

2nd Life with IPv6 experiment @ WIDE プロジェクト3月研究会

JANOG 30

「IPv6時代のIPv4を考える～第2章～」

事前資料

櫛山寛章(奈良先端科学技術大学院大学)

事前資料の趣旨

- パネルセッションでは Open Mic Time を多めにして、JANOG 30 参加者の皆さんと色々議論したい
- 一方で、検証実験の内容は複雑で、パネルセッションの前半だけで皆さんと十分議論できる説明を行うのは困難
- そこで、事前資料に実験概要、詳細な実験構成、各技術の解説、パネルディスカッションでは触れない他の実験結果などを掲載
- 予習の上、色々と質問、議題を考えてパネルディスカッションに参加していただけると幸いです

事前資料一覧

- WIDE合宿での実験全体の概要
 - 本資料
- 4RDに関する技術解説(事前資料版)
 - III 末永さんの資料
- 464XLATに関する技術解説(事前資料版)
 - NECアクセステクニカ 川島さんの資料
- SA46Tに関する技術解説(事前資料版)
 - 富士通 松平さんの資料
- 共存/移行技術とP2P対戦ゲームの相性(事前資料版)
 - コナミデジタルエンタテインメント 佐藤 良さんの資料

パネルディスカッションの タイムライン(予定)

14:10 - 14:15	検証実験の概要
14:15 - 14:25	検証結果報告
14:25 - 14:40	検証結果への各実装者による技術解説 +7月までの間のアップデート
14:40 - 15:10	次回実験の告知 + Open Mic Time

Life with IPv6 experiments

- 趣旨

- IPv6 only のユーザアクセスネットワーク環境を構築して、構築・運用ノウハウを取得
- 実際に生活してみても、OS、アプリケーションのIPv6対応状況やビジネスユースに耐えられるのかどうかを調査
- IPv6 移行、IPv4/IPv6共存技術を実際のユーザトラフィックを載せて、同様に検証
- 一般的な問題を洗い出し、IETFなどで報告
 - 可能ならば、Work Around や BCP を作成

Life with IPv6 experiments

- 過去2回(2011年9月、2012年3月)、WIDEプロジェクト3月/9月研究会(通称 WIDE合宿)での実験の一部として実施
- 検証結果は IETF v6ops BoFにて発表
 - 1回目
 - <http://tools.ietf.org/html/draft-hazeyama-widencamp-ipv6-only-experience-00>
 - 日本語はこちら
 - <http://member.wide.ad.jp/tr/wide-tr-hazeyama-ipv6-only-network-00.pdf>
 - 2回目
 - <http://tools.ietf.org/html/draft-hazeyama-widencamp-ipv6-only-experience-01>
 - 日本語は(まだ)ありません

第2回 Life with IPv6 experiment @ 2012年3月WIDE合宿

- JANOG 29 LT やInternet Week のBoFなどで協力者を募集
 - http://www.janog.gr.jp/meeting/janog29/_downloads/janog29-wide-after-hazeyama-01.pdf
- 様々な方に参加していただきました！
 - ご協力感謝いたします

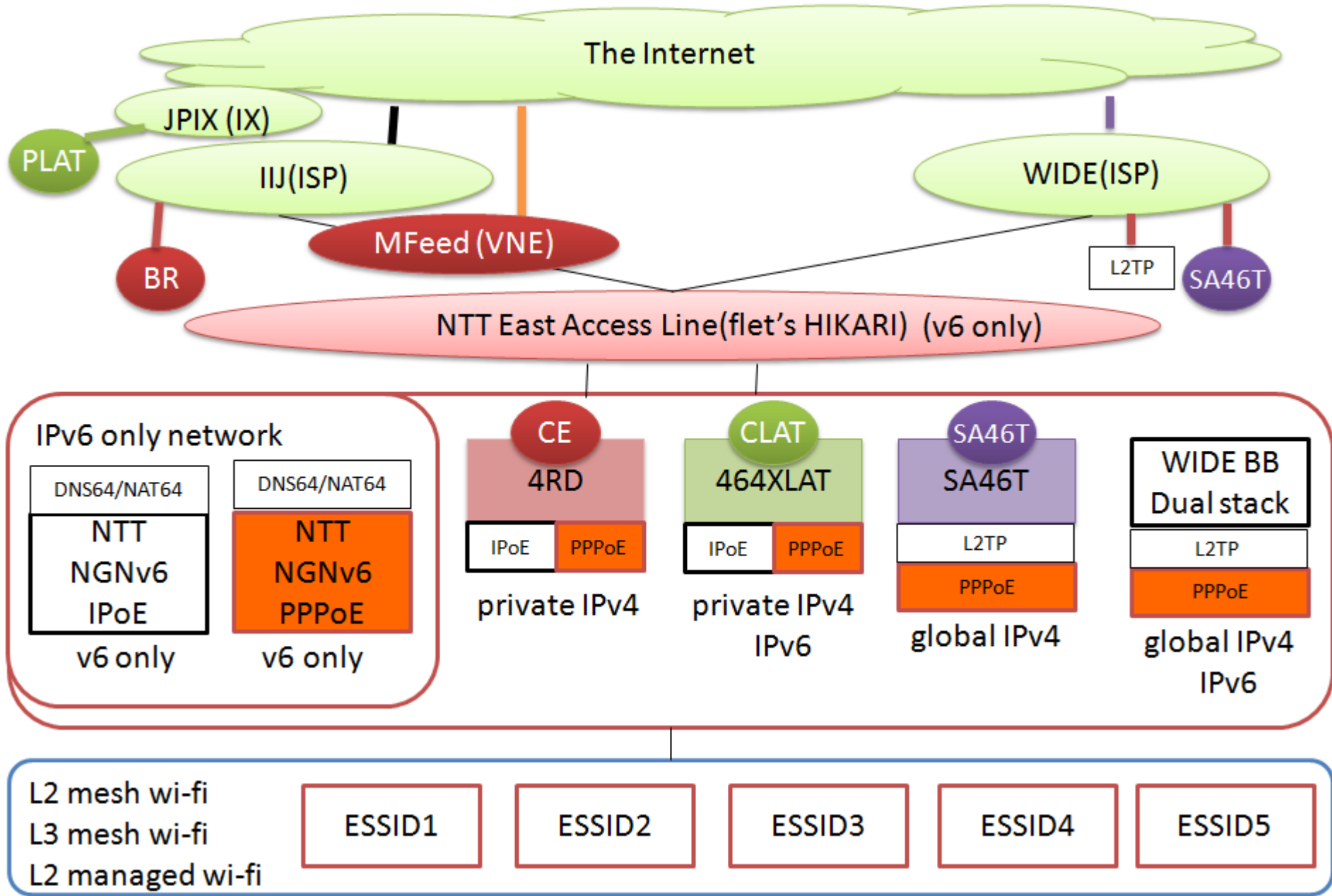
2012年3月WIDE合宿概要

- 開催日時: 2012年3月5日～8日
- 開催場所: 長野県松代ロイヤルホテル
- 参加人数: 171名
 - 持ち込みのデバイス数: 297 個
 - OSの種類: 30種類以上
- 実験
 - Life with IPv6 experiment (IPv6 / IPv4)
 - LISP / VXLAN migration (IPv6 only)
 - Layer 2 Wifi Mesh Network (-)
 - OSLR based Layer 3 WiFi Mesh Network (IPv6 only)

Life with IPv6 experiment 内の 各種実験

- コナミデジタルエンタテインメント(コナミDE)さんによるP2P対戦ゲームとIPv4/v6移行・共存技術実装との相性検証
 - 詳細はJANOG 30 のパネルディスカッションにて！
- 参加者による負荷テスト、運用実験、不具合報告、利用動向調査、原因追究など
 - 参加者全員にヒアリング
 - ほか、参加者有志による詳細報告
 - 時間の問題でパネルディスカッションでは報告しません
 - 代わりに本事前資料で簡単に紹介

IPv4/IPv6 ネットワーク 概要



ネットワーク基本情報 (IPv6)

- 対外線はフレッツ光ネクストとIIJ Mio IPv6 サービスを契約
 - IPoE 方式とPPPoE方式の2回線を契約
 - それぞれの回線を収容するルータ (DHCPv6-PD クライアント) としてIIJ さんから提供していただいたSEIL SOHOルータを使用
 - IPv6 のみを対外線接続では利用

ネットワーク基本情報 (IPv6)

- それぞれの方式ごとにIPv6 only ネットワークを参加者に提供
 - DNS64/NAT64、DHCP6 と組み合わせて提供
- サーバセグメントはWIDE BB (AS2500) にIPv6でL2TPによるトンネルを張りデュアルスタックで構築

ネットワーク基本情報 (IPv4)

- IPv6移行・共存技術検証に参加した各実装によるIPv4ネットワークの提供
 - 464XLAT (private IPv4, global IPv6)
 - 4RD (private IPv4)
 - SA46T (global IPv4)
 - それぞれ、別のISPにBRが存在する形式
 - 各方式の詳細に関しては各方式の事前資料を参照してください
- IPv6 only 環境用にNAT64 / DNS64 をサーバセグメントに設置
 - NAT64 には linuxnat64 と F5 Big IP の NAT64 機能を利用

ネットワーク基本情報 (DNS/DHCP)

- 4RD はCE ルータがDHCP4、DNSプロキシとして動作
- 464XLATはDNSプロキシ (IPv4、IPv6)として動作
- SA46T はCE 側のレイヤ2に DHCP4 サーバを設置
- v6 only 環境は会場側のレイヤ2にDHCP6サーバを設置
- v6 only 環境のDNSはF5 Big IP から合宿ウェブページ以外をDNS64 に転送
 - DNS64にはBINDとUnbound を用意し、比較評価
 - 合宿ウェブページは F5 Big IP のIPv6ロードバランス検証に利用

サブネットとDNSの一覧

Label	Subnet IP Prefix	DNS
Native IPoE	2409:150:8000:10::/64	2001:200:0:ff60::58 (DNS LB)
Native PPPoE	2001:240:2002:6d10::/64	2001:200:0:ff60::58 (DNS LB)
4RD/IPoE	192.168.12.0/24	210.130.1.1 (via proxy)
4RD/PPPoE	192.168.22.0/24	210.130.1.1 (via proxy)
464XLAT/IPoE	192.168.13.0/24 2409:150:8000:30::/64	2404:1a8:7f01:b::3 2001:4860:4860::8888
464XLAT/PPPoE	192.168.23.0/24 2001:240:2002:6d30::/64	2001:240::13 2001:4860:4860::8888
SA46T-FA, SA46T-FK	203.178.156.0/25	203.178.159.58 (DNS LB)

ネットワーク基本情報(他)

- 無線LAN
 - CISCO Wireless L2 mesh で基本的に提供
- サーバ
 - KVM 上の仮想マシンで作った合宿情報Web サーバ 13台
 - コナミDEさんのSTUN サーバ
 - DHCP(4/6)兼DNS(4/6/64)サーバ
 - Syslogサーバ
 - デバッグ用TCPDUMPサーバとストレージ
 - ロードバランサ (F5 Big IP)

機材

(パネルディスカッションに関係あり)

- ONU
 - NTT東日本からIPoE方式、PPPoE方式それぞれでレンタル
- DHCPv6-PD対応ルータ
 - IIIさんからお借りした(普通の)SEIL 2台
- L2TP ゲートウェイ
 - 慶應大学 上野幸杜君が作成したソフトウェア実装

機材

(パネルディスカッションに関係あり)

- 464XLAT
 - PLAT
 - JPIX IPv6v4エクスチェンジトライアルサービスにて検証しているPLAT 機能を備えたルータ
 - CLAT
 - NECアクセステクニカの試作機
 - JPIX IPv6v4エクスチェンジトライアルサービスにて使用中
 - 詳細はNECアクセステクニカ 川島さんの事前資料を参照
- 4RD
 - BR, CE とともに IJ SEIL ベースの試作機
 - 詳細はIJ 末永さんの事前資料を参照

機材

(パネルディスカッションに関係あり)

- SA46T
 - 3種類の実装を用意
 - SA46T-KO : 慶應大学実装、WIDE BB 上に設置
 - SA46T-FK, SA46T-FA : 富士通実装、ホテル側に設置
 - 詳細は富士通 松平さんの事前資料を参照
- STUN サーバ、STUN クライアント
 - 詳細はコナミDE 佐藤良さんの事前資料を参照

機材(その他)

- L2 メッシュ WiFi AP とコントローラ
 - CISCO Systems さんから借用して実験
- OSLR ベース L3 メッシュ WiFi ルータ
 - 東京大学とNTTによる試作機
- F5 Big IP
 - NTT アドバンスドテクノロジーさんから提供
 - ロードバランサとNAT64で利用
- Pivot3 vBank
 - NTT アドバンスドテクノロジーさんから提供
 - TCPDUMP 用サーバ兼ストレージとして利用

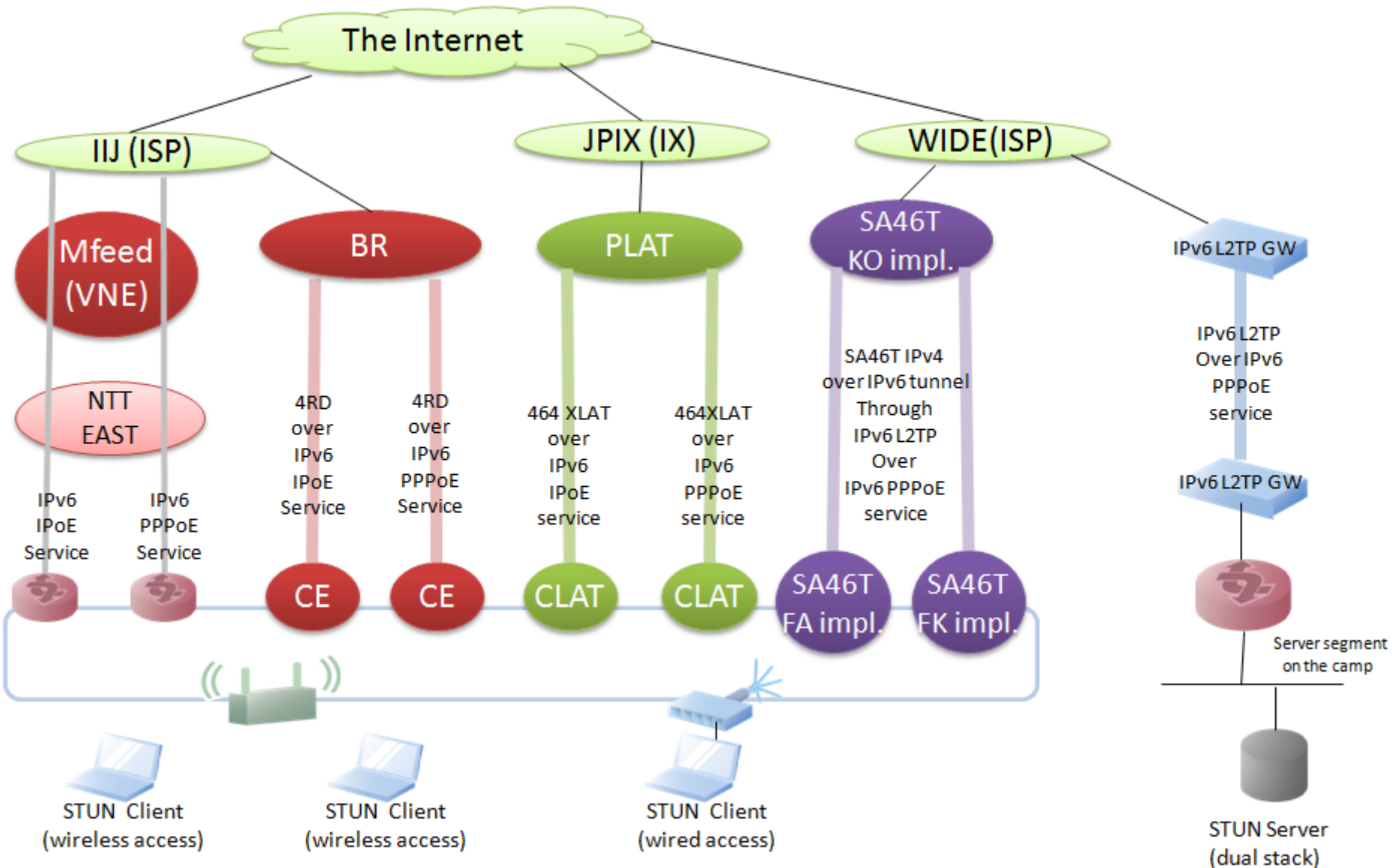
機材(その他)

- レイヤ2スイッチ
 - 東京大学物品、Juniper、Cisco など
- CISCO のAironet (普通の有線付き無線AP)
 - 国立天文台から借用
 - LISP + VXLAN 実験用に利用
- DHCP4/6
 - グローバルIPv4の提供、DNS情報提供用にISC-DHCP を利用
 - v4とv6を別々のプロセスで起動するのがハマらないコツ
- SYSLOG
 - syslog-ngのopen source edition 3.3.4 を使用

Life with IPv6関連実験

1. コナミDEさんによる P2P対戦ゲームの観点からの検証
 - パネルディスカッション前半にて報告
 - 本資料ではネットワークトポロジのみ掲載
 - 検証内容の詳細はコナミDE 佐藤良さんの事前資料を参照
2. 参加者の利用を通しての検証
 - 時間の都合により、パネルディスカッションでは割愛
 - 本事前資料にて簡単に説明

コナミDEさんによる検証トポロジ



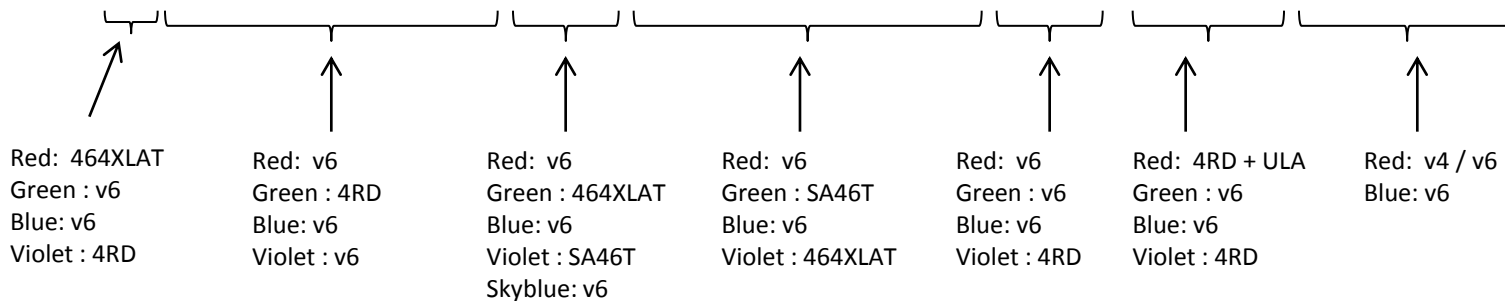
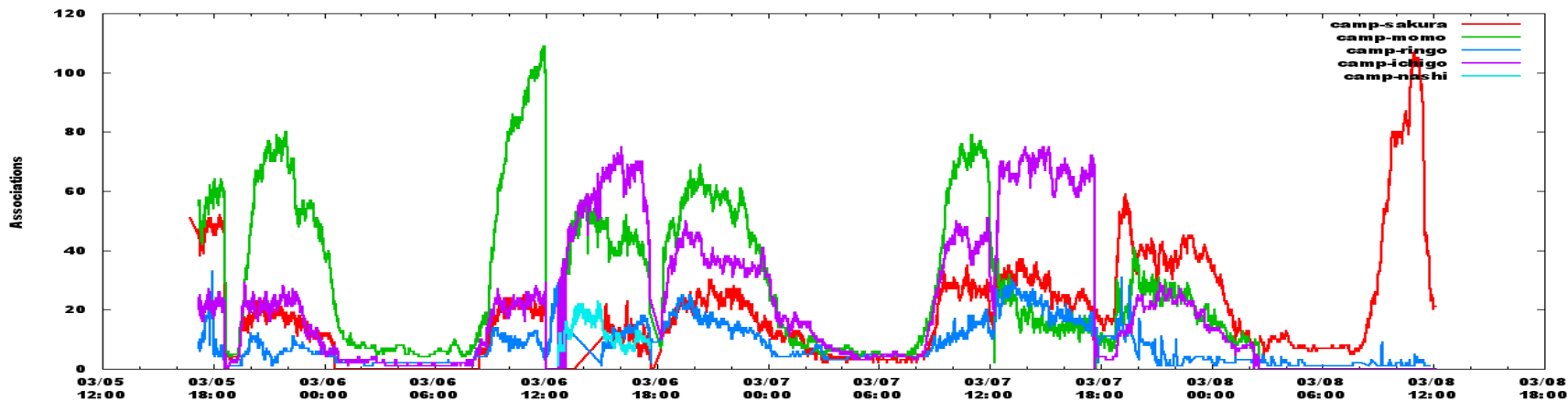
参加者の利用を通しての検証

- 参加者の利用動向把握
 - ESSID を固定で4つまたは5つ設定
 - 昼食、夕食時に提供するネットワークを変更
- 参加者が体験したトラブルの収集
 - 参加者全員にヒアリング
 - 参加者有志によるwiki への詳細報告
- トラブルの原因追究
 - 仕様上の問題なのか、実装上の問題なのか、運用上(コンフィグミス)の問題なのかの切り分け
 - 異なるOSや実装での比較検証
- ワークアラウンド・ノウハウ共有

提供ネットワークタイムライン

time SSID	3/5		3/6			3/7			3/8
	午後	夕食後	午前	午後	夕食後	午前	午後	夕食後	午前
ESSID 1 (sakura)	464XLAT/ IPoE	Native IPoE		L3mesh (IPv6)	Native PPPoE		Native PPPoE	4RD/PPP oE + ULA v6 addr.	Dual stack
ESSID 2 (ichigo)	4RD/IPoE	Native PPPoE		SA46T- FA	464XLAT/IPoE		4RD /IPoE	4RD IPoE through IEEE 8021.11 b	N/A
ESSID 3 (momo)	Native IPoE	4RD/PPPoE		464XLAT/ PPPoE	SA46T-FK		Native IPoE	Native IPoE	N/A
ESSID 4 (nashi)	N/A	N/A		Native IPoE	N/A		N/A	rogue	N/A
ESSID 5 (ringo)	LISP / VXLAN (IPv6)								

参加者の利用動向



- 多くの方がIPv4 アドレスが使えるネットワークを探して移る
- 輻輳してきた気配がすると、別のIPv4が使えるESSIDに移り始める
- 20名 から 60名程度の参加者が定常的に IPv6 only ネットワークで生活
 - 多くは自動設定してくれるWindows 7 とMac OS Lion ユーザ
 - 前回は最大20名程度

参加者の持ち込みハード

- ノートPC : 140台
- タブレットPC : 23台
- スマートフォン : 114台
- ゲーム機 : 19台
- Open WRT 入り無線AP : 1台

参加者持ち込みハードのOS

- PC
 - Windows 7 : 35台
 - Windows Vista : 4台
 - Windows XP : 11台
 - Mac OS X Lion : 48台
 - Mac OS X Snow Leopard : 31台
 - Linux 系 : 8台
 - BSD系 : 2台
- スマートフォン、モバイル
 - Android : 46台
 - iOS : 92台
 - Windows Mobile : 3台
 - Black berry : 1台
 - Nokia 5800 : 1台
- ゲーム機、他
 - PSP : 2台
 - PS Vita : 1台
 - DS/3DS : 3台
 - Kindle : 1台
 - WiMAX Router : 1台

参加者ヒアリング

- 下記の項目をヒアリング
 - 使っているOSとバージョン
 - いつも使っているブラウザ
 - いつも利用しているVPN
 - いつも利用している音声・チャット
 - いつも使っている他のアプリ
 - 会場で試したアプリとその挙動
 - 挙動がおかしい場合は、具体的な時刻やESSIDも(覚えていれば)

参加者ヒアリング

- 3月5日の夕食後から3月7日の間でヒアリング
 - 166名から収集
 - 良い点
 - 参加者が試したアプリやブラウザ、OSが把握できた
 - 反省点
 - ESSID で何のネットワークを提供していたのかを隠していたことと、後からのヒアリングだったため、トラブルが発生した時間や接続したESSID、上手くいったESSIDを参加者が忘れている
 - ヒアリングの集計方法がまとまっていなかったせいか、上手く動いていたのか、挙動不審だったのかがわからない結果がある
 - 対向のサービスがIPv6 対応だったのか IPv4 only だったのかがわからないものが多い
 - 他のトラブル(無線LAN の輻輳や設定変更時のミスなど)

主なトラブル

- VPN関連
 - 実装やバージョン、OS、対向のサーバの設定の違いでv6 only、各種共存技術配下で動いたり、動かなかったり
 - 仕様に関する知識不足もあり、切り分けが大変
- v6 only 環境でのトラブル
 - ネットワーク設定周りで、同じOSでもトラブルに出会う人と出会わない人がいる
 - TCP fallback（特にMac OSの人）

VPN関連のヒアリング結果

- PPTP-GRE
 - 対向のゲートウェイは基本IPv4のみ
 - 利用できたのはSA46Tのみ
 - 464XLAT、4RD は仕様上利用できないが、実装次第で対応は可能
 - SA46TでもWindows 7 では問題ないが Mac OSではうまく動かない(TCPセッションが途中で詰まる)現象を確認
 - 原因は、SA46T実装間、Windows 7、Mac OS 間のフラグメント処理周りの差異だと検証結果から推定

VPN関連のヒアリング結果

- IPsec
 - CISCO VPN Client はOKという人とダメだったという人が存在
 - OS やVPN Client のバージョンによって違う模様
 - どこでもMy Mac (Lion) で繋がる人と繋がらない人が存在
 - サーバ側の設定問題？
 - OpenVPN はv6 only で使えなかったという人が数名
 - サーバ側がv6対応かどうかは不明
- SSL VPN
 - ヒアリングでは問題ない模様

VPN関連のトラブル

- SSH-FS
 - シラバスページがv4 認証らしく、v6 only 環境からだとダメ
- Kerberos
 - v6 only 環境でチケット取れなかった人が存在
 - サーバがv6対応かどうかは不明
- SSH port forward
 - サーバがv6 対応してなかったり、v6対応していても名前登録してなかったりで、設定に苦労していた人が数名

VPN関連とIPv4/IPv6移行・共存技術 のトラブルシュートに関する教訓

- まず、各IPv4/IPv6移行・共存技術の思想を理解する必要がある
 - 移行を促進したいのか、共存し続けたいのか
 - 移行の初期に使うのか、移行の末期に使うのか
 - しっかり状態管理したいのか、状態管理しなくてもいいのか
 - MTUを変えたくないのか、別に変えてもいいと思ってるのか
 - PMTUD が動くことを想定しているのか、していないのか

VPN関連とIPv4/IPv6移行・共存技術 のトラブルシュートに関する教訓

- アドレス変換なのか、トンネル(カプセル化)なのか、複合技なのかを理解する必要がある
 - トンネルだけなら、多くの場合PMTUDの失敗、フラグメント処理周りの実装問題、アプリでのMTUサイズ無視のどれかに絞られる
 - アドレス変換やアドレスマップだと色々ややこしくなる
 - 組み合わせ技術だと、掛け算でややこしくなる
- 仕様上動かない、利用が制限されるプロトコルをまとめる必要がある
 - 星取表あれば、検証するまでもなく動かないと言えるので楽

VPN関連とIPv4/IPv6移行・共存技術 のトラブルシュートに関する教訓

- 検証の際はMTU、フラグメント周りはハマるポイントなので気をつける
 - ネットワーク機器
 - OS
 - アプリ
- v4, v6通してtraceroute できると便利
 - NECアクセステクニカ川島さんに464XLATのtraceroute を見せてもらって感動

VPN関連とIPv4/IPv6移行・共存技術 のトラブルシュートに関する教訓

- 設定上の問題なのか、実装上の問題なのかを切り分けるには、同じ機能を提供する別のOS、別の実装を2つ以上用意して比較評価する
 - 両方ダメならネットワーク側の実装・設定が怪しい
 - 片方ダメならネットワーク側の実装とダメな方のOS側の実装の両方で実装・設定が怪しい
 - OSの問題なのか、アプリの実装問題なのかは、別のアプリで同じことを試して挙動の違いを調べる
 - 全バージョン試すのは大変なので、最新バージョンでまず試す

MTU問題

- IPv4/v6 移行・共存技術検証に関わった全員の悩み
 - アプリ開発者側としてはOS、ネットワーク機器側でよろしくフラグメントしてほしい
 - IPv4/v6 移行・共存技術の実装者側では
 - フラグメント処理は大変
 - 特に小箱だとCPUをすっごく消費する
 - PMTUD はすべからく動いてほしいが・・・
 - 現実には ICMP Packet Too Big を返してくれないノードがいる
 - TCPはMSS clamp で頑張れるが、UDP はどうしたものか・・・

v6 only 環境でのトラブル (ネットワーク自動設定周り)

- v6 only 環境でv4 をOFF にするか、適当な IPv4 アドレス(デフォルト経路なし)を設定していないと無線に接続できない現象
 - Ubuntu
 - iOS (iPhone)
 - 現象に出会った人と、出会っていない人が存在
 - iOS のバージョンで異なる？
 - ThinkPad の Access Connection
 - 繋がる人と、繋がらない人が存在
 - バージョン問題？

v6 only 環境でのトラブル (ネットワーク自動設定周り)

- Lion でv6 only 環境に接続すると
/etc/resolv.conf がおかしくなる、GUI で見える
設定と /etc/resolv.conf の内容が違う
 - 数名、現象に遭遇
- v6 がそもそも付かない
 - 古いAndroid
 - 2.3.6 ではアドレス機種と取れない機種が存在
 - ゲーム機 (3DS, Vita も)

v6 only 環境でのトラブル (ネットワーク自動設定周り)

- ESSID を移り変わると、前の設定が微妙に残ってインターフェースのオン・オフやアプリの再起動が必要になる
 - Windows 7 や Mac OS Lion ユーザで数名報告

v6 only 環境でのトラブル

(TCP フォールバック関連のトラブル)

- Mac OS Lion で v6 only 環境に接続しているときに1分くらい待たされる
 - Windows 7では待たされない
 - 5月末日のWIDE 研究会にて再検証したところ、Mac OS Lionではv4 リンクローカルが付いたインターフェースをdefault route にしている模様
 - RFC3927 Section 3.3に MAY で記述されている項目
 - 結果、IPv6しかない環境で、ARP がタイムアウトするまで待ってしまう

v6 only 環境で上手く動かないアプリ、 サービス

- Skype, MSN Messenger
 - 2012年3月時点ではまだダメ
- Dropbox
- DNS64経由で名前がうまく引けなかったWebページ
- 一部のソーシャルゲーム、twitter クライアント
- Firefoxのプラグイン
 - ひどいものだとブラウザがクラッシュする

NAT64に関する知見

- NAT64はNATテーブル更新・延長周りの実装が重要
 - linuxnat64 のNATテーブルの更新・延長周りの実装があまり上手くない
 - どうも一定時間でリフレッシュしてしまうため、時間のかかるダウンロードが途中で止まってしまう
 - F5 Big IP のNAT64 機能ではリフレッシュの問題は特に報告されず

DNS64に関する知見

- Unbound の方がBIND より FORMERR に寛容
 - FormErr でNXDOMAINが返されると、A Record が登録されていたとしても、DNS64 の AAAA から A への fallbackが発生しないため、NAT64 が機能しない
 - かといって、Unbound ではNXDOMAIN が返されてもBINDでは返されないケースも存在
 - A record が登録されているにもかかわらず、BIND、Unbound 両方でNXDOMAIN が返された名前は15 個存在
 - 合宿中の運用では、2日目にBINDからUnboundへ切り替え