ネットワークの高度利用サポート

田中　仁
KDDI/JP-NOC
加藤　朗
東京大学/WIDE Project

ADVNET2007 in Hiroshima

2007年 1月16日(火)

Agenda

1. 研究活動と R&E ネットワーク
2. ネットワークサポートの必要性
3. サポート時の具体的な問題点
4. より良い成果を導くために
5. Discussion

1. 研究活動と R&E ネットワーク
2. ネットワークサポートの必要性
3. サポート時の具体的な問題点
4. より良い成果を導くために
5. Discussion

研究活動

- 研究者は
  - 情報のやりとりは非常に重要
  - 論文投稿、論文データベース、査読、Google…
  - 基本的に国境は関係ない
    - 必然的に Activity は国際的なもの
- ネットワークへの要求
  - そこそこの帯域と安定した接続性
  - 多くの場合には十分事足りる
  - 通信相手は広範囲である

共同研究活動

情報のやりとりがさらに重要に
- 研究者間の密接な連携
- 大量なデータ転送
- 実時間のデータ転送
  - 遠隔会議、遠隔ワークショップ、遠隔共同作業

- 先進的ネットワークを要求
  - 定常的な接続性では十分ではない場合も多い
  - 高速、広帯域
  - 遅延やパケットロス、ジッタも重要な要因に

国際的なネットワークでの活動

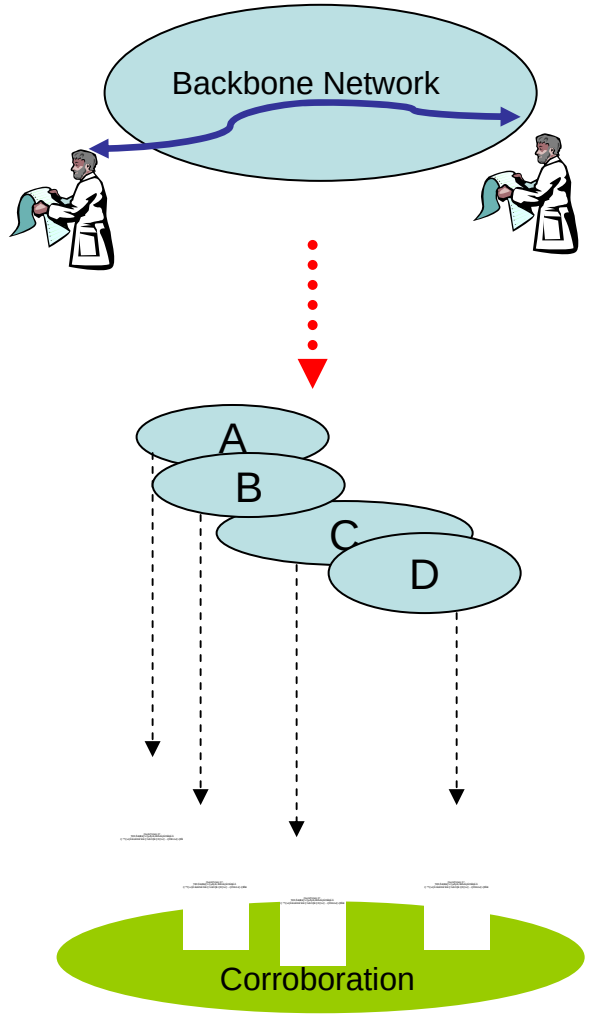
- 国際ネットワークを使った研究実験
  - 長距離間データ転送
    - 宇宙/地球観測、e-VLB I、グリッド、高エネルギー
  - ネットワーク相互接続の研究
    - UCLP、GMPLS、ユニバーサルアクセス
  - 映像系アプリケーション
    - 非圧縮 HDTV、デジタルシネマ、遠隔授業、遠隔医療
- 多くの場合には 2点間の point-to-point 通信
- 最近そうでない場合も増加
- 様々な接続形態、L2 も要求される場面も

日本を中心とした R&E ネットワークトポロジ

1. 研究活動と R&E ネットワーク
2. ネットワークサポートの必要性
3. ┬サポート時の具体的な問題点
4. より良い成果を導くために
5. Discussion

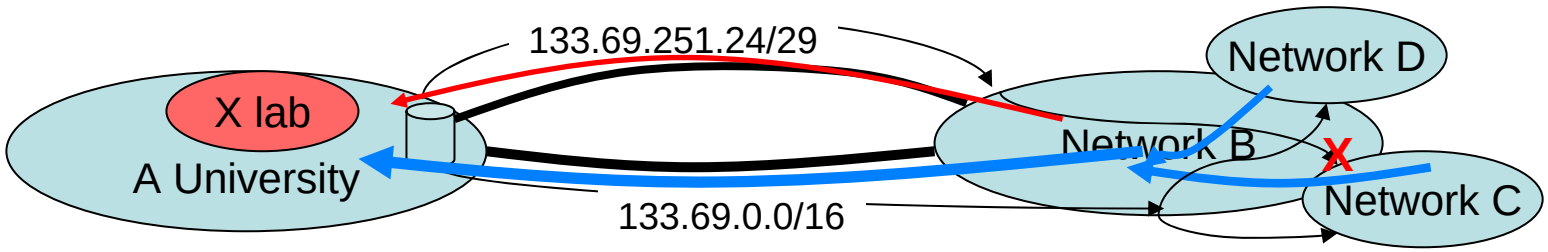
ネットワークドメインと人の把握

- 大まかなネットワーク構成を考え、「どのパスが研究・実験に最適か」を提案
- 必要なサポートのドメイン
  - 発信元キャンパスネットワーク
  - 発信元国内ネットワーク
  - 国際ネットワーク1、国際ネットワーク2、・・・
  - 受信先国内ネットワーク
  - 受信先キャンパスネットワーク
- 必要なサポートのコンタクトポイント
  - ユーザは当然
  - 連携できるエンジニア
  - NOC オペレータ



接続性の確保 -1-
（経路制御）

- 帯域や品質を遅延を気にしない場合には面倒ではない
- 経路が往復とも「好ましい」かどうかの確認は必要
- 定常的な経路が好ましくない場合
  - ネットワーク運用者間の協調による解決
  - 特定の経路を別扱い
    - 一般には発信元にはよらない、影響を最小にするためには
      特別扱いの経路の prefix 長は可能な限りに長く
      e.g. 133.69.0.0/16 ではなく、130.69.251.24/29
    - 広報に関する制御、漏洩に関する対処
    - 国際ネットワーク間では複数箇所での調整が必要
    - 反対向きの経路の調整も必要



接続性の確保 -2-
（IP パラメータ）

- IPv4 or IPv6
- Unicast? Multicast? Anycast? Xcast?
- IP address
  - どのアドレスブロックから切り出すか
- バックボーンネットワークへの接続方法
  - どのようなルーティングプロトコルで接続するか？
  - BGP? OSPF? Static? Connected?
- トランスポート
  - UDP? TCP? UDT? XTP?
- MTU サイズ
- VLAN ID
  - 利用できる範囲
  - Conflict

接続性の確保 -3-
（アプリケーションとトラヒックパターン）

ネットワーク屋がアプリケーションの特徴を知りたがる時代に

- どんなアプリケーションなのか？
  - 対話が重要ならば Latency を考慮した設計
  - 長距離 TCP データ転送であればパケットロスは致命的
  - Multicast であればリバースパスフォワーディングチェックを考慮
  - VoIP であれば Queue Buffer は短い方が良い
    - 特徴にあわせネットワーク機器のQueuing までも操作するようになってきた
- どのようなトラヒックパターンか？

  e.g. DVTS : 33Mbps

  この数字はあくまでも「平均」

  Microscopic に見ると非常に busty なものが多数ある
  - busty なトラヒックはパケット損失に繋がりやすい

  送信元の pacing : ネットワークに優しい
  - それを熟知しているユーザは多くない

1. 研究活動と R&E ネットワーク
2. ネットワークサポートの必要性
3. サポート時の具体的な問題点
4. より良い成果を導くために
5. Discussion

情報共有の不足による問題

1. 日本側研究者⇔米国側研究者+米国側オペレータだけでコーディネイトされデータ転送実験を開始
   - 期待されたパフォーマンスが出ない
     - パケットロスポイントが分からない
   - 1ヶ月後に全員でテレカンファレンスを実施
   - 日本側機器構成変更
     - 10G-> 100M から 10G -> 1G)構成へ変更
   - 期待されたパフォーマンス、本番データを転送できる環境に

   日米のネットワークエンジニア、オペレータの間の相互理解により問題解決

2. 日中国間で突然遠隔講義を実施するとの知らせが
   - 会場にはプロジェクト担当のエンジニアが配備
   - リハーサルを実施
     - パケットロスにより音と画像に乱れ
   - 本番直前にバックボーン NOC オペレータに問題の申告が届く
   - 中国内に輻輳回線を発見
     - 日本からの経路広報を調整
   - 輻輳ポイントを回避するよう通信路を迂回し本番直前に問題解決

   遠隔講義を実施する情報がもっと早く NOC に届いていれば・・・

障害切り分け時の問題

1. 日本側ノードと米国側ノード間のJumbo Frame が通らない
   - 日本国内、日米間スイッチの設定を確認
   - 拠点スイッチにVirtual Interface を作成
     - オペレータ自ら Jumbo Frame の正常性を確認
   - 時差により2日のタイムロス
   - 結局エンドユーザ側、米国設置スイッチの問題であるということが判明

   出来ればエンドホスト（ユーザ自ら）側を調査した結果を添えて申告してほしい
2. 研究者とベンダとの間で
   - 予想された速度の 5% しかパフォーマンスが出ないと研究者から申告
   - 研究者と日米オペレータが問題解決に挑む
   - ネットワークか？マシン側の問題か？
   - マシン側の細かな質問に対しての反応が鈍い
   - マシン側担当のベンダと直接やり取りを開始
   - TCP スタックの改善を提案

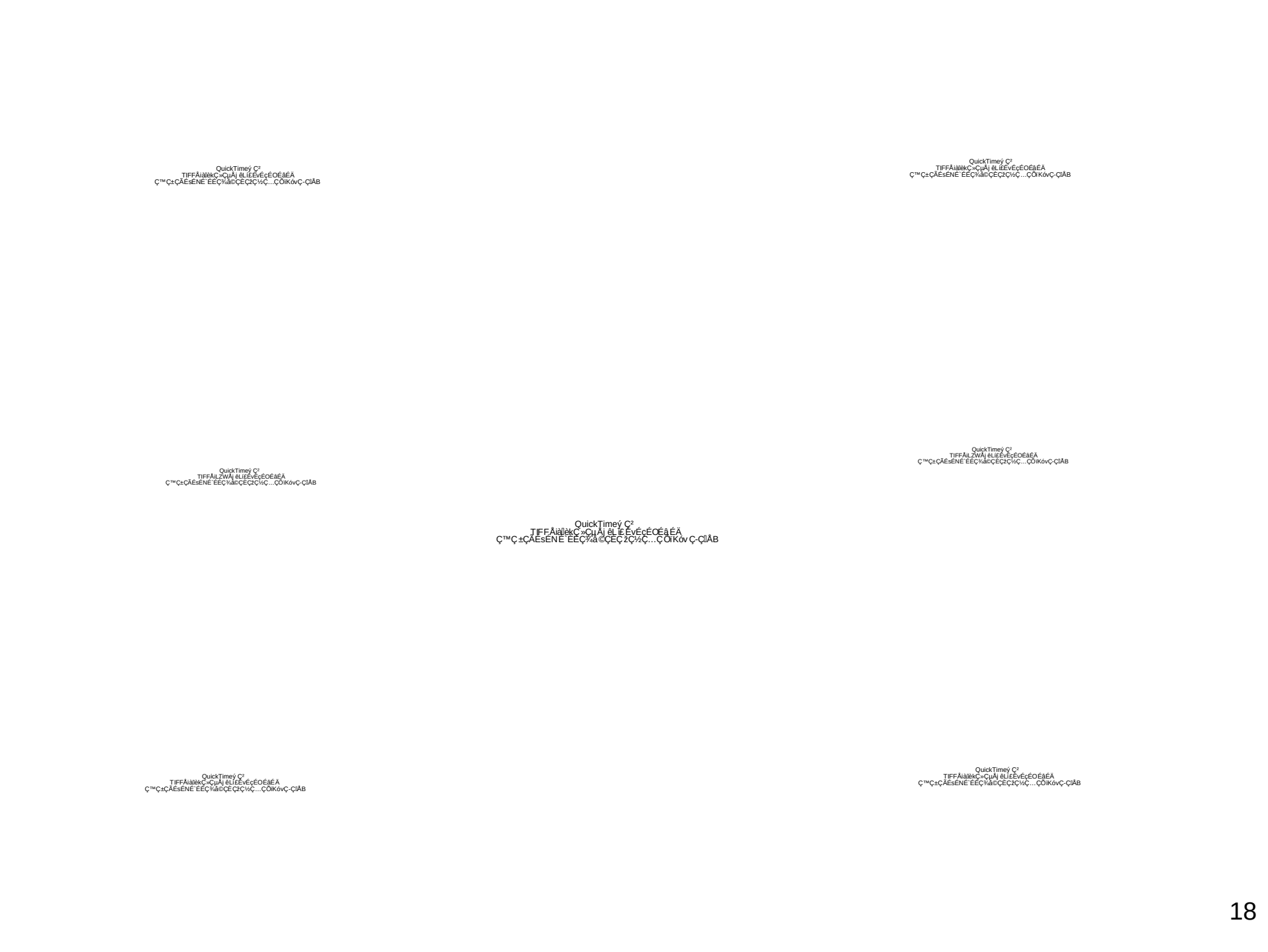
   関わりのあるベンダと連携できる体制が望ましい

1. 研究活動と R&E ネットワーク
2. ネットワークサポートの必要性
3. サポート時の具体的な問題点
4. より良い成果を導くために
5. Discussion

SC06 における

- 複数ドメインを利用した重要なデモやイベント時に活動
  - SC05 で活動開始
  - SC06 で本格的に活動
- 日本の R&E ネットワークの運用窓口を一元化
  - 異なるサービス、品質、ポリシー
  - ユーザの混乱を配慮
  - R&E ネットワークの横の繋がりが強化され迅速に対応
- Members

K.Hasebe(WIDE/JGN2/NTT-c), M.Hirabaru(NICT), T.Ikeda(JGN2/APAN-JP/KDDI), Y.Kanaumi(JGN2/NEC), A.Kato(WIDE/Univ.of Tokyo), Y.kitamura(NICT/APAN-JP), K.Kobayashi(AIST), K.Konishi(APAN-JP), J. Matsukata(SINET/NII), Y.Tahara(JGN2/APAN-JP/KDDI), J.Tanaka(JGN2/APAN-JP/KDDI), APAN-JP NOC, JGN2-NOC, TEIN2-JP NOC
- アルファベット順 -



R&ENOG
Research & Education
Network Operators Group

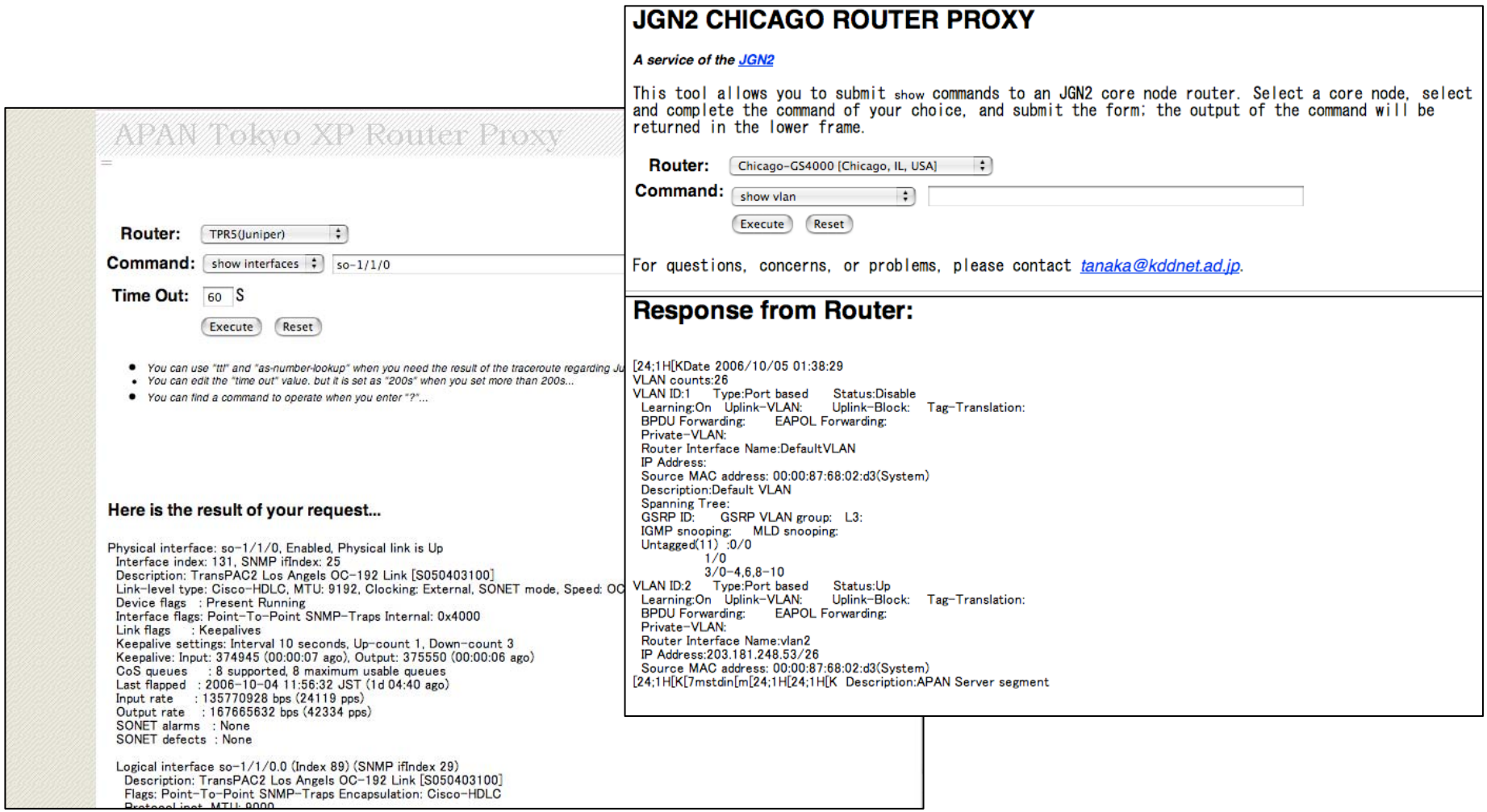
- 世界中の R&E ネットワークオペレータが参加
  - 現在 70 名ほどがメーリングリストに登録済み
  - http://www.renog.org/
  - GlobalNOC(USインディアナ大)、GEANT(EU)、ESnet(US)、Internet2(US)、AARNet(AU)、JGN2(JP),SINET(JP), WIDE(JP) APAN-JP(JP)など
- 現在議論している話題
  - 世界規模のルーティング問題
  - インタードメインでのL2 ネットワークの障害切り分け
  - NOC ツールの紹介
- Internet2 及び APAN Meeting に併せ face-to-face Meetingを開催

運用ツールの公開

- 研究者がオンデマンドにネットワークの稼働状況を把握できる
- ネットワークの運用情報公開も R&E NOC の重要な役割り
- 商用 ISP ではなかなかできないこと
- 共通なデータを提供
  - オペレータ同士が分かり易い
  - Free-soft
  - Standard

    e.g. RRDdata, Iperf, Ticket System

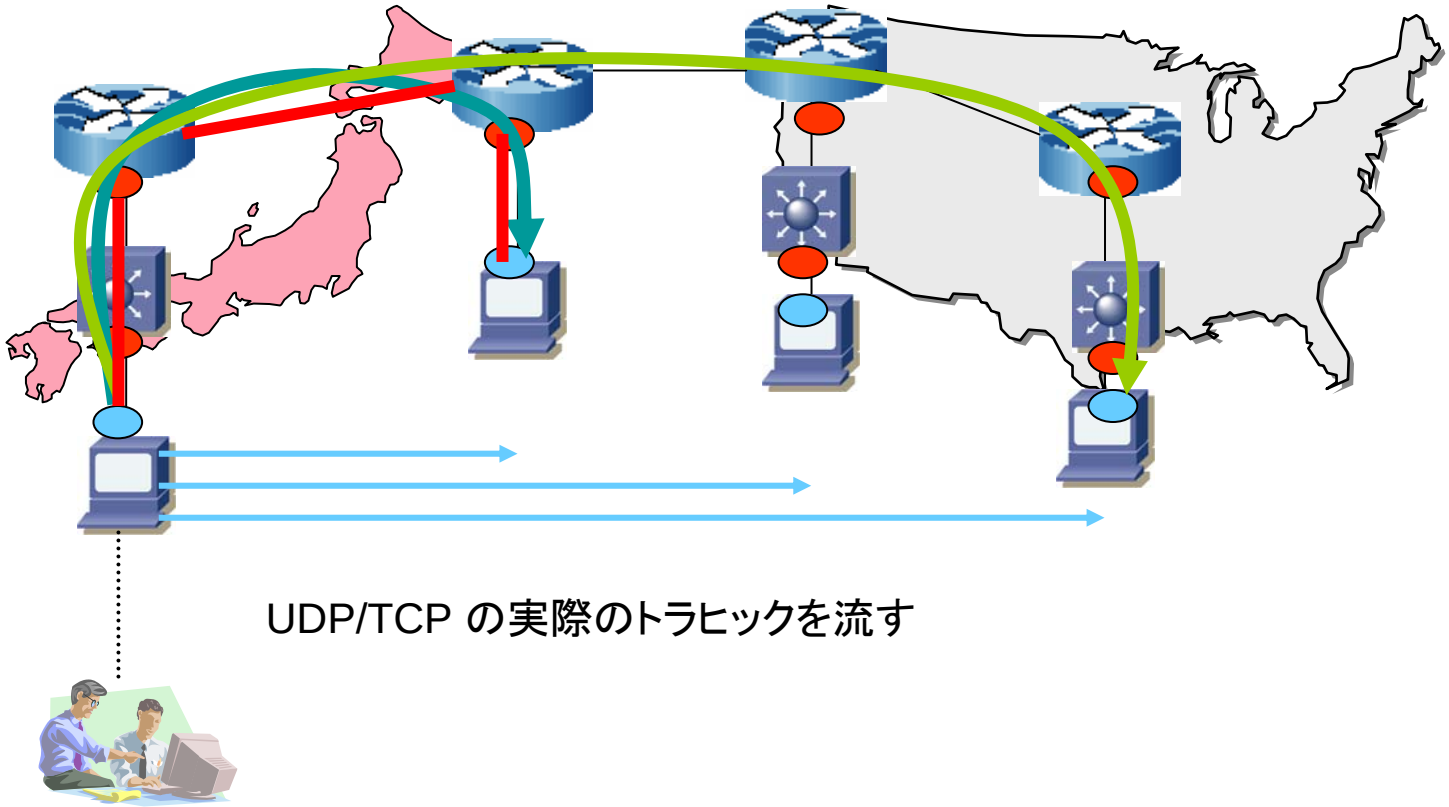
Router Proxy



http://tools.jp.apan.net/rp/
http://192.26.87.202/proxy/
http://routerproxy.grnoc.iu.edu/abilene/

BWCTL / Iperf

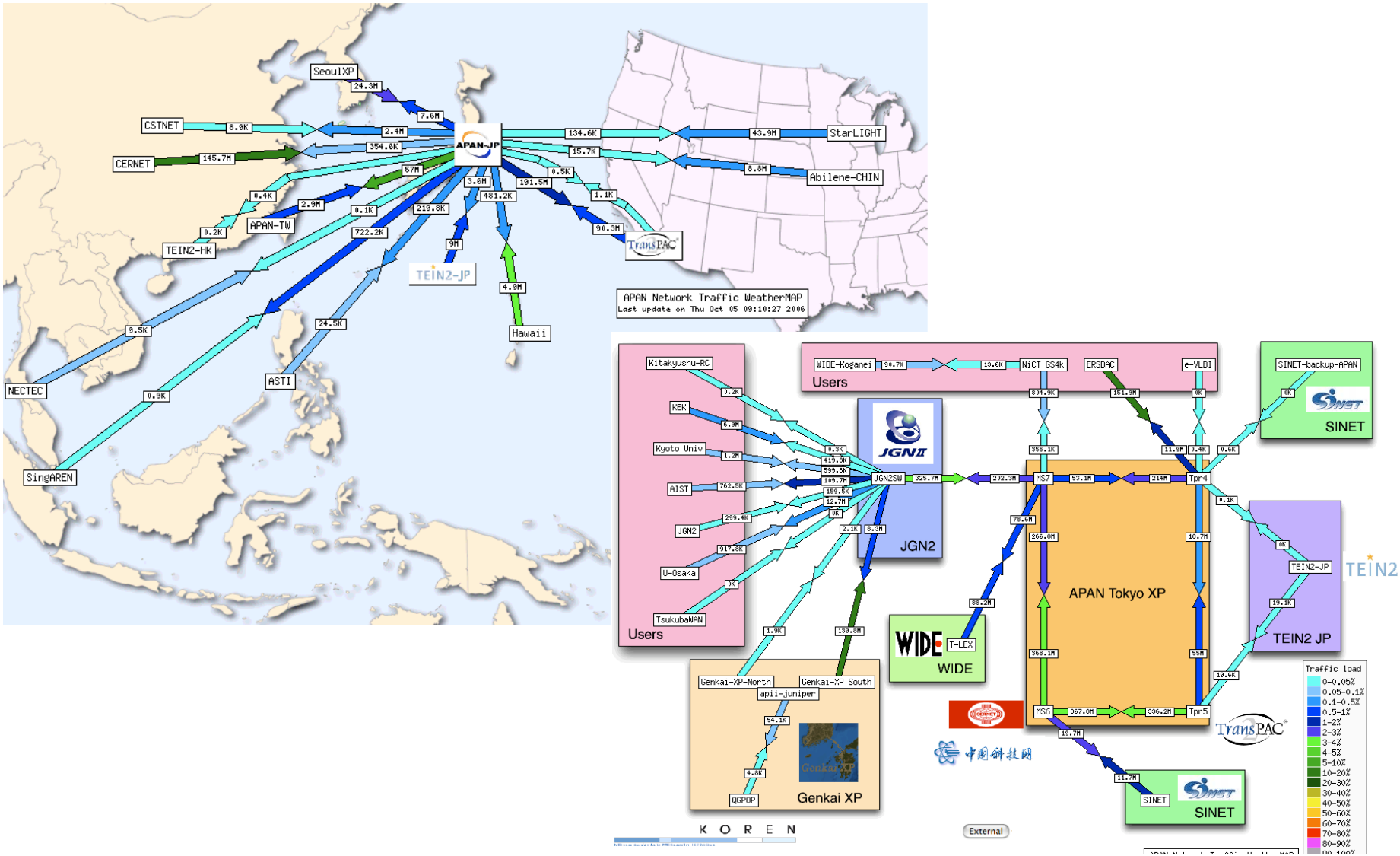
ユーザ自身がバックボーン区間のパフォーマンスを測定出来る仕組み



UDP/TCP の実際のトラヒックを流す

http://www.jp.apan.net/NOC/bwctl/

Traffic Weather MAP



http://mrtg.jp.apan.net/weathermap/index.php
http://mrtg.jp.apan.net/weathermap_domestic/index.php

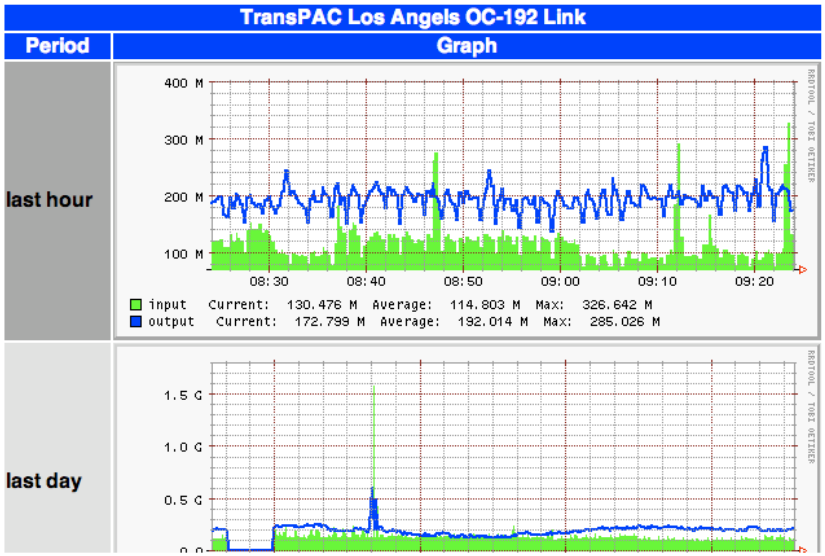
Traffic Graphs

APAN Asia-Pacific Advanced Network

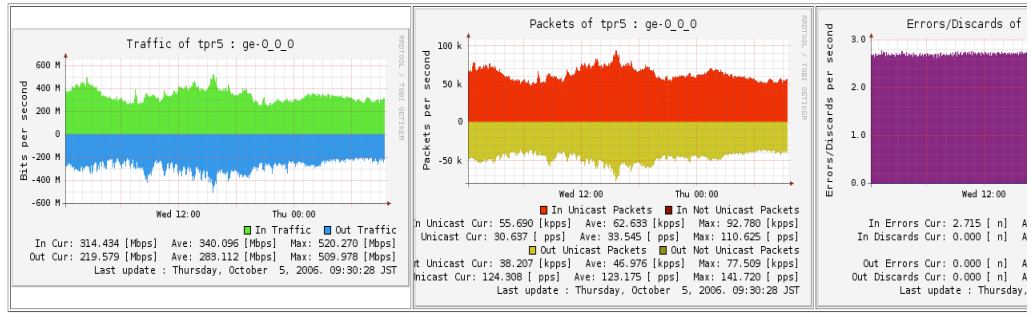
SNAPP
SNMP Network Analysis and Presentation Package

Standard Graphs for tpr5-so-1-1-0_0

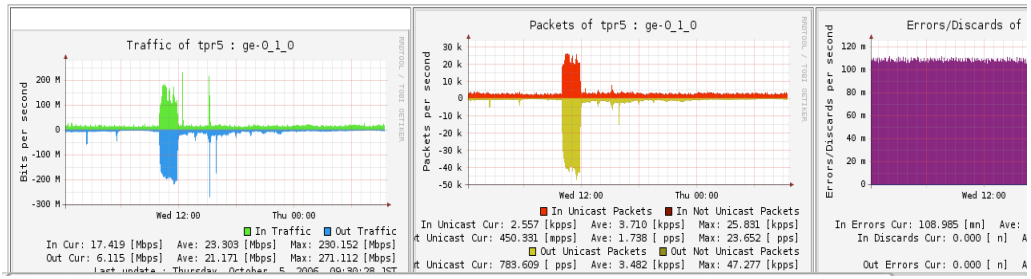
Create a custom graph



tpr5 : ge-0_0_0 (to MS6:tagging) 10 Gbits/sec RAW Data (.rrd)



tpr5 : ge-0_1_0 (to TPR4 10G 0/1/0:untagged) 10 Gbits/sec RAW Data (.rrd)



http://mrtg.jp.apan.net/cricket/router-interfaces/index.html
http://nms2.jp.apan.net/cgi-bin/snapp

お願いしたいこと -1-

- 新しい利用の場合
  - 実時間アプリケーション
  - 1sec 分の TCPDUMP を預けると
  - 通信相手は Local で良い
  - Port-mirroring で手軽で取得できる
    - 時間的にはかなり jitter が加味されるが
- ネットワーク屋（キャンパスネットワーク）との協調
  - 研究者は研究をし、ネットワーク屋がサポートする体制
  - 研究用機器ベンダーも重要であるが、ネットワーク屋も重要
  - 成功例：九州大学病院
    - 清水先生（医者/アプリケーション）
    - 岡村先生（ネットワーク研究者/ネットワーク）

お願いしたいこと -2-

- 情報の共有
  - なるべく早い時期に
    - 各種準備には時間がかかることが多い
    - 国際的習慣、時差、祝日等の問題もある
  - 我々が求めるのは
    - 日時、通信相手、関係者の連絡先(E-mail、電話番号等)、双方の正確な IP アドレス情報、通信形態、利用パターン、特別なパラメータ
- デモンストレーションの場合には
  - リハーサルを早期から入念に実施してから本番に臨む体制
  - 人手が必要な場合、まずはご連絡ください
- バックアッププランの策定
  - もし動かなかったら(各種事故を想定)

お願いしたいこと -3-

- 可能であればトラヒック測定環境の準備
  - パケット損失問題には不可欠
    - トラヒックパターンの傾向を掴む事が重要
      - 散発的、busty、時刻などとの関連性
  - NOC も協力できるところは協力
    - NTP
    - SNAPP(10秒間隔のトラヒックグラフ)
- ネットワークリソースの共有
  - 基本的には NOC が中立的な立場で調整を行う
  - しかし最終的には研究者同士の相互理解
- フィードバック
  - ネットワークの状態
  - 経験(失敗)は次に繋げたい

1. 研究活動と R&E ネットワーク
2. ネットワークサポートの必要性
3. サポート時の具体的な問題点
4. より良い成果を導くために
5. Discussion