Vol.2014-IOT-27 No.19
Vol.2014-SPT-11 No.19
2014/10/10

仮想ルータを使ったスイッチレス・サーバクラスタリングの考察

松本直人[†1]

現在私たちを取り巻くサーバ環境は，すでに Ethernet では 10Gbit/sec から 56Gbit/sec に達し InfiniBand では 100Gbit/sec にまで広帯域化しています。さらにサーバに導入する NIC(Network Interface Card)や HCA(Host Channel Adapter)の低価格化･普及も進んでいます．しかし広帯域のネットワークスイッチは高価であり，システム導入の障害となっています．

本稿では，仮想マシンマネージャ上に仮想ルータを導入しサーバ間をリング状に直結することで，広帯域のネットワークスイッチを導入せずコストを抑えるシステム設計についての考察を紹介します．

Switch-less server clustering using virtual router

NAOTO MATSUMOTO[†1]

This paper is introduce to analysis a Switch-less server clustering using virtual router and sharing best current practice of network design pattern.

1. はじめに

今日，サーバを取り巻く環境では，ネットワークの広帯域化が進んでいます．現在利用可能なネットワークインターフェイスカード conceptual model (NIC: Network Interface Card)の種類には 10Gbit/sec から 40Gbit/sec さらに 56Gbit/sec の Ethernet が存在しており低価格化も進んでいます．しかしながら広帯域なネットワークスイッチは高価格に推移しており，数台からの小規模なサーバクラスタリングを導入する障害となっています．

本稿では VMware vSphere ESXi, Microsoft Hyper-V, Linux KVM など仮想マシンマネージャ(VMM: Virtual Machine Manager)上に仮想ルータ(VR: Virtual Router)を導入することで高価な広帯域のネットワークスイッチを導入することなく，仮想化環境に数台からの比較的小規模なサーバクラスタリングの広帯域化を行う手法についての考察を紹介します．(図 1)

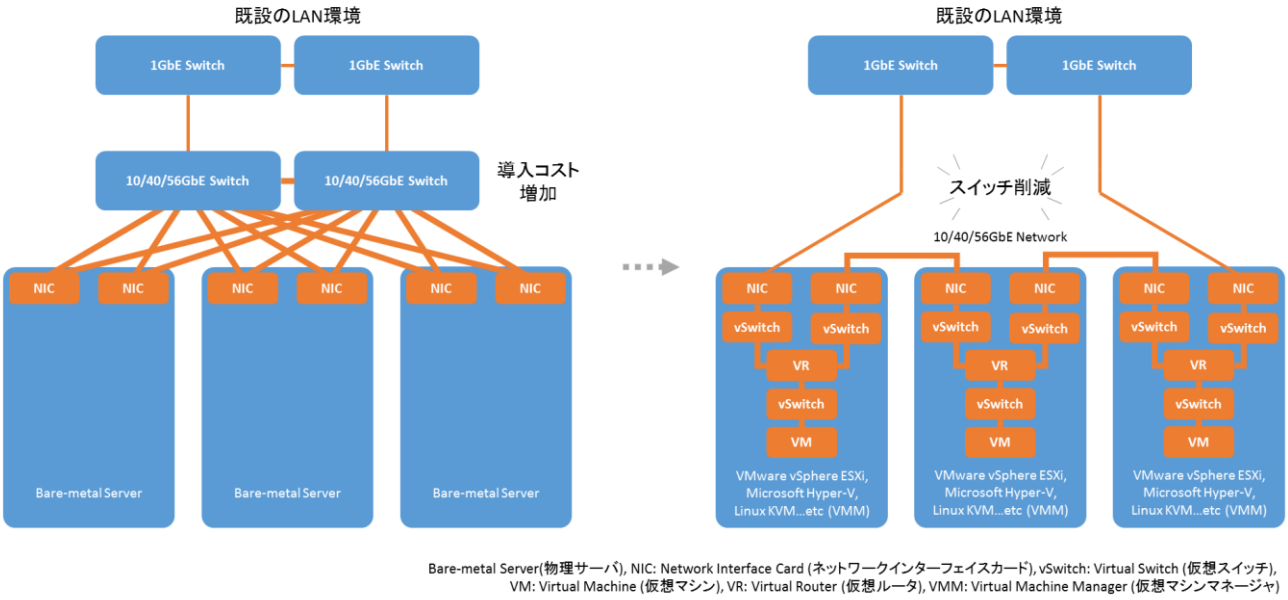

図 1 仮想ルータを使ったスイッチレス・サーバクラスタリングの概念モデル
Figure 1 Switch-less server clustering using virtual router: Conceptual model.

†1 さくらインターネット株式会社 さくらインターネット研究所
SAKURA Internet Research Center, SAKURA Internet, Inc.

Vol.2014-IOT-27 No.19
Vol.2014-SPT-11 No.19
2014/10/10

2. スイッチレス・サーバクラスタリングの考察

仮想ルータを使ったスイッチレス・サーバクラスタリングに導入するにあたり，大きく二つの異なる考え方に分かれます．それはネットワーク冗長化を行うか，行わないかです．

図 2 は，スイッチレス・サーバクラスタリングの理解を深めるために前述の図 1 を整理した抽象化モデルです．

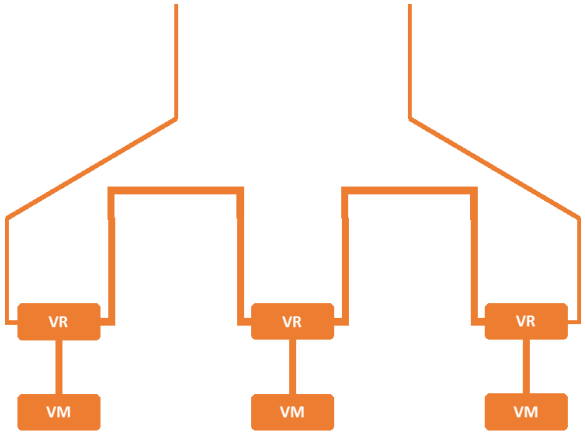

図 2 スイッチレス・サーバクラスタリングの抽象化モデル
Figure 2 Switch-less server clustering: Abstract model

VMware vSphere ESXi，Microsoft Hyper-V，Linux KVM など仮想マシンマネージャ(VMM: Virtual Machine Manager)上に仮想ルータ(VR: Virtual Router)を導入することで構築された仮想化環境であっても，IP(Internet Protocol)を用いたネットワークには変わりないことが分かります．

図 3 は，仮想ルータを使ったスイッチレス・サーバクラスタリングのネットワーク設計パターンについて整理したものです．

ネットワーク冗長化を行わない場合，仮想ルータは一本の幹に連なる形となり，ネットワークのルーティング設計も静的ルーティングのみで良くシンプルな構成になります．しかし，仮想ルータの一つで障害が発生した場合には，一本の幹に連なるすべてのネットワークに影響が及ぶため適用範囲に考慮が必要となります．システムの停止が許容できる小規模な実験環境などには適用可能ですが，システムの停止が許されないサービスに適用することは難しいでしょう．

ネットワーク冗長化を行う場合，仮想ルータは一本の輪のような形となり，ネットワークのルーティング設計も動的ルーティングが必要となります．仮想ルータの一つで障害が発生した場合には，一本の輪に連なった異なる仮想ルータを経由して通信を迂回でき，ネットワークの全体に影響が及ぶことはありません．

図 3 では，さらに仮想ルータを使ったスイッチレス・サーバクラスタリングにおいてネットワーク冗長化を行う場合の概念理解を深めるため，三台から五台の仮想ルータを用いた小規模なネットワーク設計パターンについて記述しています．

それでは，仮想ルータの設定と考え方について，仮想ルータを五台構成としたものを例としてみていきます．

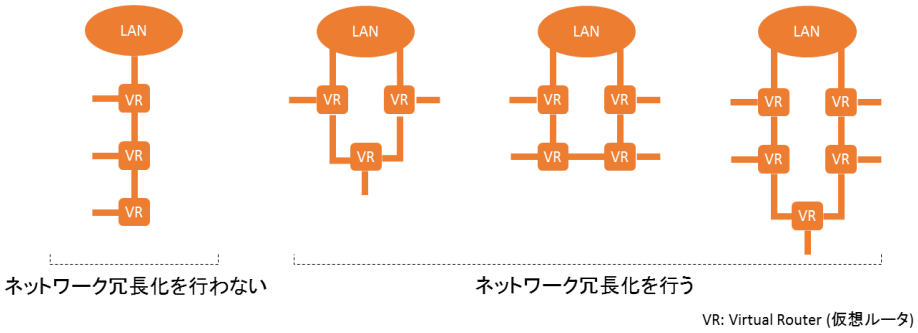

図 3 仮想ルータを使ったスイッチレス・サーバクラスタリングのネットワーク設計パターン
Figure 2 Switch-less server clustering using virtual router: Network design pattern

Vol.2014-IOT-27 No.19
Vol.2014-SPT-11 No.19
2014/10/10

3. 仮想ルータのルーティング設計

仮想ルータを五台構成したスイッチレス・サーバクラスタリングでは，ネットワークのルーティング設計として静的ルーティングと動的ルーティングを併用します．

静的ルーティングは LAN(Local Area Network)に隣接した仮想ルータのみデフォルトゲートウェイ設定を行い，その他の仮想ルータはデフォルトゲートウェイ設定を行いません．(図 4)

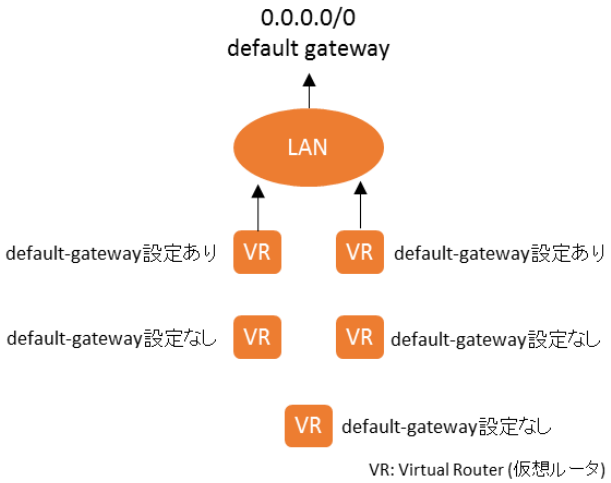

図 4 仮想ルータのデフォルトゲートウェイ設定
Figure 4 Virtual Router: default gateway config

つぎに動的ルーティングとして OSPF(Open Shortest Path First)プロトコルにより経路制御を行います．

前述の図 4 でデフォルトゲートウェイ設定を行った LAN(Local Area Network)に隣接した仮想ルータでのみ，OSPF による default-information originate による経路広告を行います．

これにより，デフォルトゲートウェイ設定を行わなかった仮想ルータに OSPF プロトコルを通じてデフォルトルートが伝搬します．(図 5)

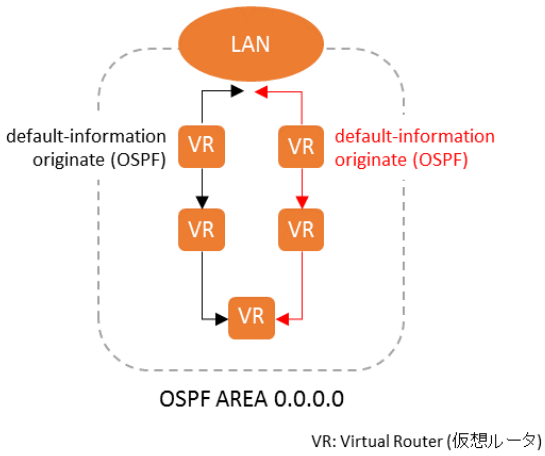

図 5 OSPF default-information originate 設定[1]
Figure 5 OSPF default-information originate config

この際異なる二つの仮想ルータから OSPF プロトコルを通じたデフォルトルートが，デフォルトゲートウェイ設定を行っていない仮想ルータへ伝搬するため，すべての仮想ルータ上に上流向けインターフェイスとバックアップ用インターフェイスに OSPF におけるコスト設定を行い，ルーティング情報を整理します．(図 6)

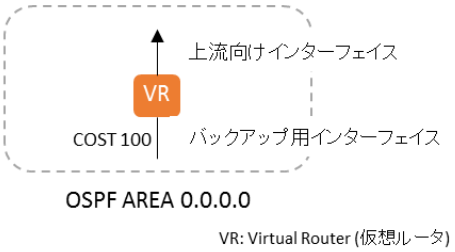

図 6 OSPF コスト・パラメータ設定
Figure 6 OSPF Cost parameter config

図 7 は仮想ルータを五台構成したスイッチレス・サーバクラスタリングにおける OSPF コスト・パラメータ設計を表したものです。

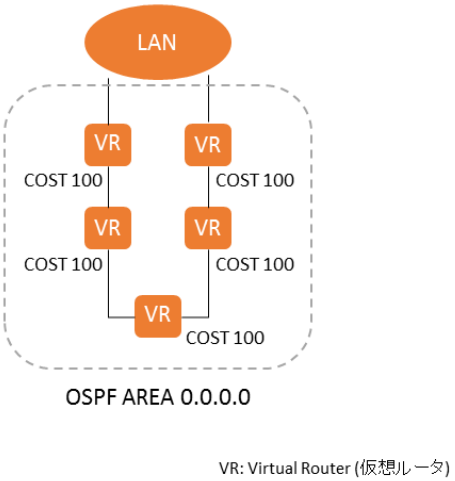

図 7 OSPF コスト・パラメータ設計
Figure 7 OSPF Cost parameter design

最後にすべての仮想ルータから管理する仮想化環境上のネットワークを OSPF プロトコルで経路広報します．(図 8)

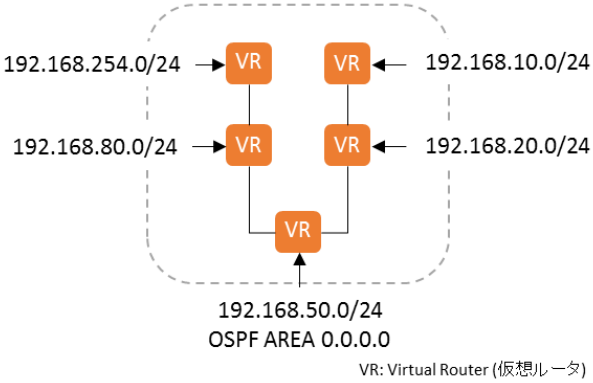

図 8 OSPF ルーティング設定
Figure 8 OSPF Routing Config

Vol.2014-IOT-27 No.19
Vol.2014-SPT-11 No.19
2014/10/10

4. 仮想ルータの冗長化

ネットワーク冗長化のために，LAN(Local Area Network)に隣接する二つの仮想ルータで VRRP(Virtual Router Redundancy Protocol)の設定を行います．仮想ルータそれぞれを MASTER と BACKUP として設定し，VIP(Virtual IP Address)を共有します．(図 9)

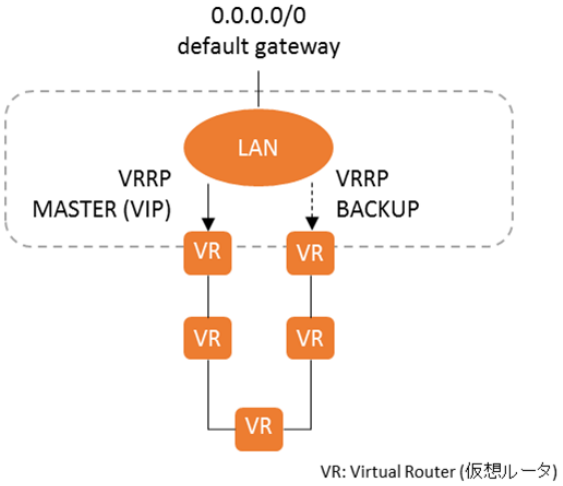

図 9 VRRP 設定 [2]
Figure 9 VRRP Config

この際、仮想ルータに LAN の上位デフォルトゲートウェイとして設定されたルータでは，VRRP で設定された VIP(Virtual IP Address)を宛先として，すべての仮想ルータが管理するルーティング情報に基づいた静的ルーティング設定を行います．

図 10 は，VRRP 設定の理解を深めるために前述の図 9 を整理した抽象化モデルです．LAN に隣接する二つの異なる仮想ルータが VRRP により，一つの仮想ルータとして機能することが分かります．

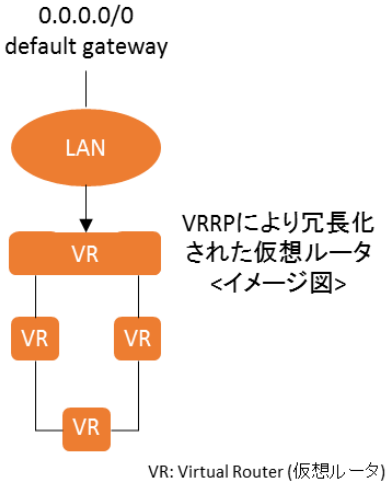

図 10 VRRP の抽象化モデル
Figure 10 VRRP: Abstract model

つづいて，仮想ルータで構成されたスイッチレス・サーバクラスタリングの有効性について動作検証を見ていきましょう．

5. 動作検証

動作環境として，一台の物理サーバで仮想マシンマネージャ(VMM: Virtual Machine Manager)として VMware vSphere ESXi 5.1 を用意し，さらに仮想ルータとして Brocade Vyatta 5400 vRouter 6.6R5 を五台と仮想スイッチを 10 台用意し動作検証を行いました．

本動作検証では，スイッチレス・サーバクラスタリングのネットワーク設計におけるネットワーク冗長化の有効性についてのみ目的としており，広帯域のネットワークインターフェイスカードを用いた検証は行っていないことにご注意ください．

図 11 は，スイッチレス・サーバクラスタリングにおいて，仮想ルータが管理するすべてのルーティング情報に関する経路制御を確認したものです．

あらかじめ OSPF プロトコルによりコスト・パラメータ設定されたインターフェイスが優先されるように経路制御が行われていることが分かります．

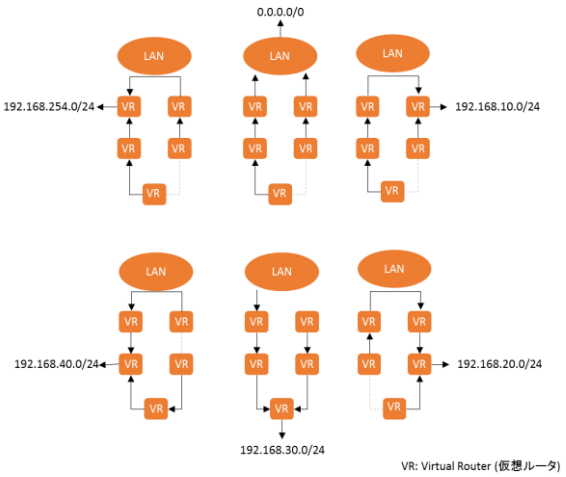

図 11 スイッチレス・サーバクラスタリングの動作検証
Figure 11 Switch-less server clustering: Benchmarks

Vol.2014-IOT-27 No.19
Vol.2014-SPT-11 No.19
2014/10/10

つづいて，仮想ルータの障害時における経路制御の動作検証についてみていきましょう．

VRRP MASTER として設定された仮想ルータが停止もしくは物理サーバのケーブル切断された際，VRRP BACKUP として設定された仮想ルータが昇格し経路制御を行います．

この際，VRRP MASTER として設定されていた仮想ルータの OSPF default-information originate のルーティング情報も，VRRP BACKUP として設定された仮想ルータの OSPF default-information originate のルーティング情報に切り替わります．（図 12）

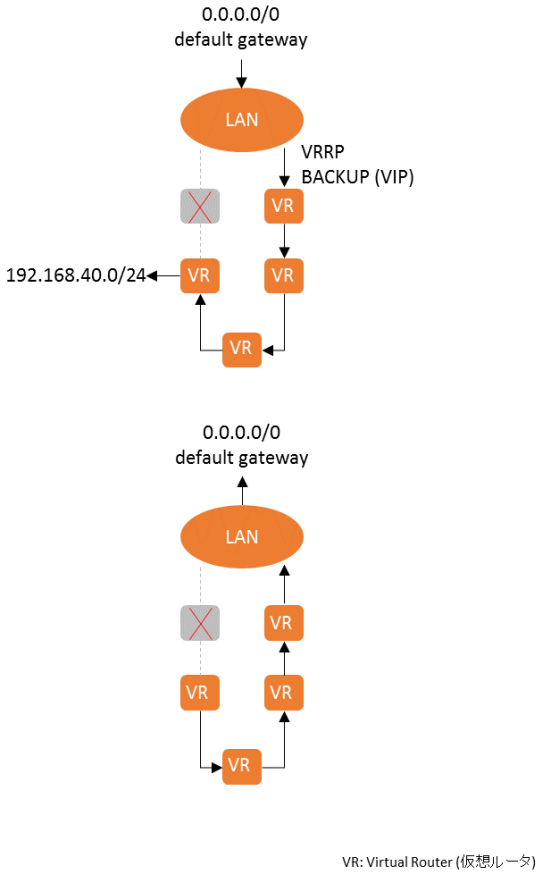

図 12 仮想ルータの障害(VRRP MASTER)
Figure 12 Virtual router failure (VRRP MASTER)

さらに，異なる仮想ルータの障害時における動作検証についてみていきましょう．

図 13 は LAN に隣接しない仮想ルータで障害が発生した際，経路制御の動作検証についての結果です．

あらかじめ OSPF プロトコルによりコスト・パラメータ設定されたインターフェイスが優先されるよう経路制御が行われるため，隣接した仮想ルータを迂回路として用いられます．

この際，VRRP MASTER として設定されていた仮想ルータの OSPF default-information originate のルーティング情報と，VRRP BACKUP として設定された仮想ルータの OSPF default-information originate のルーティング情報が停止した仮想ルータを経由した経路で変化するため，障害時の迂回路として機能することが確認できます．（図 13）

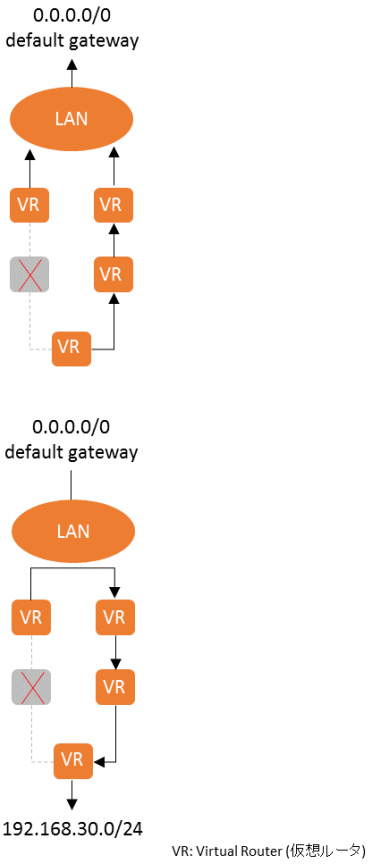

図 13 仮想ルータの障害(Leaf)
Figure 13 Virtual router failure (Leaf)

このように動作検証から，スイッチレス・サーバクラスタリングのネットワーク設計におけるネットワーク冗長化の有効性について確認されました．

Vol.2014-IOT-27 No.19
Vol.2014-SPT-11 No.19
2014/10/10

6. まとめ

本稿では VMware vSphere ESXi, Microsoft Hyper-V, Linux KVM など仮想マシンマネージャ(VMM: Virtual Machine Manager)上に仮想ルータ(VR: Virtual Router)を導入することで高価な広帯域のネットワークスイッチを導入することなく，仮想化環境に数台からの比較的小規模なサーバクラスタリングの広帯域化を行う手法についての考察を紹介しました.

また動作検証からスイッチレス・サーバクラスタリングのネットワーク設計におけるネットワーク冗長化の有効性についても確認されました.

本件手法によりサーバクラスタリングを小規模から段階的に拡張され，今後も発展するネットワークの広帯域化に即時的に対応できると期待しています. (図 14)

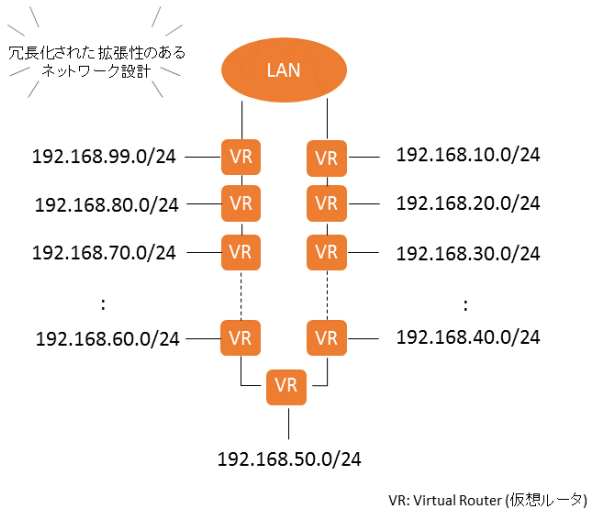

図 14 スイッチレス・サーバクラスタリングの拡張性

Figure 14 Switch-less server clustering: Expanded model

参考文献

[1] OSPF Version 2 (RFC2328) https://www.ietf.org/rfc/rfc2328.txt
[2] Virtual Router Redundancy Protocol (VRRP) (RFC3768) http://www.ietf.org/rfc/rfc3768.txt