Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

ネットワーク機器の省電力化のための仮想マシン移送を考慮したトポロジ

白柳広樹[†1]　山田浩史[†1]　河野健二[†1]

データセンタではサーバ機器が消費する電力の 10〜20% の電力がネットワーク機器によって消費される．現在のネットワーク機器はネットワーク帯域の使用量に関わらず一定の電力を消費してしまう．そのため，消費電力削減のためにはネットワーク機器の電源を切る必要がある．本研究ではネットワーク機器の省電力化を実現するネットワークトポロジ，*Honeyguide* を提案する．Honeyguide は 2 つの手法: 1) 仮想マシンおよびトラフィックの集約，2) 木構造のトポロジの拡張，を組み合わせることで不要なネットワーク機器を作り出し，その電源を切ることでネットワークの消費電力を削減する．Honeyguide は耐障害性などのデータセンタの制約がある場合でも消費電力を削減でき，また既存のトポロジへの導入が容易である．Honeyguide の消費電力削減効果を示すためにシミュレーションによる測定を行ったところ．fat tree を用いた場合と比べて，最大 7.8% のネットワーク機器の消費電力を削減できた．

A VM Migration-aware Network Topology for Saving Energy Consumption in Data Center Networks

Hiroki Shirayanagi,[†1] Hiroshi Yamada[†1]
and Kenji Kono [†1]

Current network elements consume 10-20% of the total power in data centers. They consume a constant amount of energy regardless of the amount of traffic. Thus, we need to turn off unused network switches for reducing the network energy consumption. This paper presents *Honeyguide*, an energy optimizer for data center networks that increases the number of inactive switches for better energy efficiency. To this end, Honeyguide combines two techniques: 1) virtual machine (VM) and traffic consolidation, and 2) a slight extension to the existing tree-based topologies. Honeyguide has the following advantages. The VM consolidation can handle severe requirements on fault tolerance. It can be easily introduced into existing data centers. Our simulation results demonstrate that Honeyguide can reduce the network energy consumption better than the conventional VM migration schemes, and the savings are up to 7.8%.

1. はじめに

データセンタは大量のネットワーク機器で構築され，それらはサーバ機器全体で消費される電力の 10〜20% にのぼる量の電力を消費する[1]．2006 年の調査では， U.S. のデータセンタ全体でネットワーク機器は年間 1.9 億ドルの電力を消費することが分かっており，その額は年々増加している[2]．

しかし，現在のネットワーク機器はトラフィック量に応じた電力制御ができるとは言えない．理想はトラフィックが流れていない時は電力を消費せず，トラフィック量に比例して消費電力が増加することであるが，アイドル時でもかなりの電力を消費してしまう．この問題を解決するために電力効率の良いネットワーク機器を作り出す研究が行われている[3),4)]．しかし，現在のところ電力効率は理想的なものではないため，柔軟な管理を行う必要がある．

本研究では，ネットワーク機器の電力を削減するために稼働するネットワーク機器を減らすことを考える．もし，ネットワークスイッチにトラフィックが流れていなければ，電源を切ることで消費電力を削減できる．本論文ではネットワーク機器の消費電力を削減する手法， *Honeyguide* を提案する．Honeyguide はただ単に不要なネットワーク機器の電源を切るのではなく，稼働中のネットワーク機器でも 2 つの手法: 1) 仮想マシン (VM) およびネットワークトラフィックの集約，2) 木構造のネットワークトポロジの拡張を組み合わせることにより電源を切ることができるネットワークスイッチを増やす．本手法は以下のような特徴を持つ．

- **VM の移送によりネットワークの消費電力を削減**　Honeyguide は電源を切ることができるネットワークスイッチを増やすために VM の移送を利用する．VM の移送は不要な物理マシンを作り出すことで物理マシンの消費電力を削減することに多く利用されているが，ネットワーク機器の電力削減にはあまり利用されていない．ElasticTree[5)] のようなネットワーク機器の省電力化手法はネットワークトラフィックを集約することでネットワーク機器の電力を削減するが，VM の移送は活用していない．
- **冗長性を損なわない**　一般的にデータセンタでは予期せぬ障害に対処するためにネットワーク機器は冗長に構成される．そのため，電源を削減するために冗長性を犠牲にする

†1 慶應義塾大学
Keio University

Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

ことは望ましくない．Honeyguide はネットワークスイッチの電源を切っても冗長性を損なわないように不要なネットワークスイッチを作り出す．

- **既存のトポロジへの適用が容易**　Honeyguide は既存のデータセンタに簡単に拡張できるように設計されている．データセンタでは一般的に fat tree に代表される 2N の木構造のトポロジが用いられる Honeyguide はそのようなトポロジに対していくつかのリンクを追加するだけの簡単な拡張をするだけであり，導入を容易に行うことができる．

Honeyguide のネットワークの省電力効果を測定するために実際のデータセンタのワークロードを模擬したワークロードを使用してシミュレーションを行った．シミュレーションではあらゆる構成のデータセンタにおける省電力効果を調べるために，VM の台数やネットワークトポロジの規模など様々な設定を変えながらそれぞれの省電力効果について検証を行った．シミュレーションにより $k = 12$ の時の fat tree では，Honeyguide を用いずに通常通り集約を行った場合と比べて，最大 7.8% の消費電力を削減することができた．

本論文の構成を以下に示す．2 章では現在のデータセンタのネットワークについて説明し，3 章では提案システムについて説明し，4 章ではシミュレーションによる実験および評価について述べる．5 章で関連研究を紹介し，そして，6 章でまとめを述べる．

2. データセンタのネットワーク

2.1 ネットワークスイッチの消費電力

現在のネットワークスイッチはトラフィック量に応じた電力制御ができるとは言えない．すなわち、ネットワークトラフィック量に関わらず一定の電力を消費してしまう．理想的な電力制御はネットワークスイッチがアイドル状態のときには消費電力がほとんど 0 となり，トラフィック量に比例して消費電力も増加していく形である．しかし，現在のネットワークスイッチはプロセッサやファンなどの様々な機器で構成されており，アイドル状態であってもかなりの電力を消費してしまう．スイッチの消費電力量を理想的な変化に近づける研究もされているが，そのようなハイエンドな機器を導入するには非常にコストがかかるため，大規模なデータセンタに導入することは現実的ではない．

実際に 48 ポートある CISCO Catalyst 3750G のネットワークスイッチのネットワークスイッチの消費電力を測定した結果をを図 1 に示す．ネットワーク消費電力を評価するために以下の二つの方法で測定した．1) アイドル状態：接続された物理マシン間でトラフィックが全く流れていない．2) ビジー状態：接続された物理マシンのペア間でトラフィックを最大容量まで流す．図 1 から分かるようにアイドル状態でも 90W 以上の電力を消費しており，また，アイドル状態とビジー状態の時の消費電力の差が 0.25 W 程度とほとんどないため，トラフィック量に応じた電力制御ができているとは言えない．

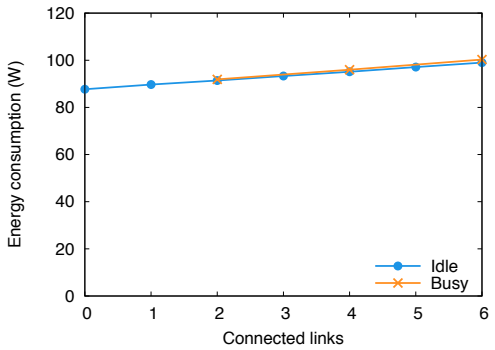


図 1　ネットワークスイッチの消費電力

そこで，本研究ではトラフィックが全く流れていないネットワークスイッチの電源を切ることで消費電力を削減する．ただ単にネットワークスイッチの電源を切るだけでなく，二つの手法をうまく組み合わせることで不要なネットワーク機器を作り出す．ひとつは VM およびトラッフィックの集約，もうひとつは木構造のネットワークトポロジにバイパスリンクを追加することである．2.2 節では VM およびトラフィックの集約について詳しく述べ，2.4 節で二つの手法をネットワーク機器の電力削減にどう当てはめるかを述べる．

2.2 VM およびトラフィックの集約

データセンタは一般的にピーク時のワークロードにも耐えられるように設計される．また，ネットワークのリンクやスイッチは障害発生時でも通信を継続できるように冗長に構成される．そのため，通常のサービス稼働時の物理マシンやネットワークのリソース使用率は非常に低く，一部の物理マシンやネットワーク機器でサービスを提供することができる．また，本研究ではデータセンタが仮想化されていることを想定する．データセンタ全体の 85% は仮想化を行っているため[6]，この想定は妥当であると言える．

データセンタのネットワークの例を図 2 に示す．図 2 は $k = 4$ の時の fat tree トポロジであり，3 つのネットワークのレイヤ：1) core, 2) aggregation, 3) edge によって構成される．この fat tree は各レイヤで冗長なリンクを持つように構成されているため，もしどれか一つのリンクやスイッチに障害が発生したとしても，別の経路を通ることで継続してサービスを提供できる．

Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

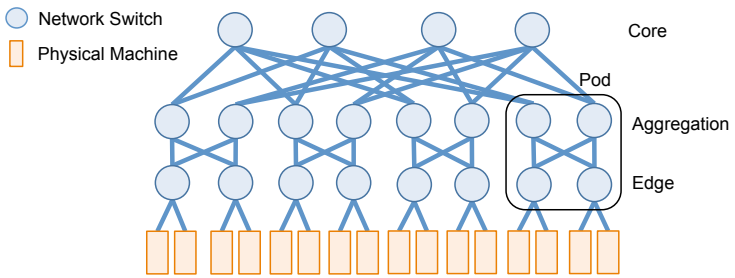
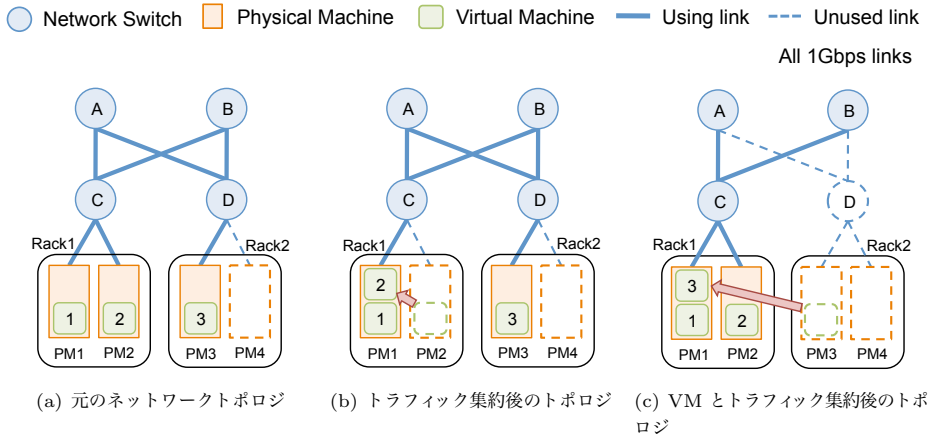


図 2 Fat tree Topology



(a) 元のネットワークトポロジ　(b) トラフィック集約後のトポロジ　(c) VM とトラフィック集約後のトポロジ

図 3 集約による消費電力削減

データセンタではネットワークだけでなく，物理マシンも同様に冗長に構成される．図 3(a) は VM が複数台配置されている場合である．図 3(a) の例では PM1 から PM4 まで 4 台の物理マシンがあり， PM1, 2, 3 には VM が各 1 台ずつ配置されている．PM4 はワークロードが増加した場合でも対処できるように VM が配置されずに稼働したままである．この場合，edge スイッチ C, D は配下に VM が配置されているため電源を入れたままにする必要がある．

VM の移送を行うことによって，データセンタのネットワークの消費電力を削減することができる．VM の移送は必要な物理マシンの台数を減らすことができ，物理マシンの消費電力削減にもつながる．本研究では VM の移送によって物理マシンおよびネットワーク機器の消費電力の両方を削減する．

本手法ではネットワークのトラフィックフローとトポロジを考慮して VM の配置を決定する．図 3(a) において PM1 には VM2 または VM3 を配置できるだけの十分なリソースが余っているとする．もし，図 3(b) のようにネットワーク機器の消費電力を削減することを考慮しないで VM2 を PM1 に移送した場合，電源を切ることができるのは PM2 のみである．しかし，図 3(c) のようにネットワークのトポロジまで考慮して VM3 を PM1 に移送した場合，PM3 だけでなく，ネットワークスイッチ D にはトラフィックは流れないためスイッチ D およびそれに接続されるリンクの電源も切ることが可能となる．

2.3 データセンタの冗長構成

ネットワークトラフィックおよび VM を集約するにあたって冗長構成を考慮しなければならない．データセンタでは，障害に対処するために VM の配置に関して様々な制約がかかる．例えば，ある VM に対して少なくとも 1 台のレプリカの VM を配置している場合を想定する．たとえどれかの VM に障害が発生しても，レプリカ VM が代わりに処理を行うことでサービスを継続して提供できる．この際，データセンタで起こりうる様々な障害に対処するためにレプリカ VM の配置先を慎重に選ばなければならない．物理マシンの障害に対処するために，異なる物理マシンに配置しなければならないことや edge スイッチの障害に対処するために異なるスイッチ配下に配置することなどが挙げられる．また，データセンタでは電源はラックごとに管理されていることが多く，電源障害に対処するために異なるラックに配置する必要がある．

VM の配置制約だけでなく，ネットワークスイッチの冗長性も考慮しなければならない．消費電力削減のために不要なネットワークスイッチの電源を切っており，ネットワークのトラフィックが増加した場合，対処するためにはスイッチの入れなければならない．もし，ネットワークトラフィックが急増した場合には，電源を入れるために時間がかかるため素早く対処することができない．また，あるネットワークスイッチに障害が発生した時にも新たなスイッチが必要となり素早く対処する必要がある．

2.4 Honeyguide の構成

本研究ではネットワーク機器の省電力化のためのネットワークトポロジ，Honeyguide を提案する．Honeyguide は 2 つの手法: 1) VM およびトラフィックの集約，2) バイパスリンクの追加，を組み合わせることでネットワークの消費電力削減する．VM を移送する際には，データセンタの冗長構成を考慮しなければならない．冗長性を損なわないようにすると制

Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

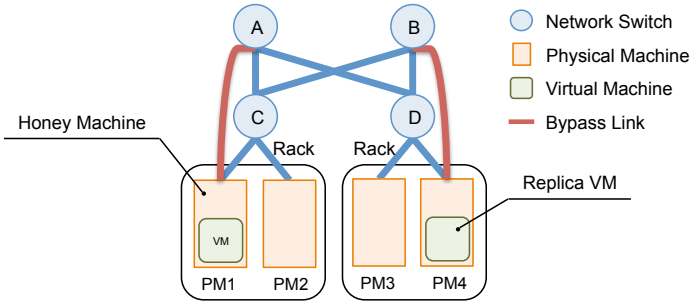


図 4 Honeyguide Overview

約によって VM を移送できない場合が生じる．そこで，Honeyguide では，制約がある状況でも電源を切ることができるスイッチを作るために，バイパスリンクを新たに接続する．

VM の移送はロードバランスや障害回避や物理マシンの電力削減に用いられる．しかし，ネットワークの消費電力の削減には用いられていない．例えば，Hermeniner ら[7] は稼働に必要な物理マシン数を減らすために制約プログラミングにより準最適な VM の配置を決定する．Verma ら[8] は物理マシンの電力効率と移送のコストを考慮して VM を再配置する．電力効率の良いネットワークトポロジ[9] は電力削減のために VM の移送は用いていない．

Honeyguide は既存のネットワークトポロジを再利用できる．既存のトポロジを新たなトポロジに置き換えることは非常にコストがかかってしまう．なぜなら，データセンタは非常に多くのネットワーク機器が複雑に接続されていたり，ネットワークのルーティングや管理システムはそのトポロジに適したもので構成されていたりするためである．そのため，適用の容易さを考えて，Honeyguide は既存の木構造のトポロジを簡単に拡張したものとした．

3. Honeyguide

3.1 Overview

Honeyguide は一般的に用いられている木構造のトポロジの物理マシンと，通常よりも上層の aggregation スイッチを接続するバイパスリンクを追加するシンプルな拡張である．図 4 に Honeyguide の全体像を示す．バイパスリンクが接続された物理マシンを *honey machine* と呼ぶ．バイパスリンクを接続することで，制約がかかってしまう状況でも VM やトラフィックを集約することができるようになる．詳細についてはこの章の後で説明を行う．

バイパスリンクが効果を発揮する状況について述べる．2.3 章で述べたようにレプリカ VM は図 4 のように異なるラックに配置する必要がある．そのため，図 4 のような状況ではレプリカ VM を元の VM がある PM1 や PM2 へは移送できないので，ネットワークスイッチ D の電源を切ることはできない．もし，バイパスリンクがあった場合ネットワークスイッチ D へのトラフィックをスイッチ B へ切り替えて，通信を行うことでスイッチ D にトラフィックは流れなくなる．トラフィックが流れなくなったネットワークスイッチ D は不要になるため電源を切ることで消費電力を削減できる．

例では，VM の配置制約がかかる場合であったが，配置制約がかからない状況でも Honeyguide はネットワークの電力を削減することが見込める．例えば，集約を行ったが計算資源が足りなかったためにラック内に VM を保持する物理マシンが数台残ってしまったような場合である．ネットワークスイッチの電源を落とすことを考えなければ，集約は行われていることになる．しかし，集約しきれなかった VM 分の空きがある honey machine が存在すれば，制約を満たしつつ，VM を honey machine に配置してバイパスリンクに切り替えることで，edge スイッチの電源を切ることができる．

このとき，edge スイッチの電源を切ってもネットワークの冗長性は損なわれない．図 4 の例では，もしネットワークスイッチ C または D に障害が発生した場合，元の VM とレプリカ VM は通信を継続できない．バイパスリンクを利用している場合について考えると，同様にネットワークスイッチ B または C に障害が発生した場合には通信を継続できない．もし，スイッチ B と D の稼働率が同じである場合，障害により通信できなくなる確率は同じになる．データセンタでは上層のネットワークスイッチになるにつれリンクが集約され重要度が高くなるため，より性能の良いものを利用するので稼働率は上層のスイッチの方が高い．そのため，ネットワークの冗長性は損なわれないと言える．

また，バイパスリンクを追加することはそれほど問題とはならない．近年，ネットワークスイッチは多くのポートを持ち，また，上層のスイッチは使われていないポートが残っている可能性が高い[10),11)]．もし上層のスイッチに空きポートがなかったとしても，そのスイッチをよりポート数の多いスイッチに入れ替えれば良いため，容易に Honeyguide を適用できる．少量のスイッチを入れ替えるコストはトポロジ全体を再構築するコストよりはるかに少なく済む．

3.2 VM とトラフィックの集約

Honeyguide は最適な VM の配置を決定するために定期的にデータセンタのトラフィックおよび VM のリソース使用量を監視し集約を行う．再配置の際に利用する情報として仮想マシンモニタによって取得できる CPU やネットワーク，メモリなどを利用する．最適な

Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

配置を計算するために定期的に情報を更新し，再配置を行う．この時，物理マシンのリソース容量を満たしつつ，冗長性の制約も満たすような VM の配置を決定する必要がある．その中で，できるだけ多くの不要なネットワーク機器を作り出せるような配置を選択する．

本研究では，再配置アルゴリズムとして広く使われている first fit アルゴリズムを利用した．これらの指標を利用して，どの物理マシンへ移送を行っていくかを決定していく．Honeyguide は移送する VM として，リソース使用率の小さい物理マシンを選択する．元々の first fit アルゴリズムは移送元としてリソース使用量が一番少ない物理マシンの VM を，移送先として移送する VM を格納できる物理マシンの中でリソース使用量が一番多い物理マシンを選択する．この行程を移送できる VM がなくなるまで繰り返す．

Honeyguide は元の first fit アルゴリズムを拡張して，できるだけ honey machine に VM を集約するようにする．なぜなら，もし honey machine 以外の VM が物理マシンに集約されていて電力を削減できる状況になった時，それらの VM をわざわざ honey machine へ移し替える必要があり，ネットワークスイッチの電源を切るまでに時間がかかってしまうためである．あらかじめ，honey machine に VM を全て集約しておけば，バイパスリンクを切り替えるだけで済み，素早く edge スイッチの電源を切ることができる．

また，Honeyguide は VM の配置制約も考慮して移送を行わなければならない．そこで，移送先の物理マシンを決定する際に，その物理マシンが属するラックにレプリカ VM があるかどうかを調べる．もし，レプリカ VM があった場合にはその物理マシンを移送先として選択せずに次の物理マシンへ移り同様に移送できるかどうか確認していく．

Honeyguide は VM の再配置手法として既存の様々な手法を適用することが可能である．なぜなら，再配置の際に honey machine に優先的に VM を移送していくように変更を加えるだけで良いためである．もし，全ての VM が honey machine に配置されている場合，バイパスリンクにより上層のスイッチを通って通信が可能なので，全ての edge スイッチの電源を切ることができる．

4. 実　　験

Honeyguide の電力削減効率を測定するために既存のデータセンタで調査されたワークロードを用いてシミュレーションによって実験を行った．ネットワーク機器の電力の削減効率を示すために VM の台数や fat tree における k の値を変更しながら，既存のトポロジで VM の集約のみをおこなった場合と比較した．ネットワークの削減効率を示すための式を以下に示す．

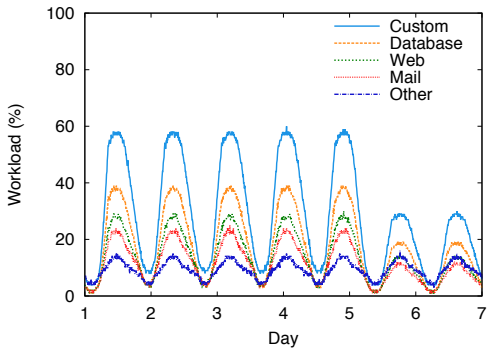


図 5　実験で用いたワークロード

$$= 100 - \frac{\text{Honeyguide を用いた場合のネットワーク機器の消費電力}}{\text{Fat tree を用いた場合のネットワーク機器の消費電力}} \times 100$$

通常のトポロジで VM の集約を行った時のネットワークの消費電力に対する Honeyguide を用いた場合のネットワークの消費電力の割合である．ネットワークの消費電力はネットワーク機器全体の消費電力の合計で表され，各ネットワークスイッチは (ベース電力＋ポート電力×稼働しているポート) × 時間 の式によって計算される．ベース電力とポート電力については 2 章で測定した値 (ベース電力: 100 W, ポート電力: 2 W) を使用した．

4.1 実験におけるワークロード

測定に際し，実際のデータセンタにおいて測定したワークロード[12)–14)] のデータを元に 5 種類のワークロード： Custom (C), Database, (D), Web (W), Mail (M), Other (O) の VM を生成した．Custom は各データセンタで稼働する独自のメインサーバ，Database はデータベースサーバ，Web はウェブアプリケーションサーバ，Mail はメールサーバ，Other は例えば LDAP のようなその他の機能のサーバである．ワークロードの時間による変化を図 5 に示す．ワークロードは時間によって周期的に変化し，昼に多く，夜は少なくなるような sin カーブに近い変化をする．

各ワークロードの時間による量の変化については実際のデータセンタで測定されたデータ[12)–14)] を参考に設定した．例えば，Database は DB2 や SQL2000 の測定データ[12)] を参考にピーク時に約 40% のリソース使用率となり，そうでない場合は約 5% の使用率とした．同様に Web は Apache の測定データ[12)] を元にピーク時に約 30%, そうでない時に約 5% となるように設定した．Mail に関しては Hotmail の測定データ[13)] を元にピーク時に

Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

表 1 実験におけるサービスとレプリカの台数の割合

	base	VM-6	VM-12	VM-24	VM-36	VM-72
C	0	6	12	24	36	72
D	0	3	6	12	18	36
W	0	3	6	12	18	36
M	0	0	3	6	9	18
O	0	0	3	6	9	18

20%，そうでない時に約 5% となるようにした．Custom と Other に関しては実データセンタで測定されたネットワーク使用量の割合[14] を考慮してそれぞれ設定した．

4.2 評価：サービス VM の数とレプリカ VM の台数を変化させた場合

Honeyguide によってどれだけネットワークの消費電力を削減できるか調べるために，まずサービス VM の数とレプリカ VM の台数を変更しながら消費電力を測定した．サービス VM とは固有のサービスを提供し，異なるサービス VM ならどのラックにも移送することができる．レプリカ VM はあるサービス VM のレプリカであり，VM の配置制約を受けるため，同じサービス VM が稼働している VM があるラックには移送できない．Heller らの実験[5] と同様の $k = 12$ の fat tree を構成して実験を行った．$k = 12$ の fat tree は 72 ラックあり各 6 台の物理マシン，合計 432 台の物理マシン，180 台のネットワークスイッチで構成される．物理マシンの台数の 2 倍の 864 台の VM を配置し，各ワークロードごとの VM の台数はデータセンタの調査[14] におけるネットワーク使用量の割合を考慮して Custom 432 台，Database 144 台，Web 144 台，Mail 72 台，Other 72 台とした．レプリカ VM の台数は網羅的に調べるために表 1 に示すように 0 台から 72 台まで変化させ，サービスの種類は各ワークロードの VM の台数をレプリカ数で割った値である．レプリカ VM が無い時を base とし，Custom のレプリカの台数にあわせてそれぞれ VM-6, VM-12 のようにした．また，honey machine の台数を 12, 24, 36, 72 と変化させて計測を行った．

結果を図 6 に示す．x 軸は honey machine の台数，y 軸は通常の集約に対する Honeyguide のネットワークの消費電力の削減率である．図 6 から分かるように Honeyguide を用いることで通常通り集約した場合よりも消費電力を削減できることが分かる．base の時に 0.8〜3.9%，VM-12 の時では 0.5〜3.9%，VM-72 の時には 2.0〜7.8% 多く消費電力を削減することができる．

4.3 評価：VM の台数を変えた場合

次に，配置されている VM の台数の違いによる Honeyguide の電力削減効果を調べるために VM の台数を変えながら測定した．レプリカの台数を固定してサービスの台数を増やすと，VM の総数も同様の割合で増加してしまい膨大な数になってしまうため，比較的規模の小さい $k = 4$ の場合で実験を行った．$k = 4$ のときトポロジは，20 台のネットワークスイッチ，8 ラック，各 2 台の物理マシン，合計 16 台の物理マシンで構成される．また，honey machine の台数を 2, 4, 6, 8 台と変えながら計測を行った．

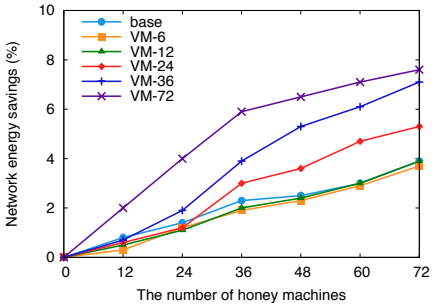


図 6 サービスの種類とレプリカの台数を変えた場合

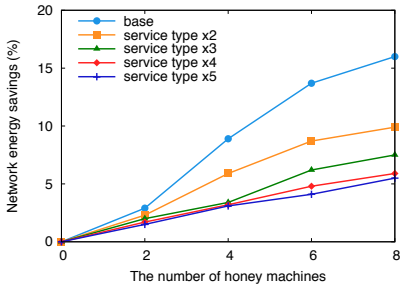
物理マシンの総数と同数のサービス数の VM を配置，各ワークロードの台数の割合は Custom 8 台，DB 3 台，Web 3 台，Mail 1 台，Other 1 台とし，これを base とした．そして，base をもとに様々なデータセンタの構成について測定するために以下の 3 種類の実験を行った．1) サービスの種類数を 2, 3, 4, 5 倍と増やした場合 2) レプリカ VM の台数を 2, 4, 6, 8 倍と増やした場合 3) サービスとレプリカ VM の数両方を増やした場合（例えば，サービスの種類を各ワークロード 1 種類増やし，レプリカを 2 倍にした場合を replica 2 & service type +1 のように表す）

図 7 に結果を示す．Honeyguide は全ての場合において電力を削減できることが分かる．サービスの数を 5 倍に増やした場合は 1.5〜5.5%，レプリカ VM の数を 8 倍増やした場合 2.1〜6.5 %，サービスの種類を各ワークロード 4 種類増やし，レプリカを 8 倍にした場合 1.1〜1.5% 消費電力を削減することができた．ネットワークの消費電力削減効率は VM の台数が増加するにつれて減少しているが，これは，VM が増えたことよりリソース使用量も増加したために honey machine に集約することが難しくなったためである．そのため，この場合，VM の台数が一番少ない base の場合が最大の効率となっている．

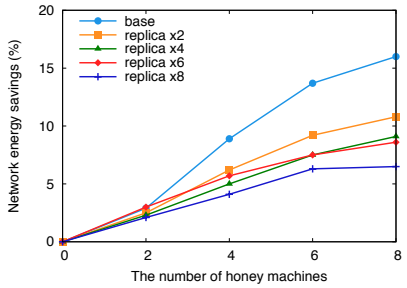
4.4 評価：データセンタの規模を変えた場合

データセンタの規模の違いによるネットワークの消費電力の削減効率を測定するために，

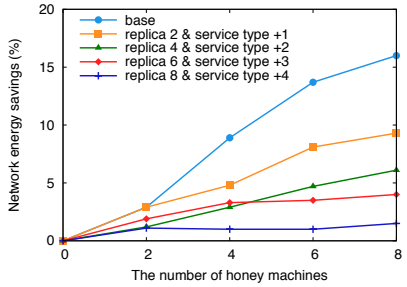
Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8



(a) サービスの種類を変化させた場合



(b) レプリカの数を変化させた場合



(c) サービスの種類とレプリカの台数を変化させた場合

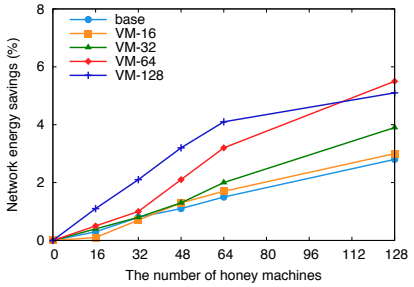
図 7 VM の台数を変化させた場合

fat tree の k の値を変更しながら実験を行った．k の値を 16, 20, 24 と変え，稼働する VM の台数を物理マシン数の 2 倍の台数とした．4.2 章の実験と同様に，サービス VM の台数とレプリカ VM の台数を変えながら実験を行った．各ワークロードごとの VM の台数は Custom $\frac{k^3}{8}$ 台，DB $\frac{k^2(k-4)}{16}$ 台，Web $\frac{k^2(k-4)}{16}$ 台，Mail $\frac{k^2}{4}$ 台，Other $\frac{k^2}{4}$ 台である．また，honey machine の台数を k, $2k$, $3k$, $\frac{k^2}{4}$, $\frac{k^2}{2}$ と変えて実験を行った．

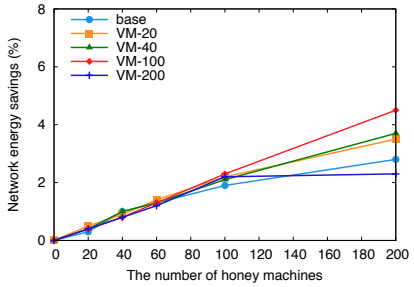
結果を図 8 に示す．データセンタの規模が異なる場合でも同様の削減効率が得られることが分かる．しかし，Honeyguide は規模が大きくなるにつれて削減効率は減少する．例えば，$k = 12$ の場合は 0.3〜7.8%，$k = 24$ の場合は 0.3〜4.6% の削減効果を得られた．これは，規模が大きくなるとラックの数も多くなり，レプリカ VM の無いラックが存在しやすくなることで，集約効率が上がるためであると考えられる．

表 2 サービスの種類とレプリカの台数の割合

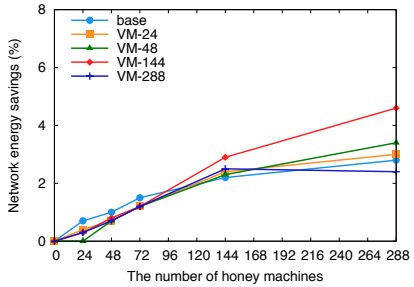
	base	VM-k	VM-$2k$	VM-$k^2/4$	VM-$k^2/2$
C	0	k	$2k$	$k^2/4$	$k^2/2$
D	0	$k/2$	k	$k^2/8$	$k^2/4$
W	0	$k/2$	k	$k^2/8$	$k^2/4$
M	0	$k/4$	$k/2$	$k^2/16$	$k^2/8$
O	0	$k/4$	$k/2$	$k^2/16$	$k^2/8$



(a) $k = 16$



(b) $k = 20$



(c) $k = 24$

図 8 k の値を変えた場合

4.5 Over-subscription

Over-subscription とは edge スイッチに通常より多くの物理マシンを接続することであり，データセンタで一般的に行われている．通常よりも多くの物理マシンを接続できるのは，トラフィックはピーク時のワークロードを想定するため通常時は非常に低く，集約して

Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

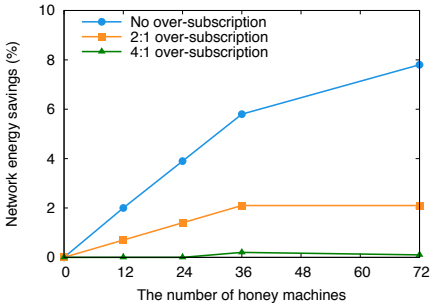


図 9 Over-subscription を行った場合

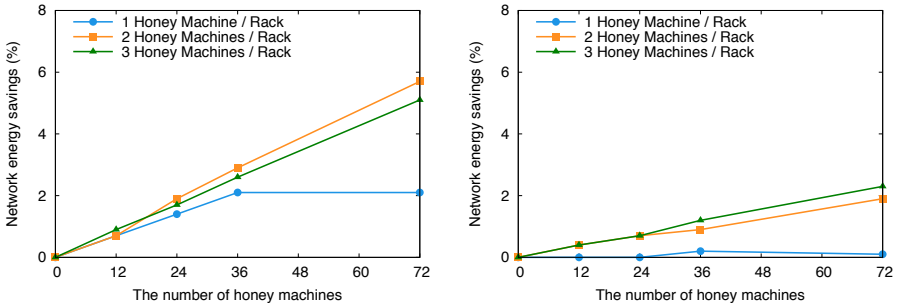
も容量内に収まるからである．例えば，通常よりも 2 倍の台数の物理マシンを接続している場合，2:1 over-subscription と表す．

Over-subscription が行われている時の省電力効果を測定するために edge スイッチに接続される物理マシンの台数を 2 倍，4 倍と増やして実験を行った．4.2 節の実験と同様に $k = 12$ の場合のデータセンタにおいて同様の割合のワークロード，honey machine の台数を 12, 24, 36, 72 と変えながら実験を行った．

結果を図 9 に示す．over-subscription の比率が上がるにつれて，ネットワークの消費電力の削減率は下がってしまっていることが分かる．over-subscription を何もしていない場合は 2.0〜7.8 % の消費電力を削減できたのに対し，2:1, 4:1 over-subscription の時はそれぞれ 0.7〜2.1%, 0〜0.2% であった．これは，over-subscription の比率があがるにつれてラック内の VM 台数も多くなるため，honey machine 1 台に集約しきれなくなったことでネットワークスイッチの電源が切れない状況が増えたためである．

この問題を解決するために，ラック内に配置される honey machine の台数を増やした．そうすることで，VM が集約しきれなくなってしまう状況が増加してしまうことを減らす．ネットワークスイッチのポート数は限られているので，honey machine の台数を 1 台増やした場合，honey machine があるラック数を 1 つ減らす．例えば，もともと 1 ラックあたり 1 台，合計 72 台接続されていて，1 ラックあたり 2 台に増やした場合，honey machine を持つラックは 36 ラックとなる．その他の構成に関しては，同様に実験を行った．

結果を図 10 に示す．ラックあたりの honey machine の台数を増やしたことで集約できずに消費電力を削減できなくなってしまう状況を解消でき，電力の削減率を上げることができていることが分かる．例えば，2:1 over-subscription の時，ラックあたり 2 台に増やした



(a) 2:1 Over-subscription　　(b) 4:1 Over-subscription

図 10 ラックあたりの honey machine の台数を変えた場合

場合の消費電力を削減率はそれぞれ，0.7〜5.7%, 0.9〜5.1% に増加した．4:1 の時ラックあたりの honey machine の台数が 2 台から 3 台に増やしたとき電力の削減率が下がってしまっている．これは，ラックあたりの honey machine の台数を増やすと，honey machine を持つラック数が減ってしまうために，honey machine が足りなくなってしまったためであると考えられる．そのため，ワークロードの割合や VM の構成によってラックあたりの honey machine の台数は調整する必要がある．

5. 関 連 研 究

データセンタのネットワークの省電力化手法として ElasticTree[5] が挙げられる．ElasticTree はネットワークのトラフィックを動的に監視し，集約することで不要な電源を作り出す．しかし，電力削減のために冗長なネットワークスイッチの電源を切るため，データセンタの冗長性を犠牲にしてしまう．本手法は，ネットワークの冗長性を損なわずに電力を削減する．また，ElasticTree と本手法は補完関係にある．

また，ネットワークの電力効率が良いネットワークトポロジが提案されている．Abts らは flattened butterfly と呼ばれる，通常より使用するスイッチの台数が少なく済む多次元のネットワークトポロジを提案している[9]．しかし，flattened butterfly のような新しいトポロジを適用することは非常にコストがかかってしまう．最悪の場合，データセンタを一から再構築しなければならない．本手法では既存の一般的な木構造のトポロジを少し変更をするだけで適用することができる．

Vol.2012-ARC-200 No.12
Vol.2012-OS-121 No.12
2012/5/8

VM の再配置に関する研究として，pMapper[8] と Entropy[7] などが挙げられる．pMapper は各物理マシンのリソースの使用率を監視し，電力，移送コスト，パフォーマンスが最小となるように VM の配置を決定する．Entropy も同様に各物理マシンの CPU 使用率，メモリ量を監視し，メモリ割当量と移送時間の関係から移送時間がが最小となるような VM の再配置方法を決定する．このような VM の再配置手法は 3.2 で述べたように集約の際に Honeyguide に優先的に配置するよう変更するだけで本手法に適用できる．

ネットワークスイッチ自体の省電力化手法も提案されている．Guranatne らは，スイッチのリンクレートをトラフィック量に応じて動的に変更することで消費電力を削減する手法を提案している[15]．また，Nedevschi らは，ネットワークスイッチがアイドルな時間を作り出し，ネットワーク機器をスリープさせることで消費電力を削減する[4]．これらの手法と本手法は補完関係にある．

6. ま と め

本研究では，冗長構成に伴うデータセンタの制約がある場合でもネットワーク機器の消費電力を削減できる Honeyguide を提案した．Honeyguide は 2 つの手法: 1) VM およびトラフィックの集約，2) バイパスリンクの追加，を組み合わせることにより冗長性を損なわずにネットワークの消費電力を削減する．シミュレーションによる実験では Honeyguide は fat tree を用いて通常通り集約を行った場合と比べて，最大 7.8% のネットワークの消費電力を削減することができた．今後の課題としては，Honeyguide を他のネットワークトポロジ，例えばメッシュやトーラスなどに適用していくことが考えられる．

参 考 文 献

1) Albert Greenberg, James Hamilton, DavidA. Maltz, and Parveen Patel. The cost of a cloud: research problems in data center networks. *SIGCOMM Comput. Commun. Rev.*, pp. 68–73, 2008.
2) Richard Brown, Eric Masanet, Bruce Nordman, Bill Tschudi, Arman Shehabi, John Stanley, Jonathan Koomey, Dale Sartor, and Peter Chan. Report to congress on server and data center energy efficiency: Public law 109-431. Technical report, Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US), August 2007.
3) M.Gupta and S.Singh. Using low-power modes for energy conservation in ethernet lans. In *The 26th IEEE INFOCOM*, pp. 2451 –2455, 2007.
4) Sergiu Nedevschi, Lucian Popa, Gianluca Iannaccone, Sylvia Ratnasamy, and David Wetherall. Reducing network energy consumption via sleeping and rate-adaptation. In *Proc. of the 5th USENIX Symp. on Networked Systems Design and Implementation*, pp. 323–336, 2008.
5) Brandon Heller, Srini Seetharaman, Priya Mahadevan, Yiannis Yiakoumis, Puneet Sharma, Sujata Banerjee, and Nick McKeown. Elastictree: saving energy in data center networks. In *Proc. of the 7th USENIX Symp. on Networked systems design and implementation*, pp. 249–264, 2010.
6) CIO Research. Virtualization in the enterprise survey: Your virtualized state in 2008, 2008.
7) Fabien Hermenier, Xavier Lorca, Jean-Marc Menaud, Gilles Muller, and Julia Lawall. Entropy: a consolidation manager for clusters. In *Proc. of the 2009 ACM int'l conf. on Virtual execution environments*, pp. 41–50, 2009.
8) Akshat Verma, Puneet Ahuja, and Anindya Neogi. pmapper: power and migration cost aware application placement in virtualized systems. In *Proc. of the 9th ACM/IFIP/USENIX Int'l Conf. on Middleware*, pp. 243–264, 2008.
9) John Kim, WilliamJ. Dally, and Dennis Abts. Flattened butterfly: a cost-efficient topology for high-radix networks. In *Proc. of the 34th annual int'l symp. on Computer architecture*, pp. 126–137, 2007.
10) John Kim, William J. Dally, Brian Towles, and Amit K. Gupta. Microarchitecture of a high-radix router. In *Proc. of the 32nd annual int'l symp. on Computer Architecture*, pp. 420–431, 2005.
11) Steve Scott, Dennis Abts, John Kim, and WilliamJ. Dally. The blackwidow high-radix clos network. In *Proc. of the 33rd annual int'l symp. on Computer Architecture*, pp. 16–28, 2006.
12) Vijayaraghavan Soundararajan and JenniferM. Anderson. The impact of management operations on the virtualized datacenter. In *Proc. of the 37th annual int'l symp. on Computer architecture*, pp. 326–337, 2010.
13) Minghong Lin, A.Wierman, L.L.H. Andrew, and E.Thereska. Dynamic right-sizing for power-proportional data centers. In *The 30th IEEE INFOCOM*, pp. 1098 –1106, 2011.
14) Theophilus Benson, Aditya Akella, and DavidA. Maltz. Network traffic characteristics of data centers in the wild. In *Proc. of the 10th annual conference on Internet measurement*, pp. 267–280, 2010.
15) Chamara Gunaratne, Kenneth Christensen, Bruce Nordman, and Stephen Suen. Reducing the energy consumption of ethernet with adaptive link rate (alr). *IEEE Transactions on Computers*, Vol.57, pp. 448–461, 2008.