複数システムのサーバー統合のためのセキュリティーと帯域効率化を実現するネットワーク共有

吉田 大祐

Network Sharing for Multiple Application Systems Consolidation, with Security and Bandwidth Efficiency

Daisuke Yoshida

ハードウェア・リソースの効率化を目的として分散系サーバーの統合を行う際に，多数の業務システム群を同一のサーバー筐体に統合するためには，サーバーのみならずネットワークについての構成検討が不可欠である．本稿では，IBM Power Systems™ の仮想ネットワークを Virtual LAN（VLAN）で分割してイーサーチャネルで外部とブリッジすることによって，サーバーと共にネットワーク機器も統合する構成を提案する．この構成を採ることによって複数業務システムでネットワーク機器を安全に共有するとともにネットワーク帯域を効率的に使用できるため，ネットワーク機器のハードウェア・コストを削減することが可能となる．

When you plan the consolidation of distributed computing servers, in order to improve hardware resources efficiency, it is necessary to investigate not only servers but network configuration, so that you can consolidate multiple application systems into the same server hardware. In this paper, a configuration is suggested which can consolidate network devices. In this configuration, a virtual network inside IBM Power Systems is divided by a VLAN (Virtual LAN) and bridged to the outside with EtherChannel. This configuration enables us to reduce costs for network hardware because, under this configuration, multiple application systems can share network devices securely and use network bandwidth efficiently.

Key Words & Phrases：VLAN，IEEE802.1q，イーサーチャネル，仮想 I/O サーバー，SEA
VLAN，IEEE802.1q，EtherChannel，Virtual I/O Server，SEA

1. はじめに

分散系システムにおいては，業務ごとに複数のサーバーを導入して構築することが一般的となっているが，そのためのサーバー群を集約・統合して台数を減らす，いわゆるサーバー統合が，近年多くのお客様から注目を集めている．これは，仮想化技術を取り入れてサーバー統合を実現することにより運用管理コストを削減するのみではなく「TCO 削減」「リソースの有効活用」「俊敏性の高いビジネス基盤の確立」「可用性の向上」といったメリットを享受することができるためである［1］．

複数業務システムの分散系サーバー群を同一のサーバー筐体に統合するためには，サーバーだけではなくネットワークについての構成検討が不可欠である．従来の分散系では，図 1 に示すようにシステムごとに個別のネットワーク機器を準備してサーバーと接続することが一般的であった．このトポロジーは不要なシステム間通信が許可されないという点でセキュリティー的には望ましいものである．しかし，この考え方を保ったままで複数システムのサーバー群を統合しようとすると，システム拡張の柔軟性の観点と帯域の効率性について問題が生

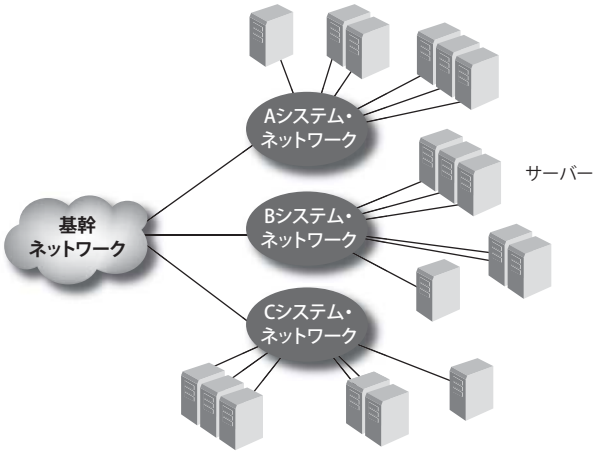


図1. 業務システムごとのサーバー導入

提出日：2008年9月8日　再提出日：2009年6月10日

じる．サーバーが物理的なネットワーク機器に縛られてしまうために仮想化の足かせとなるためである．

本稿では，IBM Power Systems（以下，POWER サーバー）の仮想化技術を使用してサーバーを統合するとともにネットワーク機器も統合する構成を提案する．この構成は，POWER サーバーの仮想ネットワークを Virtual LAN（VLAN）で分割してイーサーチャネルで外部とブリッジすることによって実現される．この構成では，複数業務システムがそれぞれのネットワークの独立性を保ったままで 1 つのネットワーク機器を共有できるためシステム構成に柔軟性が生まれ，さらにネットワーク帯域を効率的に使用することもできるようになる．またネットワーク機器のハードウェア・コストを削減することも可能となる．

2. サーバー統合とネットワークへの課題

一般に，異なるシステムのサーバー同士が同じネットワークに所属するのはセキュリティー上好ましくない．そのため，互いに異なる業務システムに属するサーバー群を同一のサーバー筐体に統合するためには，ネットワークについての構成検討が不可欠である．サーバーを統合する場合でも，ネットワークはシステムごとに独立性を保っている必要がある．

システムごとのネットワーク独立性を保証しながらサーバー統合を実現する手法の 1 つとして，図 2 に示すように，システムごとに個別のネットワーク・ポートを割り当てる構成が考えられる．この構成では，論理パーティション（LPAR：Logical PARtition）が筐体外と通信を行う際に使用するネットワーク・ポートは同システムの LPAR 同士で共有しており，別のシステムの LPAR が同じネットワーク・ポートを使用することはない．そのためシステムごとのネットワーク独立性を保ったまま複数 LPAR でネットワーク・ポートと帯域を共有することが可能である．

しかし，もし大規模なサーバー統合を考えるならば，この構成には問題点があると考えられる．その 1 つがリソース効率の問題である．一般に分散系システムでは，あるシステムの負荷が高い時間帯には，そのシステムの多くの LPAR でネットワーク帯域を消費しようとしている可能性がある．例えば負荷分散型クラスター構成を採用しているシステムでは，そのシステムの LPAR すべてにおいてネットワーク帯域使用量のピークが同じ時間帯になるであろう．しかし，図 2 の構成では同じシステムの LPAR のみでネットワーク帯域を共有しているため，帯域使用率の高いシステムが帯域使用率の低いシステムのネットワーク帯域を利用することができない．その結果，同一サーバー筐体内に帯域枯渇寸前のネットワーク・ポートとほとんど使用されていないネットワーク・ポートとができてしまう．

もう 1 つの問題はシステム追加に伴うコストである．図 2 の構成で既存サーバーに新システムを追加しようとすると，例え既存のネットワーク設備の稼働率が低くて新システムの帯域を十分賄えるほどに余裕があるとしても，新システム用にネットワークを増強する必要がある．これにはサーバー側のアダプター増強だけでなくネットワーク・スイッチ機器の増強も含まれてしまう．このように，サーバーは仮想化によって柔軟なシステム構成ができるようになってもサーバー用のアクセス・スイッチはシステムごとに物理的に準備する必要があるのでは，仮想化による「俊敏性の高いビジネス基盤の確立」のメリットが十分に生かせないことになってしまう．

3. ネットワーク共有のシステム構成

ネットワーク帯域を有効に利用し，システム全体のリソースの効率が向上させるために，また，システムの拡張に対しても柔軟に対応できるために，サーバーとスイッチの間でイーサーチャネルと Tagged VLAN を使用してサーバー内にネットワークを統合し共有する構成を提案する．図 3 にその概要を示す．

イーサーチャネルは，2 つのネットワーク機器を複数本

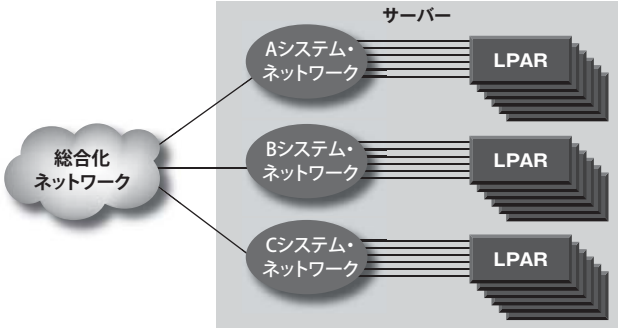


図2. ネットワーク分割のままサーバー統合

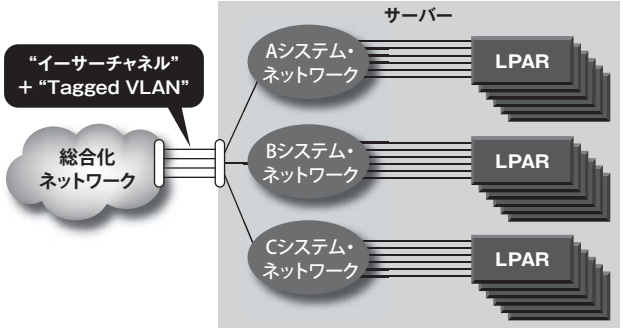


図3. ネットワーク共有構成

のケーブルで結線してロード・バランスしながら通信することで，擬似的ではあるが広帯域の実現を図る技術である．また Tagged VLAN は，ネットワーク・スイッチ間で通信されるイーサネット・フレームに対して VLAN ID を Tag として付加することで，ネットワーク機器をまたがった VLAN 分割を可能にする技術である．この 2 つが本稿の提案構成のキーとなる技術であり，提案構成ではこれを次のようにして実装する．

- POWER サーバーの仮想 I/O サーバー（以下，VIOS）によってイーサーチャネルを実装する．
- POWER ハイパーバイザーの仮想ネットワーク機能と VIOS の共用イーサネット・アダプター機能によって Tagged VLAN を実装する．
- 耐障害性のために VIOS の共用イーサネット・アダプター・フェイルオーバー（以下 SEA Failover）の機能によって共用イーサネット・アダプターの二重化を実装する．

これらの実装のための要素技術について以下に概説する．

3.1 イーサーチャネル

イーサーチャネルは主にスイッチ間接続の帯域を拡張するために使用される技術である．POWER サーバーの VIOS では，1 つのイーサーチャネルを最大 8 ポートで構成することが可能である．提案構成では，イーサーチャネルを利用して共用イーサネット・アダプターを構成する．VIOS で実装されているものであるため，LPAR からはイーサーチャネルを意識する必要は特にない．

3.2 共用イーサネット・アダプター

サーバー筐体内の仮想ネットワークと筐体外の物理ネットワークをブリッジする仮想デバイスが共用イーサネット・アダプターである．共用イーサネット・アダプターによるブリッジでは，物理イーサネット 1 ポートにつき仮想イーサネット最大 16 ポート（SEA Failover 制御チャネル用ポートを除くと最大 15 ポート）を結び付けてブリッジすることが可能である．また仮想イーサネット 1 ポートあたり最大 20 VLAN を使用することが可能である［2］．そのため，1 つの共用イーサネット・アダプターで 300 VLAN をブリッジすることが可能である．

当初，共用イーサネット・アダプターは BladeCenter® のスイッチ・モジュールと同等の機能を持つもの，という誤認識もあったが，実際には共用イーサネット・アダプターはスイッチ・モジュールとは異なり仮想ネットワークとのブリッジ機能しか持たない．

3.3 Tagged VLAN

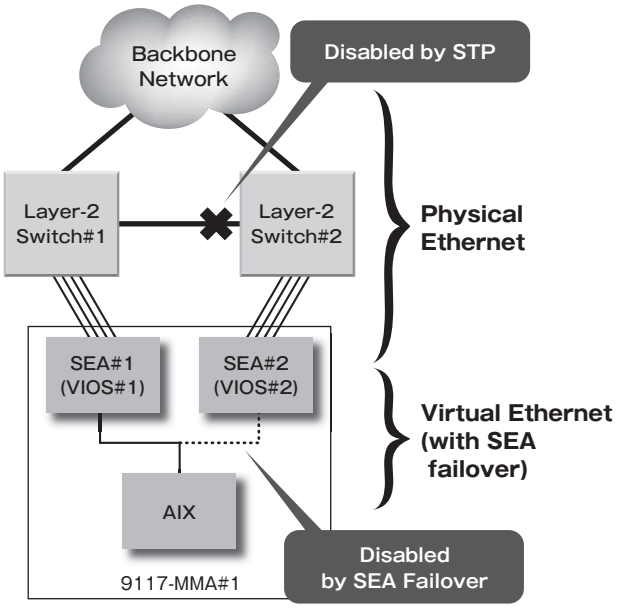
POWER サーバーの LPAR は，内部で仮想ネットワークを通じて通信を行うことが可能である．この仮想ネットワークは，IEEE802.1q で規定されている VLAN Tagging をサポートしている．これを利用して複数のネットワークにそれぞれ固有の VLAN ID を割り当てることで，システムごとに独立性を保ちながらネットワーク・ポートを共有することが可能になる．

提案構成では，仮想ネットワークの VLAN は PVID（Port VLAN ID）として設定するため，Native VLAN と同様 LPAR からは VLAN 構成を特に意識する必要はない．また IBM PowerHA™（旧 HACMP）for AIX® によるクラスター構成への影響もない．

3.4 SEA Failover

共用イーサネット・アダプターは VIOS ソフトウェアによって実装されている．そのため，共用イーサネット・アダプターを VIOS ごと冗長構成にすることで，片系 VIOS の障害により片系ネットワーク・ブリッジが働かなくなった時にも LPAR と外部との通信が停止しないようにすることができる．この共用イーサネット・アダプターの冗長化手法は SEA Failover と呼ばれる．SEA Failover によるネットワーク構成を図 4 に示す．

なお共用イーサネット・アダプターの冗長化には，SEA



注：SEA = Shared Ethernet Adapter（共用イーサネット・アダプター）

図4. 共用イーサネット・アダプター冗長化

Failoverのほかにも VIO クライアントで仮想 NIC をチーミングする方法がある．例えば，LPAR で AIX を使用している場合には Network Interface Backup 機能を使用して冗長化を行うことが可能である（参考文献［2］の 4.3 節を参照のこと）．SEA Failover が登場した当初は，まだ High Availability Cluster Multi-processing（HACMP）による高可用性クラスタリングのサポートがなかったためにこの手法が採用された過去事例も弊社内にはあるが，現時点ではすでにその制約は取り払われており SEA Failover の使用に問題はない．加えて，共用イーサネット・アダプターで Tagged VLAN を使用する場合には，Network Interface Backup は利用できないとするサポート上の制約もあるため，本稿では SEA Failover の構成を採用している．

4. ネットワーク共有構成の利点と実装・保守時の注意点

一般にシステムの非機能要件としては，セキュリティー，可用性，保守容易性，管理性といった要素が求められる．本節ではこれらの観点から考察し，構成の利点や実装・保守時の注意点について説明する．

4.1 セキュリティー

本稿の提案構成では，同一筐体に収容された複数システム間のセキュリティーはそれぞれのシステムが利用するネットワークの VLAN 分割によって実装されている．仮想ネットワーク構成や物理ネットワーク機器の設定を変更しない限り，互いに異なるシステムに属する LPAR 同士が通信できてしまうことはない．これにより複数システム間のネットワーク独立性を保ったままサーバーおよびネットワークを物理的に統合していくことが可能となる．

このセキュリティーの確保こそが，VLAN を採用した構成で主張したい主要な利点である．

4.2 可用性

複数システムが同じ共用イーサネット・アダプターや同じネットワーク機器を利用する，ということは，1 つ 1 つの機器が障害を起こした場合の影響範囲がそれだけ大きくなるということである．そのため可用性についてはサーバー統合以前にも増して十分に検討する必要がある．

今回提案している構成では共用イーサネット・アダプターにイーサーチャネルを採用している．イーサーチャネルは単一ポートにリンク・ダウン障害が発生した場合には縮退して動作するため，ポート障害が発生しても共用イーサネット・アダプターの動作を継続することが可能である．またレイヤー 2 スイッチの全体障害や VIOS の全体障害の際にも，SEA Failover の機能によりブリッジが別 VIOS に切り替わる．このようにして，大規模統合サーバーとしての用途に耐え得る高可用性を提供することが可能になっている．

ただ，筆者の担当したプロジェクトにおいて，VIOS 障害への可用性を検討している中で，SEA Failover による仮想ネットワークのトポロジー変化が物理ネットワーク機器に伝わらないという問題があることが判明したため，この点について注意を促したい．一般に，物理スイッチの障害が発生した場合には STP（Spanning Tree Protocol）プロトコルや TCN（Topology Change Notification）による通信によって新しいトポロジーが全スイッチに伝播する．しかし SEA はこれらのプロトコルに対応していないため，SEA Failover による仮想ネットワークのトポロジー変化は物理ネットワーク機器には伝わらない．このため，SEA Failover 発生直後にネットワーク機器が正しくスイッチングできずに，結果としてサーバー間通信が行えなくなる現象が発生する可能性がある．筆者の担当したお客様では，全 LPAR から定期的にブロードキャスト ping をネットワークへ送出する仕組みを作り込むことでこの問題への対応を行ったが，それ以外にもスイッチの MAC テーブルのエージング・タイマーを短くすることも有効であろう．また，もし AIX V6.1 を使用しているシステムであれば APAR IZ41516 を適用することによってこの問題への対応とすることもできる．

4.3 保守容易性

一般に，1 つのサーバーに複数システムが統合されている場合，そのすべてのシステムを同時にメンテナンスのために停止するのは大変に困難である．特にシステムの対象ユーザーや運用担当チームが異なる場合にはその困難さは大きくなる．そのため複数システムを収容するサーバー統合では，サーバーや LPAR やネットワーク機器がオンラインのままでさまざまな共通コンポーネントを保守できることが，サーバー統合を成功させる上で見過ごせない要素になる．

本稿で提案している共用イーサネット・アダプターの構成でも，サーバーの物理ネットワーク・アダプターの交換はオンラインのまま実施することができる．また VIOS のソフトウェア保守を実施する場合でも，SEA Failover を利用して共用イーサネット・アダプターの手動切り替えを行うことでやはりオンラインのまま保守することができ

る．そのため，複数システムを収容する統合サーバーとしては保守容易性の観点では問題がないと考えられる．ただし具体的な手順については幾つか注意点があるため，以下に紹介する．

（1）物理アダプター交換のためにイーサーチャネル構成を変更する際には，ethchan_config コマンドの -p オプションで SEA デバイス名を指定する必要がある［4］．これを行わないとイーサーチャネルが SEA によって使用中であるため構成変更を行うことができない．それ以外の手順は AIX のイーサーチャネル構成でのアダプター交換手順とほぼ同じなので特に注意点はない．
（2）VIOS の保守の際には，保守の前にあらかじめ手動で SEA Failover を行っておく必要がある．そうして保守対象 VIOS がネットワーク・ブリッジを提供しないことを保証してから VIOS の保守を実施する．VIOS への PTF 適用でもこの方法でローリング・アップデートを行う．手動での SEA Failover 実施は，共用イーサネット・アダプターのデバイスの ha_mode 属性を auto から standby に変更することで行う．これによって変更した共用イーサネット・アダプターは事前の優先順位設定によらずにスタンバイ・モードになりブリッジを提供しなくなる．

4.4 管理性

本稿で提案しているネットワーク共有構成では主要な仮想化機能を VIOS に頼っている．そのため運用に際しては VIOS の管理が必要となる．特に各 LPAR のネットワーク使用状況について管理するためには VIOS のリソース管理が必要になる．

VIOS の一般特性として，共用イーサネット・アダプターによるブリッジングは VIOS の CPU を消費する．しかし現在の VIOS の実装では VIOS の CPU リソースのワークロード管理を行うことができない．そのため，もしも VIOS の CPU リソースが不足したらどのシステムのどの LPAR の通信にどのような影響が出るか，ということを制御したり予測したりすることができない．これは統合サーバーとしてのサービス・レベルの低下を意味する．

このような事象が顕在化することのないように，VIOS には事前に十分な CPU リソースを割り当てておく必要がある．そのために，まずサーバー筐体全体で必要な帯域を正確に見積もり，それに十分な余裕率を見込んだ上で VIOS に割り当てる CPU リソースを決定する．このような VIOS のサイジングとしては，VIOS のマニュアル［3］には 1.65GHz の POWER5 プロセッサーにおけるサイジング・ガイドがある．

なお補足だが，LPAR ごとの仮想ネットワーク通信量や VLAN ごとの通信量は，それぞれ seastat コマンドあるいは entstat コマンドを実行することで求めることができる［5］．筆者の担当したお客様では，entstat コマンドを利用してシステムごとのネットワーク流量を常時監視する仕組みを作りこんでいる．この仕組みによって，どのシステムがネットワーク帯域を酷使しているかがすぐに分かり，また事前定義されたしきい値を超えた場合には直ちにアラートが上げられるようになっている．

5. ネットワーク共有構成の検証とその考察

前節までに述べたイーサーチャネルと Tagged VLAN を利用したネットワーク共有構成を，筆者の所属するプロジェクト・チームが実際のお客様に提案するに当たり，複数システムでネットワーク・ポートをどの程度効率的に共有できるかについて事前にパワーシステムズ・テクニカル・セールスに対して実機検証を依頼して有用な結果を得た．

5.1 検証環境

検証に使用したハードウェアは次の通り．

- IBM Power 570 モデル 9117-MMA
- CISCO Catalyst 2970G

Power 570 サーバーでは VIOS を共有プロセッサー・パーティションとして構成し，動作周波数 4.7GHz の POWER6 プロセッサー 2.0 CPU とメモリー 2GB をそれぞれ割り当てた．共用イーサネット・アダプターはギガビット・イーサネット 4 本による GEC（Gigabit EtherChannel）構成として，Catalyst 2970 と接続した．イーサーチャネルの負荷分散アルゴリズム設定は，VIOS 側が src_dst_port，Catalyst 側が src_mac とした．またチャネルのネゴシエーションは PAgP および LACP のどちらも使用せず，mode=ON で明示指定した．

このネットワーク構成の下で，サーバー内部の仮想ネットワークを VLAN によって 6 つに分割し，さらに VIO クライアントとなる LPAR を合計 18 LPAR 構築した．なお VIO クライアント LPAR の OS にはすべて AIX 5.3 を使用した．

5.2 検証内容と結果

18 の LPAR のそれぞれから共用イーサネット・アダプター経由で外部へ TCP セッションを張って，16384 バイ

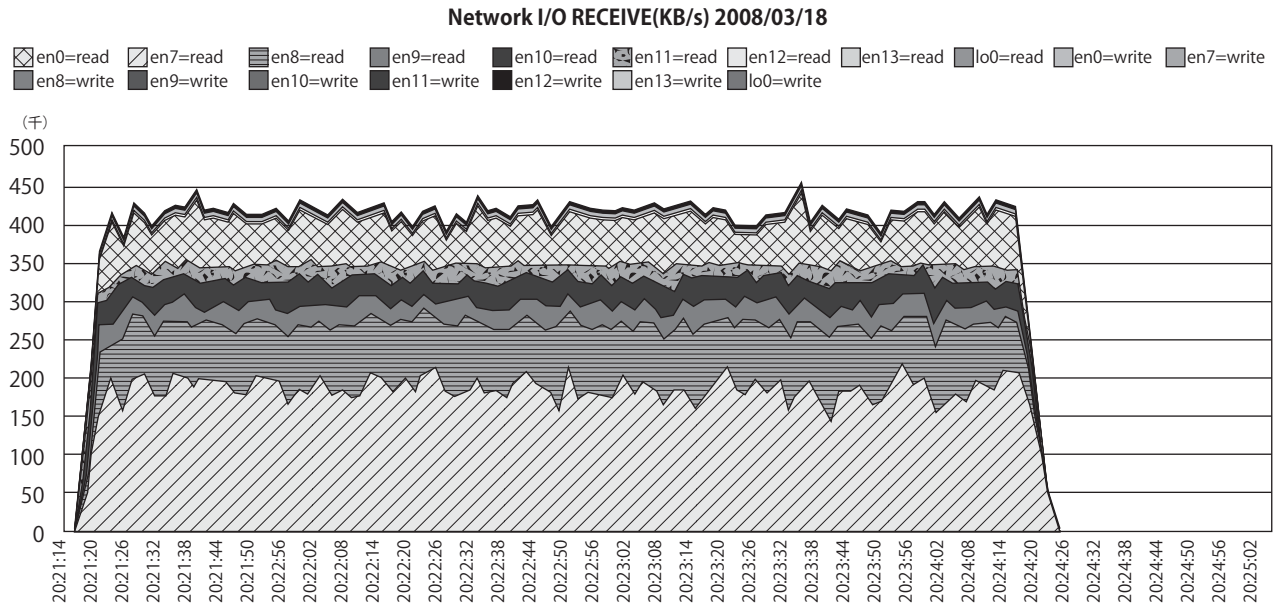


図5. 共用イーサネットの消費帯域

トのサイズのメッセージを 3 分間繰り返し送信し続ける処理を実施した．なおそれぞれの LPAR では 4 並列で処理が実施されていたので，合計で 72 本の TCP セッションが張られていたことになる．このときの帯域使用状況をグラフに表したものが図 5 である．

このグラフは，6 つの VLAN のそれぞれで測定された帯域を縦に積み上げたグラフとなっている．これによれば，この共用イーサネット・アダプター構成では全 18 LPAR の合計で約 400MB/s 以上の帯域を使用することが可能であった．これはギガビット・イーサネット 4 ポートを利用したイーサーチャネルとしては，帯域をほぼ余すところなく使用できているといえる．

なおこの時 VIOS の CPU 使用率は常に 90% 以上を記録しており，割り当てた 2.0 CPU をほぼ完全に使いきっていた．

5.3 考察

前節で紹介した検証によれば，4 本で構成された GEC では 400MB/s 程度のスループットが期待できる．従って，少なくとも 4 ポートまでの GEC ならば共用イーサネット・アダプターとして物理アダプター制限いっぱいまで帯域を有効活用することが可能であるといえる．

ただし，この構成では CPU リソースを必要とするということが注意点として挙げられる．この検証では 400MB/s を実現するために VIOS に 2.0 CPU を割り当てる必要があった．この結果は 4.4 節で述べた VIOS のサイジングに対しての有用な情報となる．もし仮に 400MB/s を超えるスループットが必要になるほど大規模な統合サーバーを構成するならば，物理アダプターを増やすと同時に VIOS に割り当てる CPU リソースも増やさなければならない．

6. ネットワーク共有の実現に向けた考慮点

大規模なシステム開発を行うプロジェクトにおいて本稿で提案している構成を採用するためには，技術的な観点ではないが，担当者の協業体制が非常に重要となる．

ネットワーク共有構成を実際に構築するためには,サーバーとネットワーク機器の両方で，イーサーチャネルおよび Tagged VLAN の設計・構築が不整合なく行われる必要がある．大規模なシステム開発ではサーバー担当者とネットワーク担当者が明確に分かれてしまっていることが多いが，その場合に本稿で提案している構成を採るためには，両担当者で十分な意思の疎通を図らなければならない．具体的には，サーバーの VLAN およびイーサーチャネルの設定をネットワークの設計に合わせるために，サーバー担当者はネットワーク担当者に対してイーサーチャネルや Tagged VLAN の構成，仮想ネットワークのトポロジーなどを説明する必要がある．また逆にネットワーク担当者は，サーバーがイーサーチャネルや VLAN トランク接続を使用する，という点についての意識を持たなければならない．ここでの意思疎通を十分にすることが統合を成功させる上でのキーポイントである．

これは筆者が実プロジェクトから得た大きな教訓の 1 つである．一言でまとめると，サーバーとネットワークを統

合するにはまず両担当者の認識も統合する必要がある，ということになろう．

7. おわりに

本稿では，POWER サーバー内の仮想ネットワークを VLAN 分割し，イーサーチャネルでブリッジする構成を提案した．これによって，複数システムを同一サーバーに統合してもセキュリティー上問題がない上に，さらにネットワーク機器を統合してネットワーク帯域をリソースとして効率的に利用することが可能になる．その上，この構成はセキュリティー，可用性，保守容易性といったサーバー統合に欠かせない非機能要件もクリアすることができており，すでに実際のお客様先での構築も進んでいる．

このようなネットワーク統合まで考慮したサーバー統合によって，今まで困難であった複数システム間でのリソース効率化，リソース共有化が実現可能となる．ネットワーク機器統合への道も開け，これによってさらなるコスト削減にもつながるであろう．

謝辞

本論文の執筆に当たっては，プロジェクト・メンバーの皆様より多くの助言をいただきました．またパワーシステムズ・テクニカル・セールスの菊地光代氏には実機検証結果を本論文で紹介することを快諾していただきました．あらためて深謝いたします．

参考文献

［1］ 濱田 正彦「仮想化の新潮流を追う」http://www-06.ibm.com/systems/jp/saiteki/library/article/vol02/（2007.06.26）．
［2］ “PowerVM Virtualization on IBM System p: Introduction and Configuration, Fourth Edition” SG24-7940-03, http://www.redbooks.ibm.com/redbooks/pdfs/sg247940.pdf（2008.05）．
［3］ 「PowerVM エディション オペレーション・ガイド」, SA88-4061-04: http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/iphdx/sa76-0100.pdf（2008.08.24）．
［4］ “IBM System p Advanced POWER Virtualization（PowerVM）Best Practices”, REDP-4194-00, http://www.redbooks.ibm.com/abstracts/redp4194.html（2006.10）．
［5］ “IBM PowerVM Virtualization Managing and Monitoring” SG24-7590-01, http://www.redbooks.ibm.com/redbooks/pdfs/sg247950.pdf（2009.03）．

日本アイ・ビー・エム
システムズ・エンジニアリング株式会社
インフラストラクチャー・アーキテクチャー
ITスペシャリスト

吉田 大祐 Daisuke Yoshida

［プロフィール］
2000 年日本アイ・ビー・エム システムズ・エンジニアリング入社．主として Power Systems のスペシャリストとして活動．