S8 ルーティング関連セッション（Ⅱ）
～ 経路爆発を考える ～ rev1.2

Shishio Tsuchiya

shtsuchi@cisco.com

© 2010 Cisco and/or its affiliates. All rights reserved.

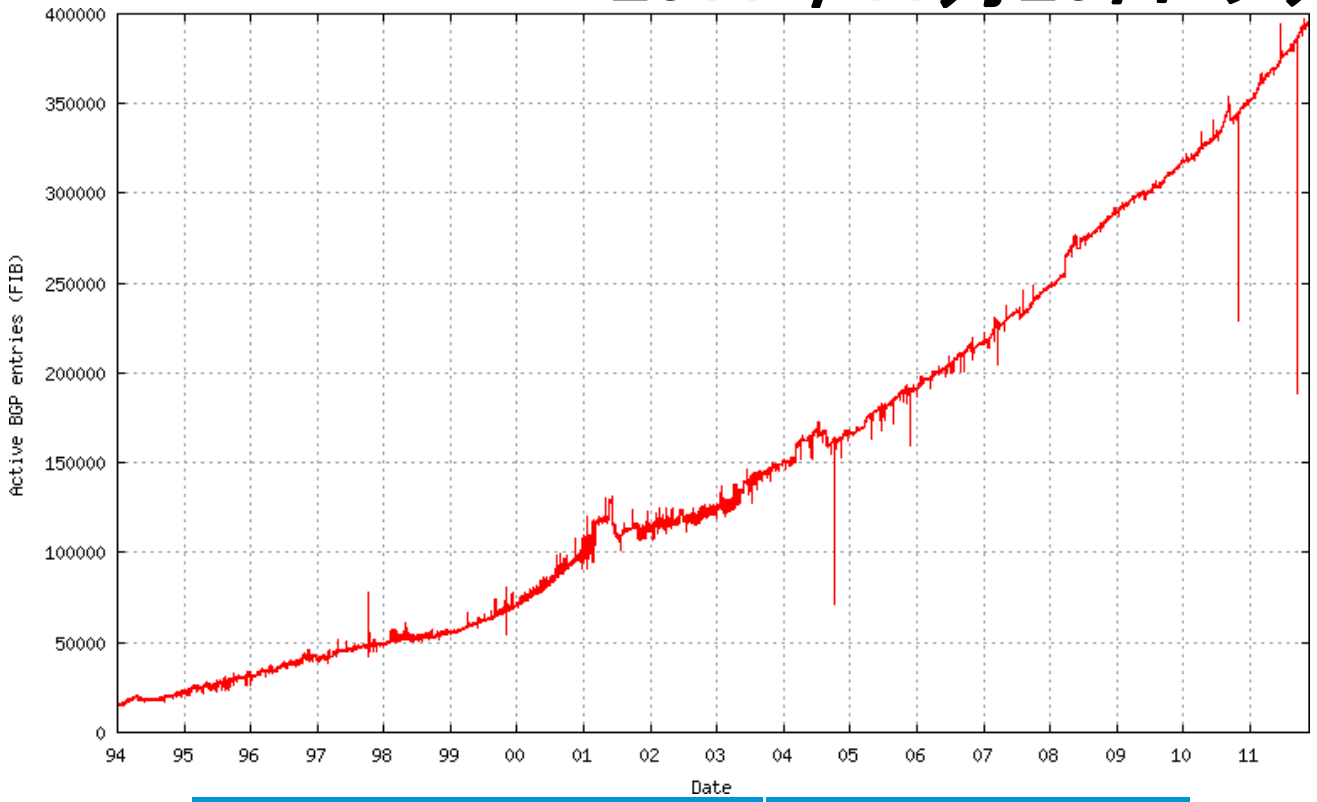
Part 2. – 経路爆発問題に対する対処は？！

- 通信事業者自身による対応
- 通信機器による対応
- 新しいプロトコルの導入

© 2010 Cisco and/or its affiliates. All rights reserved.

AS6447 BGP Routing Table Analysis Report

2011年11月20日のデータ



FIB / RIB Table	Data
Active BGP entries (FIB)	396,184
All BGP entries (RIB)	12,561,626
RIB/FIB ratio	31.7065

http://bgp.potaroo.net/as6447/

© 2010 Cisco and/or its affiliates. All rights reserved.

University of Oregon Route Views Project

http://www.routeviews.org/

- いくつかの異なるバックボーン・ロケーションから見たインターネット全体の状況をリアルタイムに観測する為のツール
- マルチホップeBGPセッション
- ネイバーからルートを受け取り、トラフィックはフォワードしない
- route view自身はアドバタイズしない
- アクティブIPv4　39ピア　IPv6 11ピア

© 2010 Cisco and/or its affiliates. All rights reserved.

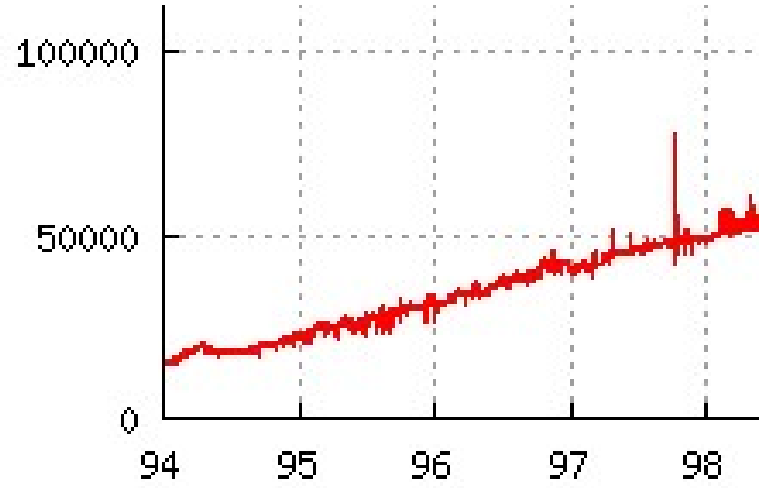
routeview via telnet

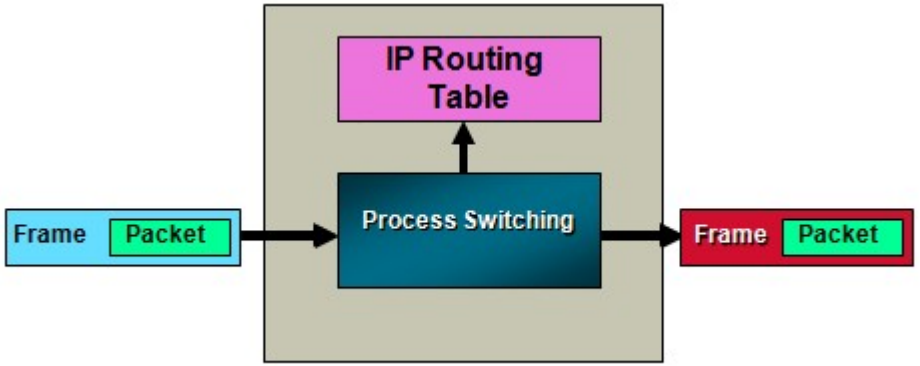
```
route-views>show bgp ipv4 unicast summary
BGP router identifier 128.223.51.103, local AS number 6447
BGP table version is 1340205878, main routing table version 1340205878
412308 network entries using 54424656 bytes of memory
12887562 path entries using 670153224 bytes of memory
2138823/74265 BGP path/bestpath attribute entries using 359322264 bytes of memory
1895813 BGP AS-PATH entries using 74990584 bytes of memory
47224 BGP community entries using 3511002 bytes of memory
115 BGP extended community entries using 3036 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 1162404766 total bytes of memory
Dampening enabled. 4770 history paths, 5215 dampened paths
BGP activity 764720/339603 prefixes, 52952232/39975714 paths, scan interval 60 secs
```

- BGPのみで1GB以上のメモリを使用

© 2010 Cisco and/or its affiliates. All rights reserved.

1998年頃までのBGPテーブルと転送方式

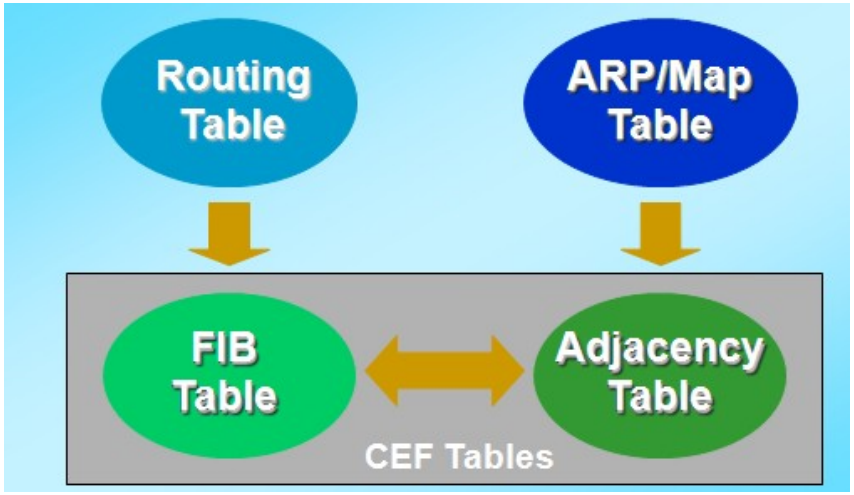


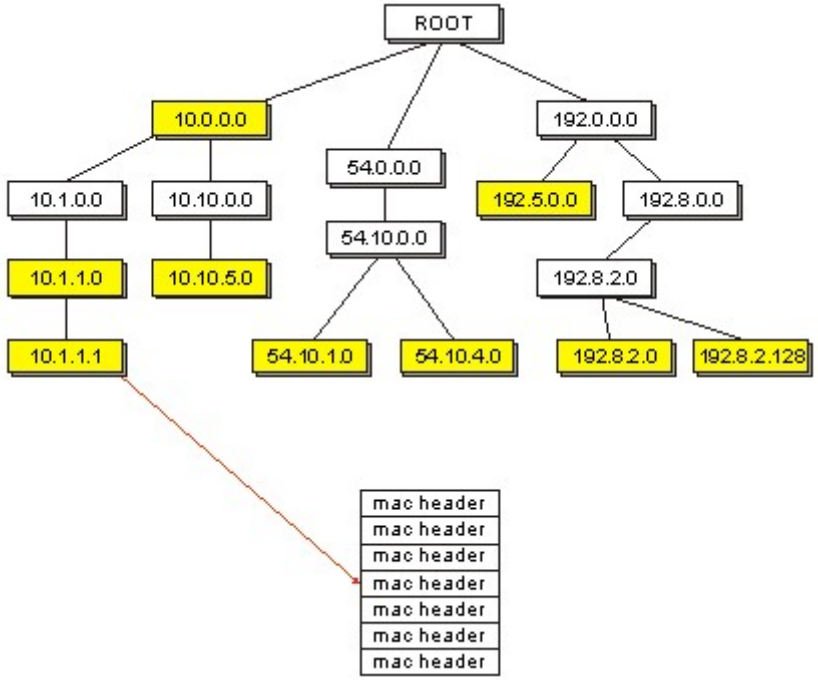


- インターネット全体でも50K程のルーティングテーブル
- 転送方式もパケットが到達するとルーティングテーブルを確認するプロセススイッチや一度確認したフローの結果をキャッシュするファーストスイッチが主流
- しかし、CIDR(RFC1518,RFC1519)やBGP4(RFC1771)により、今後ルーティングがテーブルが複雑化・肥大化する事が予想された。
- またギガビットイーサネットなどの登場により、高速なテーブルルックアップが必要になると予想された。

© 2010 Cisco and/or its affiliates. All rights reserved.

Innovation for Gigabit network

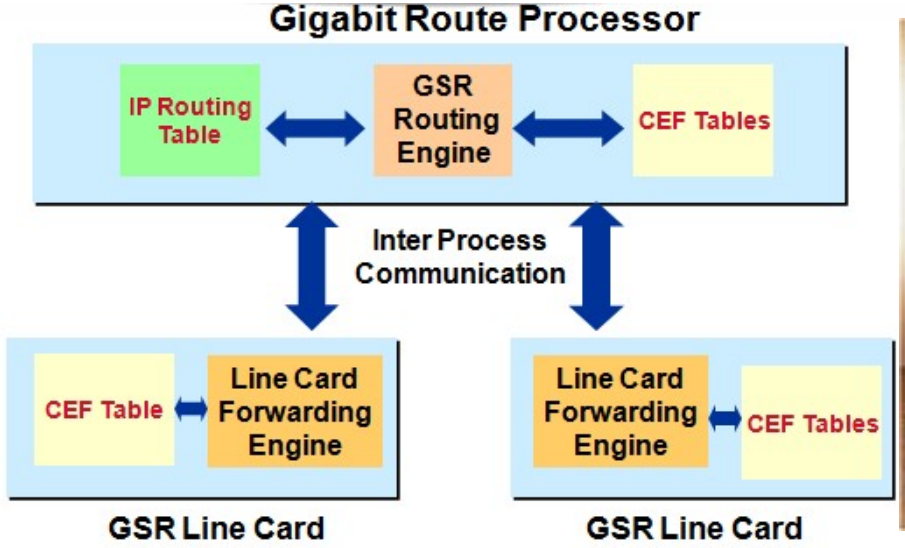


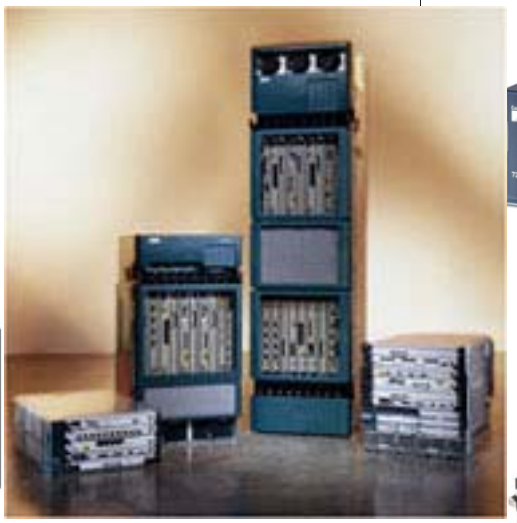


- データプレーンとコントロールプレーンの分離
- テーブルの階層化(8-8-8-8)により、高速ルックアップの実現

© 2010 Cisco and/or its affiliates. All rights reserved.

Innovation for Gigabit network *cont'd*

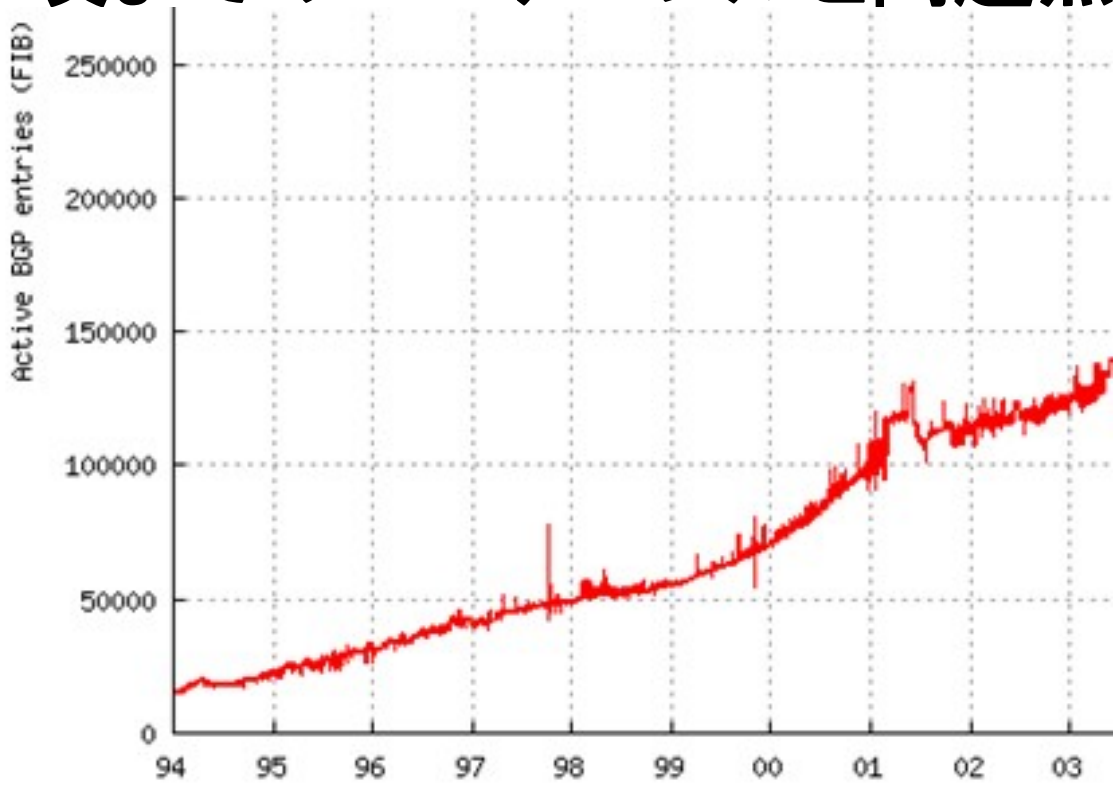




- ハードウェア化も実現 GSR/PFC
- データプレーンを分離する事で分散化も実現 GSR/DFC
- ソフトウェアルータにもスイッチング方式として共有化

© 2010 Cisco and/or its affiliates. All rights reserved.

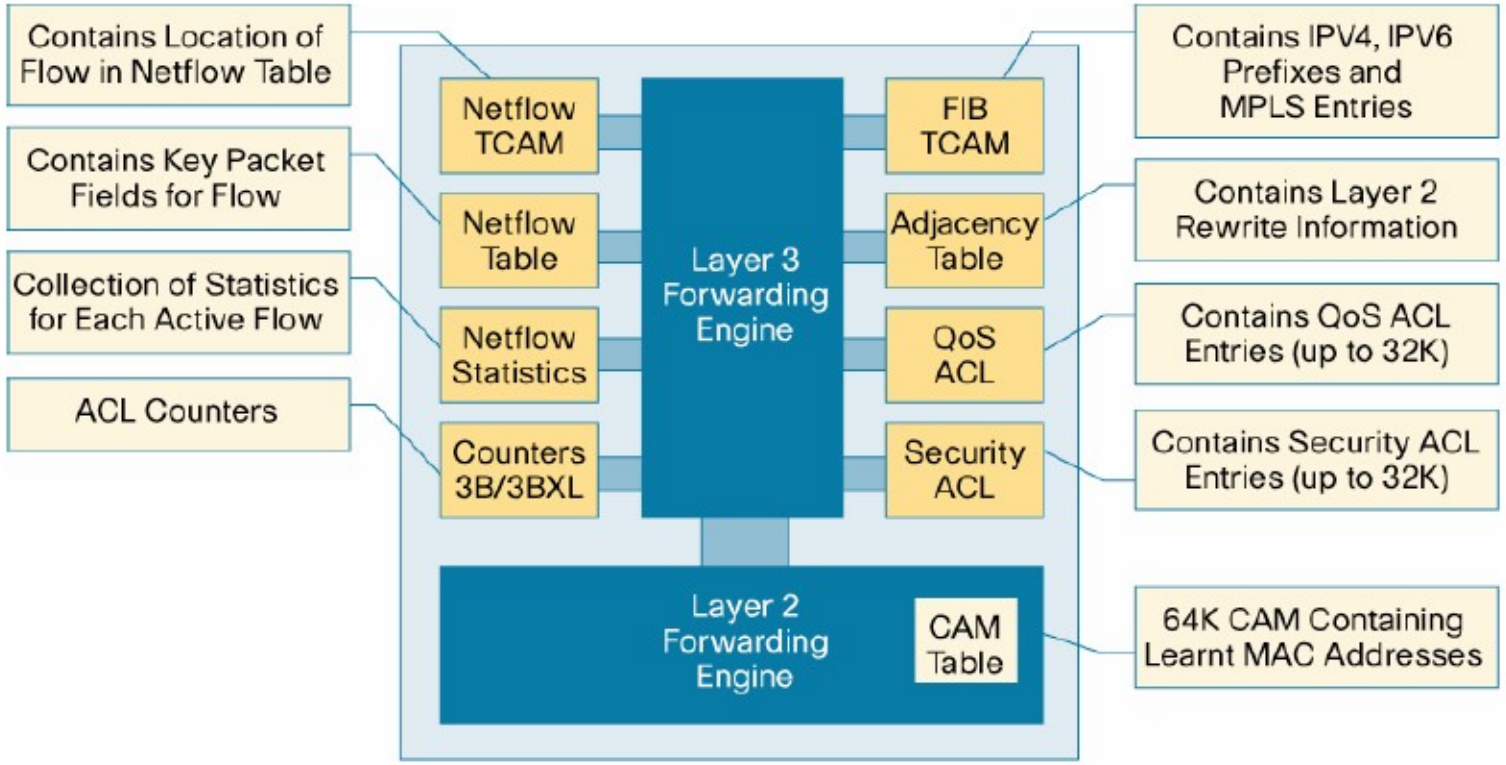
2003年頃までのBGPテーブルと問題点



- ハードウェアにおける256Kまでの制限が近づく(150K)
- IPv6の商用サービスも始まる(IPv6ハードウェア転送)
- QoS/Netflow/ACLなどのサービスのハードウェア化

© 2010 Cisco and/or its affiliates. All rights reserved.

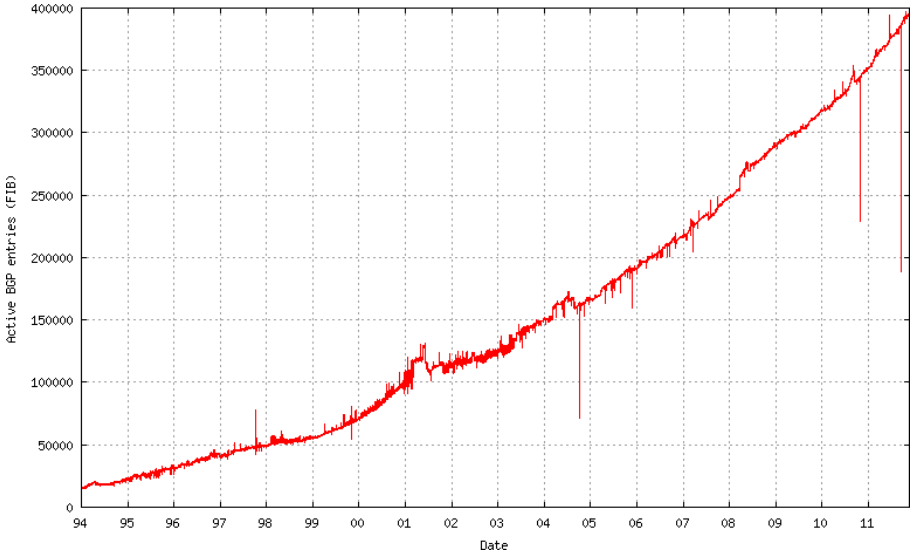
Innovation for Multiservice
TCAM (Ternary Content Addressable Memory)

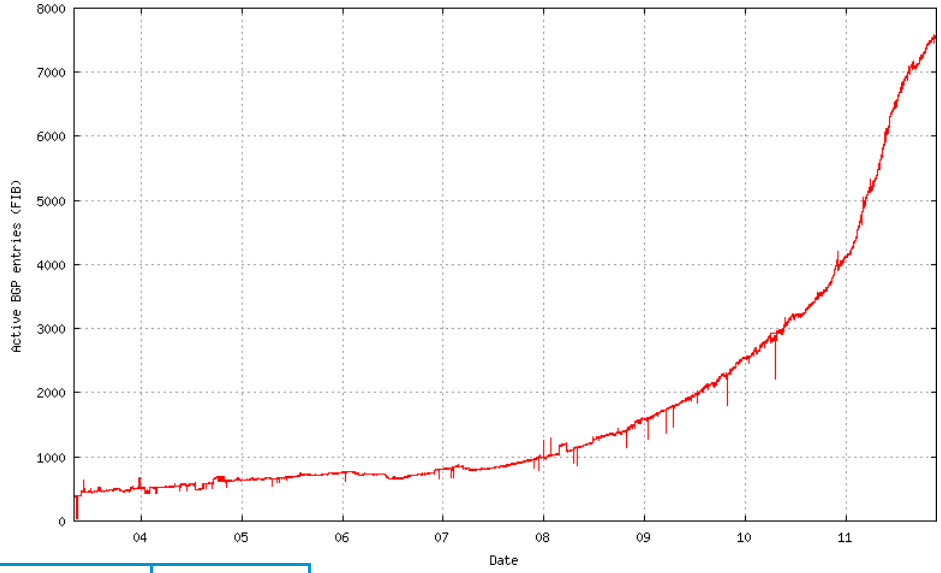


- 三値連想記憶(0,1 および Don't Care)で高速ルックアップ
- サービス適用やサブネットマスク(経路計算)に適用
- 容量を変える事で、SUP-3BXL,SUP-3B,Catalyst 3750など幅広いラインナップに適用

© 2010 Cisco and/or its affiliates. All rights reserved.

最近のBGPテーブルと問題点





	IPv4フルルート	成長率	IPv6フルルート	成長率
2009年1月	284,194	–	1,596	–
2010年1月	312,989	10.1%	2,458	54.0%
2011年1月	341,483	9.1%	4,100	66.8%
2011年11月	396,184	17.0%	7,581	94.8%

http://bgp.potaroo.net/v6/as6447/

http://www.potaroo.net/ispcol/2011-11/bgp2011.html

- IPv4　396,184 ≒400K　IPv6 7581≒8K
- 2011年2月のIANAにおけるIPv4アドレスの枯渇、World IPv6 Dayなどの影響により、IPv4/IPv6ルートが更に増加傾向にある

© 2010 Cisco and/or its affiliates. All rights reserved.

2003年頃の導入機器
SUP-720

	SUP-720	SUP720-3B	SUP720-3BXL
IPv4フォワーディング	400Mpps	400Mpps	400Mpps
IPv6フォワーディング	200Mpps	200Mpps	200Mpps
デフォルトTCAM	192,000(IPv4) 32,000(IPv6)	192,000(IPv4) 32,000(IPv6)	512,000(IPv4) 256,000(IPv6)
最大サイズ	239,000(IPv4) 119,000(IPv6)	239,000(IPv4) 119,000(IPv6)	1,007,000(IPv4) 503,000 (IPv6)

http://www.cisco.com/en/US/docs/ios-xml/ios/ipswitch/command/isw-i1.html#GUID-4E2B2432-60E7-4D95-A9EB-B90024B27229

- 2005年-(170k-200k)あたりでデフォルトのTCAMサイズではルートがあふれる事象が起こる
- mls cef maximum-routes {ip | ip-multicast | ipv6 | mpls} *maximum-routes*でそれぞれの割合を変える事が可能

© 2010 Cisco and/or its affiliates. All rights reserved.

2003年頃の導入機器
Engine 3 4GE-SFP-LC

```
R1#show controllers ISE 4 tcam
Total Tcam size = 100%

Reg#    Name     Config(%, Cells)        Default(%, Cells)
-------------------------------------------------------------
0  RX_TOP_NF    33.00 % (86507 cells)   33.00 % (86507 cells)
1  RX_TOP_72b    1.00 % (2621 cells)     1.00 % (2621 cells)
2  RX_TOP_144b   1.00 % (2621 cells)     1.00 % (2621 cells)
3  RX_TOP_288b   1.00 % (2621 cells)     1.00 % (2621 cells)
4  RX_72b        4.00 % (10485 cells)    4.00 % (10485 cells)
5  RX_144b      20.00 % (52428 cells)   20.00 % (52428 cells)
6  RX_288b      29.00 % (76021 cells)   29.00 % (76021 cells)
7  RX_IPv6_128   4.00 % (10485 cells)    4.00 % (10485 cells)
8  RX_IPv6_127   0.00 % (0 cells)        0.00 % (0 cells)
```

```
SLOT 4:00:05:08: %EE48-3-IPV6_TCAM_CAPACITY_EXCEEDED: IPv6 pkts will be software switched.
To support more IPv6 routes in hardware:
 Get current TCAM usage with: show controllers ISE <slot> tcam
 In config mode, reallocate TCAM regions e.g. reallocate Netflow TCAM to IPv6
  hw-module slot <num> tcam carve rx_ipv6_1 <prefix> <v6-percent>
  hw-module slot <num> tcam carve rx_top_nf <nf-percent>
 Verify with show command that sum of all TCAM regions = 100%
 Reload the linecard for the new TCAM carve config to take effect
WARNING: Recarve may affect other input features(ACL,CAR,MQC,Netflow)
```

- E3はデフォルトでIPv6 FIBに4% 5120prefixが保持可能
- これを超えるとソフトウェアスイッチとなる
- 2011年よりIPv6は急成長し、4K->8Kへ

© 2010 Cisco and/or its affiliates. All rights reserved.

2003年頃の導入機器
Engine 3 4GE-SFP-LC　*cont'd*

```
R1(config)#hw-module slot 4 tcam carve rx_ipv6_144b_REGION 128 36
R1(config)#hw-module slot 4 tcam carve rx_top_nf_REGION 1
R1(config)#microcode reload
```

```
R1#show controllers ISE 4 tcam
Total Tcam size = 100%

Reg#    Name      Config(%, Cells)        Default(%, Cells)
-------------------------------------------------------------
0  RX_TOP_NF      1.00 % (2621 cells)     33.00 % (86507 cells)
1  RX_TOP_72b     1.00 % (2621 cells)      1.00 % (2621 cells)
2  RX_TOP_144b    1.00 % (2621 cells)      1.00 % (2621 cells)
3  RX_TOP_288b    1.00 % (2621 cells)      1.00 % (2621 cells)
4  RX_72b         4.00 % (10485 cells)     4.00 % (10485 cells)
5  RX_144b       20.00 % (52428 cells)    20.00 % (52428 cells)
6  RX_288b       29.00 % (76021 cells)    29.00 % (76021 cells)
7  RX_IPv6_128   36.00 % (94371 cells)     4.00 % (10485 cells)
8  RX_IPv6_127    0.00 % (0 cells)         0.00 % (0 cells)
```

- デフォルトでNetflow(RX_TOP_NF)が33%、IPv6(RX_IPv6_128)が4%となっている
- 設定により、IPv6:Netflow=36:1に変更する
- 9*デフォルト≒45K prefixまで対応可能

© 2010 Cisco and/or its affiliates. All rights reserved.

現在のルータに求められている要求

- スピード！

Lookup Speed

10GE 15Mpps

100GE 150Mpps

- 大きさ

IPv4 FIB

IPv6 FIB

MACアドレス

MPLSラベル

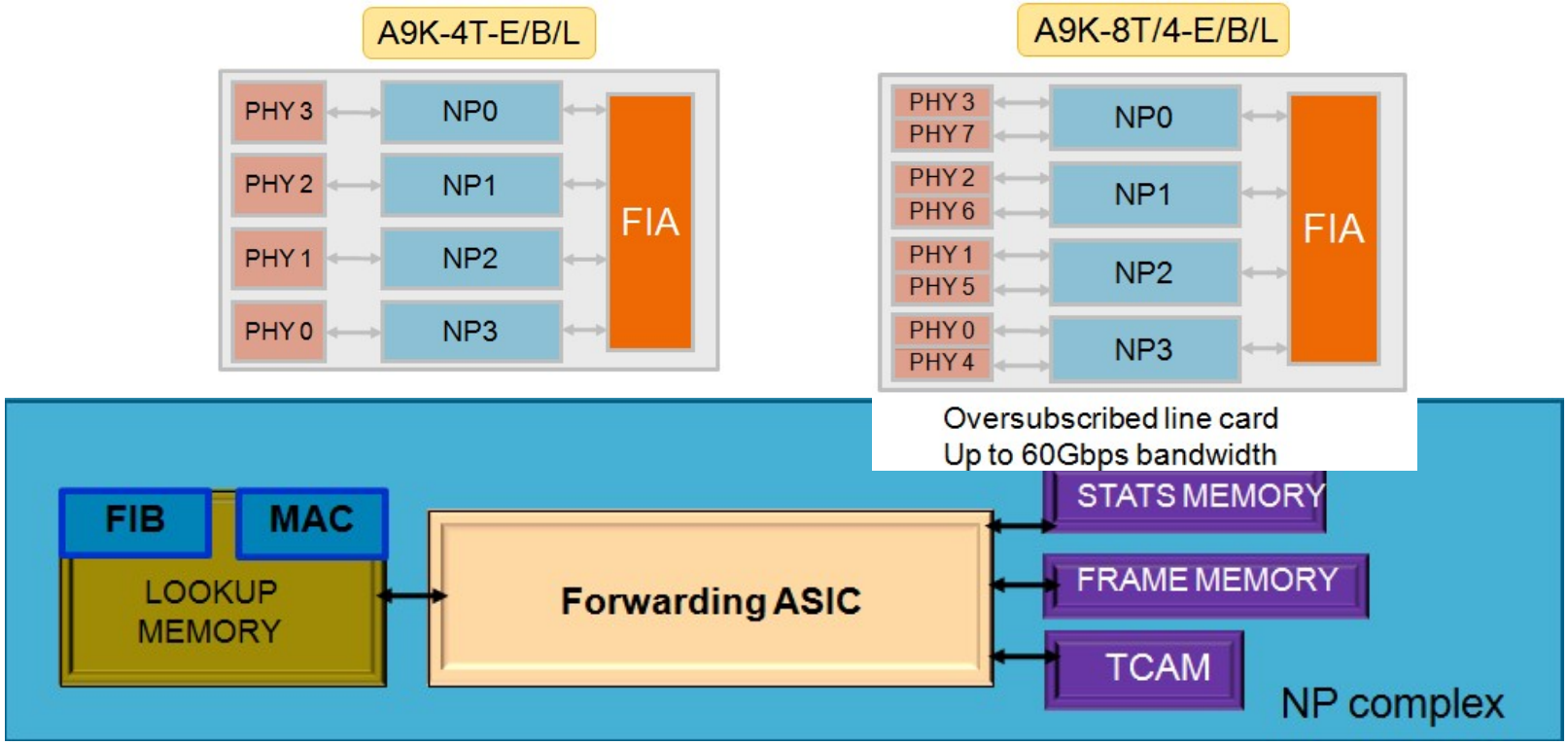
ACL,QOS

VLAN

省電力
低コスト
拡張性

© 2010 Cisco and/or its affiliates. All rights reserved.

Network Processor



- 柔軟なハードウェア構成
- TCAMの容量を極力省く事で、低コスト・省電力を実現

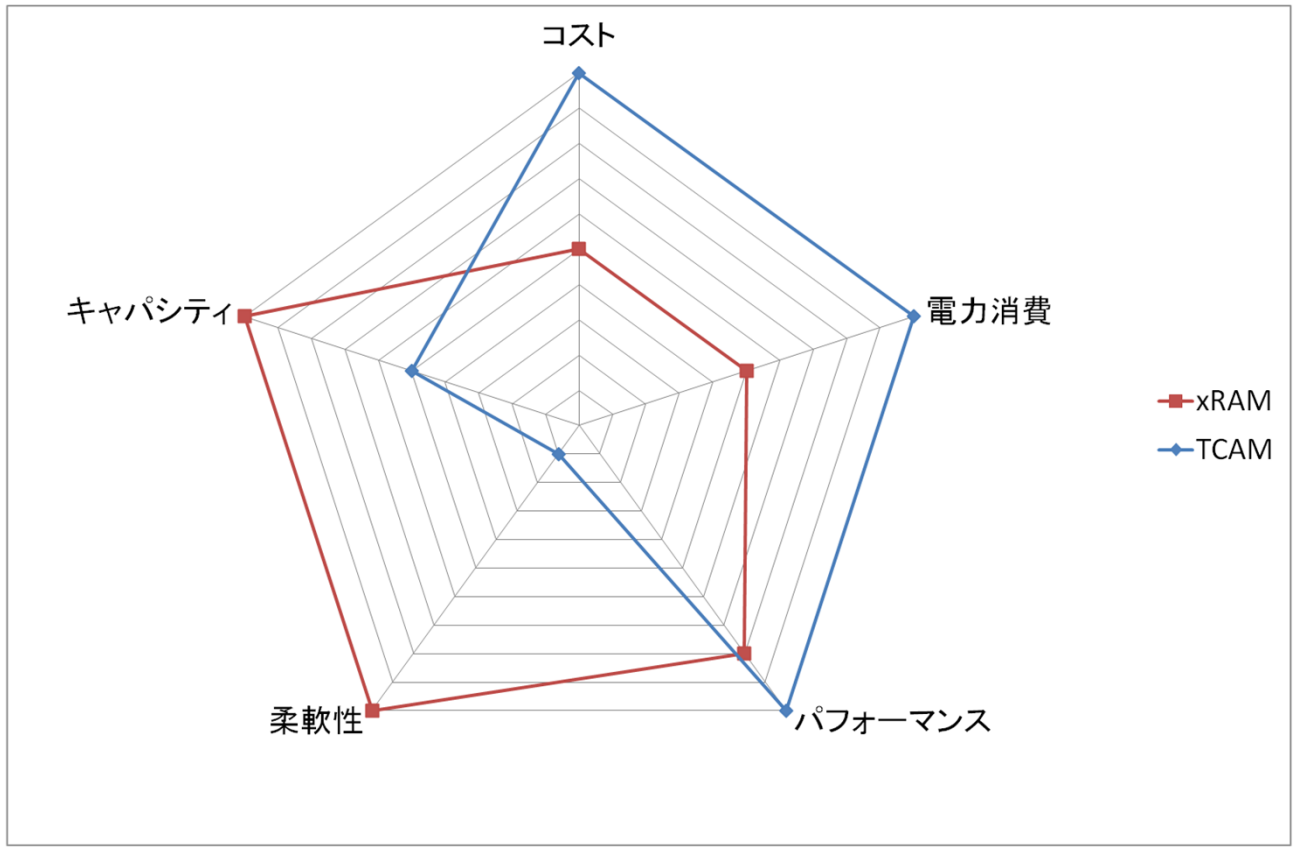
© 2010 Cisco and/or its affiliates. All rights reserved.

Network Processor *cont'd*

- それぞれのNPUは大きく4つで構成
- TCAM：VLANタグ・QoS・ACL処理
- Lookup Memory:FIB/MACアドレス/Adjacencyテーブルの保存
- Stats memory:インターフェースやフォワーディングの統計
- Frame memory:Queueメモリのバッファー

© 2010 Cisco and/or its affiliates. All rights reserved.

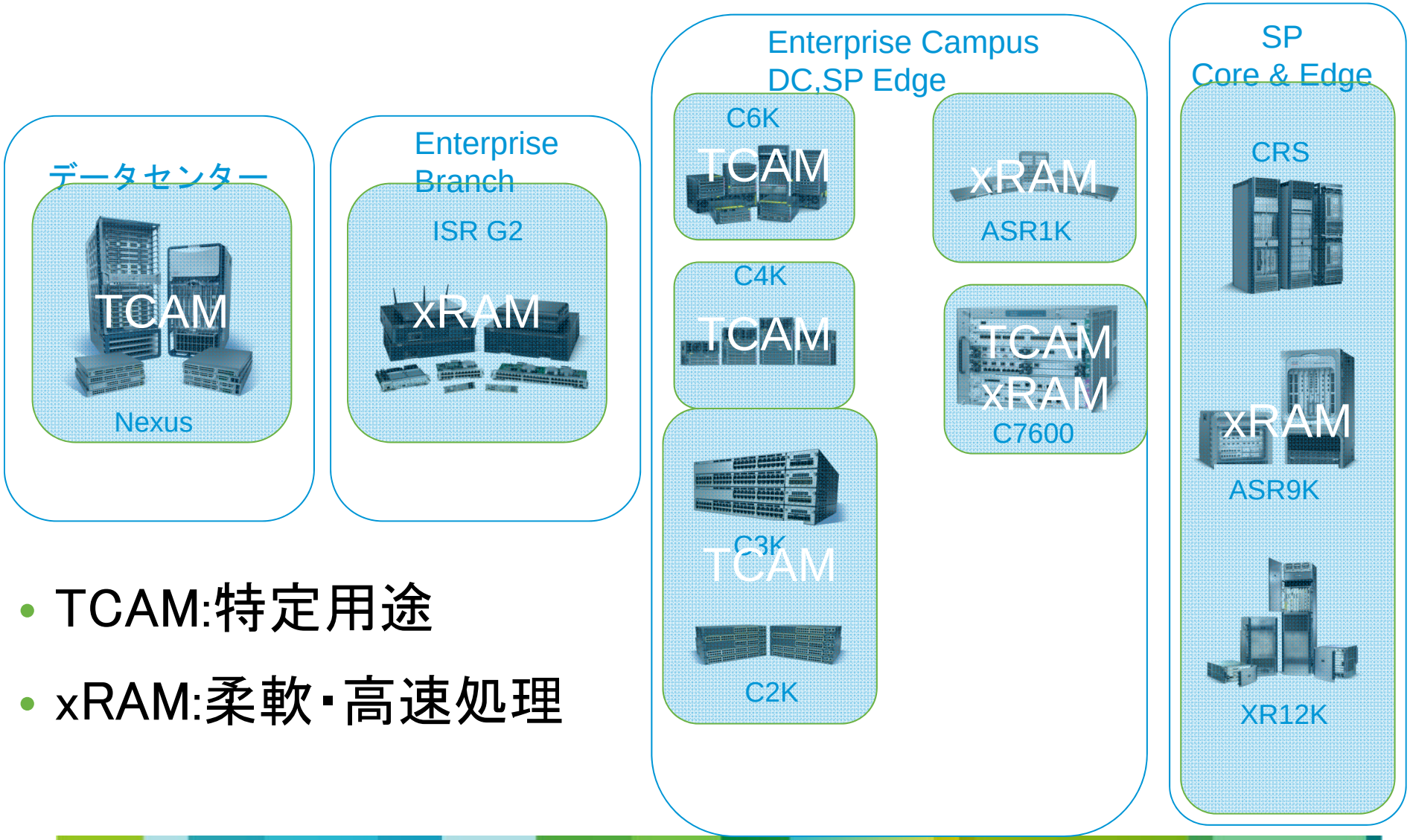
TCAM vs xRAM



- 一般的なTCAMおよびxRAMのメリット・デメリットを図示化したもの

© 2010 Cisco and/or its affiliates. All rights reserved.

Ciscoプラットホーム ルート格納方法まとめ



- TCAM:特定用途
- xRAM:柔軟・高速処理

© 2010 Cisco and/or its affiliates. All rights reserved.

まとめ

- 今後起こりうる経路爆発に対し、トラフィックを柔軟にかつ大容量、高速に処理出来るように、ネットワークプロセッサーの実装が進んでいる
- 既存のTCAMは容量を調整できる様な仕組みを実装している。
- TCAM実装機器はダウンサイジング・処理能力向上が行われた最新のTCAMを搭載
- 最新のBGP情報などを確認しつつ、実機のキャパシティを確認する事が望まれる。

© 2010 Cisco and/or its affiliates. All rights reserved.

Thank you.