，FlowFallにおけるトラフィック制御を容易にするためである．FlowFallは，インターネットから出展者収容ネットワークに向けたトラフィックの識別に，送信元IPアドレスを利用するため，NATが送信元IPアドレスを変更した場合，FlowFallのトラフィック制御で利用する送信元IPアドレスが2種類になり，OpenFlowコントローラにおける管理が複雑になる．そのため，パケットの送信元IPアドレス，ポート番号を変換するNATはCPEに実装した．

出展者ごとのサービス要求に基づいた動的なネットワーク構成変更への対応については，出展者収容ネットワークに対して利用するVNFを制御可能なWebポータルを用意した．出展者が利用するサービスをWebポータルから指定することで，トラフィック制御に利用するToSフィール

ドの値を決定し，vSRXに対してNetconfを利用して設定変更を行い，出展者収容ネットワークから出力されるトラフィックにToSフィールドの値を設定する．

5.2　運用上の課題

FlowFallは2.1節で挙げたvCPEサービスの要件を満たし，ShowNet 2015での運用を通して実用性を示すことができた．しかし，ShowNetにおいてFlowFallをより安定して運用するためには，少なくともFlowFallに特化した死活監視機能と効率的なトラフィックの分散が必要である．

（1）FlowFallに特化した死活監視機能：

FlowFallでは，スケールアウトを実現するトラフィックの分散にIPアドレスのハッシュ値を用いている．このため，外部のネットワークから任意のVMに対してPingなどを用いた一般的なIPネットワークにおける死活監視の手法を適用できない．

この課題については，Link Layer Discovery Protocol（LLDP）を用いたOpenFlowネットワークにおける死活監視手法[20]をVNFおよびリンクの監視に応用することで解決できると考えている．このような手法をFlowFallに実装するためには，OpenFlowコントローラがpacket-inおよびpacket-outメッセージを用いてOpenFlowスイッチ経由でLLDPパケットを送受信することで，リンクの状態を管理する必要がある．

（2）効率的なトラフィックの分散：

FlowFallでは，パケットの送信元IPアドレスをキーとしたハッシュアルゴリズム（MD5）を利用して，あるフローを処理するVMを決定している．このため，フローごとのトラフィック量が異なる実際のネットワークでは，同一VNFレイヤ内において均等にトラフィックを分散できない場合がある．

この課題については，ネットワーク機器に入力されるトラフィックの傾向を解析し，動的に転送先を変更するトラフィック分散手法[21]を，同一VNFレイヤのVMに対して入力するトラフィック分散に応用することで解決できると考えている．このような手法をFlowFallに実装するためには，sFlow[22]に対応したOpenFlowスイッチを利用し，OpenFlowコントローラでトラフィックの傾向を解析した結果に基づいて転送先のVMを変更する必要がある．

6. 本研究によって得られた知見

本章では，本研究によって得られたNFV-SCにおける相互接続性とスケールアウトに関するプラクティスと，それに伴うトレードオフについて述べる．

6.1 相互接続性とスケールアウトに関するプラクティス

FlowFallの設計・実装・運用から得られたNFV-SCにおける相互接続性とスケールアウトに関するプラクティスを示す．

（1）相互接続性：

NFV-SCで相互接続性を実現するには，VMのネットワークI/O，VMの連結方式，VNF間のトラフィック制御について考慮する必要がある．

（a）VMのネットワークI/O：HVとVNFの両方が対応している技術が必要である．SR-IOVやVirtio-Netなどの標準化もしくは仕様が公開された仮想デバイスの利用が有効である．
（b）VMの連結方式：VMのネットワークI/O技術と接続可能で，外部からトラフィック制御のシグナリングが可能な技術が必要である．OVSやOpenFlowスイッチなど，標準化されたコントロールプレーンプロトコルに対応したSDNスイッチの利用が有効である．
（c）VNF間のトラフィック制御：VNFとCPEを含めたすべてのネットワーク構成要素が対応しているデータプレーン技術が必要である．ToSフィールドなど標準化されたパケットヘッダのサービス識別子への利用が有効である．

（2）スケールアウト：

NFV-SCでスケールアウトを実現するには，VM間における資源競合を回避し，複数のVMでトラフィックを分散処理可能なVMの連結方式，VMの資源割り当て方法について考慮する必要がある．

（a）VMの連結方式：仮想NICや仮想スイッチなど，HV内部におけるソフトウェアパケット処理による資源競合を回避可能な技術が必要である．SR-IOVとHV外部のハードウェアスイッチを利用したHV内のソフトウェアパケット処理のバイパスが有効である．
（b）VMの資源割り当て方式：VMのネットワークI/O技術が持つキュー数に非依存で，VM内の資源競合を回避可能な技術が必要である．VMごとに必要最小限のCPUコアを割り当てる資源分散が有効である．

6.2 トレードオフ

FlowFallには，vCPEサービスの要件を満たすことを優先した結果，トレードオフとしてVNFの数，VNFの種類，VNFの順序に関して制約が発生した．

（1）VNFの数：

FlowFallでは8ビットのToSフィールドをサービス識別子に用いて，ToSフィールドの1ビットが1個のVNFを示す設計としたため，連結可能なVNFの数（最大8個）に制約がある．ShowNet 2015においてVNFは8個で十分であったが，それ以上の数のVNFを利用するにあたっては，新たなネットワーク制御手法を検討する必要がある．

（2）VNFの種類：

FlowFallではToSフィールドをトラフィック制御のサービス識別子に用いたため，IPヘッダを書き換えるネットワーク機能がVNFレイヤで利用しにくい制約がある．例えば送信元IPアドレスとポート番号を変換するNetwork Address Translation（NAT）装置がToSフィールドの値を書き換える場合，FlowFallのトラフィック制御で利用する送信元IPアドレスが2種類になり，OpenFlowコントローラにおける管理が複雑になるため，ShowNetではNAT機能はCPE側で実装した．

（3）VNFの順序：

FlowFallでは，VMのネットワークI/O技術にハードウェアによる補助（SR-IOV）を用いたため，VNFレイヤ間の接続はハードウェアスイッチを介した静的なものとなった．そのため，VM間の資源競合の排除によるスケールアウトを実現できたが，適用するVNFの順序はVNFレイヤの順序に依存し変更することができない．一方，仮想スイッチを用いれば，自由な順序でのVNFの

連結を実現できるが，性能は犠牲となる[23]．このように，ハードウェアによる補助を用いてスケールアウトを実現するか，パケットのソフトウェア処理を用いて柔軟性を実現するかは，トレードオフの関係にある．

7. 結論

本稿では，ShowNet 2015の出展者収容ネットワークにvCPEサービスを実現するため，スケールアウト可能なNFV-SCのVNF構成について検討し，検討結果に基づきFlowFallを設計・実装について述べた．また，Interop Tokyo 2015 ShowNetにおいて，FlowFallを用いて出展者収容ネットワークを構築し，20の出展者に対して3日間のvCPEサービス提供を通して課題解決の実現性を示した．

本研究から得られた知見は，ハードウェアによる補助を利用してVM間の資源競合を排除し，複数のVMを用いてトラフィックを分散処理することでスケールアウトを実現できるということである．FlowFallはこの知見にもとづき，SR-IOVとOpenFlowを用いたトラフィックの分散処理によって，相互接続性とスケールアウトを同時に実現した．しかし，FlowFallを運用した結果，VMの死活監視機能と効率的なトラフィックの分散が課題として残り，トレードオフとしてVNFの数，VNFの種類，VNFの順序に制約があることが分かった．

今後は，セキュリティ，サービス保証，耐障害性，既存ネットワークとの共存，電力効率などの要求事項についても検討を進めるとともに，NFV-SCを取り巻く最新技術を利用し，FlowFallでの制約や課題を解決したvCPEサービスの実現に取り組んでいく予定である．具体的には，DPDK[15]を利用した仮想スイッチ（vpp[24],bess[25],ovs-dpdk[26]），共有メモリを使った仮想NICの実装（vhost-net[27]），サービスチェイニングに特化したデータプレーン技術（NetworkServiceHeader[28]）やSDNコントローラ実装（OpenDay-light[29]，Nuage[30]，OpenContrail[31]）などの利用を検討している．

参考文献

1）Blendin, J., Rückert, J., Leymann, N., Schyguda, G. and Hausheer,D.：Position Paper：Software-Dened Network Service Chaining, In EWSDN Workshop (2014).
2）John, W., Pentikousis, K., Agapiou, G., Jacob, E., Kind, M., Manzalini, A., Risso, F., Staessens, D., Steinert, R. and Meirosu,C.：Research Directions in Network Service Chaining. In Future Networks and Services （SDN4FNS）, 2013 IEEE SDN for, pp.1-7 (Nov.2013).
3) Network Functions Virtualisation - Update White Paper, https://portal.etsi.org/nfv/nfv_white_paper2.pdf.
4）ETSI NFV ISG, Service Chaining for NW Function Selection in Carrier Networks, https://nfvwiki.etsi.org/images/NFVPER%2814%29000004r2_NFV_ISG_PoC_Proposal_Service_Chaining_for_NW_Function_Select.pdf.
5）Sefraoui, O., Aissaoui, M. and Eleuldj, M.：Openstack：Toward an Open-source Solution for Cloud Computing. International Journal of Computer Applications, Vol.55, No.3 (2012).
6）Interop Tokyo：http://www.interop.jp
7）ShowNet：http://www.interop.jp/2015/shownet
8）中村　遼，堀場勝広，関谷勇司：SDNを用いたクラウドサービスネットワークの実現（インターネット運用・管理，　一般），電子情報通信学会技術研究報告, IA, インターネットアーキテクチャ, Vol.113，No.200，pp.5-10（2013）．
9）ETSI NFV ISG, Network Functions Virtualization Use cases,

https://www.etsi.org/deliver/etsi_gs/NFV/001_099/001/01.01.01_60/gs_NFV001v010101p.pdf
10) 中村　遼：INTEROP Tokyo 2014 ShowNetにおけるSDN/NFV, http://www.sdnjapan.org/document_2014/30_session5_Nakamura.pdf, (Oct. 2014).
11) Nakamura, R., Okada, K., Saito, S., Tanahashi H. and Sekiya, Y. ：Flowfall: A Service Chaining Architecture with Commodity Technologies, In 2015 IEEE 23rd International Conference on Network Protocols (ICNP), IEEE, pp.425-431 (2015).
12) PCI-SIG. SR-IOV Primer: An Introduction to SR-IOV Technology. http://www.intel.com/content/www/us/en/pci-express/pci-sig-sr-iov-primer-sr-iov-technology-paper.html, (accessed 2015-2-2).
13 ） ETSI NFV ISG, Network Functions Virtualization （ nfv ） Virtualisation Requirements, https://www.etsi.org/deliver/etsi_gs/NFV/001_099/004/01.01.01_60/gs_NFV004v010101p.pdf
14） Hwang, J., Ramakrishnan, K. K. and Wood, T. NetVM: High Performance and Exible Networking Using Virtualization on Commodity Platforms, In Proceedings of The 11th USENIX Conference on Networked Systems Design and Implementation, NSDI'14, pp.445-458, Berkeley, CA, USA, USENIX Association (2014).
15）Intel Corp, Intel dpdk: Data plane development kit： http://dpdk.org/(accessed 2014-10-26).
16）Rizzo, L., Netmap: A Novel Framework for Fast Packet I/O, In Proceedings of the 2012 USENIX Conference on Annual Technical Conference, USENIX ATC'12, pp.9-9, Berkeley, CA, USA, USENIX Association (2012).
17 ） Nicira Networks. Open vSwitch: An open virtual switch ： http://openvswitch.org/ (accessed 2015-1-28).
18）Russell, R., Virtio: Towards a De-Facto Standard for Virtual I/O Devices, SIGOPS Oper, Syst, Rev., Vol.42, No.5, pp.95-103, (July 2008).
19 ） Wu, W., DeMar, P. and Crawford, M.: A Transport-Friendly NIC for Multicore/Multiprocessor Systems. Par-allel and Distributed Systems, IEEE Transactions on, Vol.23, No.4, pp.607-615 (April 2012).
20）Sharma, S., Staessens, D., Colle, D., Pickavet, M. and Demeester, P.：Enabling Fast Failure Re-Covery in Openflow Networks, In Design of Reliable Communication Networks (DRCN), 2011 8th International Workshop on The, IEEE, pp.164-171 (2011).
21）Nuaimi, K. A., Mohamed, N., Nuaimi, M. A. and Al-Jaroodi, J.：A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms. In Network Cloud Computing and Applications (NCCA),2012 Second Symposium on, IEEE, pp.137-142 (2012).
22）Phaal, P., Panchen, S. and McKee, N. : InMon Corporation's sFlow : A Method for Monitoring Traffic in Switched and Routed Networks, RFC 3176（Informational） (September 2001).
23）Ziri, S. R., Samsudin, A. T. and Fontaine, C. : Service Chaining Implementation in Network Function Virtualization with Software Dened Networking, In Proceedings of The 5th International Conference on Communications and Broadband Networking, ACM, pp.70-75 (2017).
24）Linguaglossa, L., Rossi, D., Pontarelli, S., Barach, D., Marjon, D. and Pfister, P. : High-speed Software Data Plane via Vectorized Packet Processing.
25）Niu, Z., Xu, H., Liu, L., Tian, Y., Wang, P. and Li, Z. : Unveiling Performance of Nfv Software Dataplanes (2017).
26）Jackson, E. J., Walls, M., Panda, A., Pettit, J., Pfaff, B., Rajahalme, J., Koponen, T. and Shenker, s. : Softflow : A Middlebox Architecture for Open Vswitch, In USENIX Annual Technical Conference, pp.15-28 (2016).

27）Gordon, A., Har'El, N., Landau, A., Ben-Yehuda, M. and Traeger, A. : Towards Exitless and Efficient Paravirtual I/O, In Proceedings of The 5th An-nual International Systems and Storage Conference, ACM, p.10 (2012).
28）Quinn, P. and Elzur, U. : Network Service Header.
https://tools.ietf.org/html/draft-ietf-sfc-nsh-12(May 2016).
29）Medved, J., Varga, R., Tkacik, A. and Gray, K. Opendaylight : Towards a Model-driven Sdn Con-troller Architecture, In 2014 IEEE 15th International Symposium on, IEEE, pp.1-6 (2014).
30）Ferro, G. : Packet Pushers White Paper, Nuage Networks, White Paper (2013).
31）中嶋大輔：クラウドサービスの課題とopencontrail の実装（ポストipネットワーキング，次世代・新世代ネットワーク（ngn），障害対策・bcp，ネットワークコーディング，セッション管理（sip・ims），相互接続技術/標準化，ネットワーク構成管理および一般），電子情報通信学会技術研究報告. IN，情報ネットワーク, Vol.114, No.207, pp.37-42 (2014).

脚注

☆1 Interop Tokyo[6]は毎年6月に開催されるネットワーク機器と技術の展示会である.

☆2 ShowNet[7]はInteropに出展している企業からプロモーションを目的として提供されるネットワーク機器によって構築されるデモンストレーションネットワークである. ShowNetの役割は，出展者と来場者に対してインターネットの接続性を提供すると共に，新しいネットワーク技術の実現性を示すことである.

堀場勝広（正会員）qoo@sfc.wide.ad.jp
2006年慶應義塾大学政策・メディア研究科前期博士課程修了．修士（政策・メディア）．2007年より同研究科後期博士課程．2012年より同研究科特任助教．2015年よりソフトバンク株式会社．クラウドコンピューティング，SDN，NFVに関する研究開発に従事.

中村遼（非会員）upa@wide.ad.jp
2012年慶應義塾大学環境情報学部卒業，2017年東京大学大学院情報理工学系研究科博士課程修了．2017年より東京大学情報基盤センター助教．博士(情報理工学)．オーバーレイネットワークやSDN/NFVの高速化，運用技術に関する研究・開発に従事.

鈴木茂哉（正会員）shigeya@wide.ad.jp
情報システム研究者およびエンジニア．コンピュータネットワーク，コンピュータを用いた通信，ブロックチェーン技術，サイバーセキュリティ，量子情報システム，および，RFIDを含む実空間情報システムの研究に従事．システムアーキテクチャ，ソフトウェア開発についてのエキスパート．1989年よりインターネットに基づく情報システム開発に従事．2012年に慶應義塾大学大学院政策・メディア研究科において博士号取得．現在，慶應義塾大学大学院 政策・メディア研究科特任准教授，慶應SFC研究所ブロックチェーンラボ副所長，WIDEプロジェクトボードメンバを兼任．電子情報通信学会英文論文誌(EB)編集委員.

関谷勇司（正会員）sekiya@wide.ad.jp
1997年京都大学総合人間学部卒．2005年慶應義塾大学政策・メディア研究科後期博士課程修了．博士 (政策・メディア)1999年から2000年まで米国カリフォルニア州USC/ISIにてDNSの研究に従事．2002年に東京大学情報基盤センター助手に就任．2008年同センター講師を経て2011年同センター准教授．分散サービスの計測，クラウドコンピューティングの可用性向上，SDNとNFV，ならびにサイバーセキュリティに関する研究に従事．

村井純（正会員）jun@wide.ad.jp
1984年慶應義塾大学大学院工学研究科後期博士課程修了，博士（工学）1984年東京工業大学総合情報処理センター助手，1987年東京大学大型計算機センター助手，1990年慶應義塾大学環境情報学部助教授を経て1997年より同教授2005年-2009年学校法人慶應義塾常任理事，2009年-2018年環境情報学部長，2018年-大学院政策・メディア研究科委員長，インターネット網の整備，普及に尽力．初期インターネットを，日本語をはじめとする多言語対応へと導く．

投稿受付：2017年4月5日
採録決定：2018年6月1日
編集担当：寺田真敏（日立製作所）