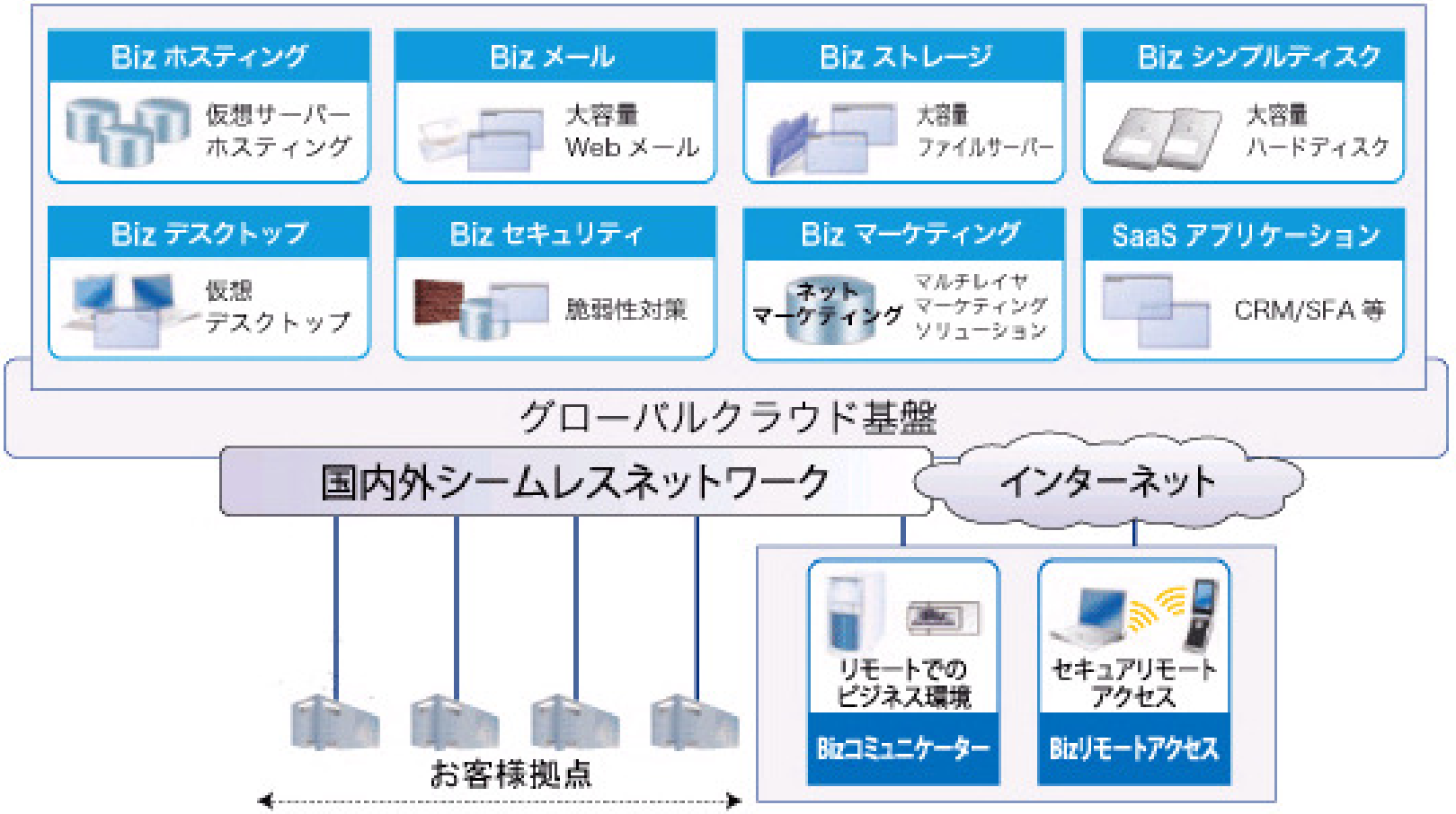
データセンタネットワーキング　Discussion

NTTコミュニケーションズ
2011年10月24日
池尻 雄一

NTTコミュニケーションズのクラウドサービス

Global ICT Partner
Innovative. Reliable. Seamless.
NTT Communications

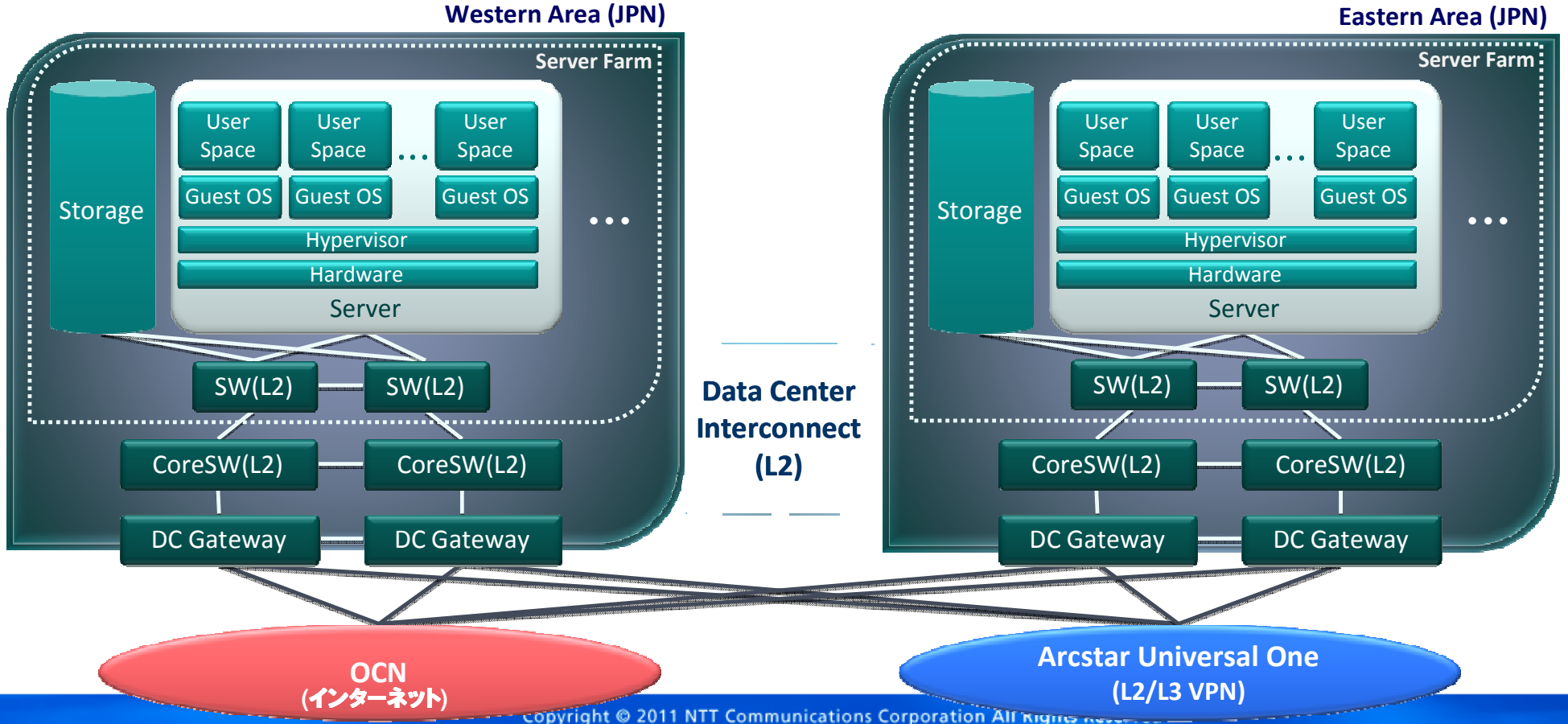
■ BizCITYのブランド名にてネットワークとセットで複数のIaaS～SaaSサービスを展開
■ その基盤としてクラウドプラットフォームを構築



クラウドプラットフォームネットワーク

Global ICT Partner
Innovative. Reliable. Seamless.
NTT Communications

- 安価なIaaSサービス（VM貸し）とPaaS/SaaS系サービスのプラットフォームとして使用
- ビル分散とインターネットとVPN双方からのアクセスを担保
- VLANセグメント分けでVM+FW+LB+DMZ（インターネット側セグメントとVPN側セグメントの接続）



クラウドネットワークの抱える課題

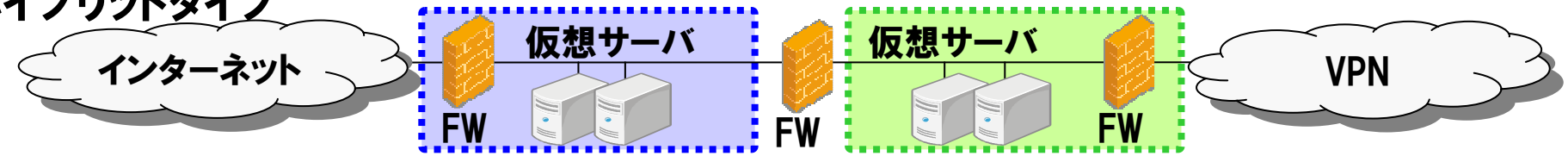
Global ICT Partner
Innovative. Reliable. Seamless.
NTT Communications

- L2セグメントの拡張(ライブマイグレーションサポート等)
- MACアドレス学習数の拡張性
- マルチテナント性の確保、拡張性
- トポロジーの柔軟さ、冗長性の確保方式
- 10Gを超える帯域のサポート
- APIサポート

L2セグメント拡張のモチベーション

- サーバ側はL2を使ったシンプルなオペレーション
- ライブマイグレーションは、G-ARPを使って同一L2セグメント内での移動
- 結果的にL2セグメントを拡張する方向になる。
- L2セグメントをひとつの物理ネットワークに重畳させながら拡大。
- マルチテナント性やセグメント分割はVLANで。

■ 例：ハイブリッドタイプ

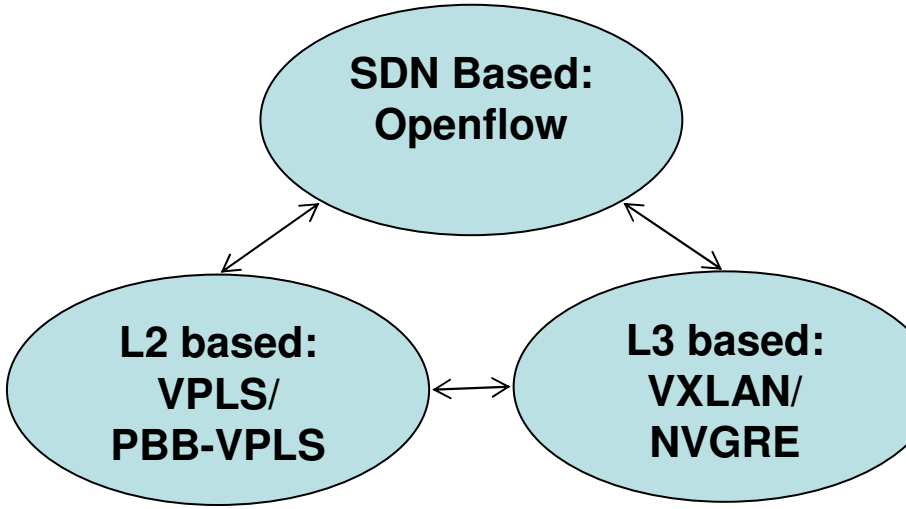


MACアドレスの拡張性

- 理論的には、クラウドプラットフォームとしては、無限のMACアドレスをサポートできるプラットフォームとしたい。
  - セグメントを分けることも含む。
- L2セグメントを極力拡張しようとするとセグメントの規模がどうしても大きくなりがち。
- L2SW(エッジToR etc.)のMACアドレス数がどこまでいるか？
  - 1Gx48等の汎用チップを使ったL2SWでは、32kサポートのSW。少ないものは16k, 8k..
  - 実際どこまで必要か。
    - 1VLANグループ(4096)であれば、4096x10vm=40k..
    - メーカの出す公式の数値と実際に持てる数値は異なる。
    - VLANが異なれば同じMACアドレスを持つことができるか。
    - 常に全VMがフルメッシュで通信するわけではないが。
  - コスト重視。

マルチテナント性確保の議論（1/2）

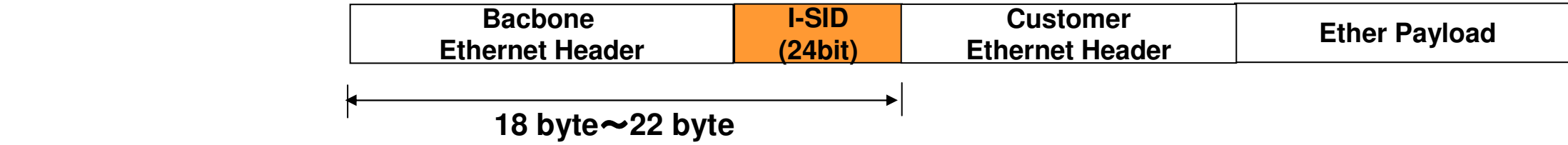
- VLAN-ID: 4096の制限。
- 最低でも4096 x nの拡張性を確保したプラットフォームとしたい、かつ共通部は論理的に束ねたい。
- これまではVLANが主流ではあったが、新しい提案も含めて乱立
  - L2系Solution
    - VPLSインスタンス拡張 × Nの実装
      - おおよそ 4096 x 8ぐらいとか。
    - PBB, PBB-VPLS(L2VPN-WG)への拡張
      - EoEによる 24bit I-SIDへの拡張
  - L3系Solution (L2 over L3トンネルの提供）
    - VxLAN(VMWare/Cisco etc.)
      - Ether over UDPカプセリングによる 24bit VNIへの拡張
    - NVGRE(Microsoft etc.)
      - Ether over GREカプセリングによる 24bit TNIへの拡張
  - Beyond L2/L3系Solution
    - Openflow(Open Networking Foundation)
      - フロー定義により既存概念にとらわれない。



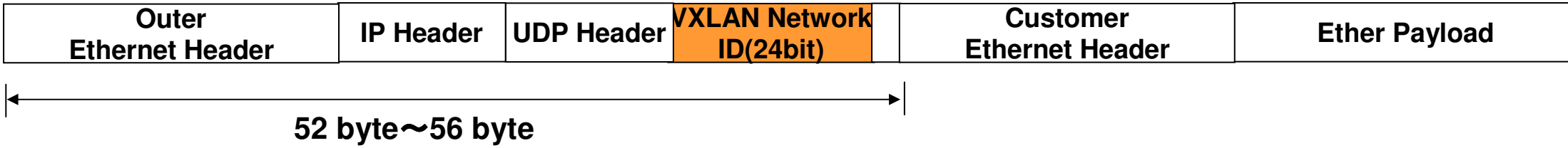
【参考】Header拡張方式の比較

■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。

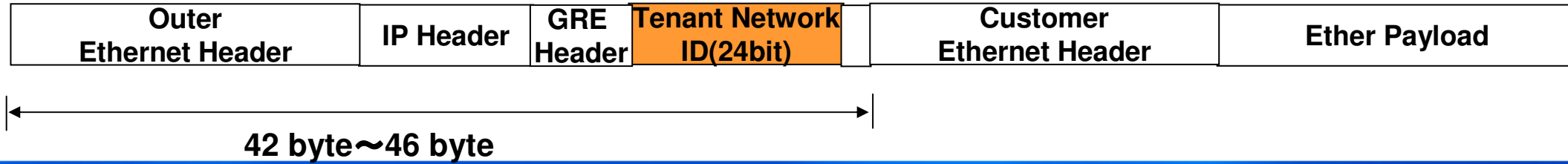
(1) PBB



(2) VXLAN



(3) NVGRE



マルチテナント性確保の議論 (2/2)

- L2系Solutionの課題
  - SW側でカプセリングを行えば、HV側は特別なものはいらない。HVに拡張機能を持たせることも可能。
  - 冗長構成のとり方はどうするか(L2の根源的問題)。
  - VPLSインスタンス　or　PBB　ISID管理
  - VLANドメイン/エリア間のL2通信　(VLAN-ID変換)
- L3系Solutionの課題
  - HV側の実装に依存。ToRに機能を持たせることも可能？
  - トンネルインスタンスの管理
  - VxLAN/NVGREもFlooding制御のためのIP Multicast運用が必要。
- Beyond　L2/L3系Solution
    - コントローラでの制御粒度の定義とコントローラソフトウェア

【参考】IETFでのDC Multi-tenancy Discussion..

- draft-mahalingam-dutt-dcops-vxlan-00
  - Cisco/VMWare etc.
- draft-hasmit-otv-03
  - Cisco
- draft-sridharan-virtualization-nvgre-00
  - Microsoft
- draft-narten-nvo3-overlay-problem-statement-00
  - IBM/Microsoft
- draft-wkumari-dcops-l3-vmmobility-00
  - Google/Ericsson
- draft-ietf-l2vpn-pbb-vpls-interop-02
- draft-ietf-l2vpn-pbb-vpls-pe-model-04
  - PBB-VPLS, L2VPN WG系
- draft-eastlake-trill-rbridge-fine-labeling-01
  - TrillでのID拡張

他にもあるかもしれません。。

トポロジーの柔軟性

- データセンタ間ネットワーク
- L3で構築した場合は特に問題はなさそう。。
- L2ネットワークの拡張で考えると物理トポロジーの制約が出る。
  - Ring構成　(G.8032, proprietary protocol)
  - Tree構成
  - MC LAG構成
- L3並の柔軟性(ルーティング)のアプローチもあるが。
  - TRILL(IETF系)・・・似て非なるTRILL実装の乱立。。
  - SPB(IEEE系)・・・実装がどこまで追いついているか。
  - 両者ともISISの拡張の点は共通だが、デファクト標準が実質的に存在していない状態。

10Gを超える帯域に向けて

Global ICT Partner
Innovative. Reliable. Seamless.
NTT Communications

- サーバIFが10Gdefaultになってくると広帯域が必要
- L2SW間で10G LAGを組む
- L2SWの40G-IFのサポートが進んできている
- DC間への適用は？
- 100Gはこの分野ではまだ実装は出ていない

APIサポート

- OpenStackなどのクラウドコントローラからネットワーク機器を制御してネットワークを同時にソフトウェア制御
- 各ベンダさんでのOpenstackのモジュール開発
- NW機器の業界共通のAPIが作れないか。
  - APIとしてのOpenflow
  - Netconf+YANG
  - ALTO-WG
  - REST-API
  - SDN BOF

  etc.

Thank You