SAKURA Internet

www.sakura.ad.jp

Vyattaによる仮想ルータの世界

2010年8月26日

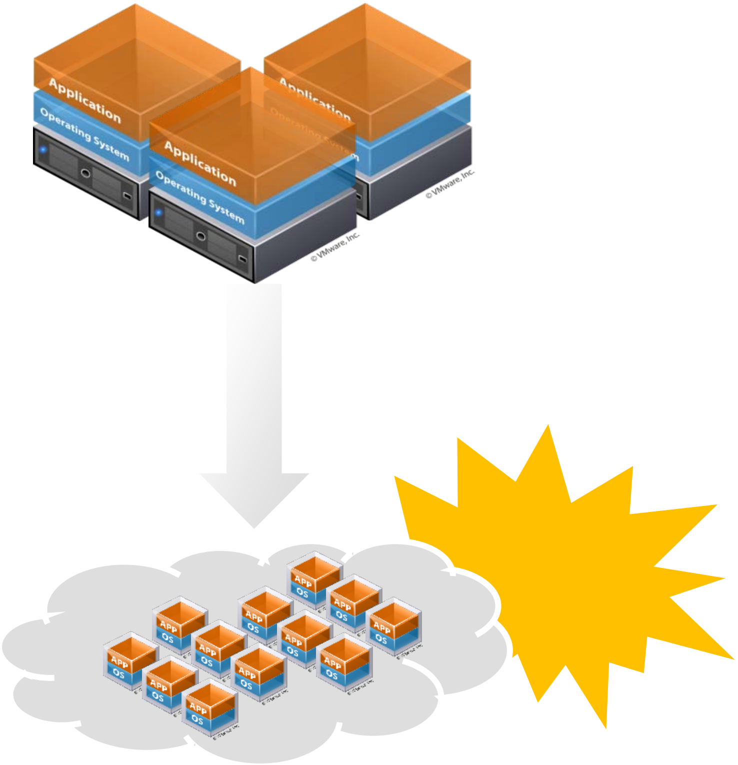
さくらインターネット研究所

上級研究員　松本直人

仮想ルータが求められる背景

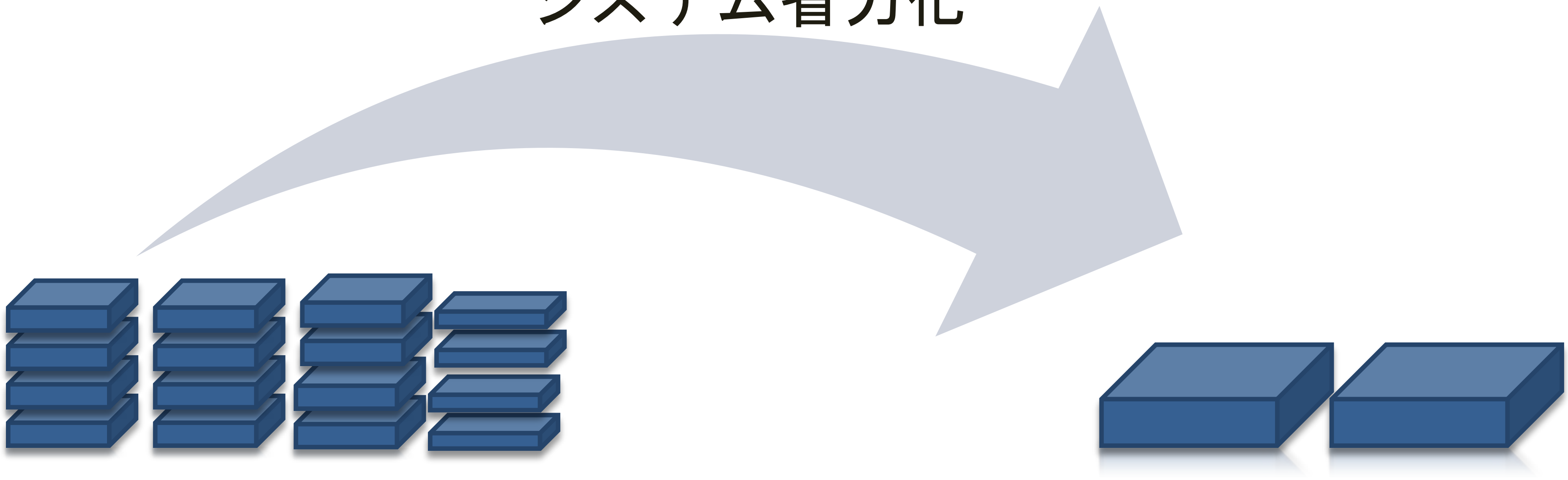
従来のコンピューティング

クラウド･コンピューティング



クラウド中で動作するネットワーク部品が必要となる

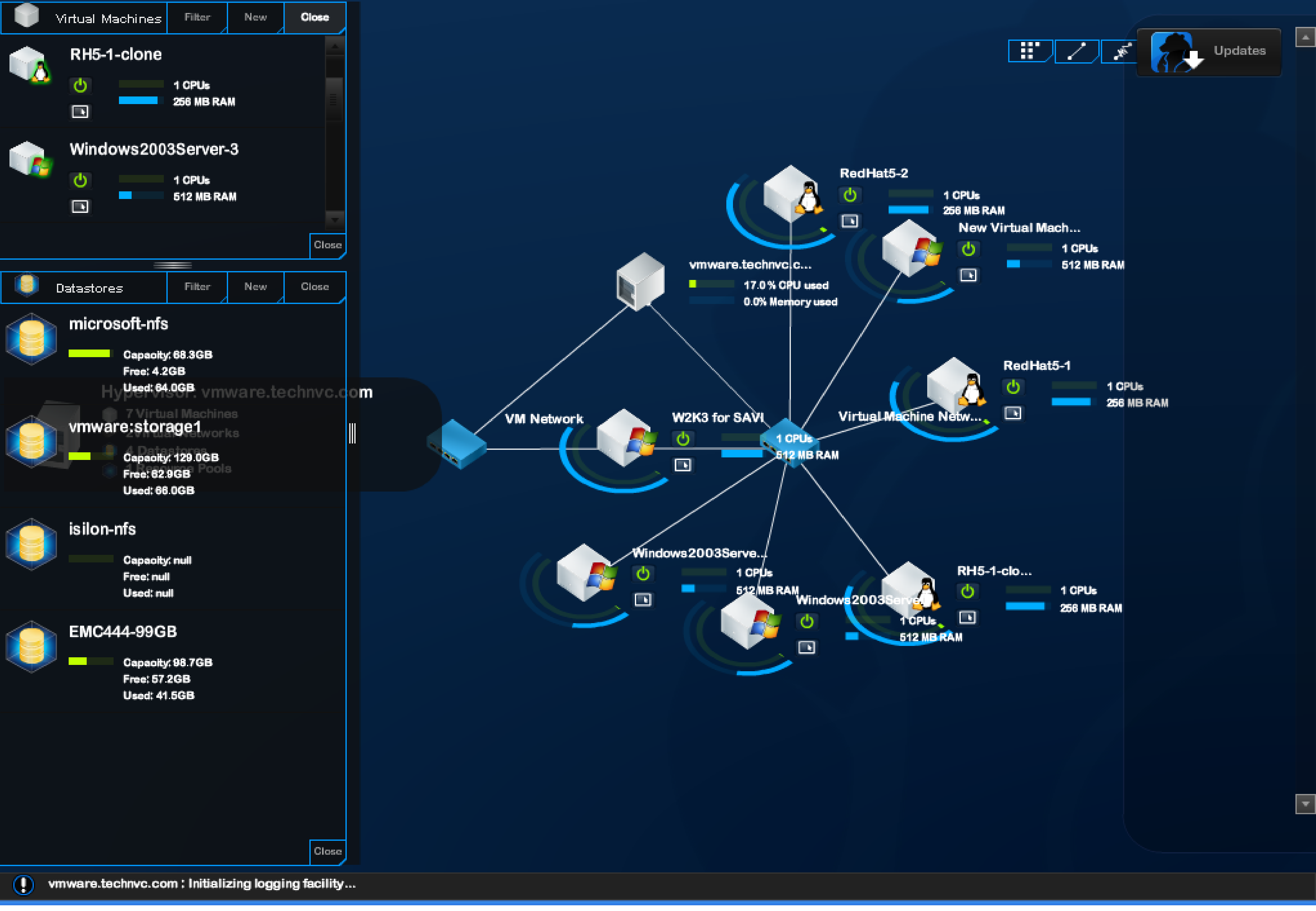
仮想ルータ導入による効用



システム省力化

顧客毎の既存ルータ

ハイパーバイザー上に
仮想ルータ集約



RH5-1-clone
1 CPUs
256 MB RAM

Windows2003Server-3
1 CPUs
512 MB RAM

microsoft-nfs
Capacity: 68.3GB
Free: 4.2GB
Used: 64.0GB

Hypervisor: vmware.technvc.com
7 Virtual Machines
2 Virtual Networks
4 Datastores
4 Resource Pools

vmware:storage1
Capacity: 129.0GB
Free: 62.9GB
Used: 66.0GB

isilon-nfs
Capacity: null
Free: null
Used: null

EMC444-99GB
Capacity: 98.7GB
Free: 57.2GB
Used: 41.5GB

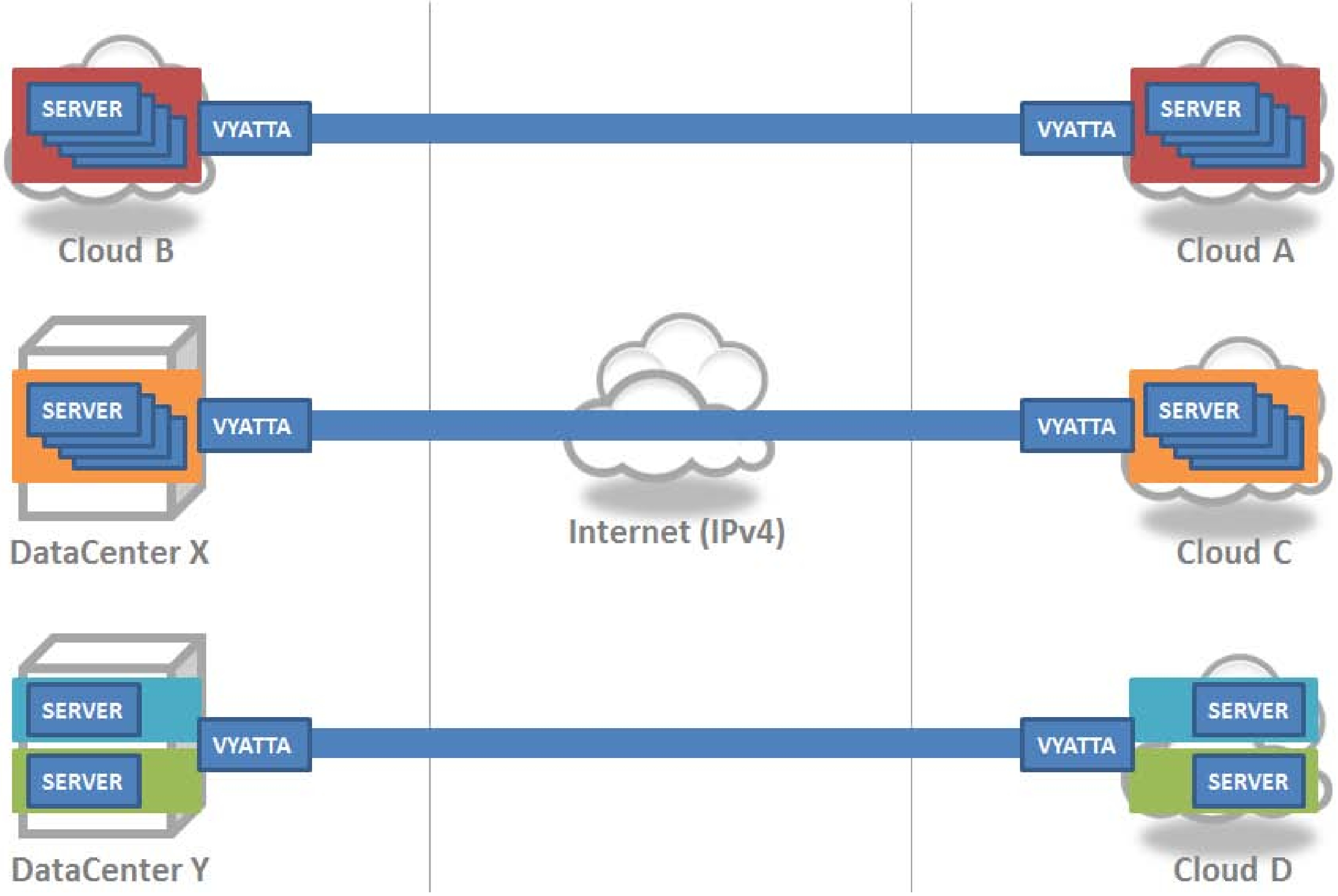
vmware.technvc.com : Initializing logging facility...

Vyattaが持つ機能

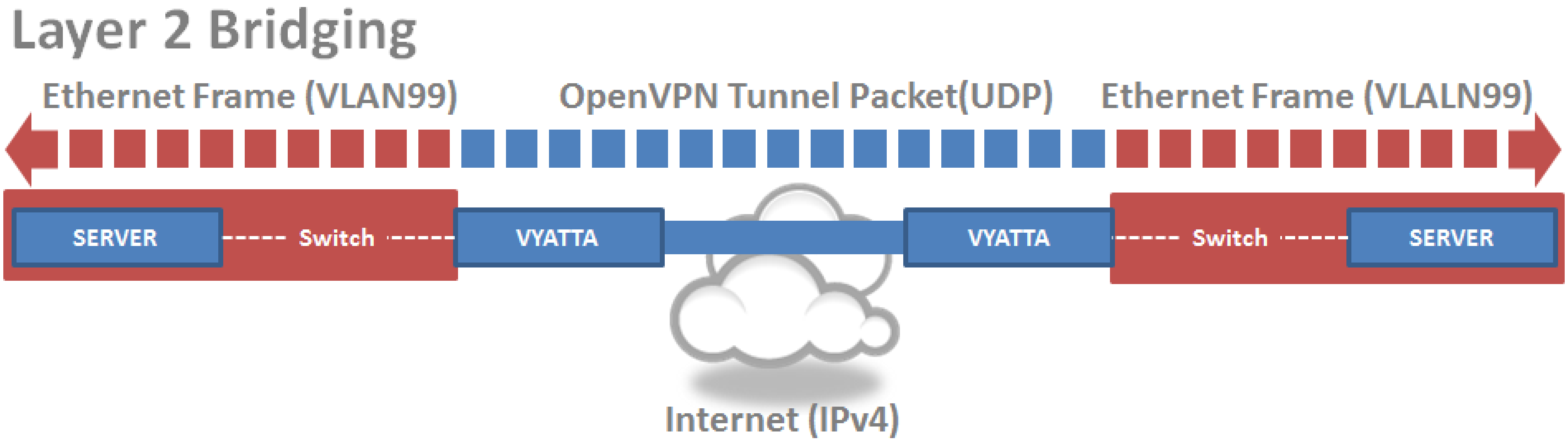
IPv4 / IPv6 Routing	» BGPv4, BGPv6 » OSPFv2, OSPFv3*	» RIPv2 » Static Routes	» IPv6 Policy » IPv6 SLAAC
IP Address Management	» Static » DHCP Server » DHCP Client	» DHCP Relay » Dynamic DNS » DNS Forwarding	» DHCPv6 Server » DHCPv6 Client » DHCPv6 Relay
Encapsulations	» Ethernet » 802.1Q VLANs » PPP	» PPPoE » IP in IP » Frame Relay	» MLPPP » HDLC » GRE
Firewall	» Stateful Inspection Firewall » Zone-based Firewall » P2P Filtering	» IPv6 Firewalling » Time-based Firewall Rules » Rate Limiting	» ICMP Type Filtering » Stateful Failover
Tunneling / VPN	» SSL-based OpenVPN » Site to Site VPN (IPSec) » Remote VPN (PPTP, L2TP, IPSec)	» OpenVPN Client Auto-Configuration » Layer 2 Bridging over GRE » Layer 2 Bridging over OpenVPN	
Additional Security	» Network Address Translation » Sourcefire VRT Intrusion Prevention » VyattaGuard Web Filtering	» DES, 3DES, AES Encryption » MD5 and SHA-1 Authentication » RSA, Diffie Helman Key Mgmt	» NAT Traversal » Role based access control
WAN / LAN Device Drivers	» WAN Device Drivers - ADSL, T1, T3 » Intel 10/100Mbps - 10Gbps	» IEEE 802.11 wireless » Drivers in 2.6.31 Linux Kernel	» Synchronous Serial - V.35, X.21, RS-422, EIA530
Performance Optimization	» WAN Link Load Balancing » Ethernet Link Bonding » Web Caching	» MLPPP » ECMP » Bandwidth Management	
QoS Policies	» Priority Queuing » Network Emulator » Round Robin	» Random / Weighted Random » Classful Queuing » Ethernet Header Matching	» VLAN Tag » IPv6 Address » Port Mirroring
High Availability	» Stateful Firewall / NAT Failover » VRRP » HA Clustering	» Configuration Replication » RAID 1	» IPSec VPN Clustering » Protocol Fault Isolation
Administration & Authentication	» Integrated CLI » Web GUI » Vyatta Remote Access API	» Telnet » SSHv2 / SSH Public Key » Binary Image Install	» RADIUS » TACACS+* » Single Configuration File
Diagnostics & Logging	» tcpdump » Wireshark Packet Capture » BGP MD5 Support	» Serial Loopback Commands » Netflow / sFlow » LLDP	» Syslog » SNMPv2c » SNMP for IPv6

クラウド・ブリッジングという使い方

Inter-Cloud Networking Model



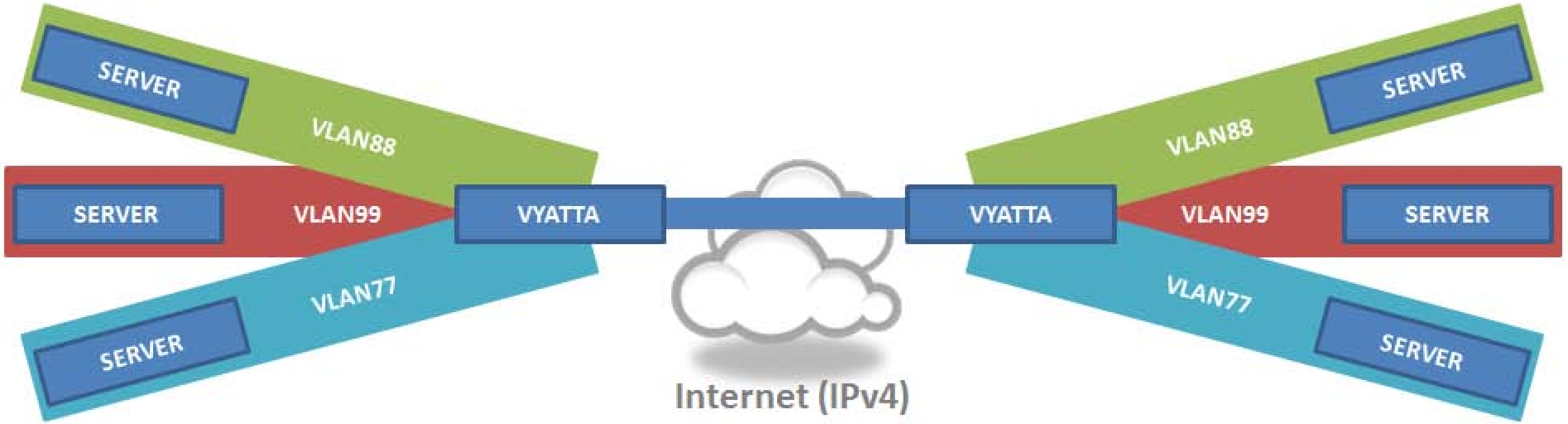
VyattaのOpenVPN機能を使った例



VLANを広げるVyattaの使い方

SAKURA Internet

VLAN Bridging

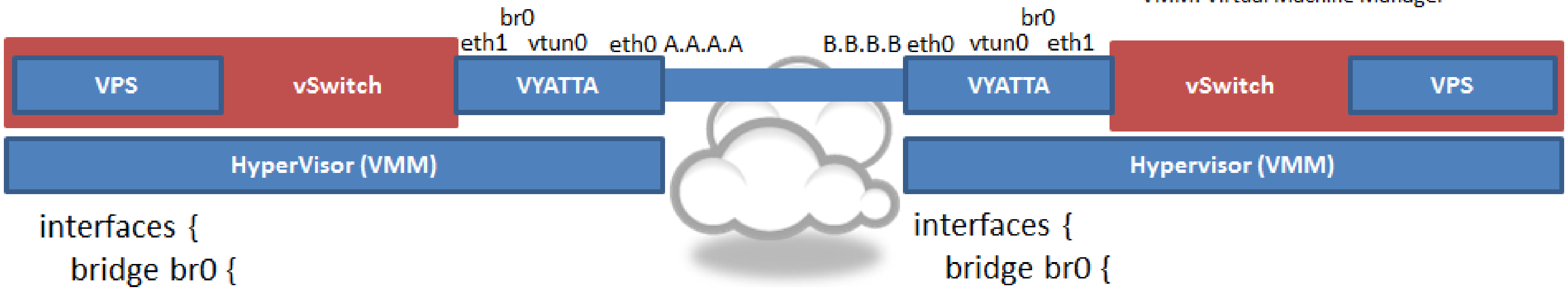


VLAN番号枯渇対策

クラウド・ブリッジングの設定例

VYATTA: Layer 2 Bridging Sample Config

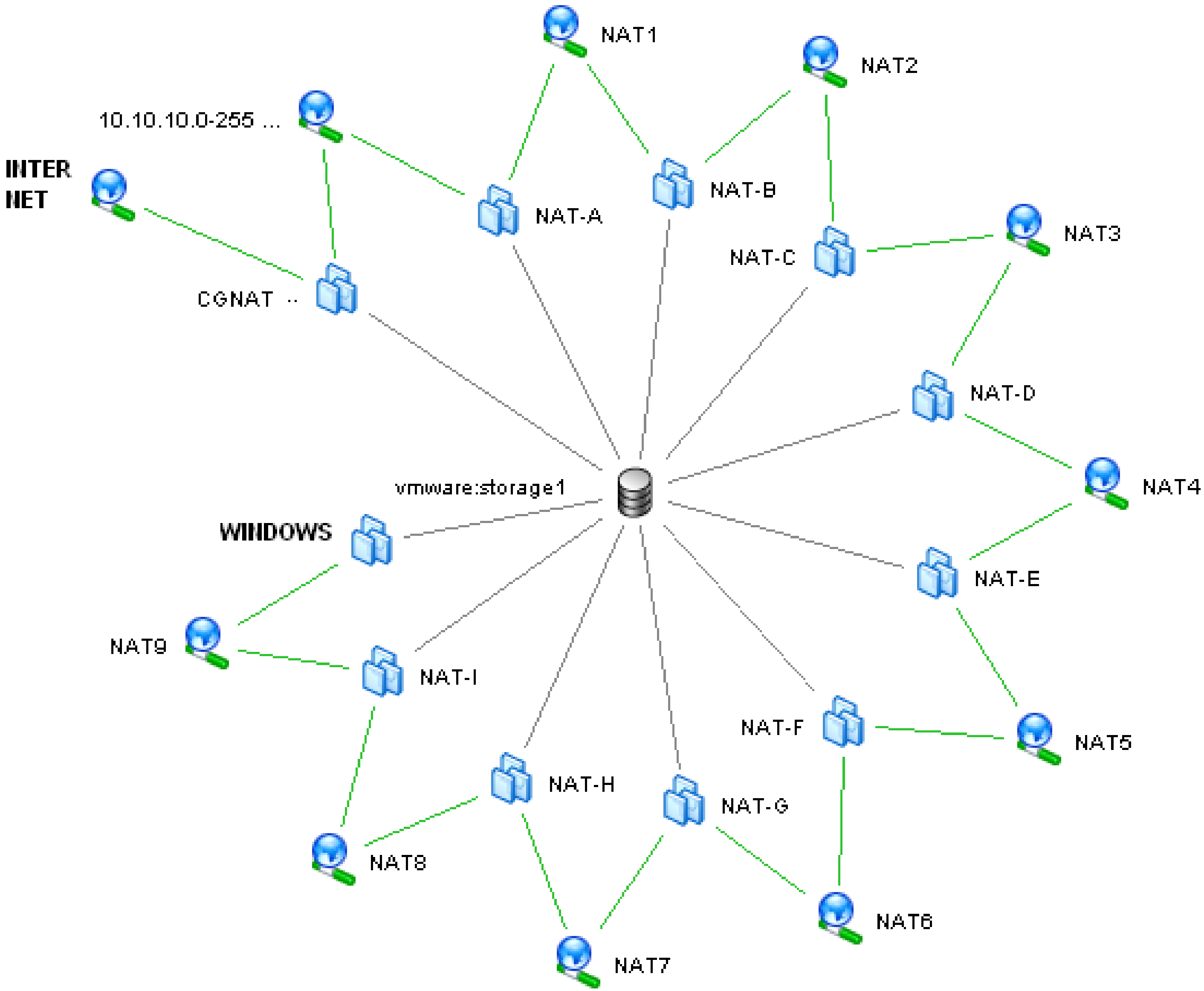
vSwitch: Virtual Switch on HyperVisor (VMM)
VPS: Virtual Private Server
VMM: Virtual Machine Manager



```
interfaces {
  bridge br0 {
  }
  ethernet eth0 {
    address A.A.A.A
  }
  ethernet eth1 {
    bridge-group {
      bridge br0
    }
  }
  openvpn vtun0 {
    bridge-group {
      bridge br0
    }
    mode site-to-site
    remote-host B.B.B.B
    shared-secret-key-file /root/key
  }
}
```

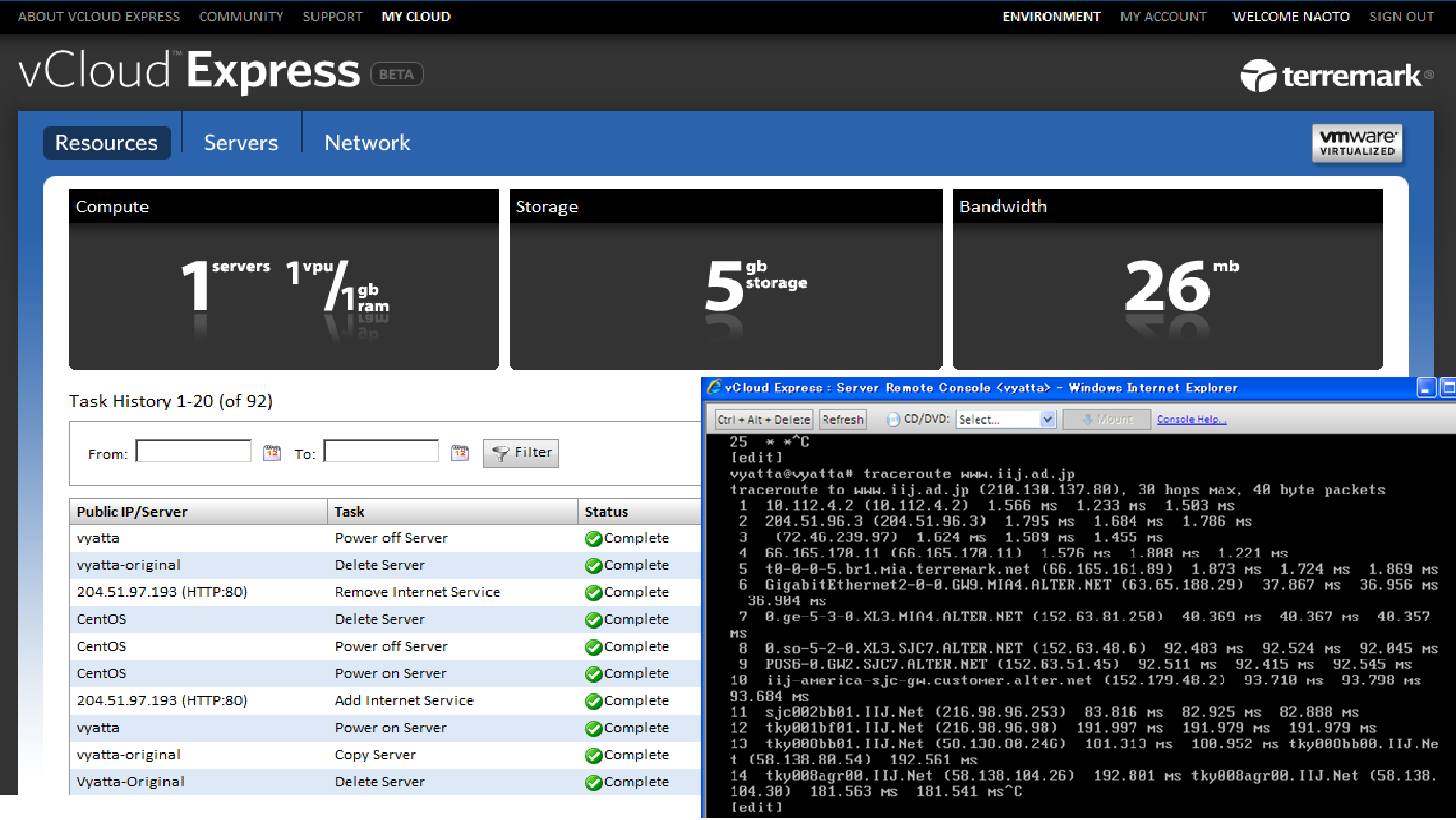
```
interfaces {
  bridge br0 {
  }
  ethernet eth0 {
    address B.B.B.B
  }
  ethernet eth1 {
    bridge-group {
      bridge br0
    }
  }
  openvpn vtun0 {
    bridge-group {
      bridge br0
    }
    mode site-to-site
    remote-host A.A.A.A
    shared-secret-key-file /root/key
  }
}
```

仮想ルータを使った10段NAT



パブリッククラウドでの利用例

SAKURA Internet



Compute

1 servers 1 vpu / 1 gb ram

Storage

5 gb storage

Bandwidth

26 mb

Task History 1-20 (of 92)

From: To: Filter

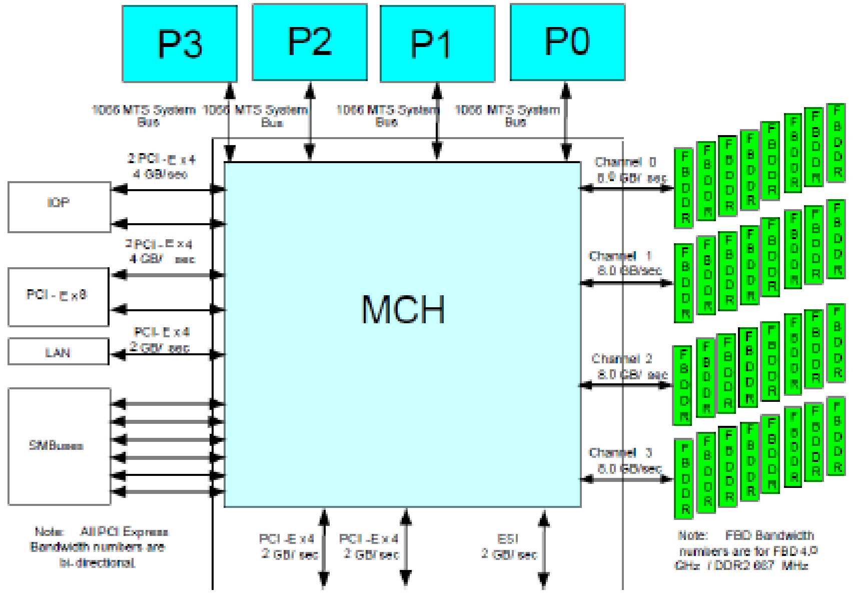
Public IP/Server	Task	Status
vyatta	Power off Server	✓Complete
vyatta-original	Delete Server	✓Complete
204.51.97.193 (HTTP:80)	Remove Internet Service	✓Complete
CentOS	Delete Server	✓Complete
CentOS	Power off Server	✓Complete
CentOS	Power on Server	✓Complete
204.51.97.193 (HTTP:80)	Add Internet Service	✓Complete
vyatta	Power on Server	✓Complete
vyatta-original	Copy Server	✓Complete
Vyatta-Original	Delete Server	✓Complete

vCloud Express : Server Remote Console <vyatta> - Windows Internet Explorer

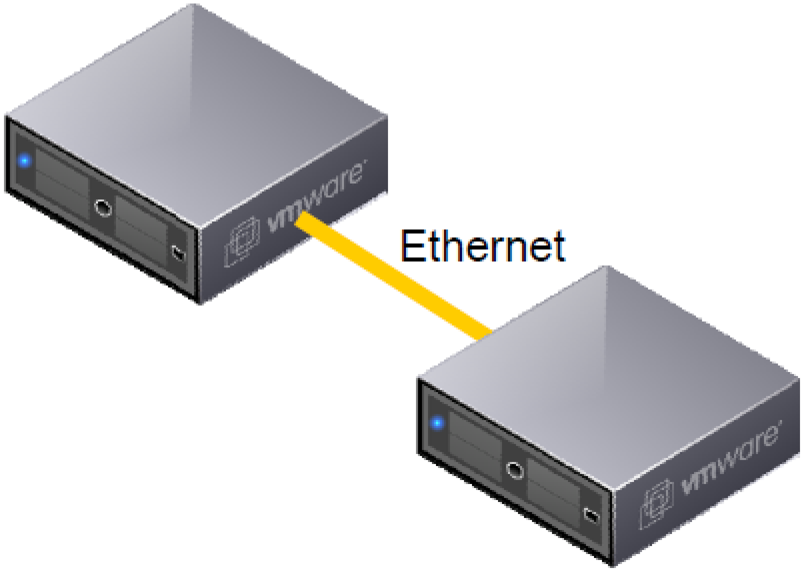
```
25  * *^C
[edit]
vyatta@vyatta# traceroute www.iij.ad.jp
traceroute to www.iij.ad.jp (210.130.137.80), 30 hops max, 40 byte packets
 1  10.112.4.2 (10.112.4.2)  1.566 ms  1.233 ms  1.503 ms
 2  204.51.96.3 (204.51.96.3)  1.795 ms  1.684 ms  1.786 ms
 3   (72.46.239.97)  1.624 ms  1.589 ms  1.455 ms
 4  66.165.170.11 (66.165.170.11)  1.576 ms  1.808 ms  1.221 ms
 5  t0-0-0-5.br1.mia.terremark.net (66.165.161.89)  1.873 ms  1.724 ms  1.869 ms
 6  GigabitEthernet2-0-0.GW9.MIA4.ALTER.NET (63.65.188.29)  37.867 ms  36.956 ms
  36.904 ms
 7  0.ge-5-3-0.XL3.MIA4.ALTER.NET (152.63.81.250)  40.369 ms  40.367 ms  40.357
ms
 8  0.so-5-2-0.XL3.SJC7.ALTER.NET (152.63.48.6)  92.483 ms  92.524 ms  92.045 ms
 9  POS6-0.GW2.SJC7.ALTER.NET (152.63.51.45)  92.511 ms  92.415 ms  92.545 ms
10  iij-america-sjc-gw.customer.alter.net (152.179.48.2)  93.710 ms  93.798 ms
93.684 ms
11  sjc002bb01.IIJ.Net (216.98.96.253)  83.816 ms  82.925 ms  82.888 ms
12  tky001bf01.IIJ.Net (216.98.96.98)  191.997 ms  191.979 ms  191.979 ms
13  tky008bb01.IIJ.Net (58.138.80.246)  181.313 ms  180.952 ms tky008bb00.IIJ.Ne
t (58.138.80.54)  192.561 ms
14  tky008agr00.IIJ.Net (58.138.104.26)  192.801 ms tky008agr00.IIJ.Net (58.138.
104.30)  181.563 ms  181.541 ms^C
[edit]
```

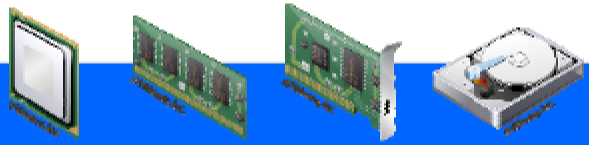
仮想ルータの性能課題を把握する

バス・ボトルネック



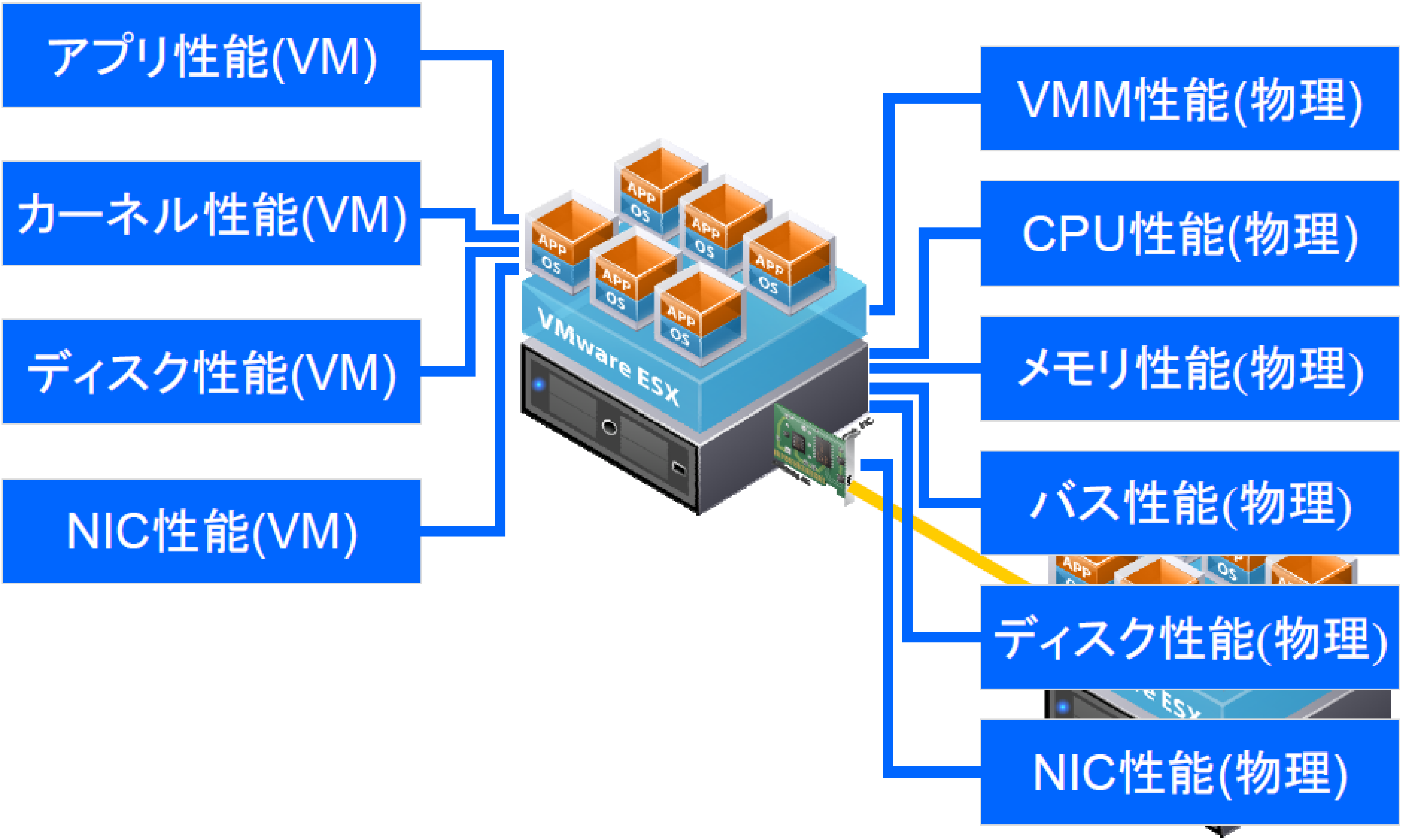
インターフェイス・ボトルネック





仮想化技術であってもパフォーマンスは、コンポーネントの性能限界につよく依存している

性能と依存関係を示すポイント



アプリ性能(VM)

カーネル性能(VM)

ディスク性能(VM)

NIC性能(VM)

VMM性能(物理)

CPU性能(物理)

メモリ性能(物理)

バス性能(物理)

ディスク性能(物理)

NIC性能(物理)

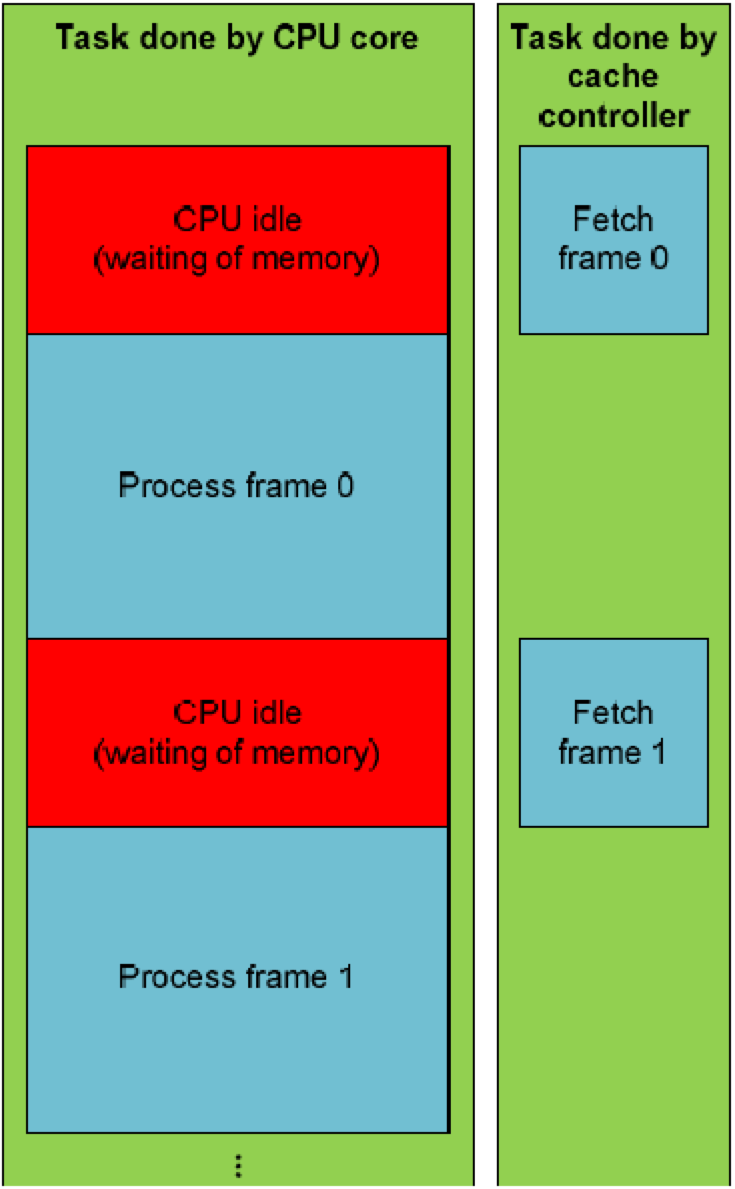
ソフトウェアである性能の壁

アプリ性能(VM)

カーネル性能(VM)

ディスク性能(VM)

NIC性能(VM)



Task done by CPU core

CPU idle
(waiting of memory)

Process frame 0

CPU idle
(waiting of memory)

Process frame 1

⋮

Task done by cache controller

Fetch frame 0

Fetch frame 1

VMM性能(物理)

CPU性能(物理)

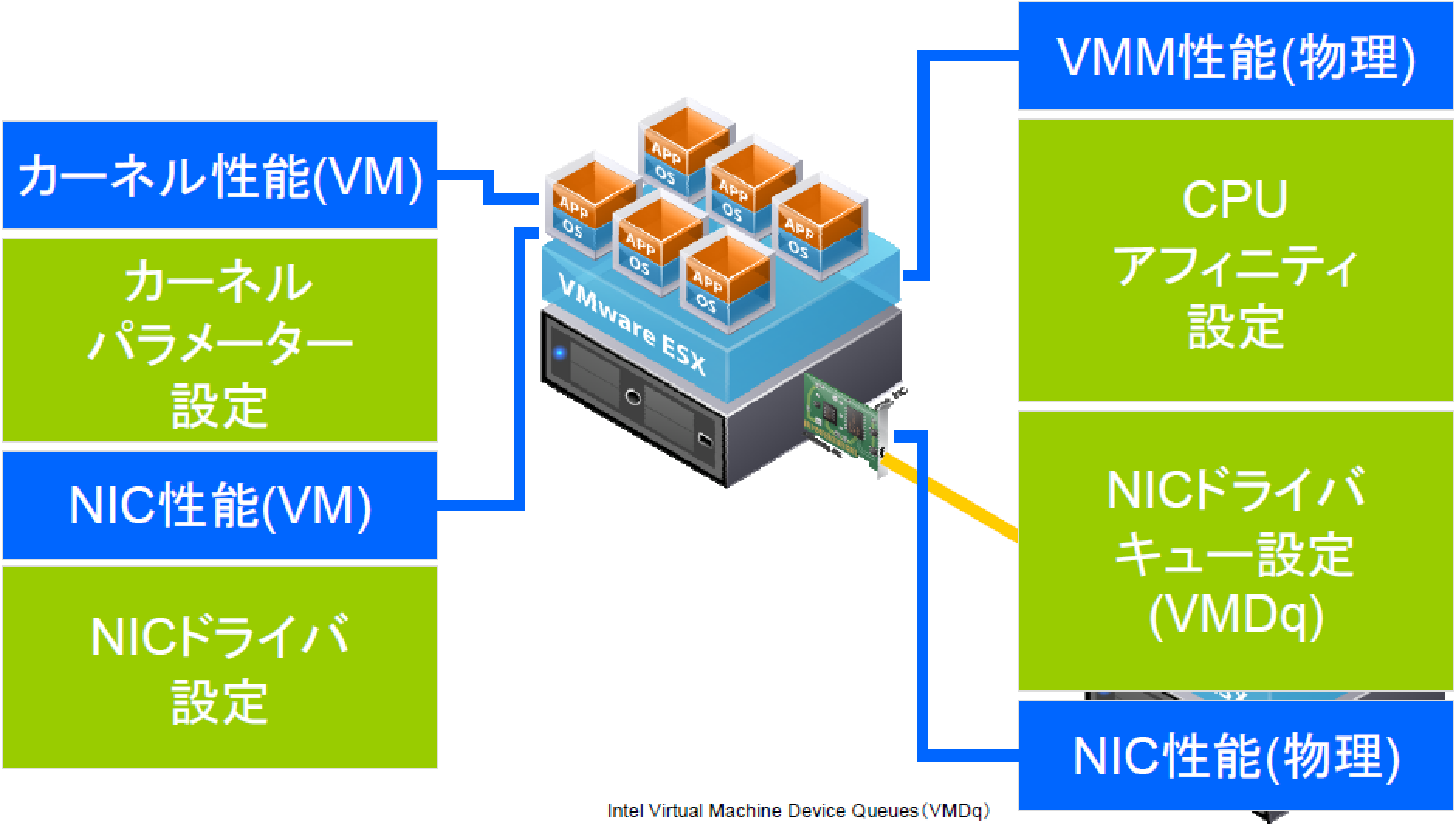
メモリ性能(物理)

バス性能(物理)

ディスク性能(物理)

NIC性能(物理)

チューニングすべきポイント



カーネル性能(VM)

カーネル
パラメーター
設定

NIC性能(VM)

NICドライバ
設定

VMM性能(物理)

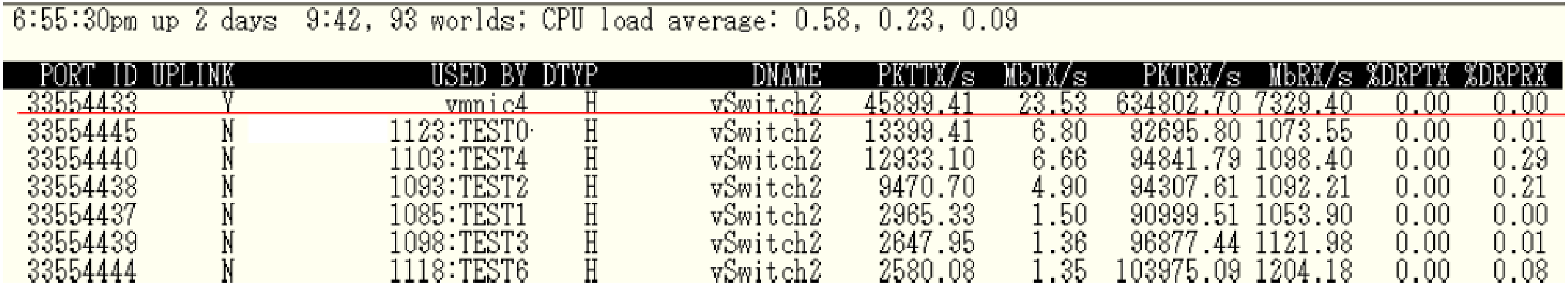
CPU
アフィニティ
設定

NICドライバ
キュー設定
(VMDq)

NIC性能(物理)

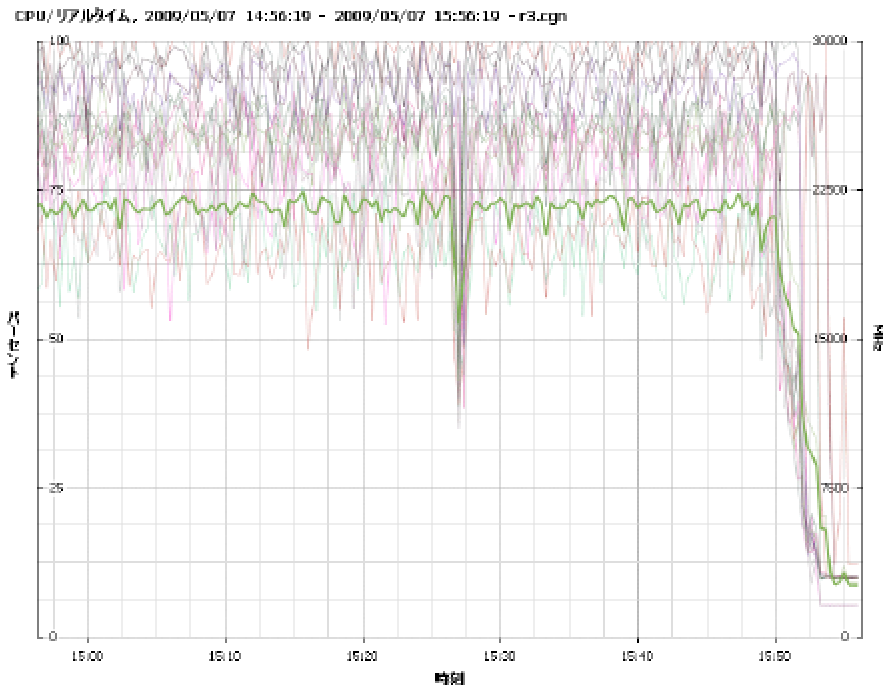
Intel Virtual Machine Device Queues（VMDq）

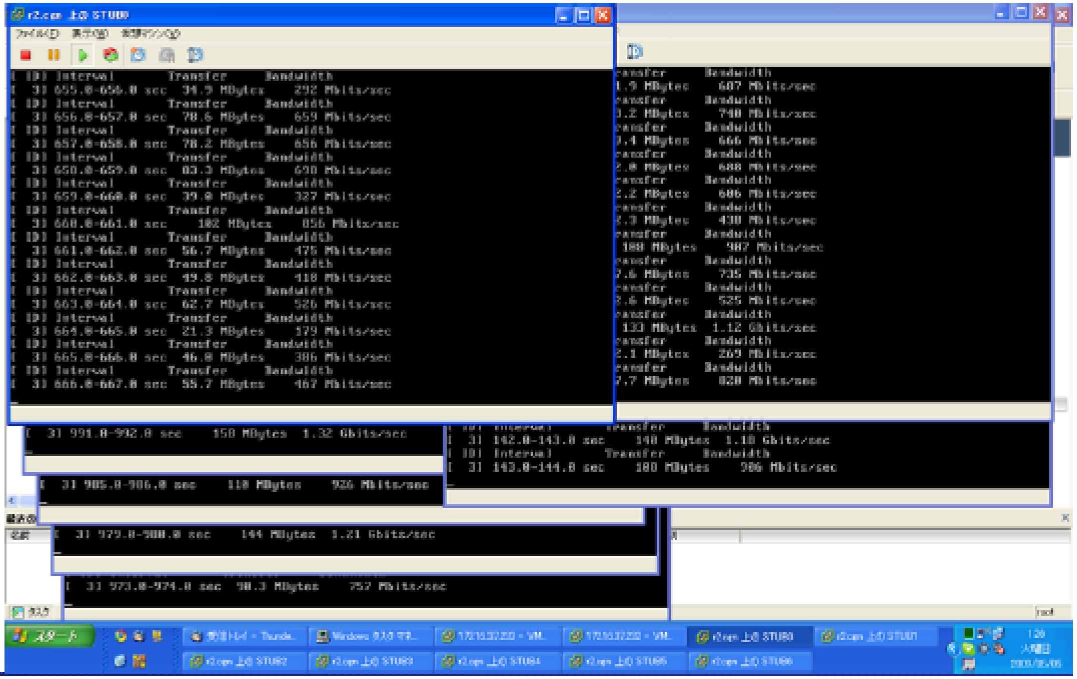
仮想化インフラ全体を外観する方法



```
6:55:30pm up 2 days  9:42, 93 worlds; CPU load average: 0.58, 0.23, 0.09

  PORT ID UPLINK                USED BY DTYP          DNAME   PKTTX/s  MbTX/s    PKTRX/s  MbRX/s %DRPTX %DRPRX
 33554433      Y                 vmnic4    H       vSwitch2  45899.41   23.53  634802.70 7329.40   0.00   0.00
 33554445      N              1123:TEST0   H       vSwitch2  13399.41    6.80   92695.80 1073.55   0.00   0.01
 33554440      N              1103:TEST4   H       vSwitch2  12933.10    6.66   94841.79 1098.40   0.00   0.29
 33554438      N              1093:TEST2   H       vSwitch2   9470.70    4.90   94307.61 1092.21   0.00   0.21
 33554437      N              1085:TEST1   H       vSwitch2   2965.33    1.50   90999.51 1053.90   0.00   0.00
 33554439      N              1098:TEST3   H       vSwitch2   2647.95    1.36   96877.44 1121.98   0.00   0.01
 33554444      N              1118:TEST6   H       vSwitch2   2580.08    1.35  103975.09 1204.18   0.00   0.08
```





```
[ ID] Interval       Transfer     Bandwidth
[  3] 655.0-656.0 sec  34.9 MBytes    292 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 656.0-657.0 sec  78.6 MBytes    659 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 657.0-658.0 sec  78.2 MBytes    656 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 658.0-659.0 sec  83.3 MBytes    698 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 659.0-660.0 sec  39.0 MBytes    327 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 660.0-661.0 sec   102 MBytes    856 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 661.0-662.0 sec  56.7 MBytes    475 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 662.0-663.0 sec  49.8 MBytes    418 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 663.0-664.0 sec  62.7 MBytes    526 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 664.0-665.0 sec  21.3 MBytes    179 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 665.0-666.0 sec  46.0 MBytes    386 Mbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 666.0-667.0 sec  55.7 MBytes    467 Mbits/sec
```

```
ransfer     Bandwidth
1.9 MBytes    607 Mbits/sec
ransfer     Bandwidth
3.2 MBytes    740 Mbits/sec
ransfer     Bandwidth
3.4 MBytes    666 Mbits/sec
ransfer     Bandwidth
2.0 MBytes    688 Mbits/sec
ransfer     Bandwidth
2.2 MBytes    606 Mbits/sec
ransfer     Bandwidth
2.3 MBytes    430 Mbits/sec
ransfer     Bandwidth
 108 MBytes    907 Mbits/sec
ransfer     Bandwidth
7.6 MBytes    735 Mbits/sec
ransfer     Bandwidth
2.6 MBytes    525 Mbits/sec
ransfer     Bandwidth
 133 MBytes  1.12 Gbits/sec
ransfer     Bandwidth
2.1 MBytes    269 Mbits/sec
ransfer     Bandwidth
7.7 MBytes    820 Mbits/sec
```

```
[  3] 991.0-992.0 sec    150 MBytes  1.32 Gbits/sec
```

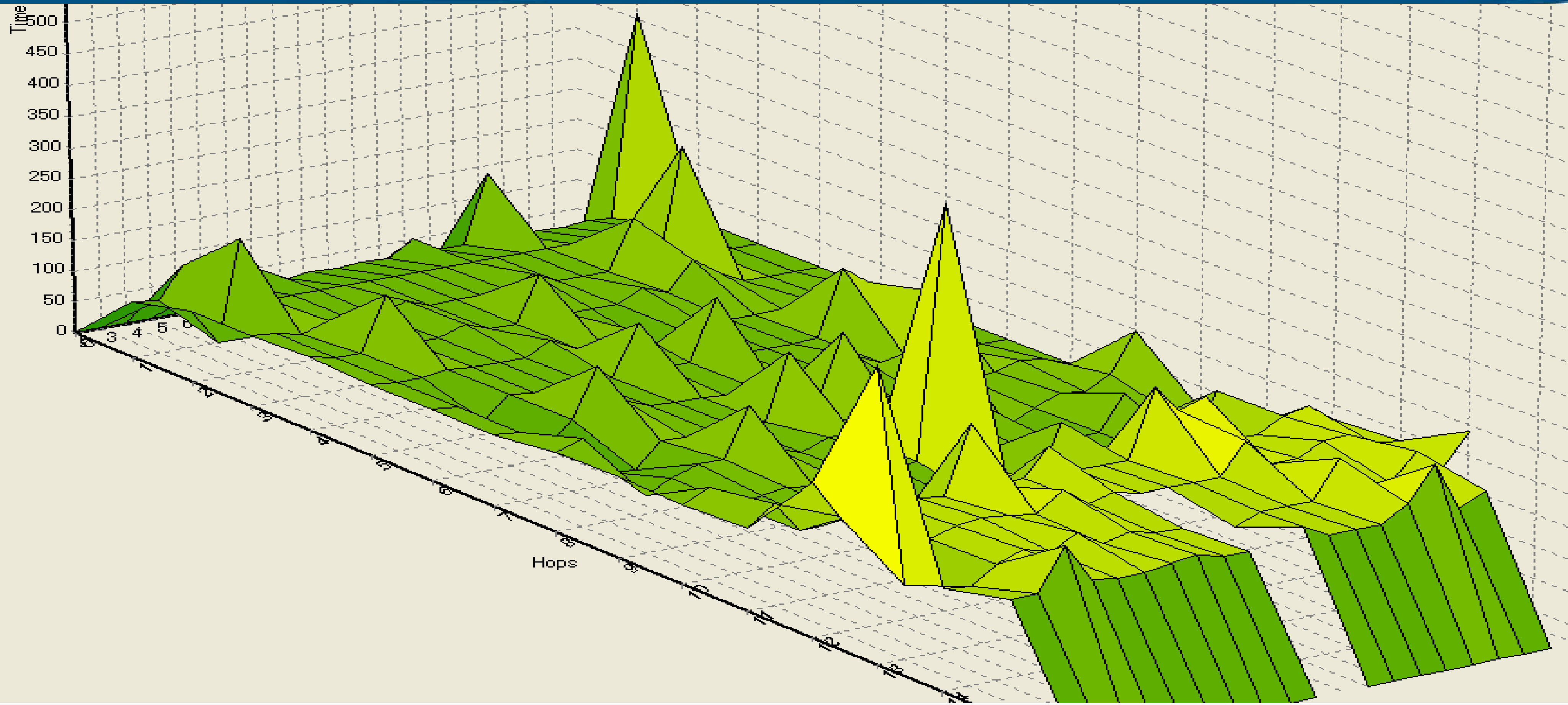
```
[  3] 142.0-143.0 sec    140 MBytes  1.18 Gbits/sec
[ ID] Interval       Transfer     Bandwidth
[  3] 143.0-144.0 sec    100 MBytes    306 Mbits/sec
```

```
[  3] 905.0-906.0 sec    110 MBytes    926 Mbits/sec
```

```
[  3] 979.0-980.0 sec    144 MBytes  1.21 Gbits/sec
```

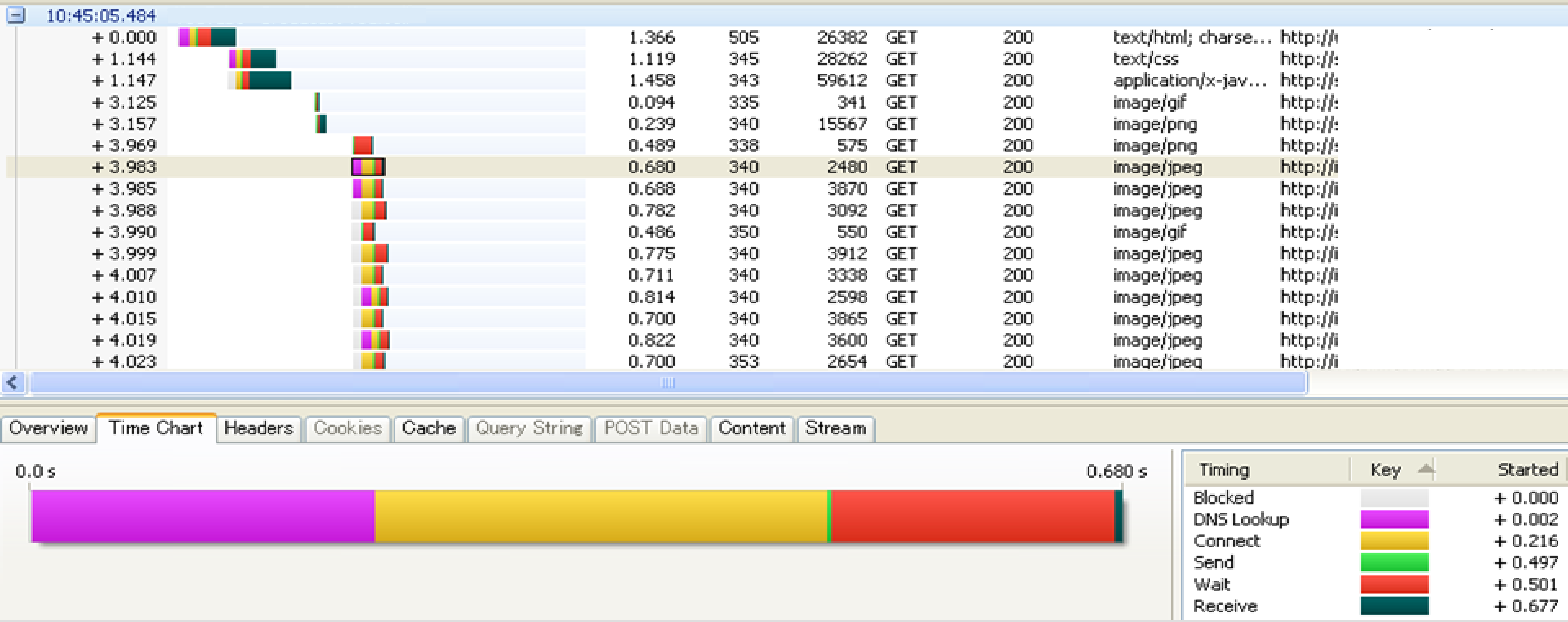
```
[  3] 973.0-974.0 sec  90.3 MBytes    757 Mbits/sec
```

仮想ルータ単体での性能確認



ネットワーク全体ではなくシステム単位での測定が基本

仮想ルータを通した全体システムの把握



10:45:05.484

+ 0.000	1.366	505	26382	GET	200	text/html; charse...	http://\
+ 1.144	1.119	345	28262	GET	200	text/css	http://:
+ 1.147	1.458	343	59612	GET	200	application/x-jav...	http://:
+ 3.125	0.094	335	341	GET	200	image/gif	http://:
+ 3.157	0.239	340	15567	GET	200	image/png	http://:
+ 3.969	0.489	338	575	GET	200	image/png	http://:
+ 3.983	0.680	340	2480	GET	200	image/jpeg	http://i
+ 3.985	0.688	340	3870	GET	200	image/jpeg	http://i
+ 3.988	0.782	340	3092	GET	200	image/jpeg	http://i
+ 3.990	0.486	350	550	GET	200	image/gif	http://:
+ 3.999	0.775	340	3912	GET	200	image/jpeg	http://i
+ 4.007	0.711	340	3338	GET	200	image/jpeg	http://i
+ 4.010	0.814	340	2598	GET	200	image/jpeg	http://i
+ 4.015	0.700	340	3865	GET	200	image/jpeg	http://i
+ 4.019	0.822	340	3600	GET	200	image/jpeg	http://i
+ 4.023	0.700	353	2654	GET	200	image/jpeg	http://i

Timing	Key	Started
Blocked	⬜	+ 0.000
DNS Lookup	🟪	+ 0.002
Connect	🟨	+ 0.216
Send	🟩	+ 0.497
Wait	🟥	+ 0.501
Receive	■	+ 0.677

個別システム・セッション単位での測定が基本

ご清聴ありがとうございました