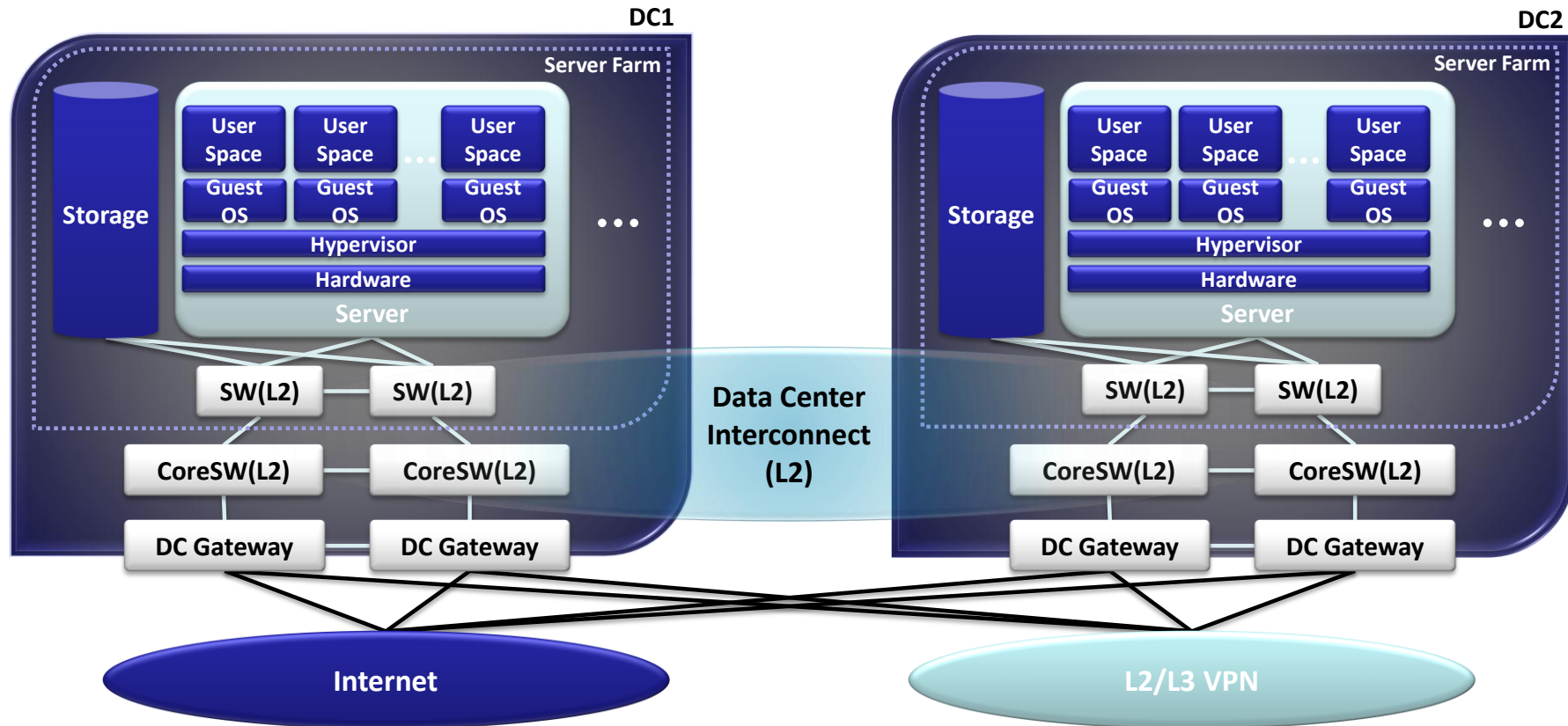


データセンタネットワーク

NTTコミュニケーションズ
2012年01月20日
池尻 雄一

クラウドネットワークの構成 (NTTコミュニケーションズでの構成を例に)

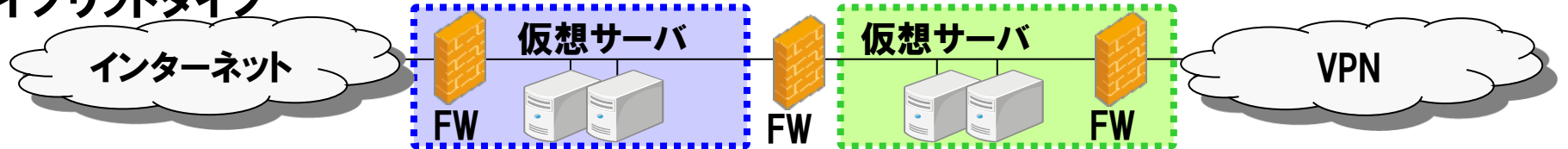
- VM (仮想サーバ) / Storageサービス等のIaaS提供プラットフォーム
- HypervisorをToR (Top of Rack) SWに接続し、CoreSWでアグリゲーション
- DC Gatewayを介してInternet/VPNと接続される



現状のネットワーク構成と運用

- クラウド内部はL2NWを基本に構成
- 複数のL2セグメントを一つの物理ネットワークに重畳
 - Multi-tenancy確保のためにVLANによるセグメント分けを利用
 - Internet接続とVPN接続のセグメント分けにもVLANを利用。
- VMのライブマイグレーションを利用
 - VMのIPアドレスは変えずに收容されるHypervisorを変更
 - G-ARPを利用した同一L2セグメント内移動
- VM数の増加に伴ってL2NWを拡大

■ 例:ハイブリッドタイプ



クラウドネットワークでやりたいことと課題

■ 拡張性/柔軟性を確保したい → 今回のメインピック

- ネットワークの拡張スピードは非常に速い。
- その中で拡張性・柔軟性を究極まで確保したい。
 - DC内・DC間など物理制限にとらわれずにネットワークを拡張したい
 - VMのインスタンス数を無限に増やしたい。
 - VMのモビリティを確保したい。
 - マルチテナント性の確保・拡張性を担保したい。
 - トポロジーの柔軟さ、冗長性をもって拡張したい。

■ その他もいろいろありますが。 → ここではScope外。

- 帯域拡張、QoS、OAM
- ネットワーク制御のAutomation化
 - APIサポート/SDN(Software Defined/Driven Network)/Openflow etc
 - クラウドバースト etc.

拡張性・柔軟性の課題（特にL2NWの拡張） 1/2

■ VMのスケール・柔軟性 → MACアドレスのスケールリング

- サーバ側HV/VMは通常のEthernetの動きを期待。
- 理論的には、無限のMACアドレスを収容したい。
- ライブマイグレーションの柔軟性を考えるとブロードキャストドメインを拡張したい。
- DC内に留まらずDC間ネットワークにも広げたい。
- L2SWでのFDBサイズ: 1000台のHV → 15VM/HV → 2vNIC/VM → 30k MAC
- DC内/DC間のARP/ND(Neighbor Discovery)関連トラフィックの増大

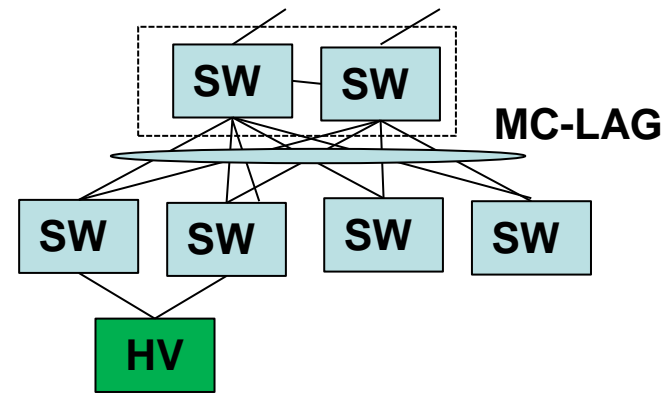
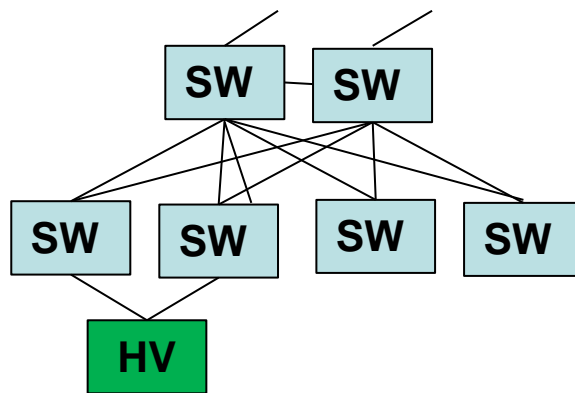
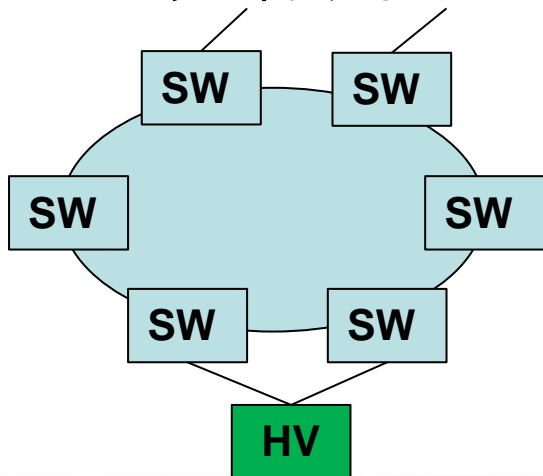
■ マルチテナンシー → VLAN IDのスケールリング

- マルチテナンシー性の確保とスケール性を両立させたい。
- 12bit = 4096 の壁。1ユーザに1VLANとしても4000ユーザ限界。
- ユーザ識別子としては小さすぎ。
- 通常はクラウド内部でも1ユーザに複数セグメントを提供するためもっと条件は厳しい。
- DC間をL2同一セグメントで伸ばそうと考えるとDC間回線で4096xNの重畳が必要。

拡張性・柔軟性の課題（特にL2NWの拡張） 2/2

■ 冗長構成などトポロジーの柔軟性

- L2ネットワークは冗長性の確保が本質的に難しい。
 - ・ Ring構成 (G.8032, その他独自プロトコル)
 - ・ Tree構成 (冗長構成部は、STP、RSTP、独自プロトコルでの制御)
 - ・ MC(Multi-Chassis) LAG構成
 - ・ etc
- これらを組み合わせながらどこまでネットワークを拡張できるのか。
- トラフィックボトルネックとの戦い



課題解決に向けて

IETF/IEEEでも様々な議論。様々な実装方式が乱立。。さてどうする？

■ MACアドレス、VLAN-ID拡張へのアプローチ

- L2VPN技術の拡張での対応
 - PBB(802.1ah), VPLS(RFC4761/4762) 及び PBB-VPLS(ietf-draft)
- L3ネットワークへのオーバーレイでの対応
 - ベースNW(HV接続NW)はL3で作リ、L2 over L3でスケールさせる
 - NVO3 (Network Virtualization over L3) : VXLAN (UDP 24bit-ID), NVGRE(GRE 24bit-ID)
 - マルチテナンシーIDとして24bitで十分か。

■ トポロジー柔軟性へのアプローチ

- L3ルーティング機構のL2ネットワークへの適用
 - ISIS適用のアプローチ: TRILL(IETF TRILL-WG)+α(独自), SPB(IEEE 802.1aq)
 - BGP適用のアプローチ: E-VPN, PBB-EVPN(individual-draft)

■ すべての既成概念にとらわれずに解決するアプローチ

- Openflow技術の適用

【参考1/2】IETF82でのDC Related Discussion.. (Problem Statement)

- L2NWのDCにおけるスケールの課題分析
 - ARMD-WG
 - draft-armd-datacenter-reference-arch-01 : Merit/Huawei etc.
 - draft-ietf-armd-problem-statement-00 : IBM
- スケール性を含むDCにおけるVM運用の課題
 - L2VPN-WG
 - draft-narten-nvo3-overlay-problem-statement-00 : IBM/Microsoft
 - draft-bitar-datacenter-vpn-applicability-01 : Verizon/Alcatel/Cisco/BT/NTT Com
 - L3VPN-WG
 - draft-so-vpn4dc-00 : Verizon/AT&T/TW Telecom etc.
 - draft-so-vdcs-02 : Verizon/TW Telecom etc.
 - SDN-BOF
 - draft-pan-sdn-dc-problem-statement-and-use-cases-01 : CA/Infinera
 - draft-nadeau-sdn-framework-01 : CA/Infinera

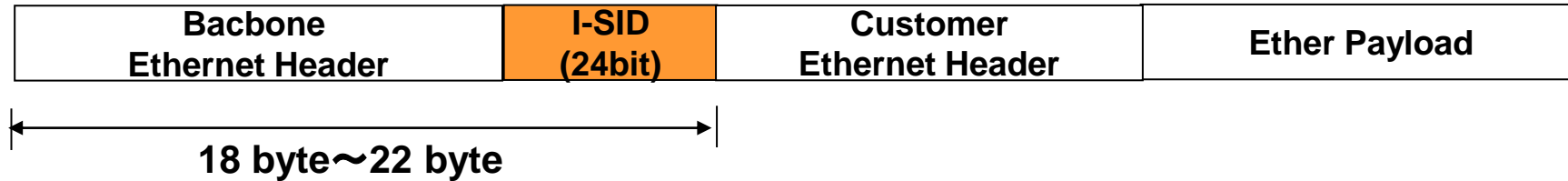
【参考2/2】IETF82でのDC Related Discussion.. (Solution Statement)

- L2VPN-WG関連
 - draft-mahalingam-dutt-dcops-vxlan-00 : Cisco/VMWare/Citrix/RedHat/Arista/Broadcom etc
 - draft-hasmit-otv-03 : Cisco
 - draft-sridharan-virtualization-nvgre-00 : Microsoft/Arista/Intel/Dell/HP/Broadcom etc.
 - draft-wkumari-dcops-l3-vmmobility-00 : Google/Ericsson
 - draft-ietf-l2vpn-pbb-vpls-interop-02 : Alcatel/Cisco/Verizon
 - draft-ietf-l2vpn-pbb-vpls-pe-model-04 : Alcatel/Cisco/Verizon
- L3VPN関連
 - draft-marques-l3vpn-end-system-03 : Cisco/Infinera/Juniper
- TRILL-WG関連
 - RFC6325
 - draft-eastlake-trill-rbridge-fine-labeling-01 : Huawei/Cisco/Broadcom etc.
- SDN関連
 - ONF (Open Networking Foundation)
 - SDN-BOF/ALTO-WG/NETCONF/NETMOD-WG

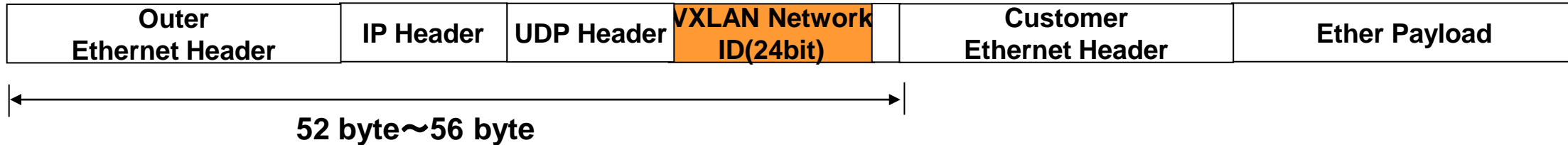
【参考】Header拡張方式の比較

- ビット拡張は24bit(1600万個余り)で共通。いずれもEtherフレームをカプセル化。
- オーバヘッドの大きさは各方式で異なる。

(1) PBB



(2) VXLAN



(3) NVGRE

