SAKURA Internet

www.sakura.ad.jp

クラウド環境に必要とされる次世代技術

2011年05月24日

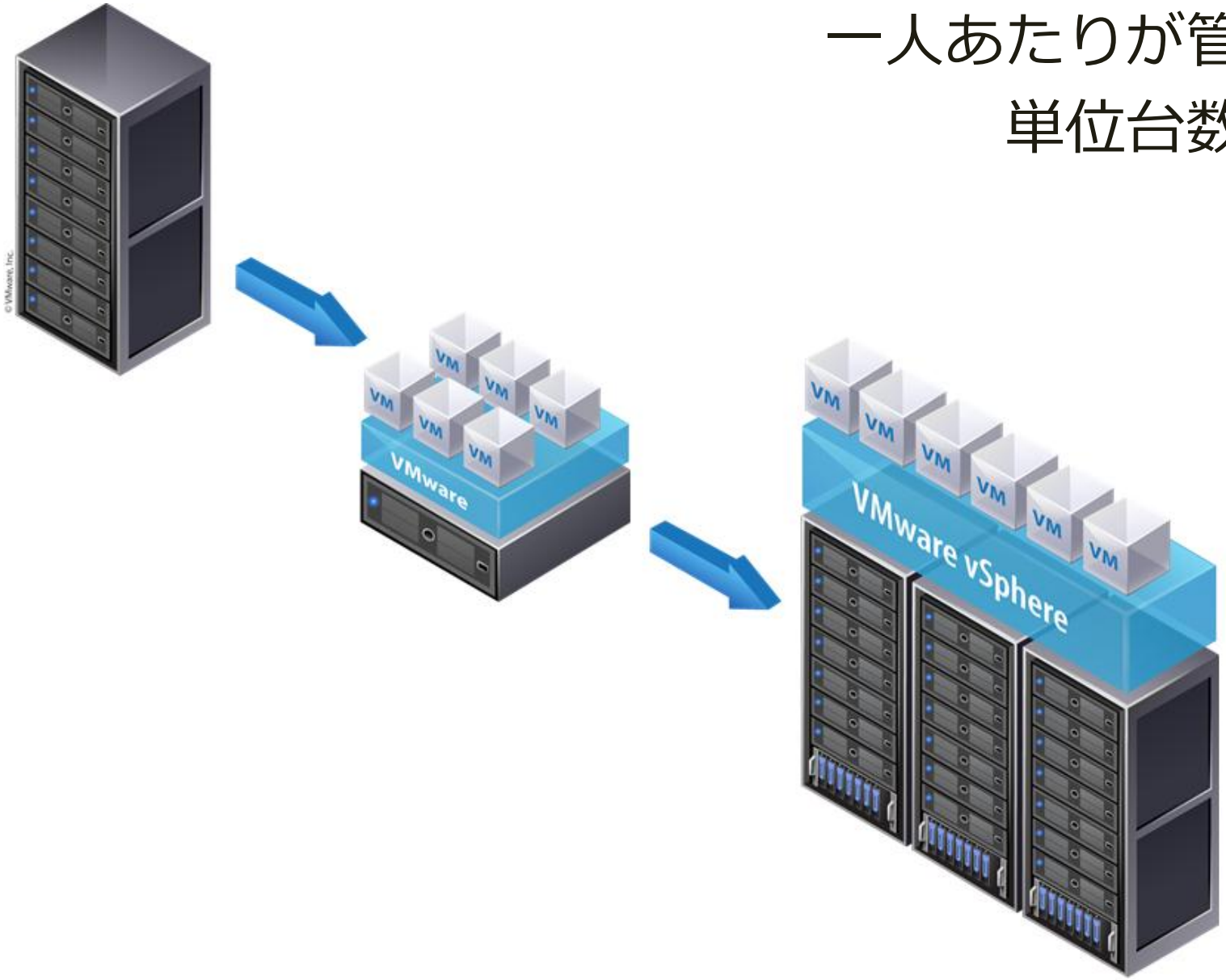
さくらインターネット研究所　上級研究員
松本直人

(C)Copyright 1996-2010 SAKURA Internet Inc.

仮想化による管理点の変化

SAKURA Internet

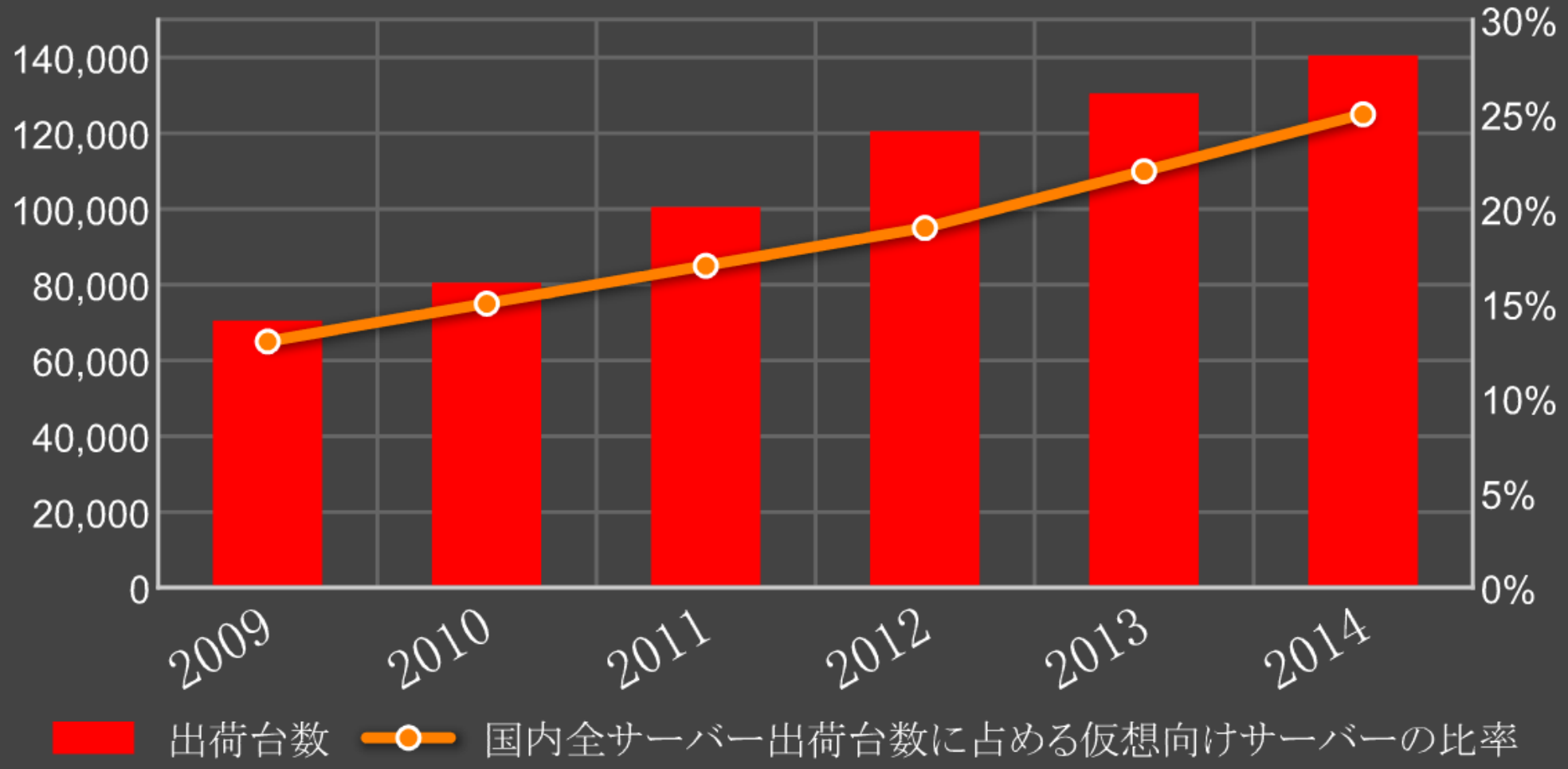
一人あたりが管理する
単位台数の増加



仮想化を取り巻く環境

SAKURA Internet

出典: 国内仮想化サーバー市場 出荷台数予測、2009年～2014年 IDC Japan株式会社 (2010年10月28日作成)

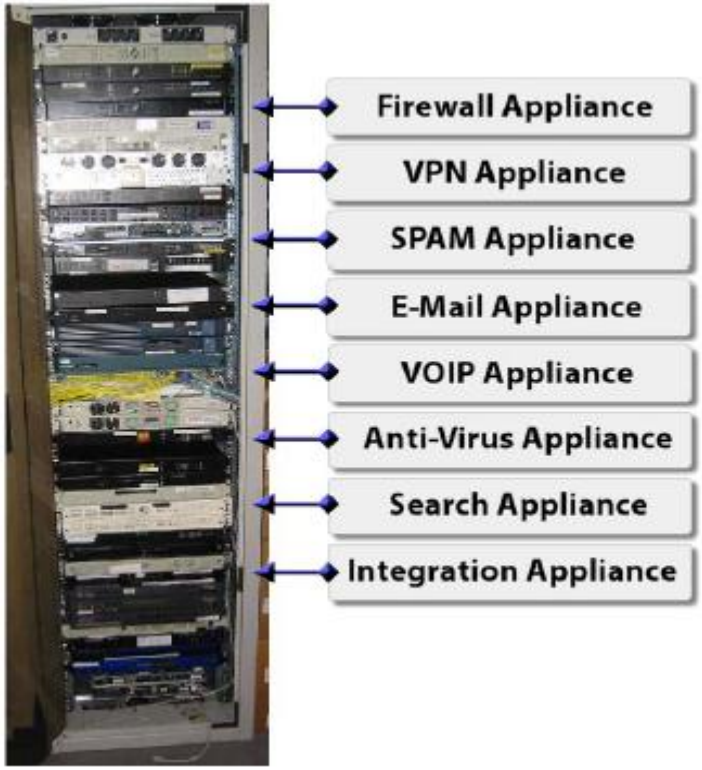


今後5年以内に国内サーバーの1/4が仮想環境となる

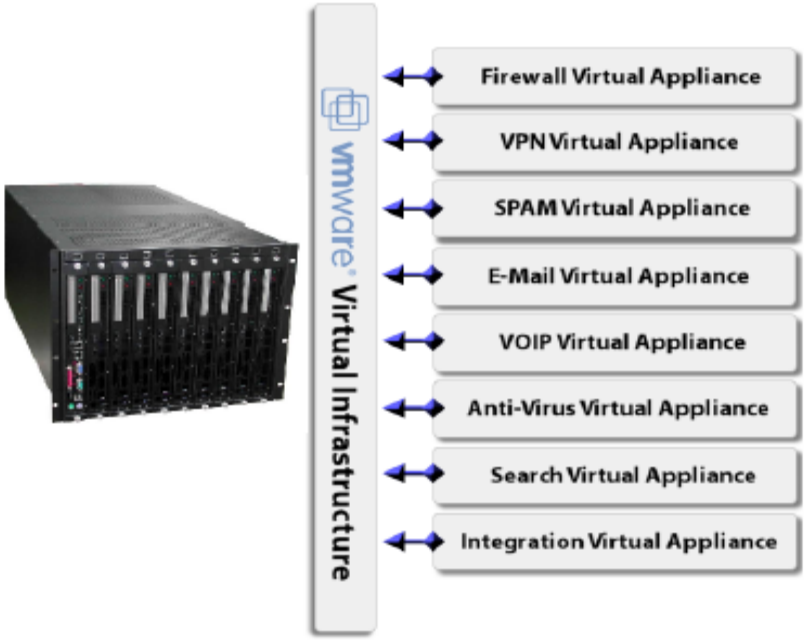
ネットワーク機器も仮想化へ

SAKURA Internet

既存のシステム構成



仮想化技術で省力化



ネットワーク機器の仮想アプライアンス化は発展途上

SAKURA Internet

www.sakura.ad.jp

増加傾向の障害

(C)Copyright 1996-2010 SAKURA Internet Inc.

セキュリティの三要素と問題の変化

SAKURA Internet

災害

不正使用

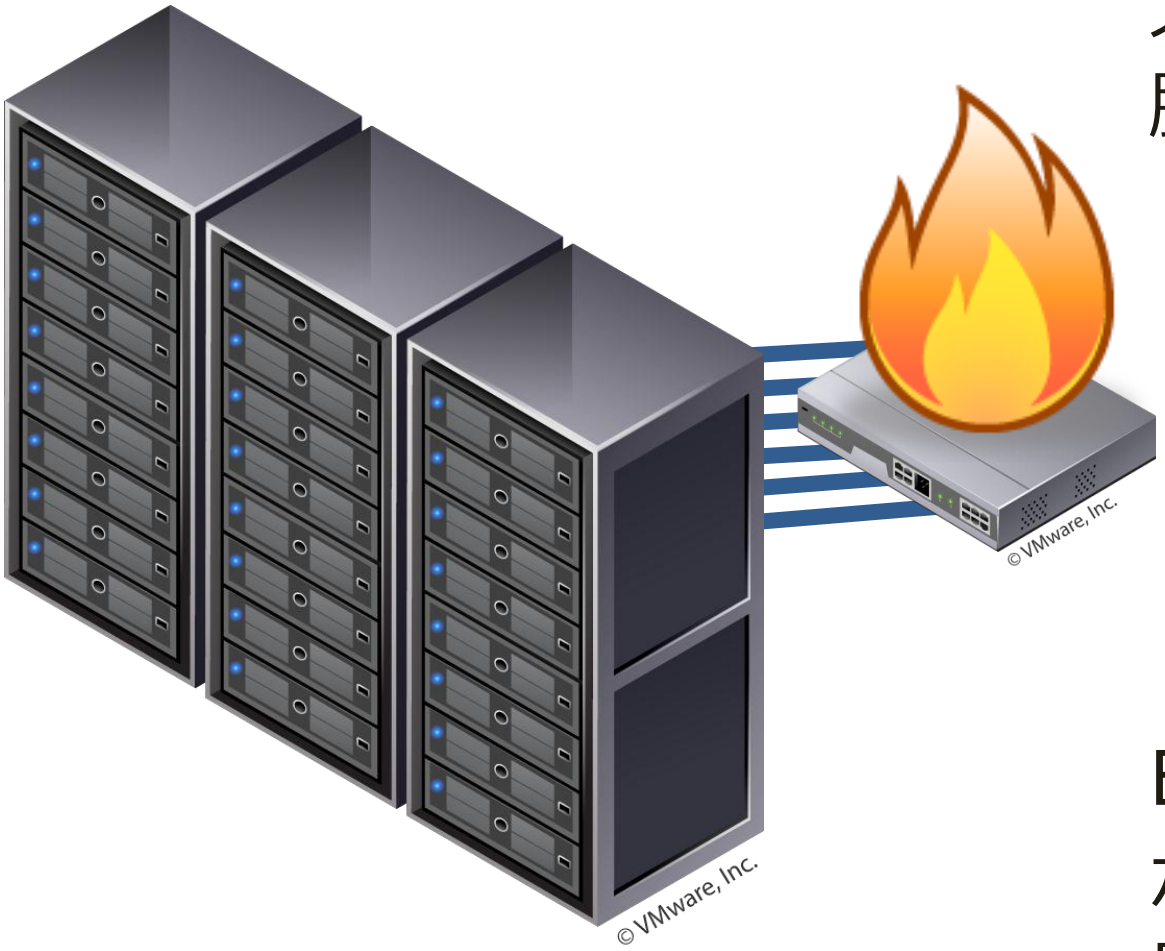
誤用

仮想化を導入 (複雑化)

↓

人災の増加

1) スイッチへの過負荷



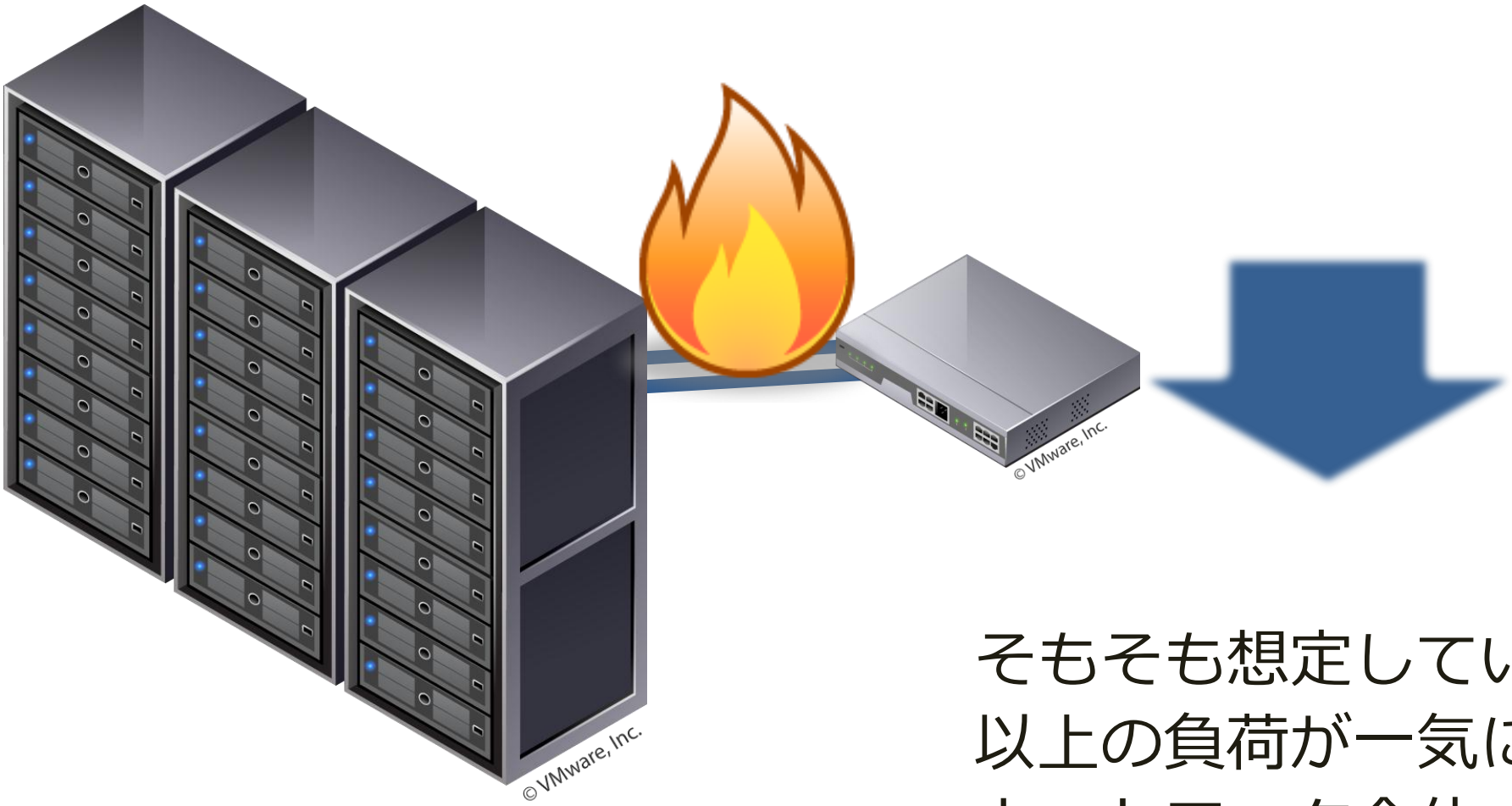
Layer2セグメント上に
スイッチの能力を超えた
膨大なサーバーが接続

Ethernetセグメント全体
が不安定にとなる。
原因の特定に時間を要す。

2) ネットワークへの過負荷

SAKURA Internet

ネットワーク帯域不足

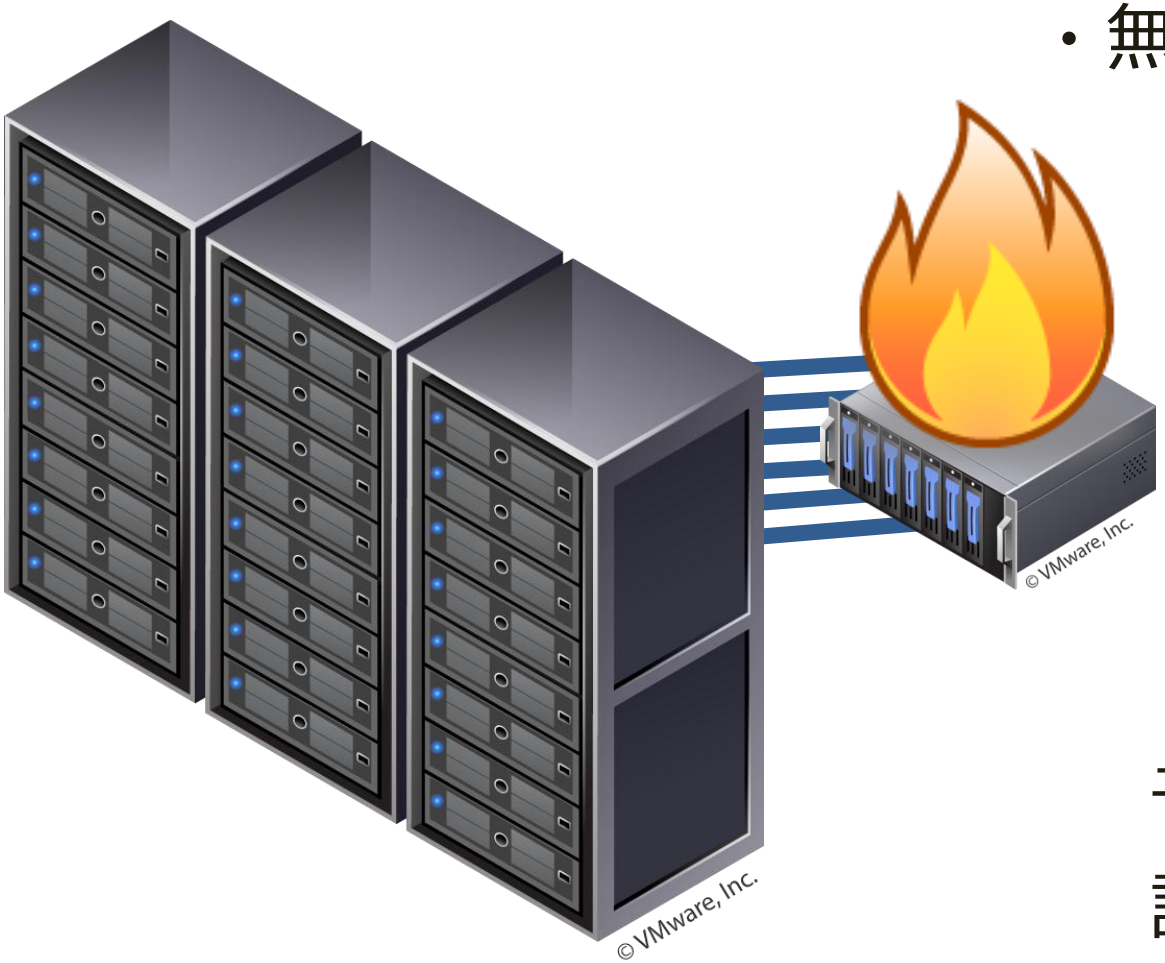


そもそも想定していた
以上の負荷が一気に
ネットワーク全体へ

3) ストレージへの過負荷

SAKURA Internet

・cronによる同時負荷集中
・無作為なVM増設と集中

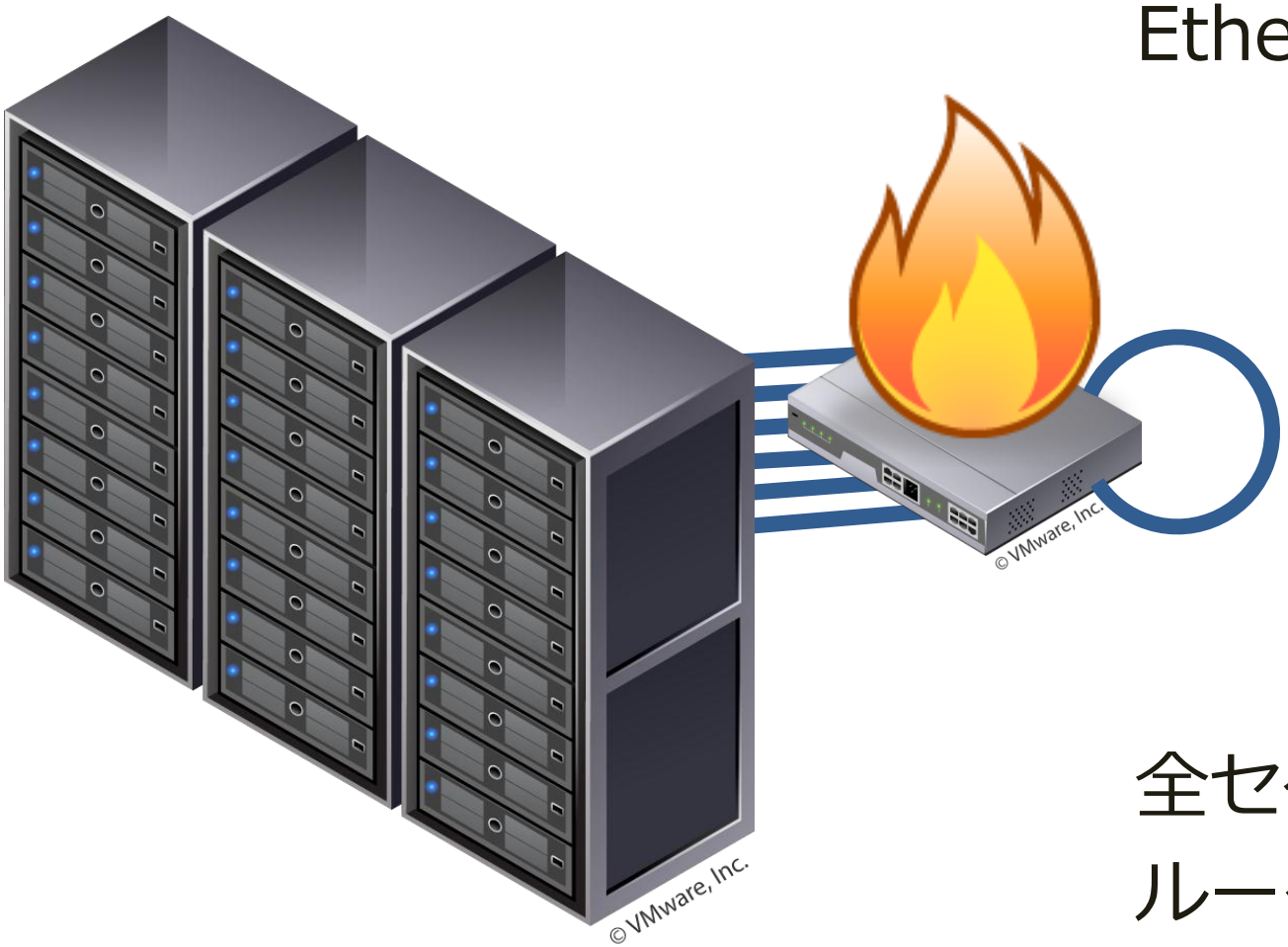


予期しないシステム動作。
設定変更の細密化。

4) 仮想スイッチのループ

SAKURA Internet

物理的に見えない箇所で
Ethernetのループが発生。



全セグメントでストーム。
ルーターの負荷も100%。

5) 仮想マシンの紛失

頻繁な仮想マシンの移動による物理マシンに対する対応が確認不能へ。

管理主体が不明確な仮想マシンが増加へ。

5) 物理マシンの失念

SAKURA Internet

そもそも物理マシンが、
どれであるかを失念する

障害時の対応が遅れる。

SAKURA Internet

www.sakura.ad.jp

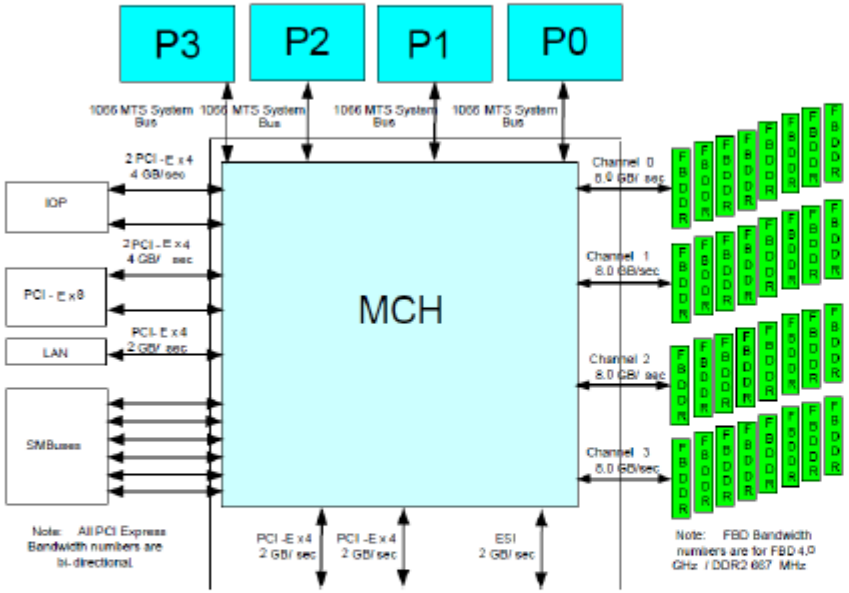
いま知っておくべき技術

(C)Copyright 1996-2010 SAKURA Internet Inc.

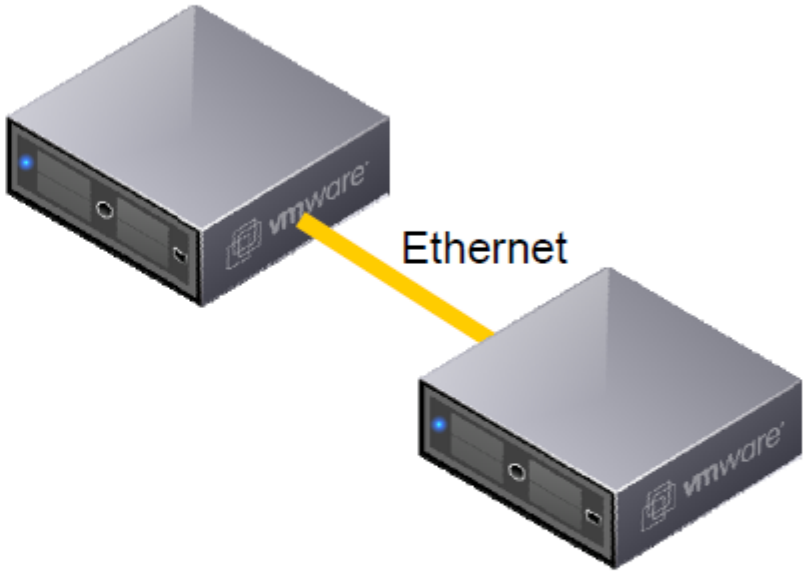
性能限界の把握する

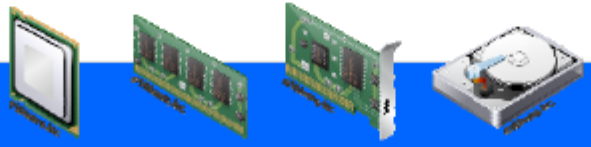
SAKURA Internet

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

仮想化との付き合い方を理解する

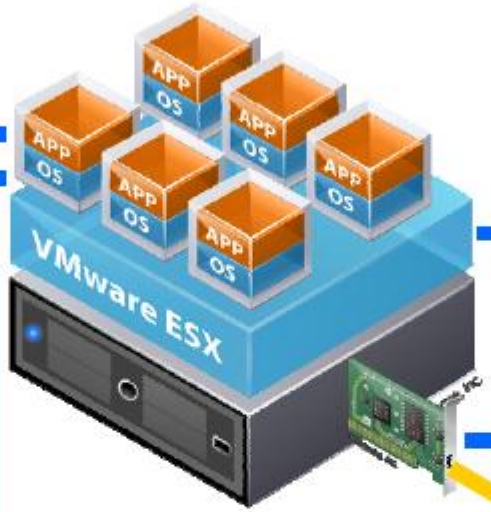
SAKURA Internet

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)

仮想アプライアンスを理解する

SAKURA Internet

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

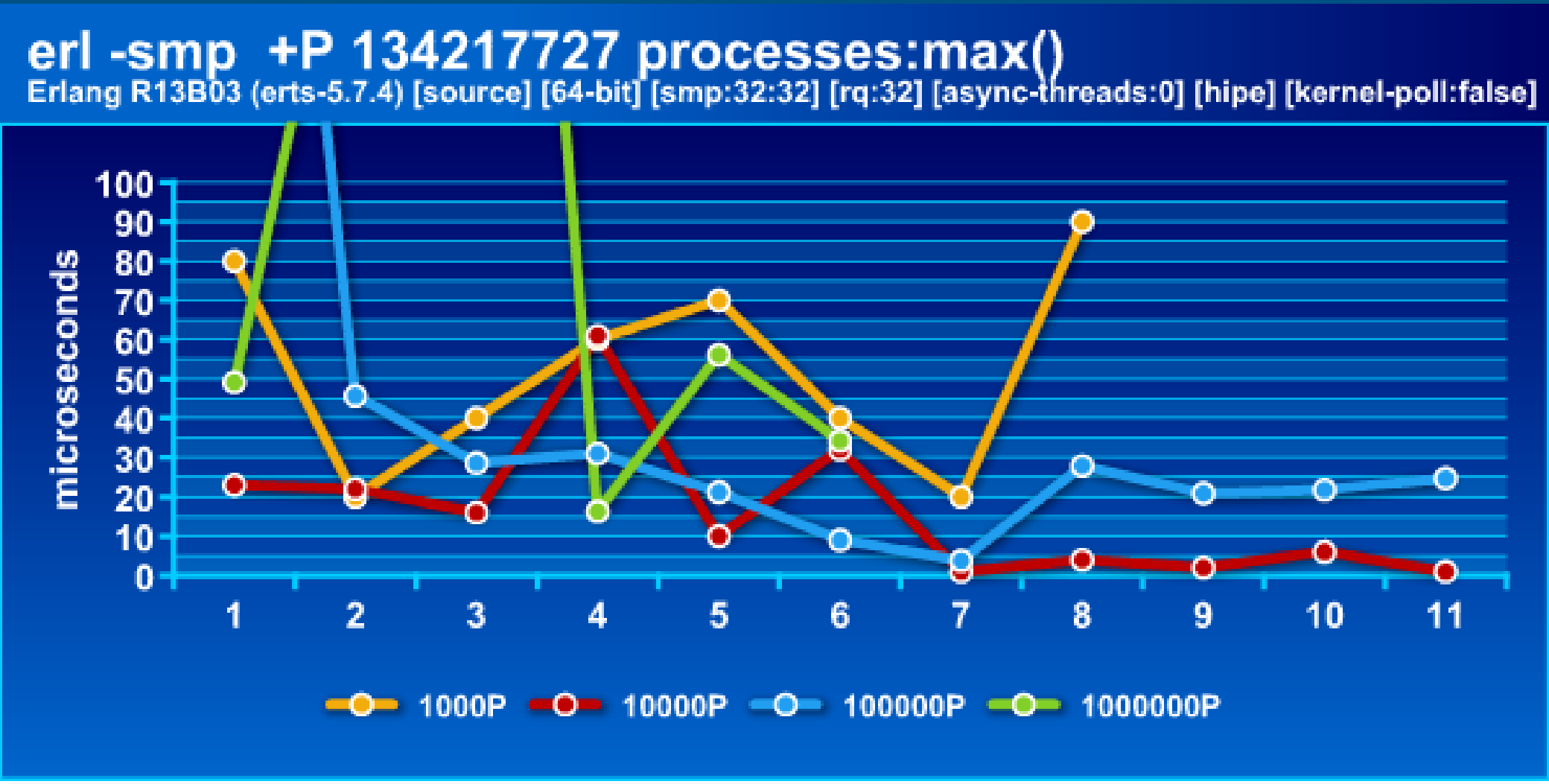
膨大な計算資源を最適に使う

SAKURA Internet

```
top - 14:16:47 up 20:30,  4 users,  load average: 3.94, 1.24, 0.72
Tasks: 643 total,   1 running, 641 sleeping,   1 stopped,   0 zombie
Cpu0  : 13.2%us, 67.9%sy,  0.0%ni, 18.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu1  :  6.7%us, 78.1%sy,  0.0%ni, 15.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu2  : 10.5%us, 82.9%sy,  0.0%ni,  6.7%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu3  :  9.4%us, 71.7%sy,  0.0%ni, 18.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu4  : 10.5%us, 73.3%sy,  0.0%ni, 16.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu5  :  9.3%us, 76.6%sy,  0.0%ni, 14.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu6  :  9.4%us, 72.6%sy,  0.0%ni, 17.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu7  :  9.9%us, 72.1%sy,  0.0%ni, 18.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu8  :  8.5%us, 84.9%sy,  0.0%ni,  6.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu9  :  9.6%us, 76.0%sy,  0.0%ni, 14.4%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu10 : 10.0%us, 71.8%sy,  0.0%ni, 18.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu11 : 19.3%us, 61.5%sy,  0.0%ni, 19.3%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu12 : 15.0%us, 66.4%sy,  0.0%ni, 18.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu13 :  8.3%us, 74.1%sy,  0.0%ni, 17.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu14 :  8.7%us, 81.7%sy,  0.0%ni,  9.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu15 : 10.6%us, 79.8%sy,  0.0%ni,  9.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu16 :  7.5%us, 83.0%sy,  0.0%ni,  9.4%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu17 : 11.6%us, 83.2%sy,  0.0%ni,  5.3%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu18 :  8.9%us, 82.2%sy,  0.0%ni,  8.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu19 :  8.7%us, 76.0%sy,  0.0%ni, 15.4%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu20 : 11.1%us, 82.4%sy,  0.0%ni,  6.5%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu21 :  8.1%us, 78.4%sy,  0.0%ni, 13.5%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu22 : 13.5%us, 74.0%sy,  0.0%ni, 12.5%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu23 : 20.2%us, 66.7%sy,  0.0%ni, 13.1%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu24 :  8.9%us, 78.2%sy,  0.0%ni, 12.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu25 :  4.0%us, 86.1%sy,  0.0%ni,  9.9%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu26 :  9.3%us, 81.4%sy,  0.0%ni,  9.3%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu27 : 31.8%us, 48.2%sy,  0.0%ni, 20.0%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu28 :  9.4%us, 76.0%sy,  0.0%ni, 14.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu29 : 10.8%us, 73.5%sy,  0.0%ni, 15.7%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu30 :  8.6%us, 82.9%sy,  0.0%ni,  8.6%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Cpu31 :  6.7%us, 78.1%sy,  0.0%ni, 15.2%id,  0.0%wa,  0.0%hi,  0.0%si,  0.0%st
Mem:  18095848k total,  4380656k used, 13715192k free,     1288k buffers
Swap:  1327100k total,   107788k used,  1219312k free,    11868k cached
```

計算機の特性を理解する

SAKURA Internet



アプリケーション特性と計算機特性を一致させる

広帯域で低価格なメディア選択する

SAKURA Internet

Infiniband vs Ethernet Performance report

Short packet (64byte)

558Mbits/sec (684kpps)

247Mbits/sec (335kpps)

Xeon E5620 2.4GHz (2P/8C)

Default packet size(1500,2044byte)

8.40Gbits/sec

6.26Gbits/sec *IP over Infiniband connected mode (RFC4755)

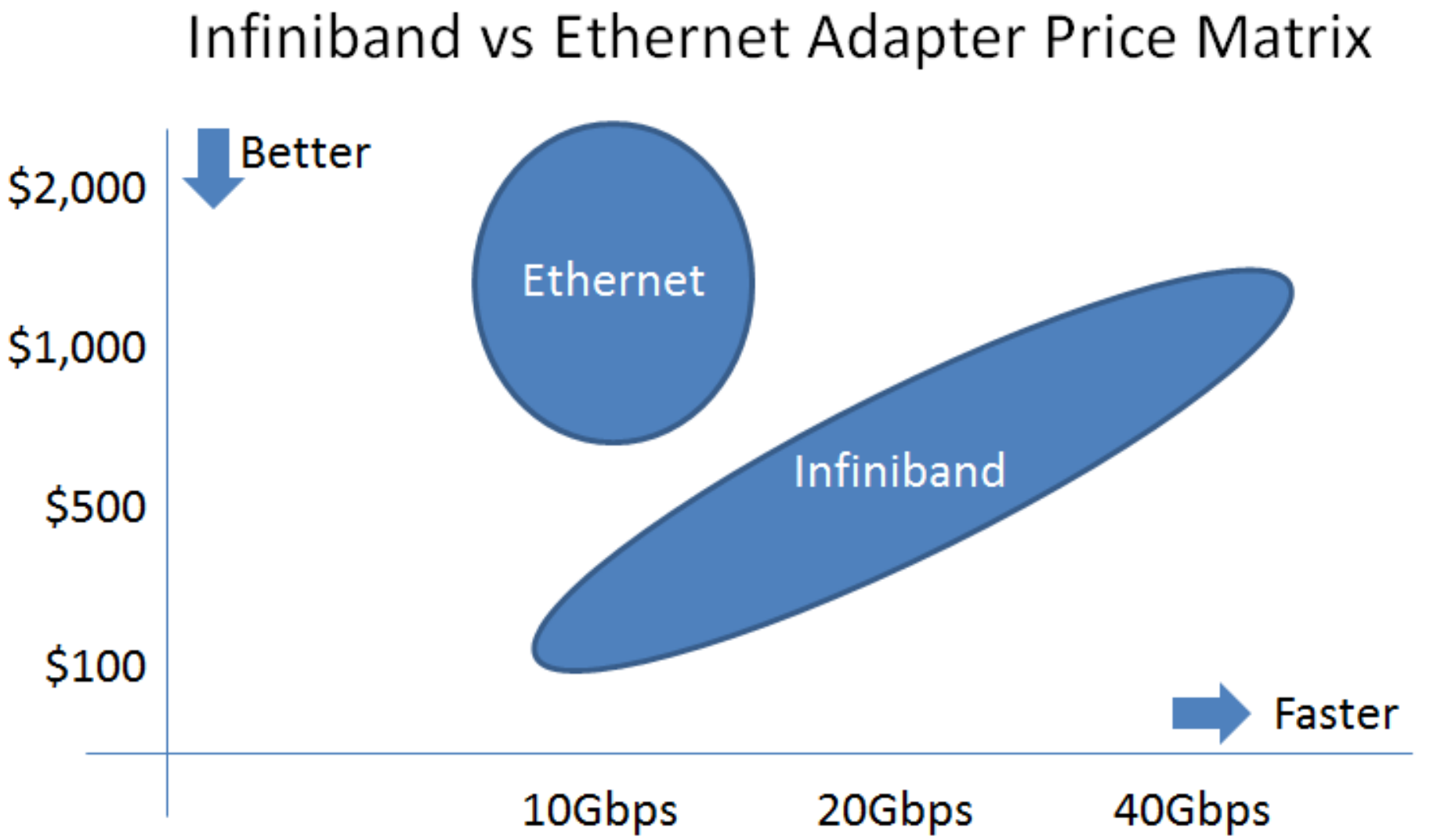
Using iperf TCP mode between 2 Servers back-to-back.

Mellanox InfiniHost III Lx(MHES14-XTC) Intel 10 Gigabit XF Server Adapter(EXPX9502AFXSR)

Source: SAKURA Internet ,Inc. 2010.DEC

広帯域で低価格なメディア選択する

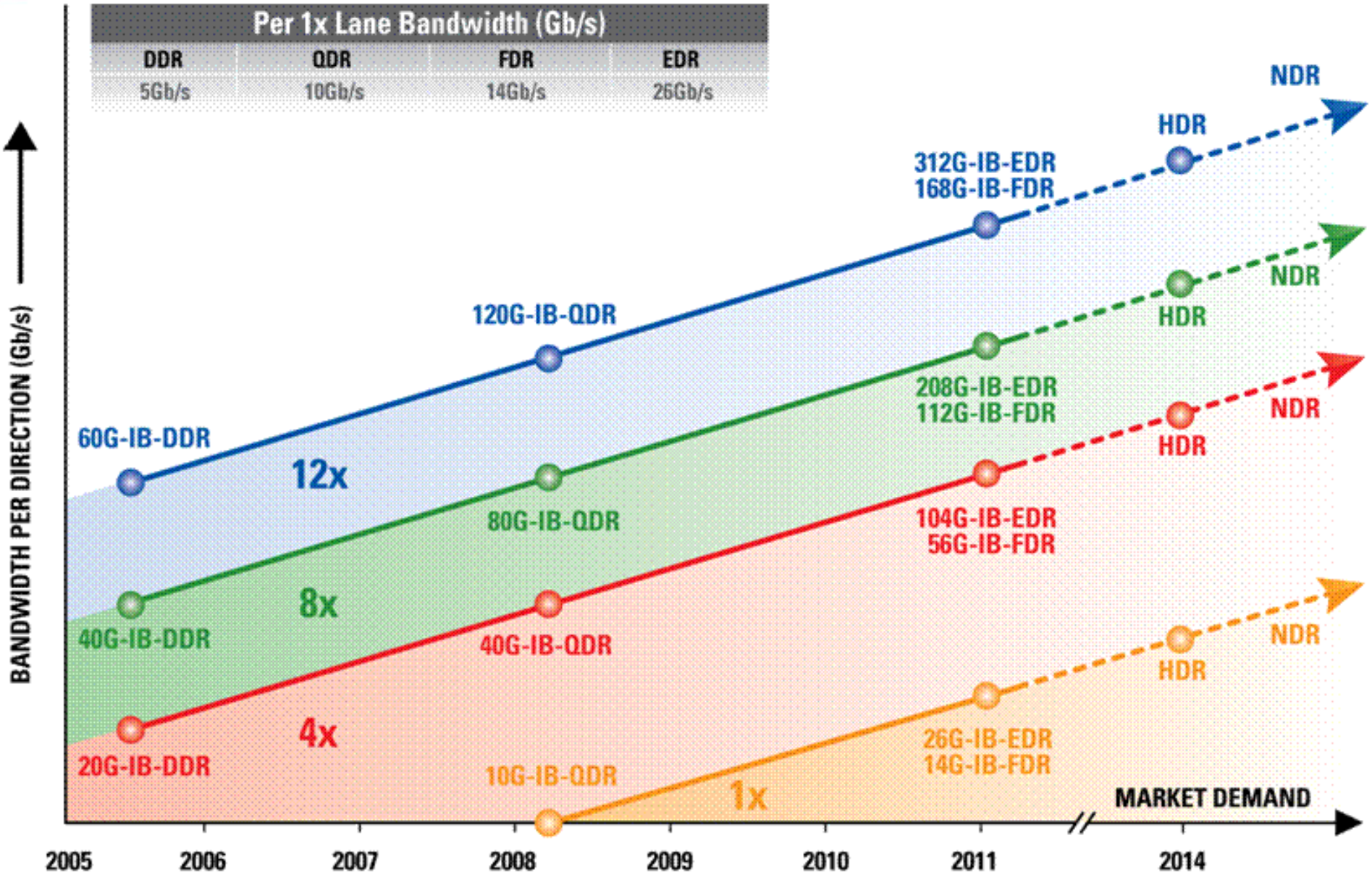
SAKURA Internet



Source: SAKURA Internet, Inc. 2010.DEC

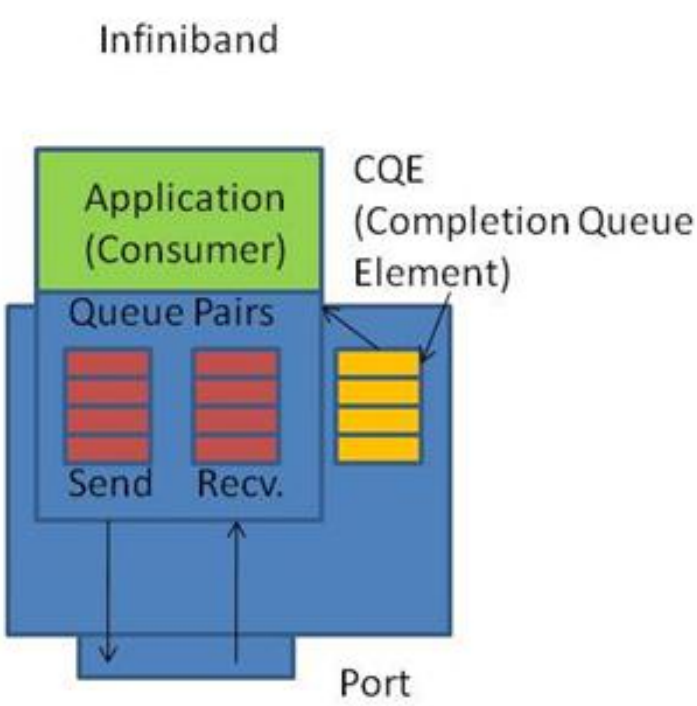
広帯域化するデータセンター内通信

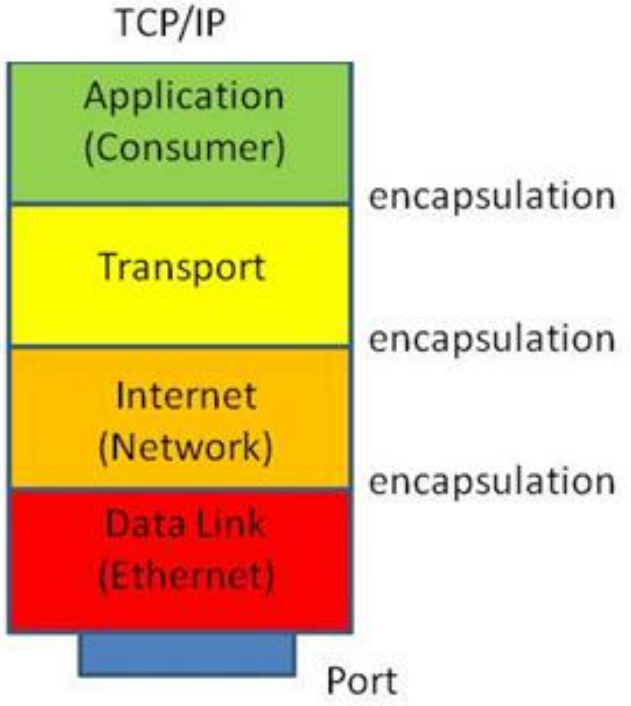
SAKURA Internet

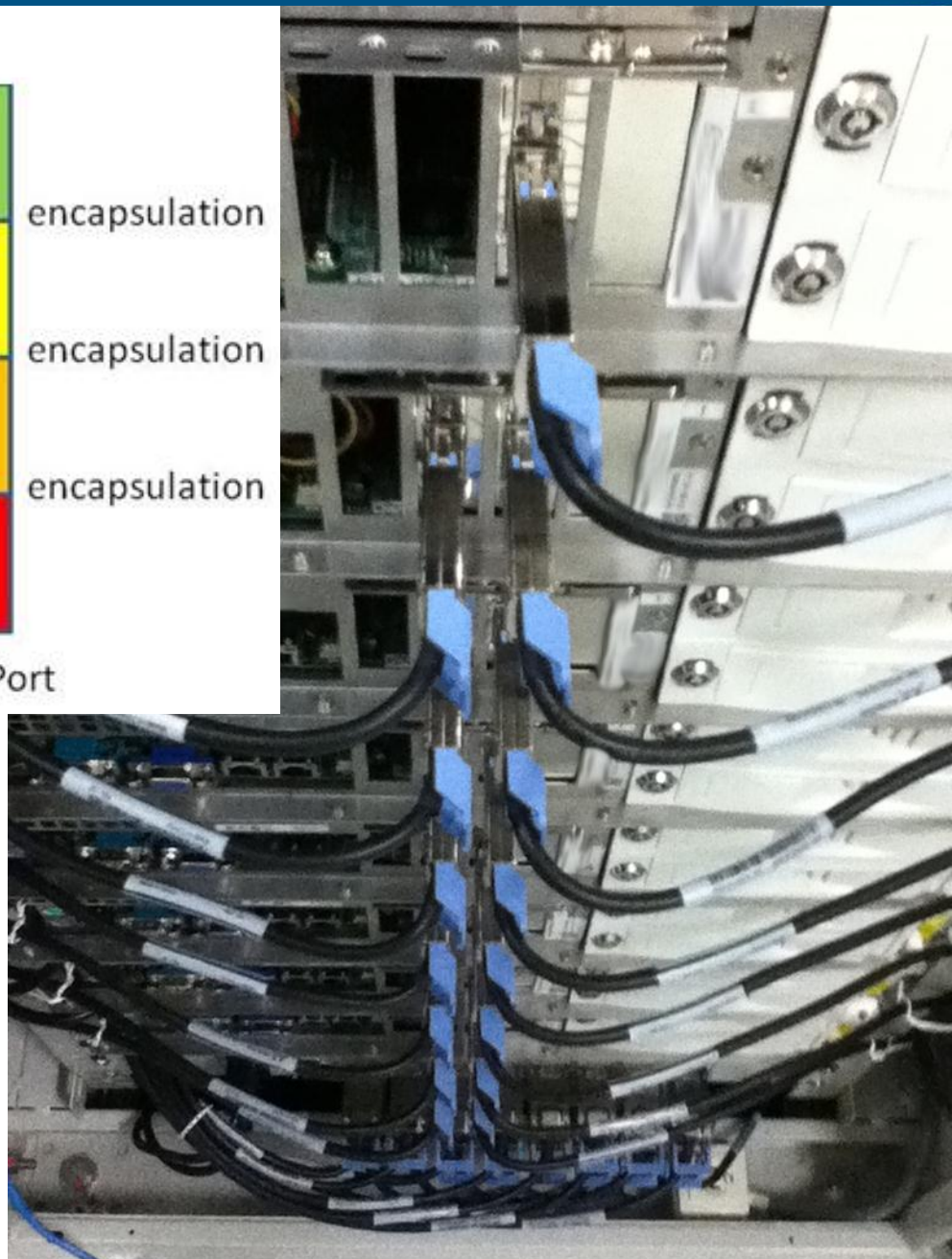


メディア毎のネットワーク特性を把握する

SAKURA Internet







ストレージの限界を理解する

SAKURA Internet

Sequential Performance with LSI 630J JBOD

System Configuration: • LSI 630J JBOD • LSI MegaRAID 9280DE-8e • 12 SATA2 Drives • Two SFF-8088 cables • One ESM	Read MB/sec 256kb Request Size	Write MB/sec 256kb Request Size	Read IOPs 4kb Request Size	Write IOPs 4kb Request Size	Applies to all results: • QD 256kb • 1 worker, 1 unit • No Read Ahead • 100% Sequential I/O • Run time 2 minutes • Ramp time 30 secs.
RAID 0	887	886	143,390	125,631	Write Through
RAID 5	810	804	87,177	136,793	Write Back
RAID 6	731	697	74,296	133,643	Write Back
RAID 10	745	443	62,288	108,987	Write Through

Tyan S7020AGM2NR motherboard w/ 2x Intel quad core Xeon 2.0GHz, 4 x 2GB DRAM, Windows 2003 Enterprise 32 bit SP2.

Sequential Performance with LSI 620J JBOD

System Configuration: • LSI 620J JBOD • LSI MegaRAID 9280DE-8e • 24 SAS2 Drives • Two SFF-8088 cables • One ESM	Read MB/sec 256kb Request Size	Write MB/sec 256kb Request Size	Read IOPs 4kb Request Size	Write IOPs 4kb Request Size	Applies to all results: • QD 256kb • 1 worker, 1 unit • No Read Ahead • 100% Sequential I/O • Run time 2 minutes • Ramp time 30 secs.
RAID 0	2,521	3,026	177,491	173,077	Write Through
RAID 5	2,426	1,323	120,527	116,252	Write Back
RAID 6	2,464	1,300	123,513	118,475	Write Back
RAID 10	2,324	1,515	170,572	149,117	Write Through

Tyan S7020AGM2NR motherboard w/ 2x Intel quad core Xeon 2.0GHz, 4 x 2GB DRAM, Windows 2003 Enterprise 32 bit SP2.

現場作業は、決して無くなりません。

SAKURA Internet

SAKURA Internet

www.sakura.ad.jp

ご清聴誠にありがとうございました

(C)Copyright 1996-2010 SAKURA Internet Inc.