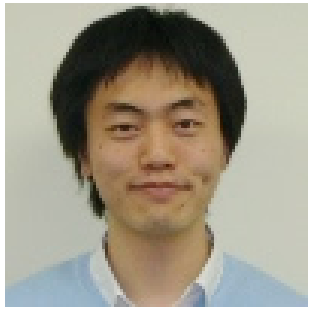
Internet Week 2013
S7 IPv4アドレス枯渇後の選択 ～IPv4アドレス移転と共有技術の最新動向～

IPv4アドレス共有技術設計方法とネットワークデザイン上の注意点

NTT コミュニケーションズ
西塚要
2013/11/28

自己紹介

西塚要



2006年 NTTコミュニケーションズ入社
2006～ OCN(AS4713) エッジNW設計・開発
2008～ 他社ISP(AS****) 提案・運用
2012～ IAC(先端IPアーキテクチャセンタ)にて、CGN関連技術の
IETF標準化活動等

社外活動
JANOG28 実行委員長
JANOG30 会場運営委員長など

CGN/Carrier Grade NAT(あるいはLSN/Large Scale NAT)とは？

Common Requirements for Carrier-Grade NATs(CGNs) [RFC6888:BCP127]

通常のNATと異なるキャリア網NATとしての要求条件

- セッション上限機能
- Fullcone NAT機能
- ポート割当手法(動的/静的)
- セッション保持時間
- NAT Logの記録

標準化とともにCGNの実装は洗練されてきました。
これからどのように展開するかが重要なフェーズです。

<Agenda>

1. CGN(Carrier Grade NAT)をめぐる背景
2. 想定するNW
3. CGN設計レシピ
   - アドレス設計編
   - ログ編
   - アプリケーション編
   - 性能評価編
   - ルーティング編
4. IPv6化によるトラフィックオフロード
5. まとめ

IPv4枯渇に対する最終的な解決策はIPv6への移行です。

しかし、現実的に進行するIPv4枯渇に対する暫定解として、IPv4アドレス共有技術(特にCGN)が選択され始めています。

1. CGNをめぐる背景

2013.05.07

UQではデフォルトのサービスをプライベートアドレス に移行。
グローバルアドレスはオプションで、追加100円/月が必要。

1. CGNをめぐる背景

2013.05.03

BTでもCGNのトライアルを開始
BT forumではアプリケーションへの影響が懸念されるという論調

1. CGNをめぐる背景

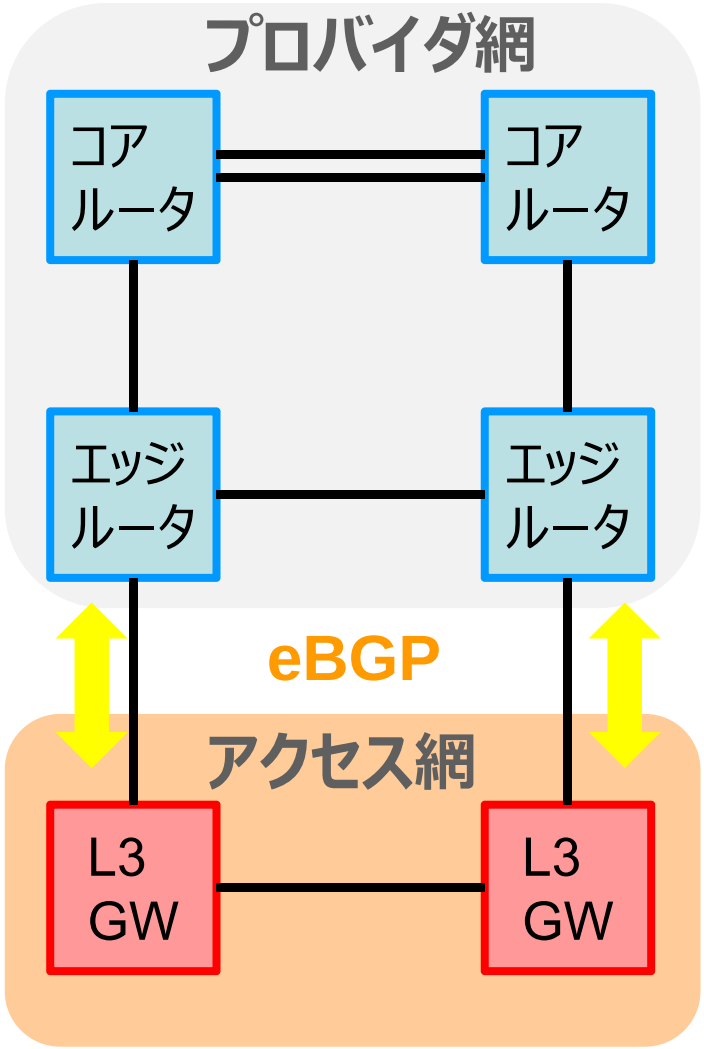
2013.07.22

MVNO事業者のワンコインSIMも、プライベートアドレスで提供

CGNを取り巻く状況は、数年前と大きく変わってきています。
～要因～
機器実装の洗練
価格の低廉化
Web利用状況およびWeb技術の変化

CGNをいざ導入するとなったときに、慌てないために、
設計のための”要所”を本日はお伝えします。

2. 今回の発表で想定するNW



<条件>
①中小規模のプロバイダ
(あるいは大規模プロバイダの1エリア)
②左記POI配下に15万ユーザ
③現状IPv4ユーザのみ

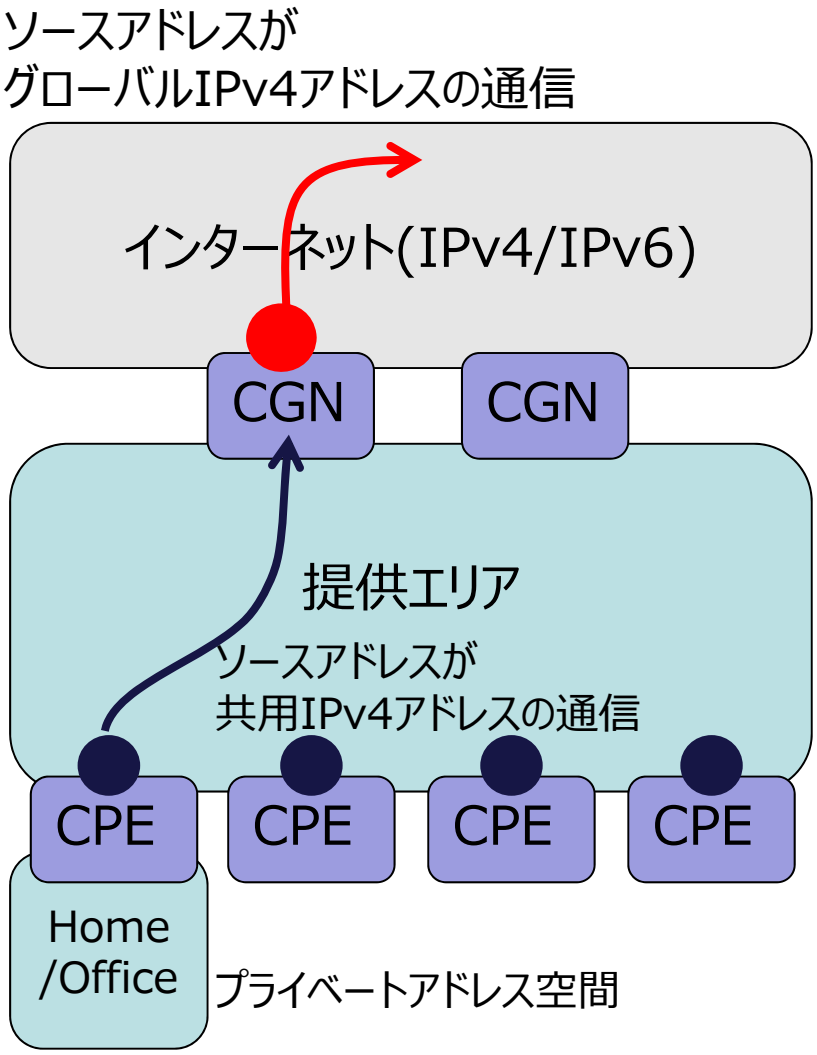
CGNの導入

<Question>
①どれだけアドレスが必要か
②CGN周辺システム(ログ等)はどのくらい必要か
③アプリケーションへの影響は
④CGNの性能はどのくらい必要か
⑤どこにCGNを入れるか。どのようにルーティングするか。

①アドレス設計編

「どれだけアドレスが必要か」

CGNのアドレッシング

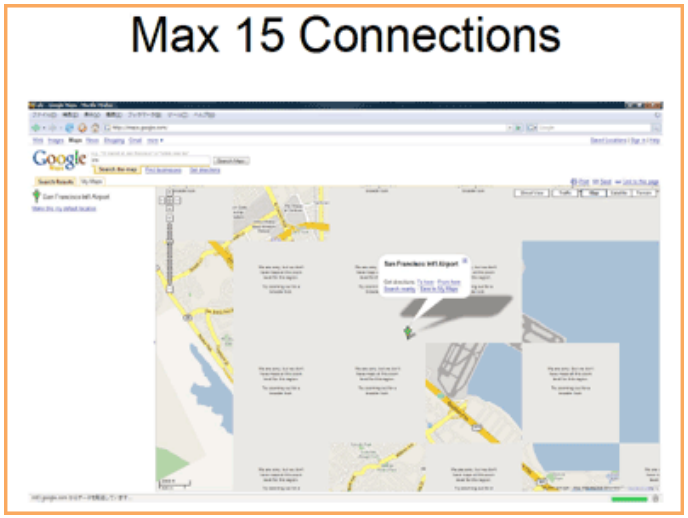


●グローバルIPv4アドレス数は、
①1ユーザアドレスあたりのポート使用数
②ポート割当手法(動的か静的か)によって決まります。

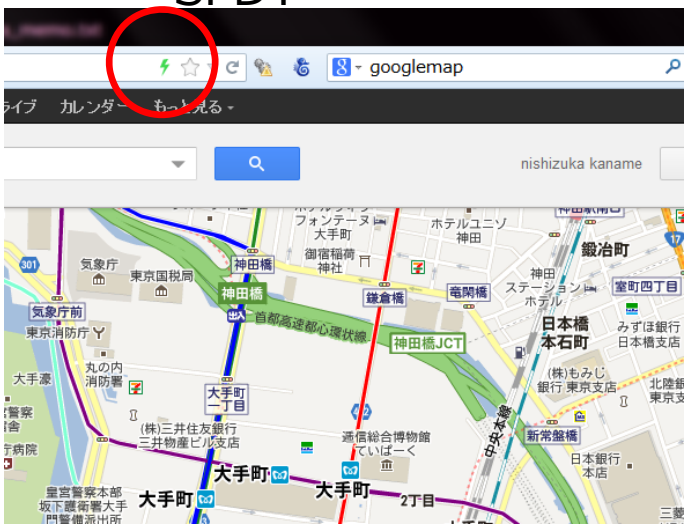
●共用IPアドレス空間は、100.64.0.0/10 アドレスを配布します。[RFC6598]

●提供エリア内の設備アドレスは、グローバルを用いた方が良いと思います。

①1ユーザアドレスあたりのポート使用数について
SPDY/WebSocketなどの次世代Web技術によってセッションが重畳されるため、1ユーザあたりのポート使用数に変化が起きています。



SPDY

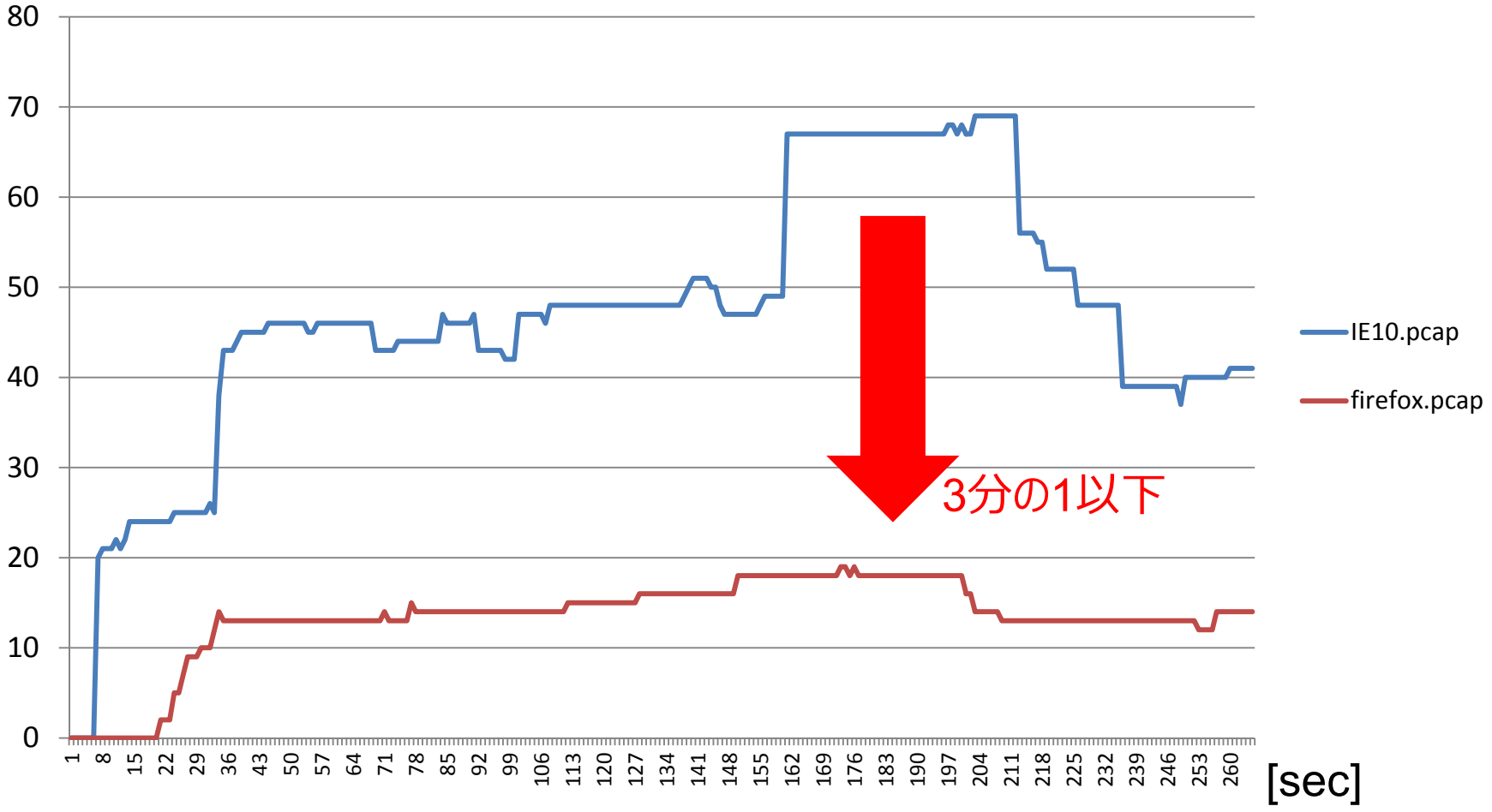


Google Mapは複数ポートを利用するアプリケーションの象徴として使われてきましたが、現在はほとんどポートを消費しません(10～15)

Google MAP 閲覧時の同時セッション数

非SPDY(IE10) vs SPDY(firefox21.0)

[同時セッション数]



代表的なウェブサイトやアプリケーション、オンラインゲーム等のセッション数をそれぞれ測り、通常の利用シーンなどを想定しながら、１ユーザ当たりのセッション数を明らかにします。

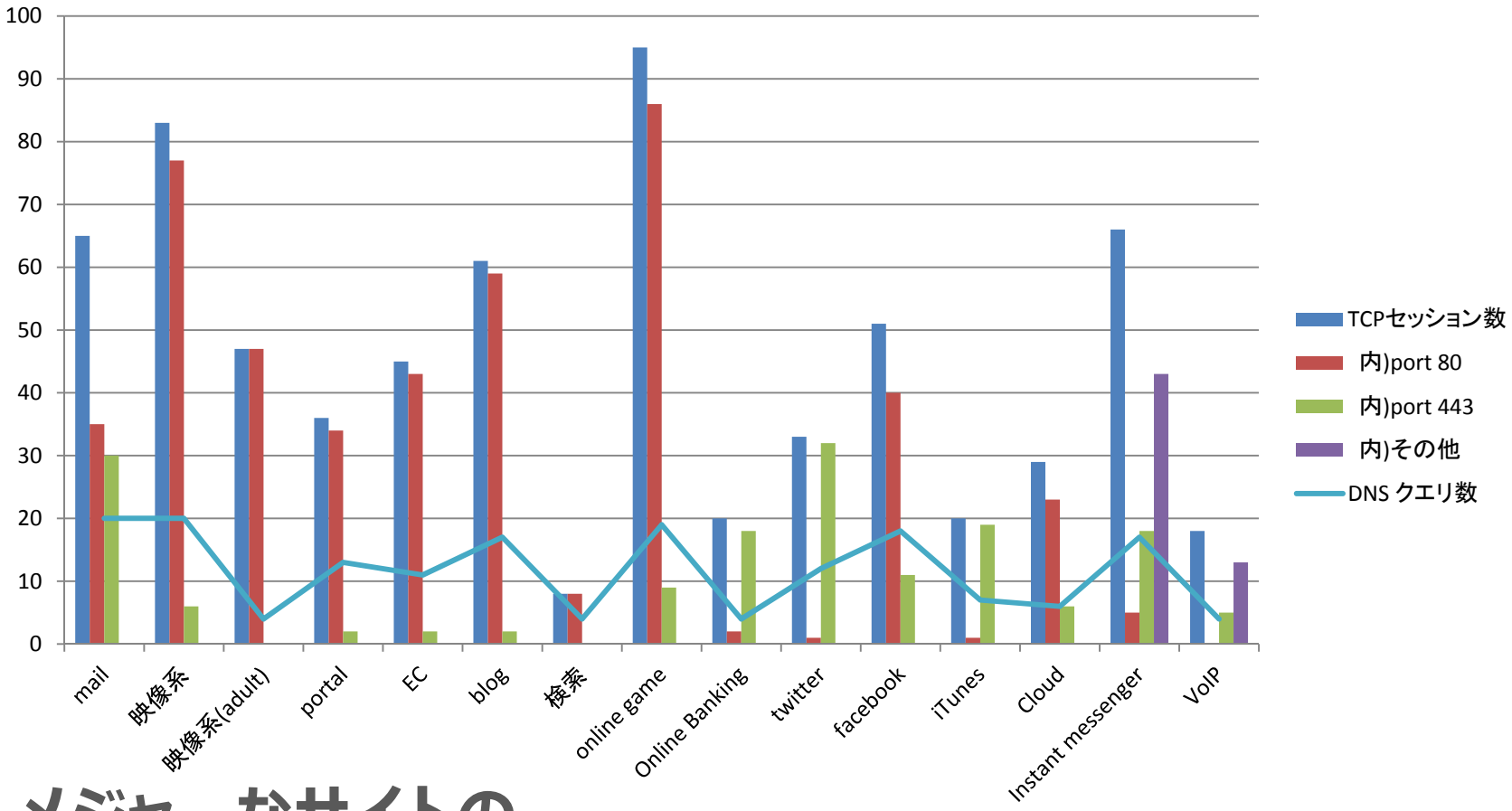
<手法>

- パケットキャプチャによって、TCP/UDPセッション数を計測
- クライアント: Windows7 (最新のchrome/firefox)
- サイト: 日本のTop100サイトおよび有名サービスから抽出

対象サイト/サービス (抜粋)

Webmail	gmail, yahoo mail, hot mail
映像/テレビ系	ustream, youtube, ニコニコ動画, Hulu, daily motion, daum, qq
映像/テレビ系(adult)	fc2, dmm.co.jp, xvideos
portal	yahoo.co.jp
EC	rakuten, amazon.com, apple.com
blog	livedoor blog, ameba blog
検索	google
Online PC game	aeria games, ameba pig, nexon, 777town, hangame
Online Banking	みずほ銀行，DCカード

各サービス分類における平均セッション数



メジャーなサイトの
https対応および省セッション化の傾向が反映

①1ユーザアドレスあたりのポート使用数　(観測に基づく仮定)

	平均セッション数	(許容)最大セッション数
1ユーザ当たり	100	1000

②ポート割当手法(動的か静的か)

・動的割り当て

1ユーザあたりの平均ポート利用数(=100)がkey factor

1グローバルアドレスあたり約600ユーザ

・静的割り当て

1ユーザの最大ポート利用数(=1000)がkey factor

1グローバルアドレスあたり約60ユーザ

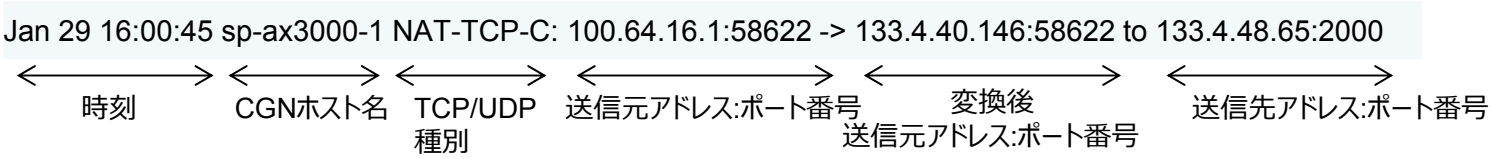
15万ユーザ収容の場合

	動的割り当て	静的割り当て
プールアドレス数	250	2500

②ログ編

「CGN周辺システム(ログ等)はどのくらい必要か」

NATログの見積もり



ASCII format: 120byte/record
Binary format: 26byte/record (20～25%)

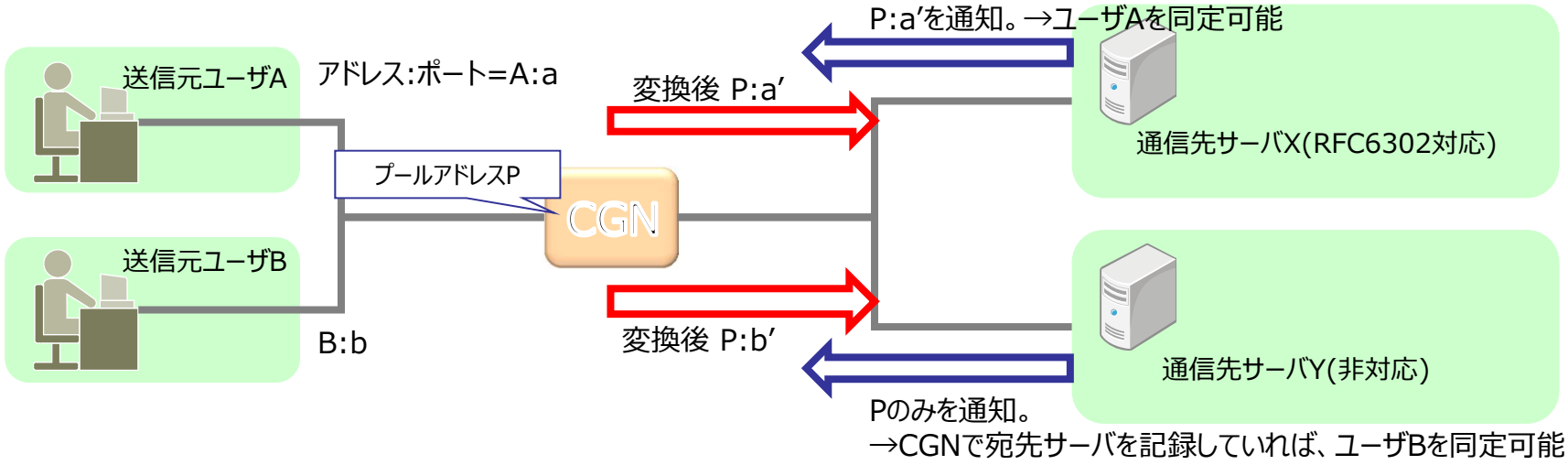
15万ユーザ収容の場合

	1日	1か月	2年
NATログ量	1TB	30TB	720TB

NATログの収集はストレージ価格の低廉化によって、非現実的なレベルではありません。
また、ポートブロック割り当てなど、ログの量を軽減する手法も発達しています。

【要注意】静的割り当ておよびポートブロック割り当てでは、警察対応できないケースがあります。

『通信先サーバがポート情報を保持(RFC6302対応)していない場合、CGNで通信先サーバ情報をログ取得しないとユーザの特定が不可能』



※送信元ポート情報のない申告に対して法的対応をするには、送信先ログの取得が必要となります。しかし、送信先ログを取得しようとすると、静的割り当てやポートブロック割り当ての旨みは出ません。

③アプリケーション編

「アプリケーションへの影響は？」

以下に挙げたタイプの挙動をするアプリケーションはCGNを超えることができません。

	ＶＰＮ型	Ｐ２Ｐ型	ＳＩＰ型
代表的アプリケーション	L2TP/IPsec, PPTP	skype, P2P file share	VoIP(050plus等)
CGN(NAT)を超えられない理由	TCP/UDPのポート番号も暗号化されるため、NATでポート変換ができない。	NAT外部から不特定多数の接続を待ち受けるアプリケーションであり、NATテーブルが存在しないため。	IPアドレス情報がペイロードに含まれており、NATで変換されないため。
CGN側対策実装	ALG	Fullcone NAT	ALG
アプリケーション側対策実装	IPsec NATトラバーサル(UDPカプセル)	STUN/UDP hole punching, TURN	TURN

ALG (Application Level Gateway) : CGNにおいてアプリケーションごとにNAT越えをサポートする機能

<VPN型>
L2TPは、クライアント側のIPsec NATトラバーサル(UDPカプセル)の実装が進んでおり、NAT越えができるケースが多い。
PPTPに関しては、例えばDocomo　SPモード網はALGをOFFにしているとみられる。

<P2P型/SIP型>
CGNをFullcone NATで動作させることによって、STUN/UDP hole punchingを利用したP2P通信をすることが可能になる。しかし、アプリケーション側の作りによってはFullcone NAT であっても、P2P通信を利用せずに、TURN(サーバ経由通話)にフォールバックする(Skypeの例)。

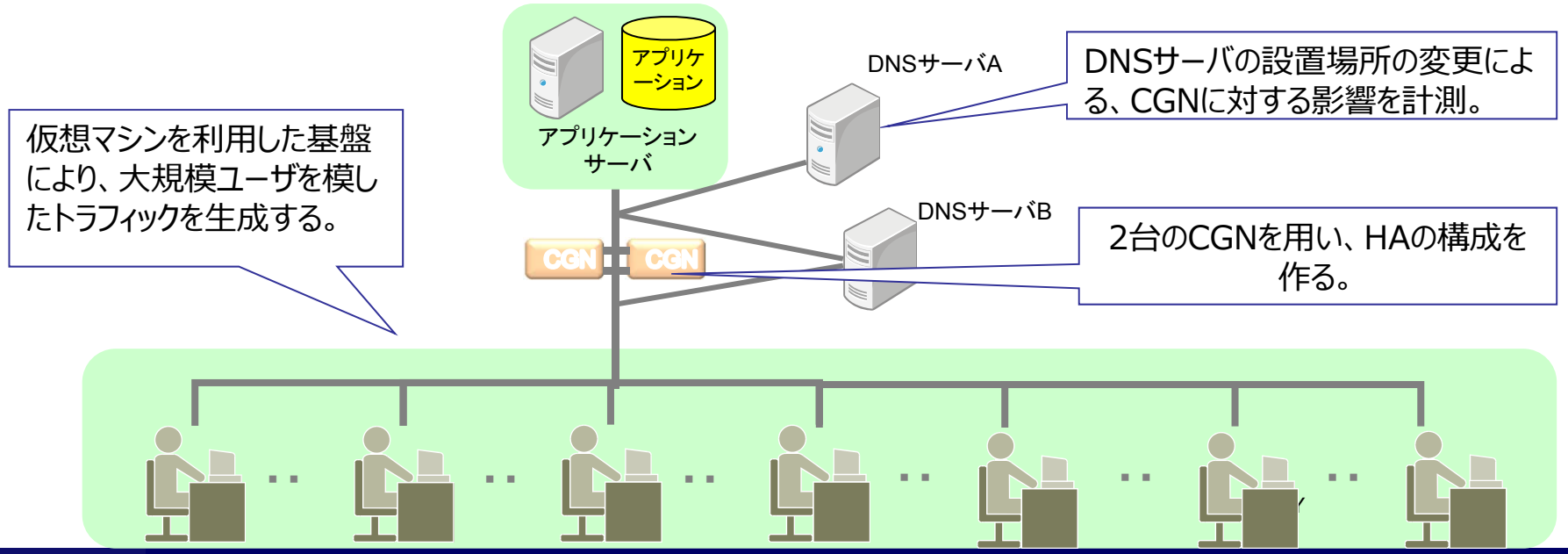
CGNにはトランスペアレンシを高めるためにALGやFullconeNATの機能が実装されています。しかし、処理が非常に重く、救済対象とするアプリケーションを選択しなければなりません。

④性能評価編

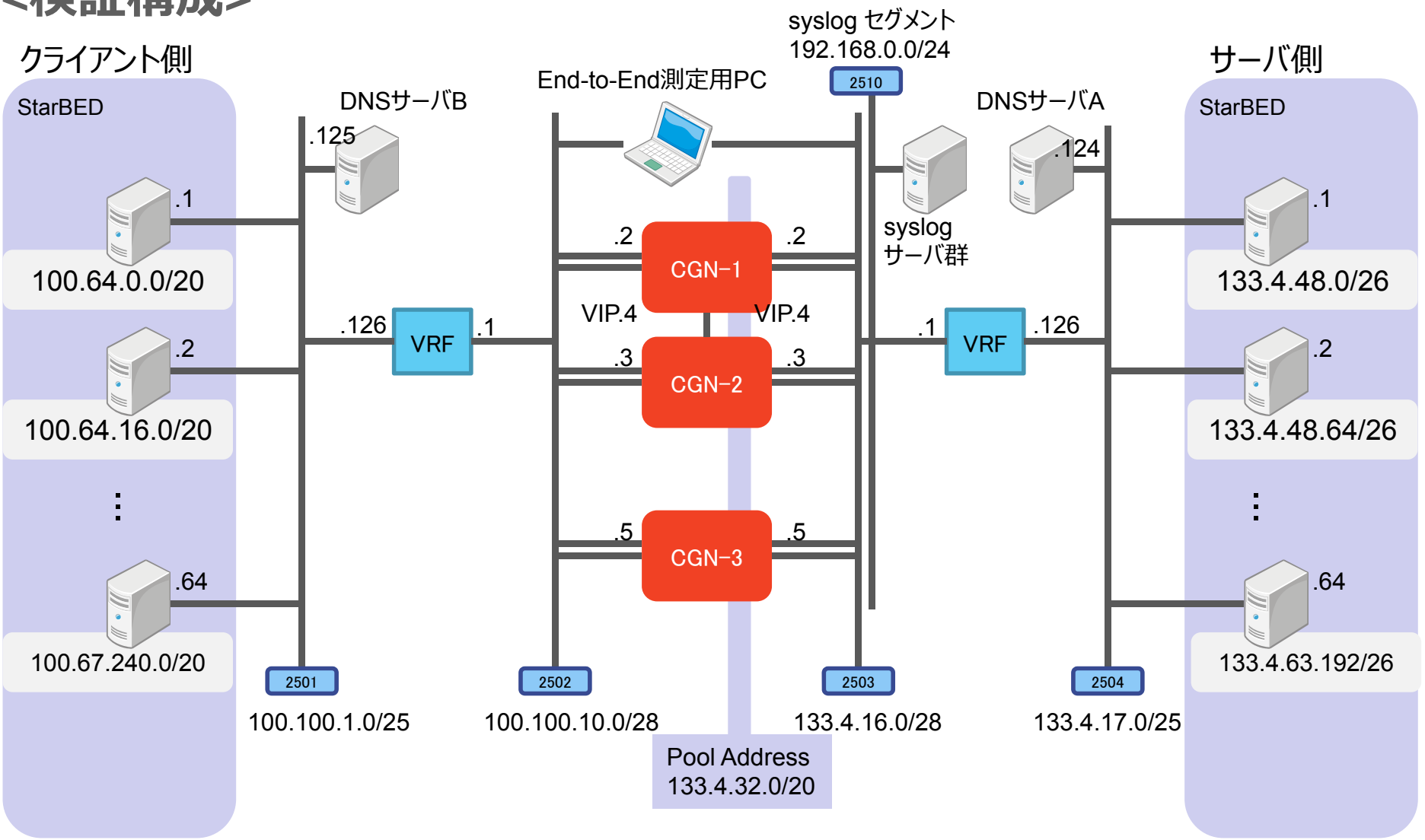
「 CGNの性能はどのくらい必要か？」

<CGNの性能評価>

- 仮想マシン群から、ISP規模 ：1万人～100万人を模したトラフィックを生成し、負荷をかけたCGNの性能値を測定した
- DNSサーバについて、複数の設置パターンを構成し、DNSクエリがCGNに与える影響を調べた
- 障害を模擬した切替実施により、HA(High Availability)構成の有効性を確認した



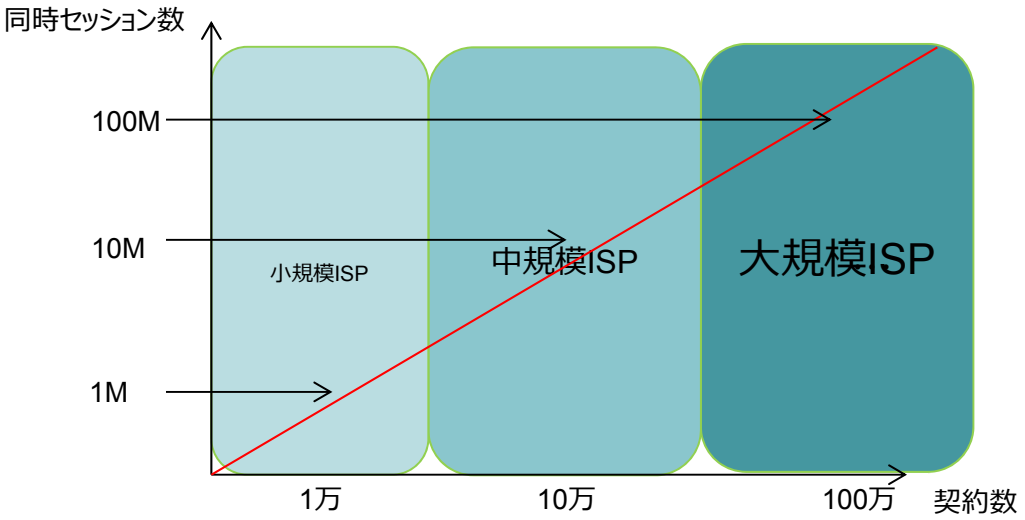
<検証構成>



３．CGN設計レシピ～性能評価編

<顧客規模と発生セッションについて>
顧客の契約数から最大同時セッション数の導出は、実際のネットワークごとの特徴があり観測が必要となるが、以下の仮定に基づいて、検証で発生させる負荷を算出した。
契約数 × Active率(25%) × 同時セッション数(400)
＝ ネットワーク上で想定される最大のセッション数

検証ケース	検証想定数		
	ユーザ数	アクティブユーザ数	同時セッション数
小規模ISP	10000	2500	1000000
中規模ISP	100000	25000	10000000
大規模ISP	1000000	250000	100000000



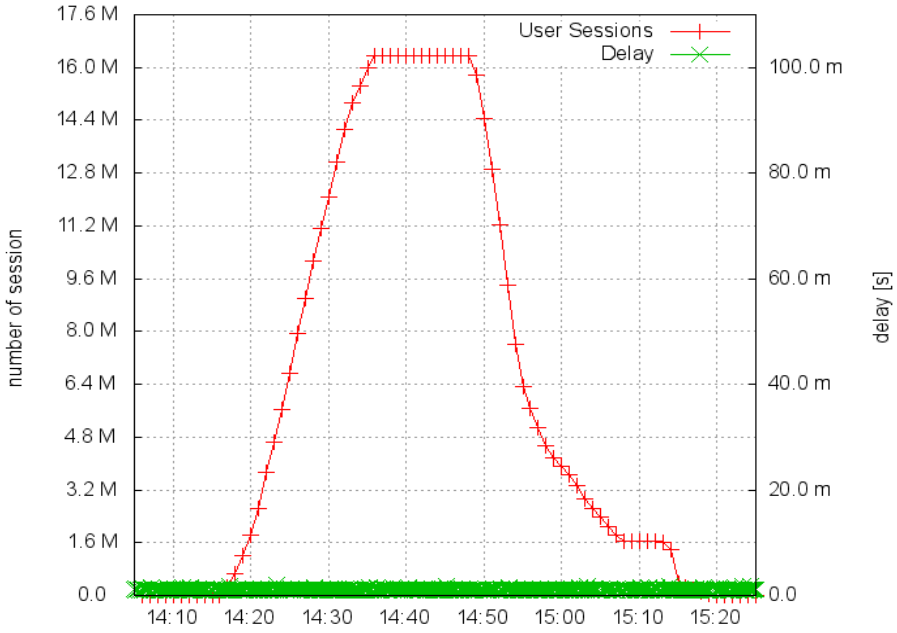
3．CGN設計レシピ～性能評価編

A10 AX3000-11を例として性能評価の結果を記載する。
AX3000-11 では16Mセッション(＝16万ユーザを想定)で限界を迎えた。

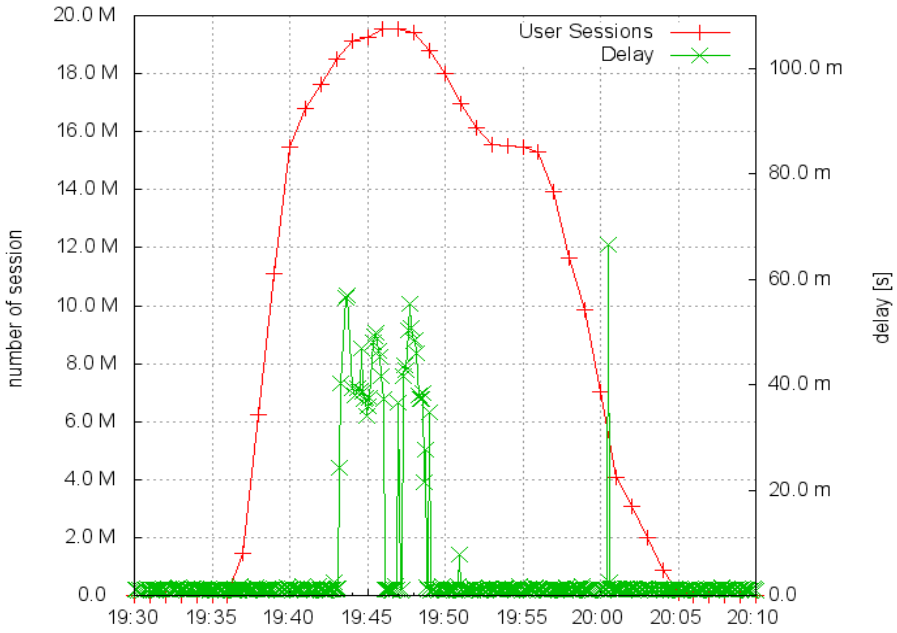
- 右の表は実証実験で用いた装置のカタログスペックである。
- 実際には16Mセッション、20Kセッション/秒程度で性能限界に達することがわかった。

	CGNカタログスペック
同時セッション数	67M（67万ユーザ規模）
セッション / 秒	440K

- 16万ユーザ規模を想定したセッション(16Mセッション)をCGN装置に印加した場合、全てのセッションを確立することができた(16万ユーザ規模は収容可能)
- CPU使用率は80%程度



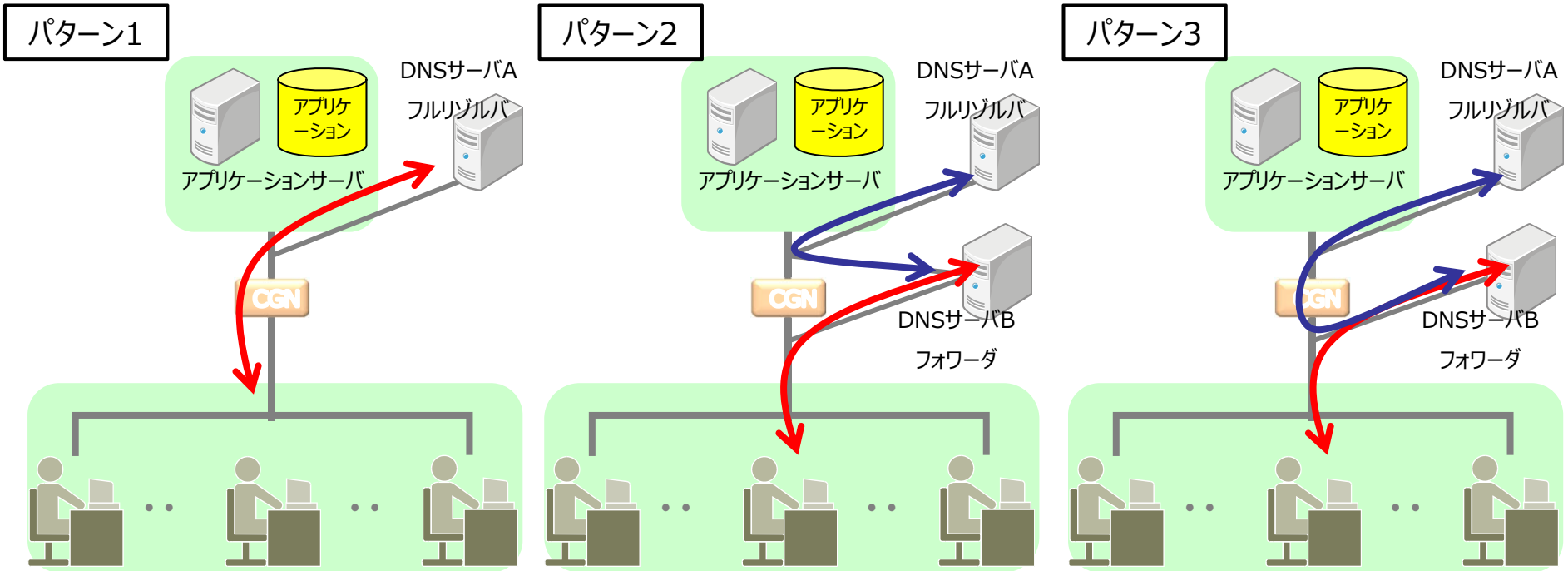
- 26万ユーザ規模を想定したセッション(26Mセッション)をCGN装置に印加した場合、20Mセッションで限界となった (26万ユーザ規模は収容不可能)
- CPU使用率は90%程度(閾値超過の警告発生)



3．CGN設計レシピ～性能評価編

以下の3パターンについて、同一の負荷を印加し、CGNの性能への影響を検証した。

	パターン1	パターン2	パターン3
ユーザに配布するDNSサーバ	CGNの外側に置かれたDNSサーバ	CGNの内側に置かれたDNSサーバ	CGNの内側に置かれたDNSサーバ
新規DNSの設置について	なし。	CGN配下のドメインに新規にDNSサーバを設置	CGN配下のドメインに新規にDNSサーバを設置
内側→外側のDNSサーバ間通信	-	CGNを通る	CGNを通らない
通過するクエリ	全てのDNSクエリがCGNを通過する	第三者の運用するDNSを利用しているユーザのDNSクエリ	フォワーダからフルリゾルバへのDNSクエリ+第三者DNSを利用しているユーザのDNSクエリ

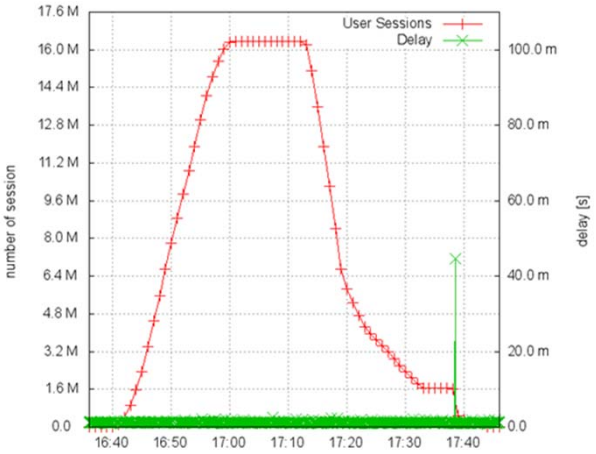


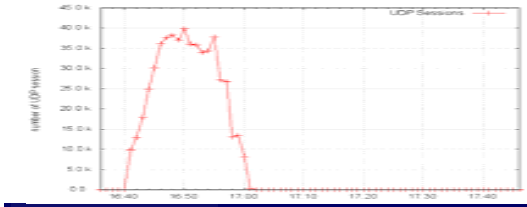
3．CGN設計レシピ～性能評価編

A10 AX3000-11を例としてDNSクエリの影響の評価を記載する。
AX3000-11 では、DNSクエリの影響がほとんどないことがわかった

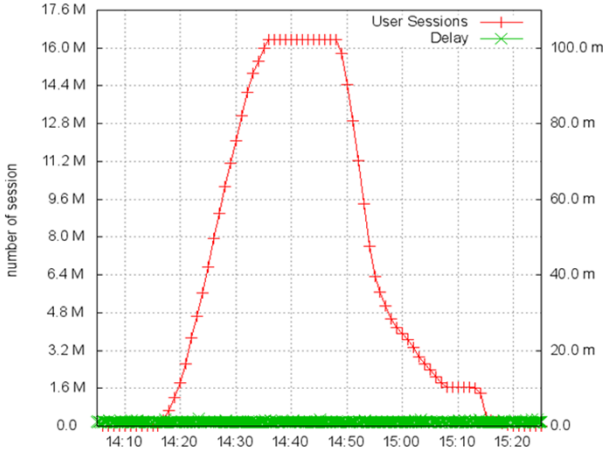
- 負荷トラフィックとして、16M同時セッションとなるTCP通信(port80)に加えて、ホスト名を解決するための16MクエリのUDP通信(port53)を発生させた。
- パターン1とパターン2を比較すると、目立った影響は見られなかった。AX3000ではDNSのセッション保持時間のが3秒と短いため、最大でも40Kセッションしかセッションエントリを消費しなかった。
- パターン1とパターン3の比較では、実際にキャッシュ効果によりCGNを通るDNSパケット数が減少することが確かめられた。この場合でも、DNSクエリによる負荷の増大は見られなかった。

パターン1

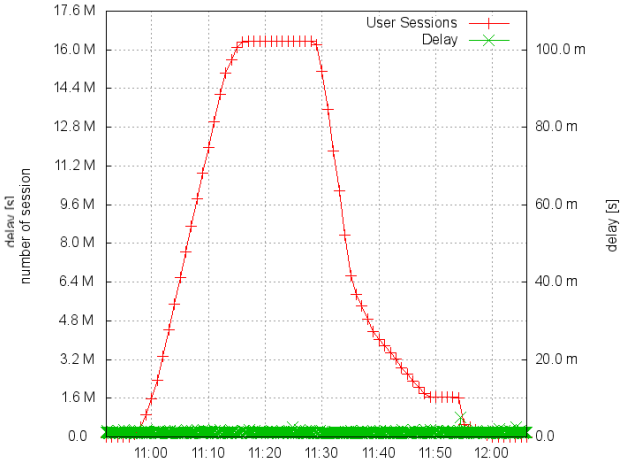


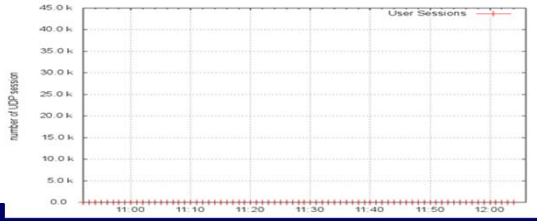


パターン2



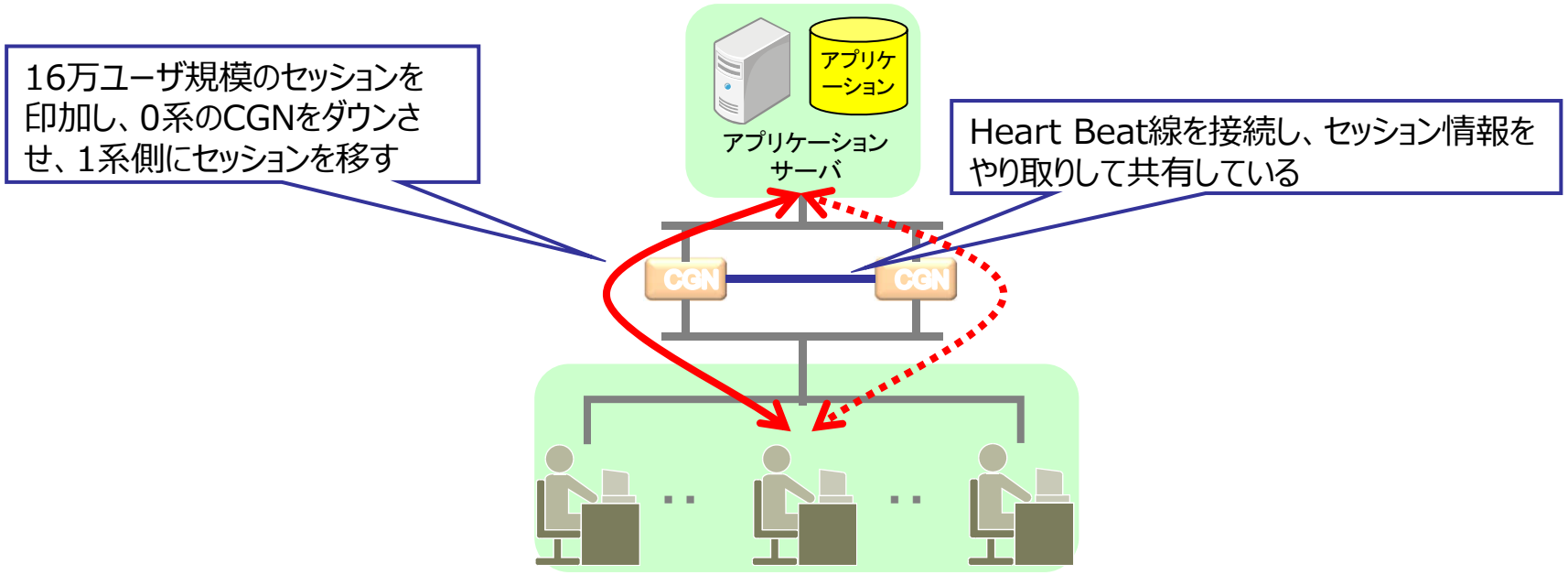
パターン3





<冗長構成検証>
トラフィックを印加した状態で、HA構成の2台目へ通信を切り替える。
切り替えによっても通信影響が発生しないことを確認する。

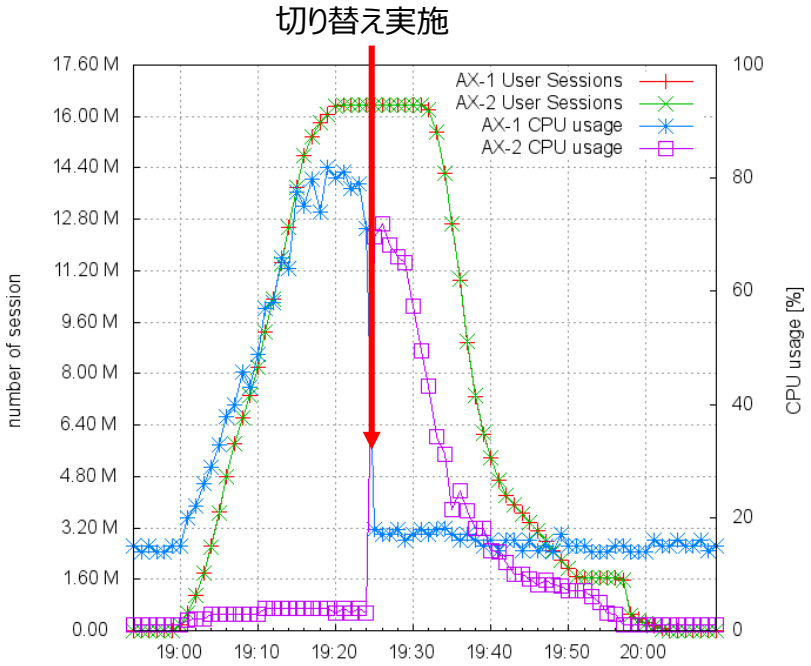
※多くのCGNは、Act-Sby構成で動作しセッション情報の同期を行う。

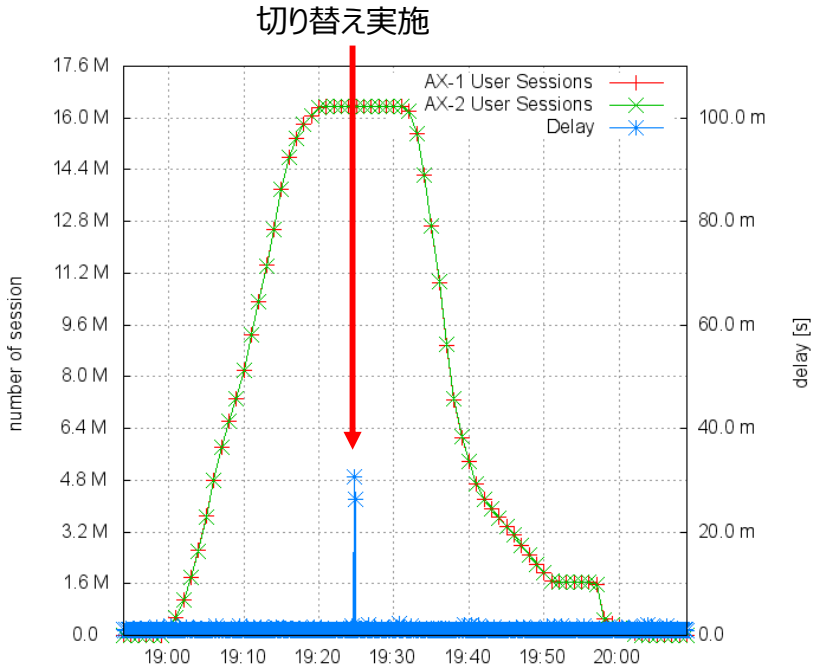


3．CGN設計レシピ～性能評価編

A10 AX3000-11を例としてAct-Sbyの切替の結果を記載する。
HA切替によって、30msec程度の遅延が発生することが確認された。

- 切り替えの前後でセッション数が減少しないのは、セッション同期が行われているため。これはCPU使用率の推移からもわかる。
- 切り替えの前後で、パケットロスは発生しなかったが、若干の遅延が発生した。
- しかしながら、サービスに大きな影響を与えるものではなく、CGNでHA構成をとることはISPが安定してサービスを提供する上で、選択し得るソリューションとなる。





⑤ルーティング編

「どこにCGNを入れるか。どのようにルーティングするか。」

3．CGN設計レシピ～ルーティング編

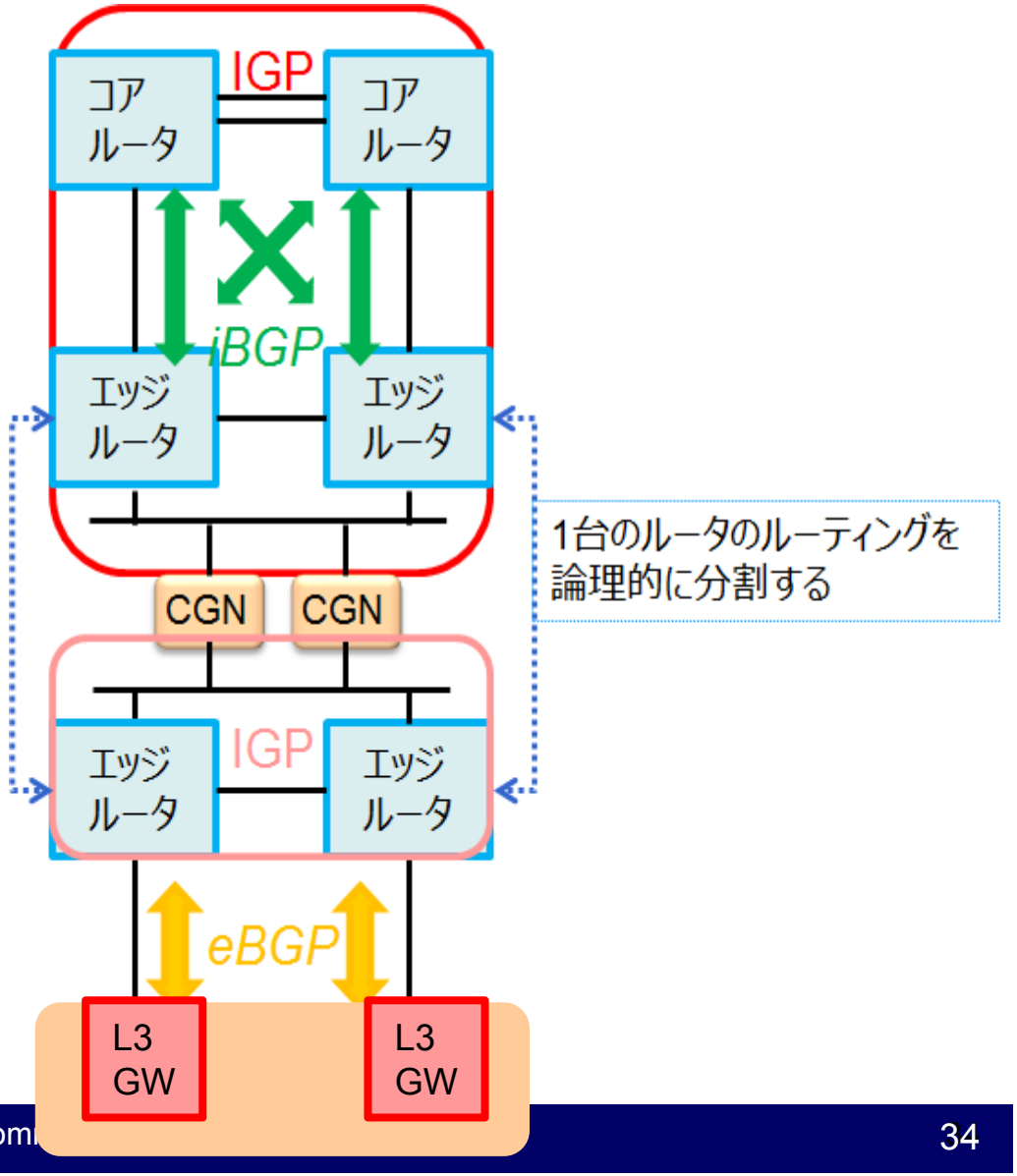
性能評価の結果より、10数万ユーザ程度を集約する階層にも設置が可能

考慮すべき点

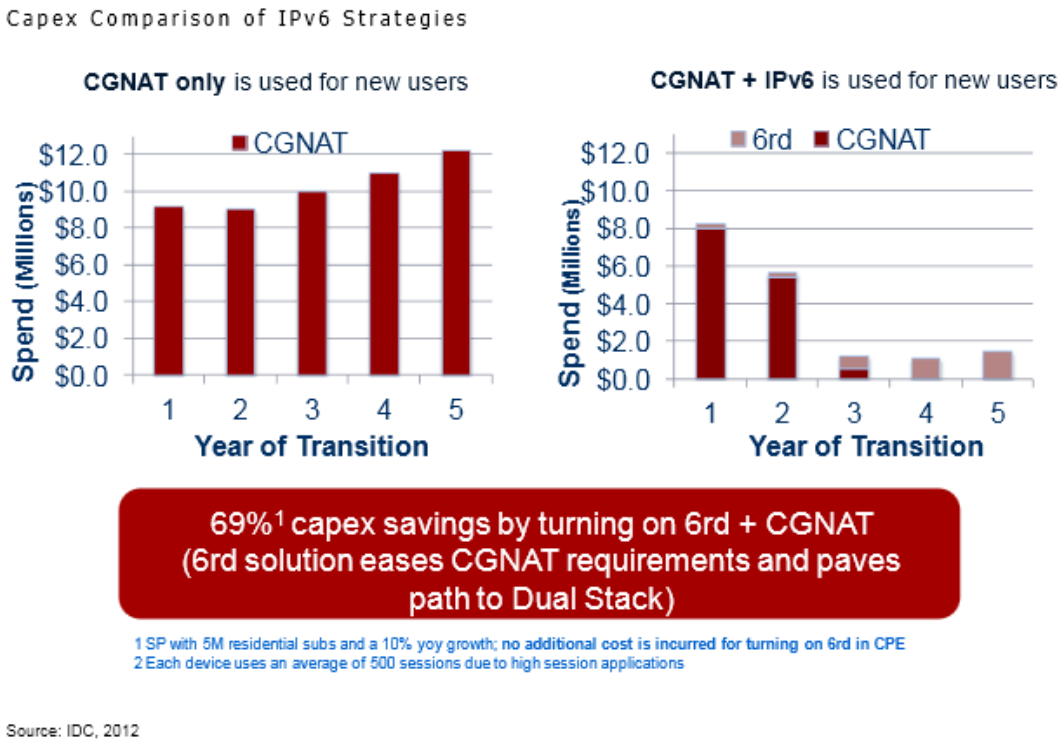
- eBGPやiBGPといったダイナミックルーティングができないCGNが多いため、以下の部分には設置できない
  - eBGPの境界
  - iBGPが動作するルータの間

ネットワーク設計例

- エッジルータのルーティングを論理的に分割し、この間に設置する(ダイナミックルーティングをエッジルータで終端できる)
- 2台1セットにする(HA構成を取り、耐障害性を高める)



CGNを提供する際には、IPv6サービスと抱合せて提供することを推奨します。



IPv6が提供されていれば、1ユーザ当たりのCGNに対する負荷は軽減されます。
さもなくば、暫定解であるCGNへの依存が継続し、高コスト(なおかつ低スペック)なサービスであり続けることになります。

◆CGNを含む全てのアドレス共有技術は暫定解
　　⇒アプリケーションへの影響をゼロにはできないため

◆暫定解として極力”使える”技術を選択したい
- NAT444+IPv6
- DS-lite
- 464xlat
- MAP-E

◆特にCGNはベンダの実装が進み、アプリケーションへの影響も詳しく調べられており、対処できるものが進んでいる。

◆CGNの適材適所
　　⇒CGN設計指針となる一連の流れを紹介しました。

謝辞

This research and experiment are conducted under the great support of Ministry of Internal Affairs and Communications of Japan.

本研究は「H24年度 IPv4アドレス枯渇に伴う情報セキュリティ等の課題への対応に関する実証実験」において総務省からの支援を基に行った成果です。