2013/9/6　IA研発表資料

IPv4/IPv6共存技術の現状とShowNetでの相互接続検証結果について

Interop Tokyo 2013 NOC
さくらインターネット研究所
大久保 修一 <ohkubo@sakura.ad.jp>

Agenda

- IPv4/IPv6共存技術の必要性
- 各種共存技術の特徴
- Interop Tokyo 2013 ShowNetにおける実装
- 検証内容と結果
- まとめ

IPv6デプロイの現状

IPv4グローバルアドレス枯渇が進行

・グローバルIPv4アドレス接続サービスの有料オプション化
・事業者間でのアドレス移転の活発化

デュアルスタック運用にも課題

・これまでと比べて2倍の運用コストが発生
・IPv4/IPv6の2重のスキルセットが必要、ドキュメント化に苦労

IPv6 Only 通信接続サービスが登場

T-Mobile
SingTel
Comcast　など

IPv6バックボーンを利用したIPv4提供技術(464技術)も標準化が進展

IPv6 Only 網の展開

T-Mobileの事例

- IPv6 only APN
- DNS64/NAT64　464XLATサポート
- 今後のSamsung Galaxy端末（Android）にCPE（464XLAT CLAT）を内蔵

APNIC 35 Conference
http://conference.apnic.net/35/program/apnic-ipv6-plenary/transcript

ShowNetで提供したIPv4枯渇対策/IPv6共存技術

各IPv4/IPv6共存技術の比較

	MAP-E	DS-Lite	464XLAT	SA46T
対象利用者	大規模ISP	中～大規模ISP、CATV事業者、モバイルオペレータ		エンタープライズ、その他
スケーラビリティ	◎	○	○	○
NATの場所	ユーザ側	キャリア側	キャリア側	NATなし
トポロジ	メッシュ ※1	ハブ＆スポーク ※2	ハブ＆スポーク	メッシュ
トレーサビリティ	△	○	○	◎
ログの量	若干増加	多い	多い	なし
パケットフォーマット	カプセル化	カプセル化	トランスレーション ※3	カプセル化
利用上の制限	NATの実装によって使用できないプロトコルがある			なし
補足事項	広大なアドレス空間が必要。キャリア側設備(BR)の冗長化が容易。サーバ側でのソースポート番号を記録する必要がある。	アドレスの利用効率が高く、アドレスの少ない事業者への導入がしやすい。ユーザ側機器(B4)の実装を単純化しやすい。	DS-Liteの特徴に加え、既にNAT64環境を構築している事業者への導入がしやすい。	プロトコル透過性が高い。1つのIPv6バックボーンで、重複したプライベートアドレス空間を持つ複数のIPv4ネットワークを接続可能。

※1 メッシュトポロジ
ユーザ同士はPeer to peerで通信可能。トラフィックはIPv6網内で折り返すため、キャリア側機器のリソースを消費せずに済む。

※2 ハブ＆スポーク
ユーザ同士の通信もキャリア側機器を経由するが、通信不良等が発生した場合の確認がしやすい。

※3 トランスレーション
元パケットのIPv4アドレスがIPv6アドレスに埋め込まれているため、IPv6バックボーン内において、元パケットのIPv4アドレスに基づいたフィルタリングやトラフィック制御を行うことができる。

ハブ＆スポークトポロジ

- 464XLAT
- DS-Lite

メッシュトポロジ

- MAP-E
- SA46T

ShowNetにおけるIPv4/IPv6共存技術

- IPv6 onlyバックボーンの登場を想定した各種技術をhall7出展社/会議棟来場者向けに提供し、実運用に挑戦

hall7出展社向けに提供した技術
- ✓ DS-Lite
- ✓ 464XLAT
- ✓ MAP-E

会議棟来場者向けに提供した技術
- ✓ SA46T
- ✓ NAT64/DNS64

Hall7編

Hall7出展社向け回線をIPv4/IPv6共存技術にて収容

MAP-Eの概要と特徴

- “Mapping of Address and Port with Encapsulation”の略
- IPv6ネイティブネットワークを越えてIPv4サービスを提供するための方式の一つ
- IPv4、IPv6共に広帯域のアドレス空間を配備可能な大規模ISPなどの環境に適している
- キャリア側機器をステートレスにできる
  - スケーラビリティが高く冗長化が容易
  - NAT処理はユーザ側機器CEで実施
- メッシュトポロジをサポート
  - ユーザ同士は低レイテンシで直接通信することが可能
  - トラフィックはIPv6網内で折り返すため、キャリア側機器のリソースを消費せずに済む

#show net
> Go to the next decade
as of 2013-6/11 16:00
INTEROP
TOKYO | 11-14 JUNE 2013
Device ID
LoopBack Address
45.0.0.X/32
P2P Address Block
Link ID = XYZ
45.0.X.(YZ×4)/30
P2P VLAN ID
XYZ(Link ID)
EQAM VID400
Tap LinkID=TapID
MGMT Address
172.16.X.Y/16 or 172.16.0.X
fd00::X:Y/64 or fd00::X
control plane
ESSID 2.4G:ShowNetFreeWiFi
5G:ShowNetFreeWiFi-a
Link ID
loopback
VIP
OSPF cost
router if
untag port
tagged link
VID
mgmt eth
exhibitor
server
va
endnode
security device
CGN/LSN, balancer
cache/WAN optimization
L3
L2
vlan-id
vswitch
LAG
MEP
DPI
DWDM pair
OpenFlow device
multimode
100G
40G
10G
1G
oil
serial server
tester, analyzer, generator
NSPIXP-23
ntt.net transit
OCN transit
JGN
dix-ie public peer
JPIX public peer
JPNAP public peer
KDDI transit
.mgmt
Top of Rack Fabric Management
.note
NTT Otemachi
Makuhari Messe
.dc
Data Center
.life
.sdnsc
.noc
.conf
Conference
.monitor
SDN Monitoring
external
core
lastresort
Pod 4,5
Pod 6
Pod 7
Hall 4,5
Hall 6
Hall 7
OpenFlow v1.3.1 Interoperability
IPv6 Migration
Hall WiFi & Location Experiment
INTEROP Pavillion
Location Experiment
Hall WiFi

MAP-E BR
古河電工 FX5000
MAP-E BR
シスコシステムズ ASR9000
MAP-E CE
インターネットイニシアティブ SEIL/x86
SEIL/x86
ShowNetで使用した
MAP-E対応機器
nm-2100-mx
spirent
studio
fx5k
ne40e-x3
ip88s36
as65013
s60
as65014
asr9k
ax86r
core
glimmerglass
AlaxalA
le1000gt
mlxe8
ax5630-2
dw7000
ex9k-2
as65000
nexus3k
as65011
nexus5k
ip88r
va2-a
va1-b
va2-b
cfx2000
qwilt
pf5248-1
pf5240-1
rtx1200-2
rtx1200-5
csr1000v
c6100
pf5240-2
pf5248-2
ns3210
seil-x86-1-8
sail-hv-1-2
aterm-1
aterm-2
aterm-3
aterm-4
aterm-5
aterm-6
aterm-7
aterm-11
aterm-12
aterm-15
sa46t
IPv6 Migration
fes2402-3.mgmt
pf5240-3
p3295
nxssp
ns3710
samurai
rex
tex
nicter
ip88s36-meijin
srx5k

IP Anycastを用いたBRの冗長化

MAP-Eパラメータ

パラメータ	設定値
Rule IPv6 prefix	2001:3e8:6000::/56
Rule IPv4 prefix	130.129.255.92/30
End-user IPv6 prefix	2001:3e8:6000:00xx::/64
Rule EA-bits length	8bit(64-56)
Port-Set ID length	6bit(8-2)
PSID offset	4
ports/user	960ports
BR IPv6 Address	2001:3e8:0:2900::1/128

MAP-Eでは自動トンネルを実現するため、ユーザに割り当てられたIPv6アドレスから、ルールに従って自身で使用可能なIPv4アドレスとポート番号の範囲を算出します。

アドレスマッピング例

あるユーザに割り当てられたIPv6アドレス

2001:3e8:6000:0042::/64

IPv4アドレスに埋め込み　　0100 0010　　ポート番号に埋め込み

130.129.255.92/30

0101 11XX → 0101 1101

130.129.255.93

XXXX 0000 10YY YYYY

XXXX = 0001 ～ 1111

YY YYYY = 000000 ～ 111111

	利用可能なポート番号
0001 0000 1000 0000 ～ 0001 0000 1011 1111	4224 ～ 4287
0010 0000 1000 0000 ～ 0010 0000 1011 1111	8320 ～ 8383
・・・省略（全15ブロック）・・・	
1111 0000 1000 0000 ～ 1111 0000 1011 1111	61568 ～ 61631

DS-Lite、464XLATの概要

- IPv6ネイティブネットワークを越えてIPv4サービスを提供するための方式
- IPv4アドレス(ポート)をダイナミックに割り当てるため、アドレス利用効率が高く、広帯域のIPv4アドレス空間を持っていない事業者でも導入しやすいのが特長
- DS-Liteはトンネルにより、464XLATはプロトコル変換によりIPv4パケットをIPv6ネットワークを越えて転送する

DS-Liteの動作

IPv4プライベート
IPv6 キャリア /
ISP ネットワーク
IPv4グローバル
DS-Lite CPE
DS-Lite AFTR
IPv4 over IPv6トンネル
NAT
Advanced SLB
192.168.1.1
2001:db8:192:168::1
2001:db8:a10:a10::99
203.0.113.11
198.51.100.99
IPv4ヘッダ
送信元
192.168.1.1
宛先
198.51.100.99
IPv6ヘッダ
送信元
2001:db8:192:168::1
宛先
2001:db8:a10:a10::99
IPv4ヘッダ
送信元
192.168.1.1
宛先
198.51.100.99
IPv4ヘッダ
送信元
203.0.113.11
宛先
198.51.100.99

464XLATの動作

IPv4プライベート
IPv6 キャリア /
ISP ネットワーク
IPv4グローバル
CLAT
PLAT
IPv6 Prefix
2001:10:10:10:10:10::/96
送信元IP Pool:70.0.0.1 ～50
NAT
NAT
Advanced SLB
192.168.1.1
Stateless
Stateful
198.51.100.99
IPv4ヘッダ
送信元
192.168.1.1
宛先
198.51.100.99
IPv6ヘッダ
送信元
2001:db8::c0a8:101
宛先
2001:10:10:10:10:10:c633:6463
IPv4ヘッダ
送信元
70.0.0.1
宛先
198.51.100.99

ShowNetで使用した DS-Lite、464XLAT対応機器

DS-Lite AFTR & 464XLAT PLAT
A10ネットワークス AX5630

464XLAT PLAT
セイコーソリューションズ iX-3740

DS-Lite AFTR & 464XLAT PLAT
ジュニパーネットワークス SRX5600

DS-Lite B4 & 464XLAT CLAT
NECアクセステクニカ Atermシリーズ

会議棟編

- 会議棟来場者向けに無線で提供
- NAT64/DNS64 (SSID: 5howNet FreeWiFi-a)
- SA46T (SSID: 5howNet FreeWiFi)

- 会期中に接続されたユニーク端末数：1402

NAT64/DNS64連携動作フロー

ShowNetにおけるSA46Tの構成

会議棟で使用した IPv4/IPv6共存技術

富士通
SA46T/SA46T-PR/SA46T-PT

SA46T
Stateless Automatic IPv over IPv6 Tunneling

NAT64
セイコーソリューションズ iX-3740

DNS64、DHCP、DHCPv6
Infoblox Trinzic 820/100

構築後の動作確認内容

- MTUの正常性
- フラグメントパケットの疎通
- TCPやUDPパケットの疎通
- ICMPパケットの疎通
- 主要なWebサイトの閲覧
- P2P系アプリケーションの動作
- 異ドラフトによる実装(MAP-Eのみ)

確認結果

- 総じて良好に接続できたが、以下のような調整が必要なパラメータが存在した。
- MTU
  - ユーザ側機器でのTCP MSS値調整を有効化
  - IPv6 MTUが1280Bytes固定設定となっている装置が存在
- フラグメントパケットの扱いの差異
  - IPv6側でフラグメントする装置とフラグメントされたICMPパケットをドロップするような実装の組み合わせ

Interop本会期中のトラブルや問合せは皆無！！

MAP-E最新ドラフトでの相互接続試験

- SEIL/x86とASRにて、MAP-Eの最新ドラフト(draft-ietf-softwire-map-07)に準拠した相互接続試験を実施
- ASRがICMPv4パケットのencapsulationに未対応だったが、TCP/UDPパケットの疎通には問題ないことを確認

参考：インターフェイスIDの変更

- draft-ietf-softwire-map-03以前

```
+--+---+---+---+---+---+---+---+---+
|PL|   8  16  24  32  40  48  56   |
+--+---+---+---+---+---+---+---+---+
|64| u | IPv4 address  |  PSID | 0 |
+--+---+---+---+---+---+---+---+---+
```

- draft-ietf-softwire-map-04以降

```
|        128-n-o-s bits            |
| 16 bits|    32 bits     | 16 bits|
+--------+----------------+--------+
|   0    |  IPv4 address  |  PSID  |
```

19:12:49.571434 2001:3e8:6000:c0:0:8281:ff5f:**0** > 2001:3e8:0:2900::1:
IP 130.129.255.95.45089 > 133.242.53.30.80: S 1937843878:1937843878(0)
win 14600 <mss 1420,sackOK,timestamp 344073592 0,nop,wscale 6>

まとめ

- 各種IPv4/IPv6共存技術は既に実用化段階に入っている
- 導入の際には、各事業者のネットワーク構成、事業体に応じて最適な技術を選択する必要がある
- 標準化の後押しをするため、インプリメントの推進、相互接続検証は重要な活動となる
- 今後も引き続き取り組みを行いたい

謝辞

ご協力いただいた、
Interop 2013 ShowNet コントリビュータ各社、
Interop 2013 ShowNet NOC チーム、
STM / CTM、Interop 2013 ShowNet 事務局、
他、Interop 2013 ShowNet の運営に関わって
頂いた皆様に厚く感謝いたします。

BB#14ラック

Photo by 徳川 義崇様

BB#04ラック

Photo by 徳川 義崇様

BB#07ラック

Photo by 徳川 義崇様

BB#12ラック

Photo by 徳川 義崇様

本資料について

- Copyright (C) 2013 Interop Tokyo NOC Team
  - 写真撮影: 徳川 義崇　株式会社ナノオプト・メディア
  - 記載されている各社の社名、商品名は各社の登録商標又は商標です。
- 本資料の文章・画像の無断転載・複製を禁止します。
- 本資料に関するお問い合わせ先
  - Interop Tokyo 2013 運営事務局
  - 株式会社ナノオプト・メディア
  - ShowNet担当 : 大嶋
  - Tel: 03-6431-7803 Email: shownet@f2ff.jp