FUJITSU
shaping tomorrow with you

L2トンネル技術によるクラウドネット

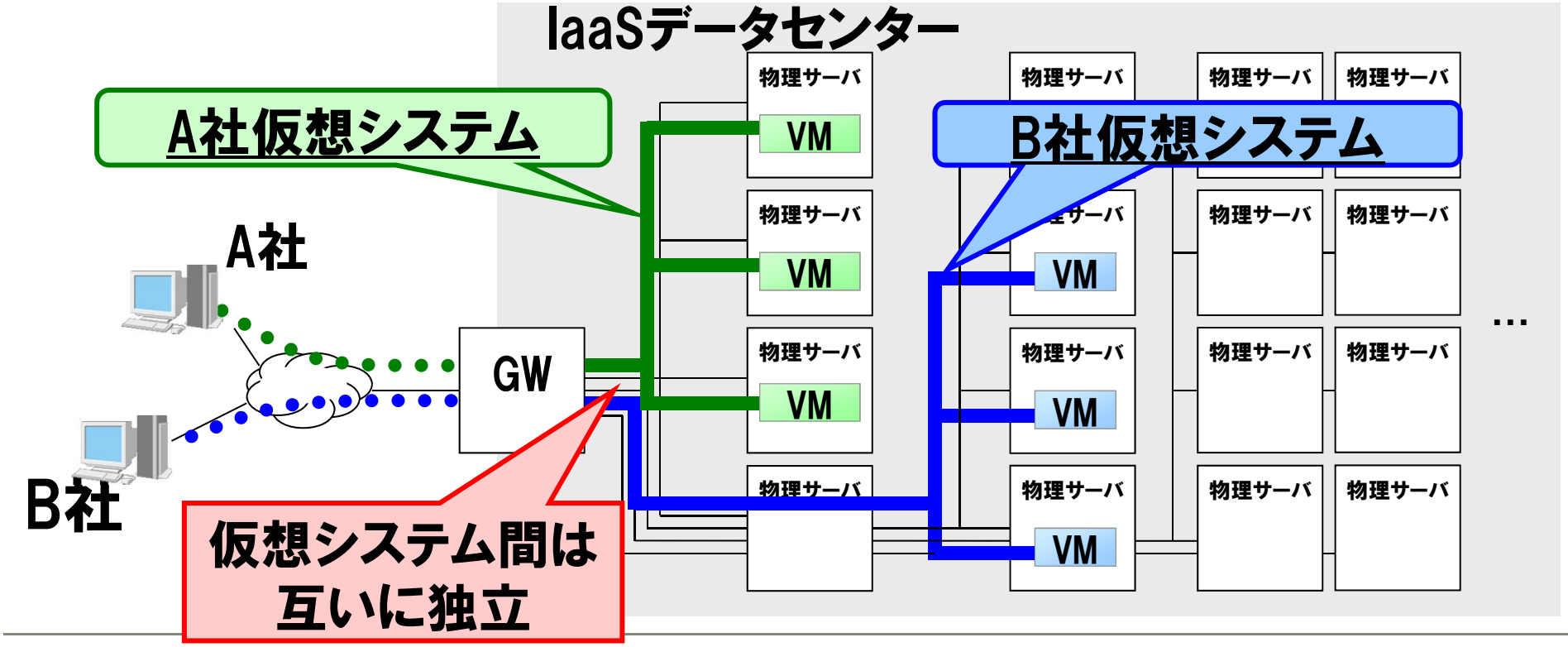
MPLS Japan 2011 パネル
「クラウド環境におけるネットワークの課題と展望」

今井 祐二

株式会社富士通研究所
クラウドコンピューティング研究センター

IaaSデータセンター

- データセンターの物理サーバ、物理ネットワーク上に、仮想マシン（VM）、仮想ネットワークからなる仮想システムを構築。
- 多数のユーザの仮想システムを集約して搭載。
- 集約によるスケールメリットを追求。



仮想ネットワークの要件

FUJITSU

- 隔離性

  ユーザ毎の仮想システムを越えてアクセスが出来てはならない。（必須）

- 柔軟性

  ネットワークの要因でVMの配置に制限が発生してはならない。

- 耐故障性

  機器故障に備えて現実的なコストで冗長構成が組めなければならない。

- スケーラビリティ性

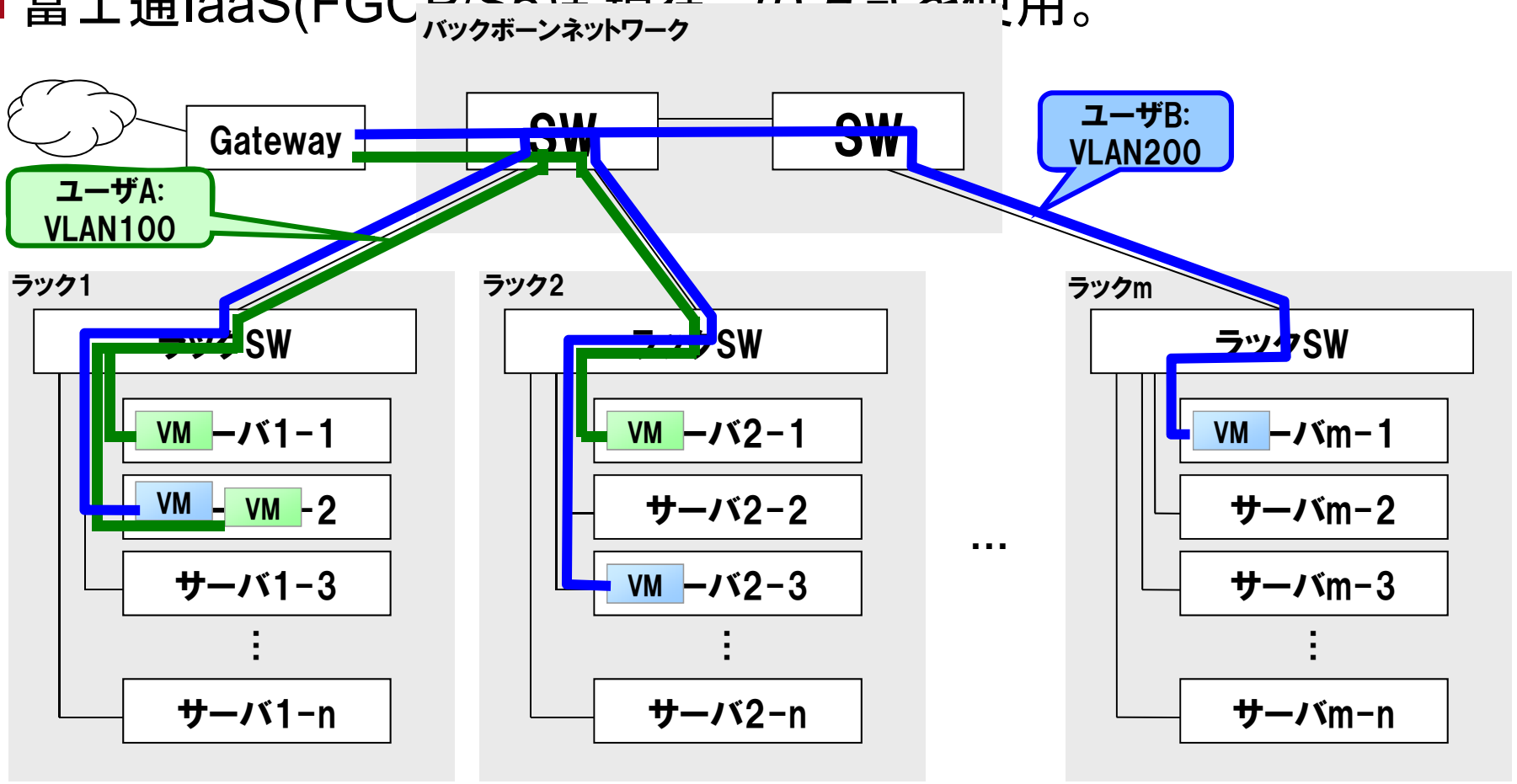
  万オーダーのVM数増加に追従できなければならない。

- コモディティ性

  ボリュームゾーンの機器を使い低コスト化でき、オペレーションコストも下げられなければならない。

VLANによる仮想ネットワーク方式

- ユーザの仮想ネットワークをVLANで構築・隔離する方式。
- データセンター内のSWやサーバ内部に、ユーザ毎のVLANを設定。
- 富士通IaaS(FGCP/S5)は現在この方式を使用。

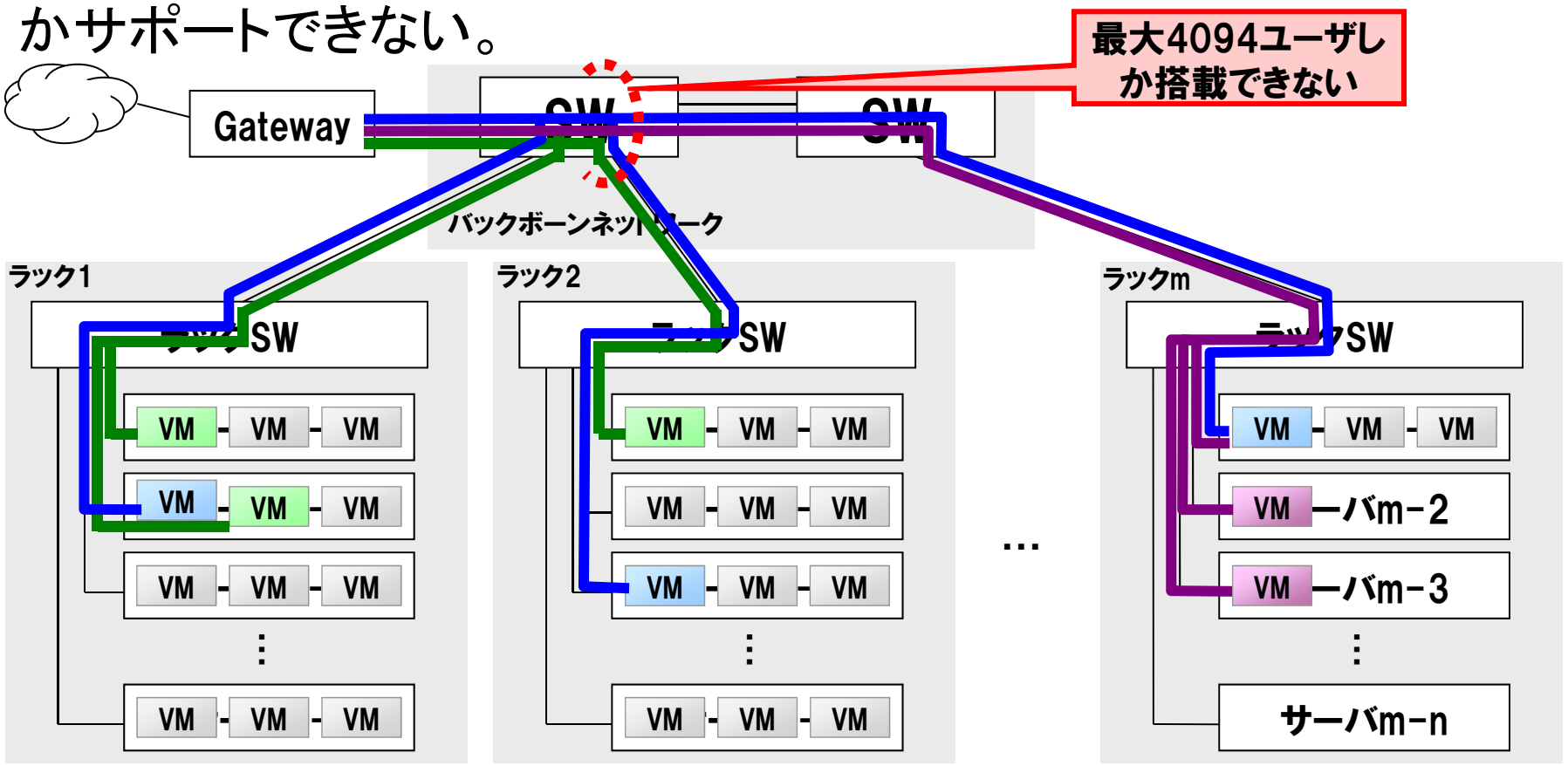


FUJITSU

クラウド内ネット基盤として VLANが抱える問題

スケーラビリティに関する問題点

- VLANのID空間は12bit(=4096)。
- データセンター内部を分割して同一VLAN-IDを複数のユーザで使ったとしても、データセンター全体としては高々4094のn倍程度しかサポートできない。

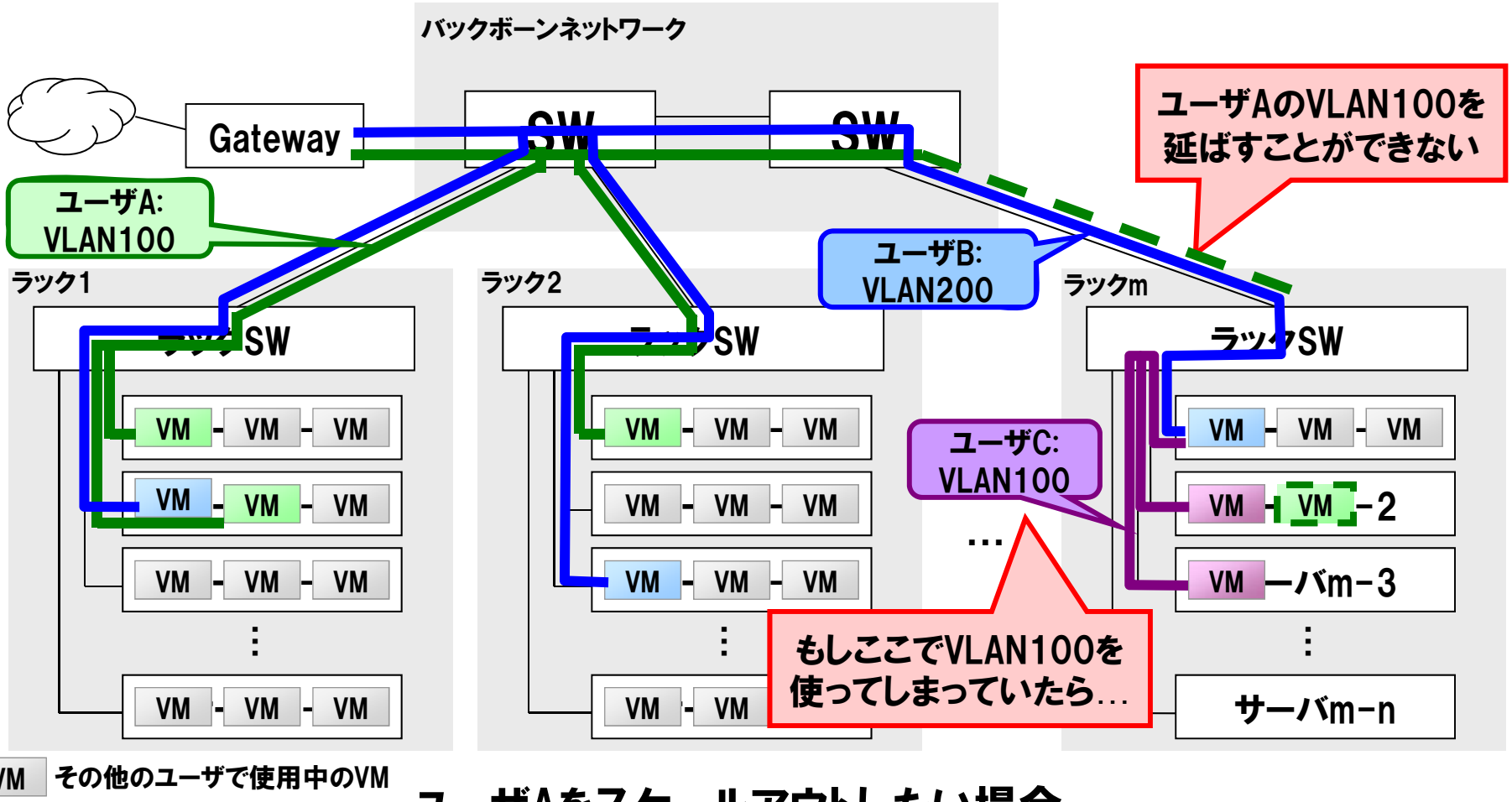


万オーダー仮想ネットに対応できない

柔軟性に関する問題点

FUJITSU

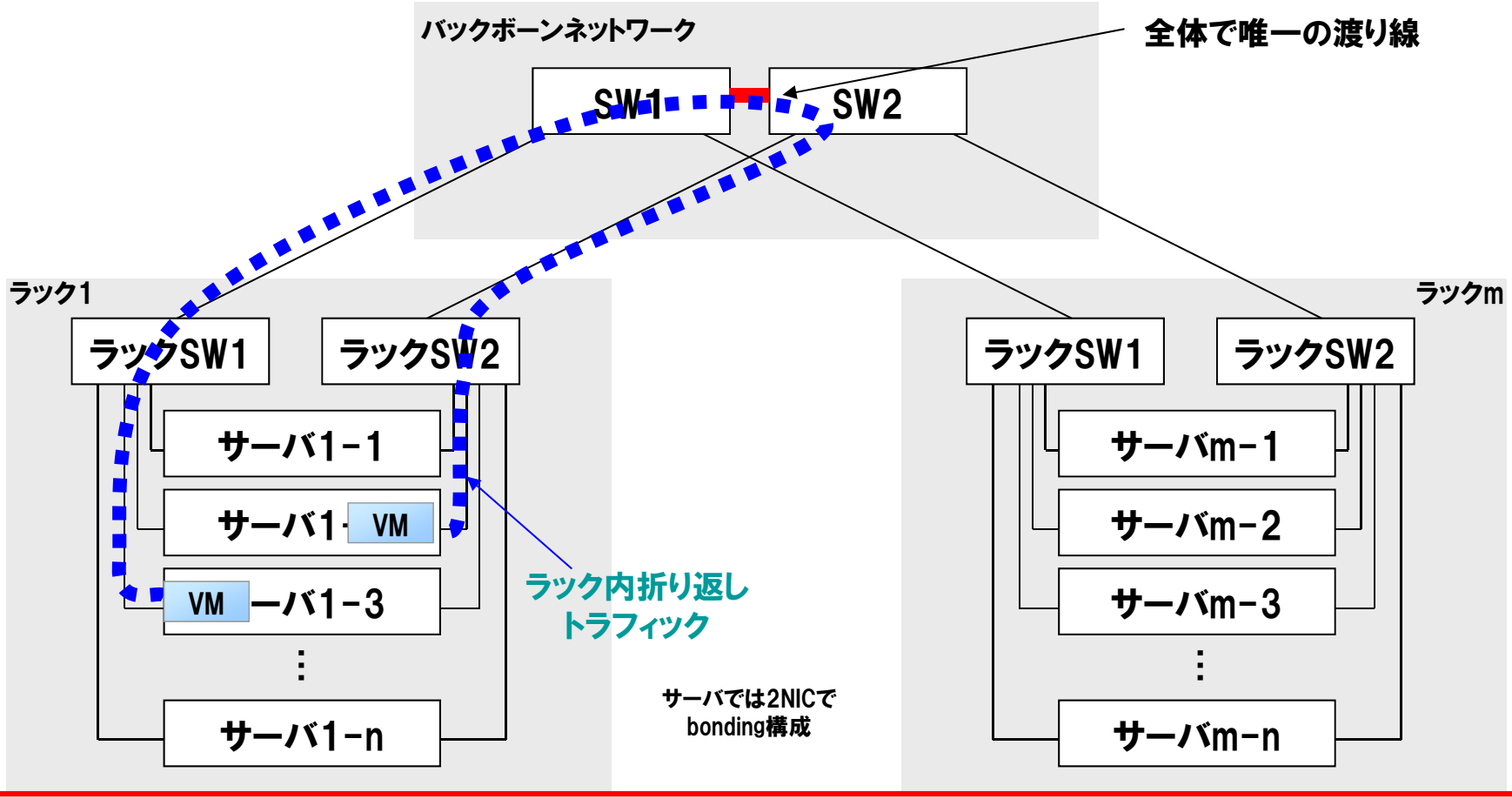
- データセンター内部を分割しVLAN-IDを重複利用する場合、スケールアウトやライブマイグレーション時に、VLAN-IDが衝突（バッティング）する場合がある。



ユーザAをスケールアウトしたい場合

耐故障性に関する問題点

- 冗長構成においてループ経路を防ぐためには、折り返し箇所（渡り線）は全体で1箇所しか作れない。
  - その一箇所の渡り線にトラフィックが集中する。



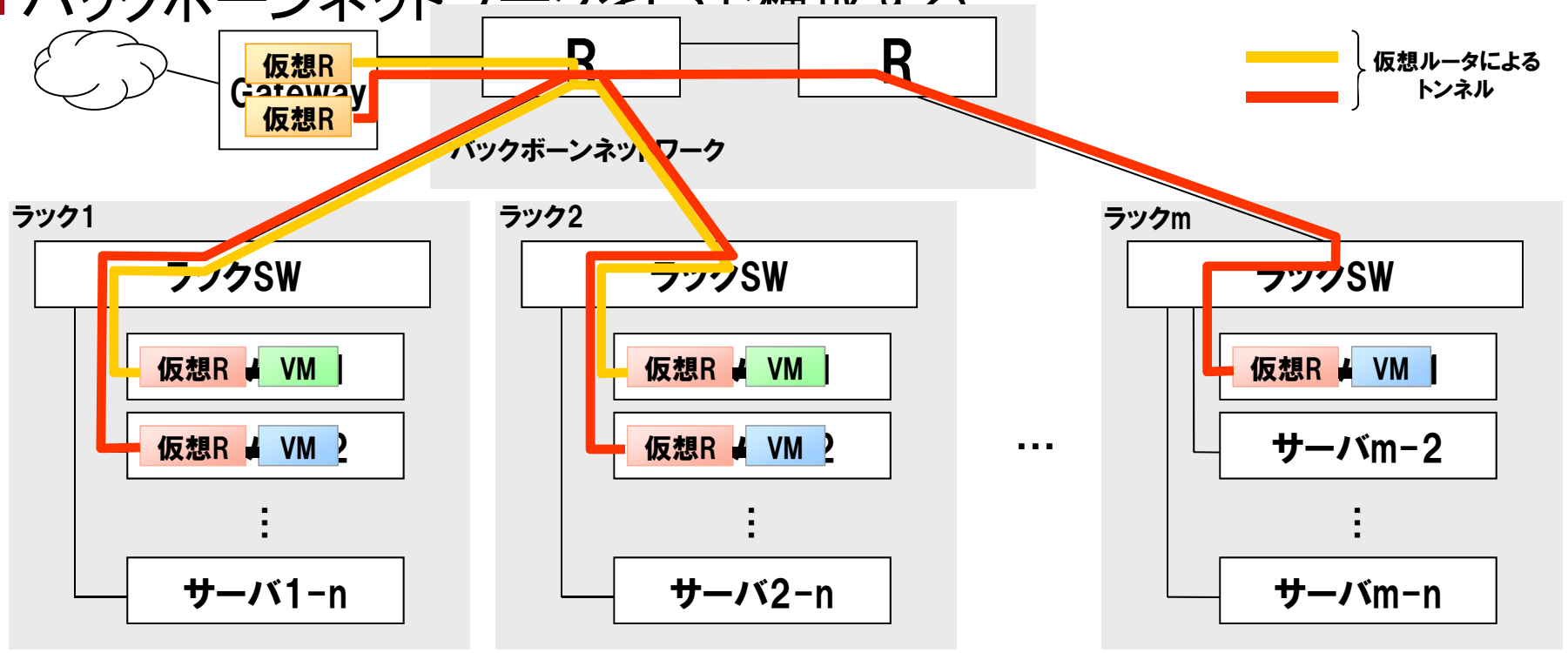
バックボーンのSWは容量の大きなものを用いなければならなくなる（＝高くなる）

FUJITSU

トンネル技術による
課題解決プロトタイプ

プロト1：VM間IPトンネル

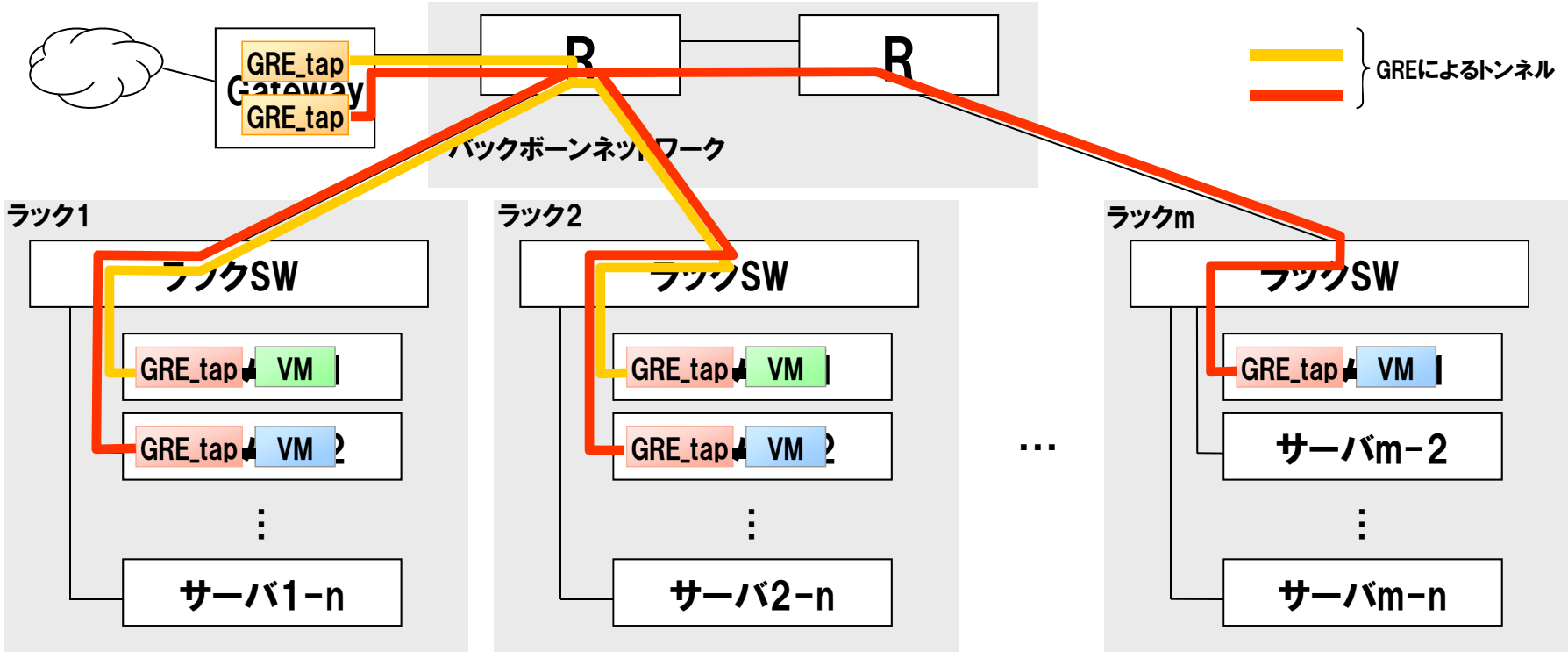
- L3 over L3トンネルでユーザVM間を接続する。
  - プロト1の時点では、LinuxにL2トンネルをブリッジする方法が無かった。
- カプセル化処理を実施する仮想ルータをVMで動かす。
- ユーザ間のL3経路混在防止のため、VMごとに仮想ルータを用意。
- バックボーンネットワークをL3で構成する



プロト2：VM間L2トンネル

FUJITSU

- L2トンネルでユーザVM間を接続する。
  - gretap(ブリッジ可能なGRE I/F)がkernelに入った。
- カプセル化処理はDom0でLinux gretapが実施する。
  - MACテーブルはbr(Linux仮想ブリッジ)ごとに持つので、ユーザ間混在なし。
- バックボーンネットワークをL3で構成する。



トンネリング方式のメリット

FUJITSU

- 柔軟性
  - 仮想リンクを自在に張れるため、ユーザVMを任意の場所に配置することが可能。
- 耐故障性
  - バックボーン部分をL3ネットワークで構築することにより、一般的な冗長化技術を用いることが可能。
- スケーラビリティ性
  - VLANタグ上限に制限されないで増強可能。
- コモディティ性
  - トラフィック量に応じたネットワーク機器を使用可能。
  - 構成を順次拡張することが可能。

各方式の評価 （定性的分析）

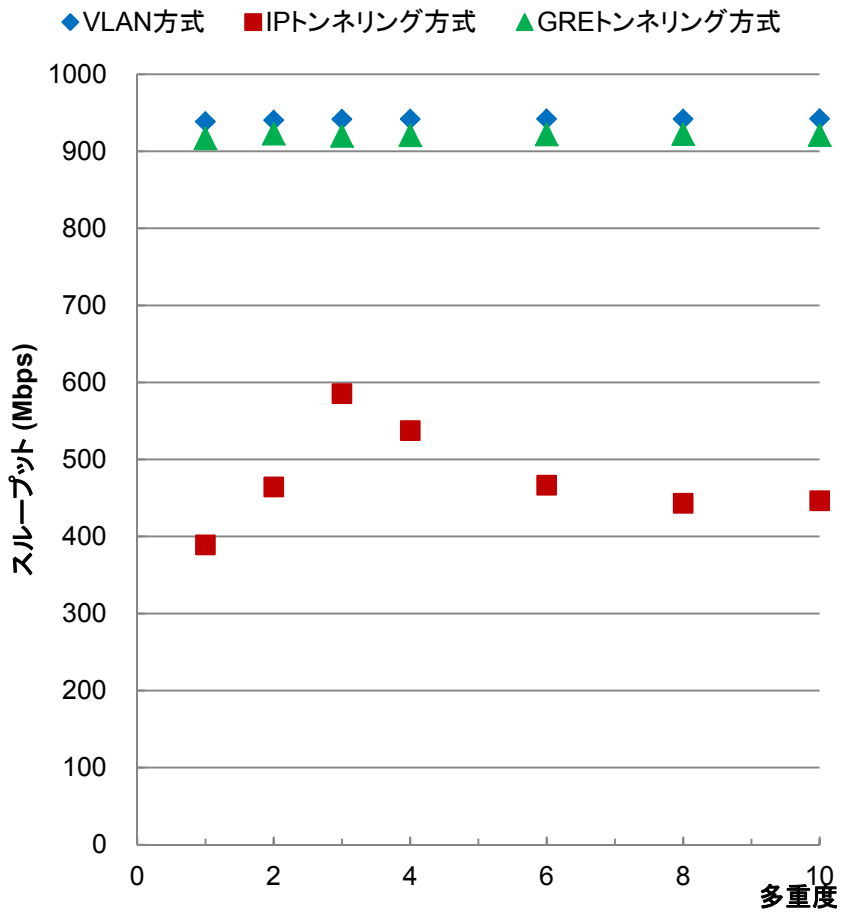
FUJITSU

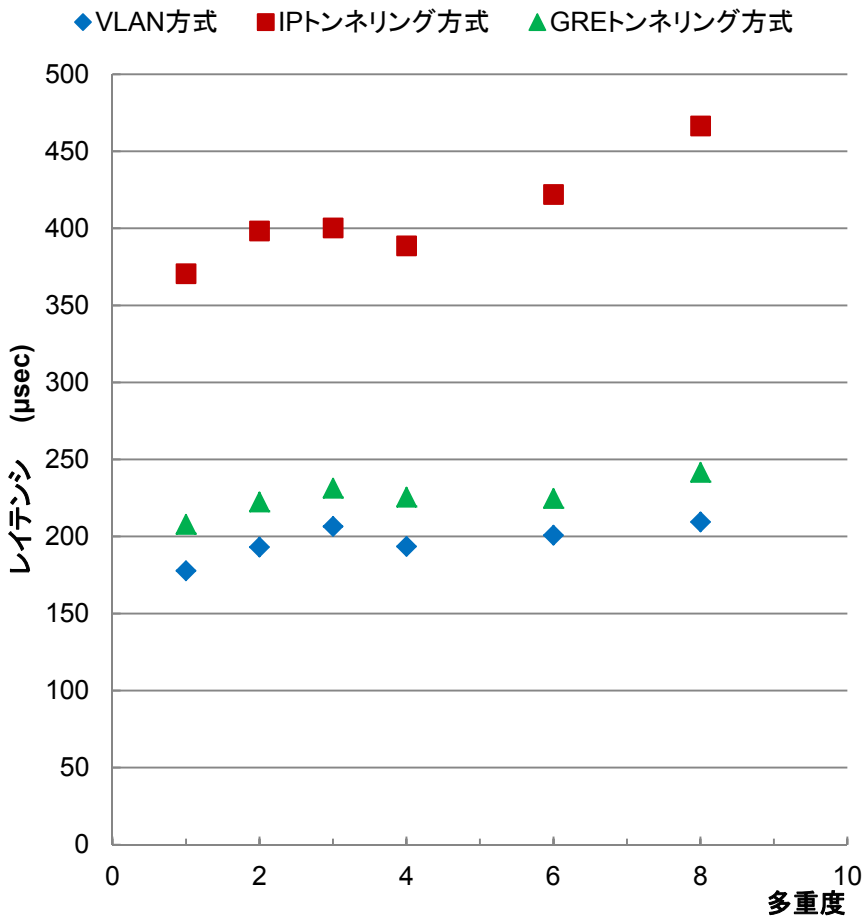
■各方式の得失の整理

評価ポイント		VLAN方式	IPトンネリング方式	GREトンネリング方式
隔離性		○ 問題なし。	○ 問題なし。	○ 問題なし。
柔軟性（VM配置の自由度）		× 高密度化を図るとVIDがバッティングする可能性あり。	○ 自由に配置可能。	○ 自由に配置可能。
耐故障性		△ 冗長構成時にトポロジーに制限がある。	○ Bonding, VRRP, OSPFによる完全二重化。	○ Bonding, VRRP, OSPFによる完全二重化。
スケーラビリティ		× VIDによる上限（4094）がある。	○ 制限なし。	○ 制限なし。
コモディティ性	機器コスト	× バックボーンのSWが高くなる。	○ 必要に応じたポート数、能力のSWでよい。	○ 必要に応じたポート数、能力のSWでよい。
	オペレーションコスト（可読性）	○ VLAN技術は一般的。	△ 構成が複雑。	○ 一般的技術のみ。

性能評価

- IPトンネル方式は、顧客L3経路区分のため仮想ルータをVM実装
  - パケットがHypervisor, ホストOSを複数回通過。約50%の性能劣化
- L2トンネル方式ならDom-0・ホストOSでパケット処理が完結。

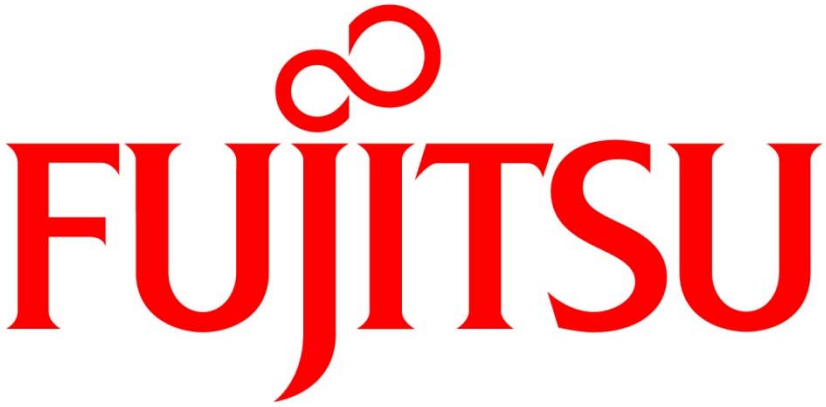




今後の注目ポイント

FUJITSU

- IETF標準化へ
  - VXLAN: draft-mahalingam-dutt-dcops-vxlan-XX
  - NVGRE: draft-sridharan-virtualization-nvgre-XX
  - NVO3 BoF(?) → L2VPNからbranchしたNVO3 WGができる方向
    - draft-narten-nvo3-overlay-problem-statement-XX
- OS, ミドルウェアの対応
  - IaaSミドルウェア
    - Eucalyptus, OpenStack, VMware, Citrix, MS, Fujitsu,,,
  - ハイパーバイザ・ホストOS
    - Linux, Xen, KVM, Hyper-V
  - 仮想スイッチ
    - Open vSwitch, Nexus 1000v
  - 監視・観測ミドルウェア
- トンネル方式の運用ノウハウ蓄積
  - 具体論はこれから。（設計・レビュー・障害対応…）



shaping tomorrow with you