Inter-domain Architecture
のこれからを考える

河野 美也 Miya Kohno, (mkohno@juniper.net)

Agenda

- *前回(IRS12)発表内容から*
- Scaleable Inter-domain routingに向けて
  - RA discussion
  - いくつかの見直し
- アーキテクチャ再考 ?!!

前回(IRS12)発表内容から (1/2)

- **BFD for the Internet e-BGP peering ?**
  - **Deployment事例見つからず**
  - **不確定要素多い**
    - 下位レイヤ(Protection/Resiliency/OAM)との組合わせ
    - IPv6
    - Graceful Restartとの組合わせ
  - **Global InternetではStabilityの方が大事**

前回(IRS12)発表内容から (2/2)

- **Global Route Oscillation**
  - 一般的に、Fast Detection/Fast Convergenceと、Stability/Scalabilityは拮抗する。

    → 共存させるために
    - Carrier delay
      for DOWN event, UP event
    - Exponential Back-off
      for IGP spf-delay, LSA generation
    - BGP RFD (*), MRAI
    - Interface Dampening

    …

(*) しかし、arbitraryなdampeningは却って害がある。
Is RFD harmful ?!
http://www.ripe.net/docs/routeflap-damping.html

Agenda

- 前回(IRS12)発表内容から
- *Scaleable Inter-domain routingに向けて*
  - RA discussion
  - いくつかの見直し
- アーキテクチャ再考 ?!!

RA discussion (1/7)

- 発端(の一つ)
  - ARINのMarla Azingerから、v6ops@ops.ietf.orgへの投稿(2006年6月28日)
    - draft-ietf-v6ops-routing-guidelines-00はmultihomeに関するguildelineが無い。v6でもv4と同様multihomeできるようにすべき。IETFにGuidelineを示すように要請。

(最初の反応)

- More specificジャンクによって、v6アドレススペースの泥沼を作るのはやめよう。
  Let's not create a swamp out of v6 address space with more specific junk. (Pekka Savola)

- RFC 4177 section 5.1: このアプローチは、マルチホーム方式の全ての目標に合致するが、一つだけ問題がある。ー それはスケーラビリティ。"This approach generally meets all the goals for multi-homing approaches with one notable exception: scaleability."

draft-baker-v6ops-l3-multihoming-analysis

RA discussion (2/7)

- **IAB workshop　18-19 Oct.2006**

  http://www.iab.org/about/workshops/routingandaddressing/

  [問題の定義]

  1. Routing Scalability
  2. The overloading of IP address semantics
  3. Routing Convergence
  4. Misaligned Costs and Benefits
  5. Others
     1. Mobility
     2. Routing Security

  [Workshopからの提言]

  1. Scalability of Routing and Addressing System is a concern
  2. More discussion is needed with broader audience
  3. Solution development should be open and transparent
  4. Short/Intermediate term solution is needed concurrently
  5. Roadmap to the solution deployment
  6. Miscs (to create the mailing list (ram@iag.org) , etc.)

  [Report]

  - http://www.potaroo.net/ispcol/2006-11/raw.html
  - http://www3.ietf.org/proceedings/06nov/slides/RRG-0.pdf
  - http://www3.ietf.org/proceedings/06nov/slides/plenaryt-5.pdf

RA discussion (3/7)

- **NANOG 39 BOF　5 Feb.2007**
  現在のルータにおけるFIBの限界という観点から

  http://www.nanog.org/mtg-0702/jaeggli.html

  Joel Jeaggli氏の呼びかけで、各ヴェンダーがFIB memoryアーキテクチャや取り組みについて報告
  (Force 10, Cisco, Foundry, Juniper, Extreme)

- **Apricot “future of routing” workshop　26-27 Feb.2007**
  IAB workshopのfollowup

  http://www.apricot2007.net/presentation/apia-future-routing/

  Dave MeyerがChair
  Jari Arkko, John Scudder, Vince FullerがRouting Scalabilityを考察

RA discussion (4/7)

- IETF68 21 Mar. 2007
  - Plenary -- 総論
    http://www3.ietf.org/proceedings/07mar/slides/plenaryw-3.pdf

[経緯]

- Historical timeline
  - Packet switching invented (1962)
  - Internet concept invented (1974)
  - IP designed (~1978)
  - BGP designed (~1988)
  - CIDR designed (1992)
  - IPv6 designed (1995)

- Growing concern about
  - scaling, transparency, multihoming, renumbering, provider independence, traffic engineering, IPv6 impact (1995-2006)

- IAB Routing & Addressing workshop (2006)

RA discussion (5/7)

- IETF68　21 Mar. 2007
  - Plenary　-- 総論
    http://www3.ietf.org/proceedings/07mar/slides/plenaryw-3.pdf

[アーキテクチャについて]

- 保持すべきアーキテクチャ原則:
  - ネットワークは、過度に他のネットワークの運用に影響を与えることなく、道理にかなった相互ネットワークの手段を選択し実装すべきである。
- アーキテクチャレヴェルの現在の問題:
  - 現在の相互ネットワークのいくつかは、この原則を脅かす方法でしか実装できない。
  - これが、ISPにまつわる問題およびend siteの不満の根本原因である。
  - ネットワークを、アーキテクチャ原則に調和させるためにはどうしたらよいか。
  - PIアドレスにまつわる「共有地悲劇」

RA discussion (6/7)

- IETF68 22 Mar. 2007
  - Internet Area
    http://www3.ietf.org/proceedings/07mar/agenda/intarea.txt
    - Background and scope (ADs 10 min)
    - Routing issues of concern where id-loc split might play a role (Ward/Scudder 20 min)
    - High-level design space (Thaler 45 min)
    - Where are current work applies, and where we need work going forward (Nikander 15 min)
    - Discussion (45 min)
    - What we do next (ADs/chairs 10 min) – (Conclude the discussion, next steps, and try to read consensus)

RA discussion (7/7)

- **IETF68 22 Mar. 2007**
  - **Routing Area**
    http://www3.ietf.org/proceedings/07mar/agenda/rtgarea.txt
    - Relationship of Inter-domain discussion with ROAP effort (ADs, 5 minutes)
    - Inter-domain Routing Trends (Geoff Huston, 15 minutes)
    - Thoughts on Improving Inter-Domain Routing (Dave Ward, John Scudder, 30 minutes)
      – -ideas for improving scaling, convergence, ...

- **Nanog40 5 Jun. 2007**
  - **LISP概説**

    http://www.nanog.org/mtg-0706/Presentations/lightning-farinacci.pdf

いくつかの見直し

- **[IDR] draft-li-bgp-stability-01**

  Dampingについての見直し

  Goal –
  - Flap damping
  - Rapid Convergence
  - Overhead削減
    - Path hunting
    - 屈折(Refraction)
      を避ける

いくつかの見直し

- **[IDR] draft-li-bgp-stability-01**

  Dampingについての見直し

  仮説 –
  - 止める (turn it off)
  - パラメータを変える (Alternate parameters)
  - Band-stop filtering
  - Path length damping
  - 最適パスヒステリシス
  - path selectionを遅らせる
  - MRAIの廃止
  - これらの組み合わせ
  - その他
    - Aggregate withdraw

いくつかの見直し

- **[RRG] draft-irtf-rrg-design-goals**

  - **Inter-domain routing architectureは、scalability, mobility, inter-domain TEといった課題を抱えている。RRGではそれに対応するアーキテクチャを検討するが、その前段となるゴールの洗い出し**

  1. **Improved routing scalability**
  2. **Scalable support for traffic engineering**
  3. **Scalable support for multi-homing**
  4. **Scalable support for mobility**
  5. **Simplified renumbering**
  6. **Decoupling location and identification**
  7. **First-class elements**
  8. **Routing quality**
  9. **Routing security**
  10. **Deployability**

Agenda

- 前回(IRS12)発表内容から
- Scaleable Inter-domain routingに向けて
  - RA discussion
  - いくつかの見直し
- *アーキテクチャ再考 ?!!*

アーキテクチャを見直すということ

- 日本では、「IPv4枯渇 -> IPv6への移行」で一部物議を醸し出しているが、アーキテクチャが変われば前提が変わる。
- 「アーキテクチャに、ビジネスは依存している」© 水越一郎

- 本当にアーキテクチャを変える必要があるのか?
- 現在出ている提案はどういうもの?
- The Internetはtransition可能か?

現在出ている提案

- **LISP　draft-farinacci-lisp**
  - Locator/ID Separation Protocol
  - Network Based Solution
    - フォワーディングの階層化
    - EID – 内側のヘッダ、Locator – 外側のヘッダ
    - ITRにおけるEID to Locator mapping
  - Variation
    - LISP1: EID = routable, EID-RLOC mapping = inband(same topology)
    - LISP1.5: EID = routable, EID-RLOC mapping = outband(different topology)
    - LISP2: EID = non routable, EID-RLOC mapping = DNS-like
    - LISP3: EID = non routable, EID-RLOC mapping = DHT-like
    - LISP2以降については、draft-lear-lisp-nerd も参照

- **iVIP**
  - Internet Vastly Improved Plumbing … 当て字
  - LISPの変種?Anycast ITR

  - GoalはLoc/ID splitによるFIB size削減　(+ multihoming, TE)

…という訳で、RoutingとAddressingのこれからを考える@Janog20

2007年7月13日

13:00-13:10 Introduction --- 河野 美也

-- Background, Motivation, Problem Statement, Scope/Non-scope

13:10-13:25 From Academic & Historical point of view – 東京大学 加藤 朗

-- Historical speculation (pre/post-CIDR, GSE, PIP, etc.), Essential acceptation

13:25-13:50 IAB/IETF/RRG update --- Robert Raszuk

-- report from IAB workshop, ROAP session and the recent discussion in IETF and IRTF/RRG

13:50-14:00 From SP point of view --- NTT-East 水越 一郎

14:00-14:10 From ISP point of view --- IIJ 浅羽 登志也

14:10-14:30 Discussion --- all

コメントや事前質問歓迎します。

おまけ

ここからは、かなり私見、試論レヴェル…

- 現在のアーキテクチャ見直し論議は、
  - RIB/FIB削減のみが主眼
  - Transition Costに見合う効果があるか疑わしい
- アーキテクチャを再考するのであれば、次のようなことも考えてもよいのではないか？
  - 品質、セキュリティ、Availability
  - 抗 – IP transportのコモディティ化

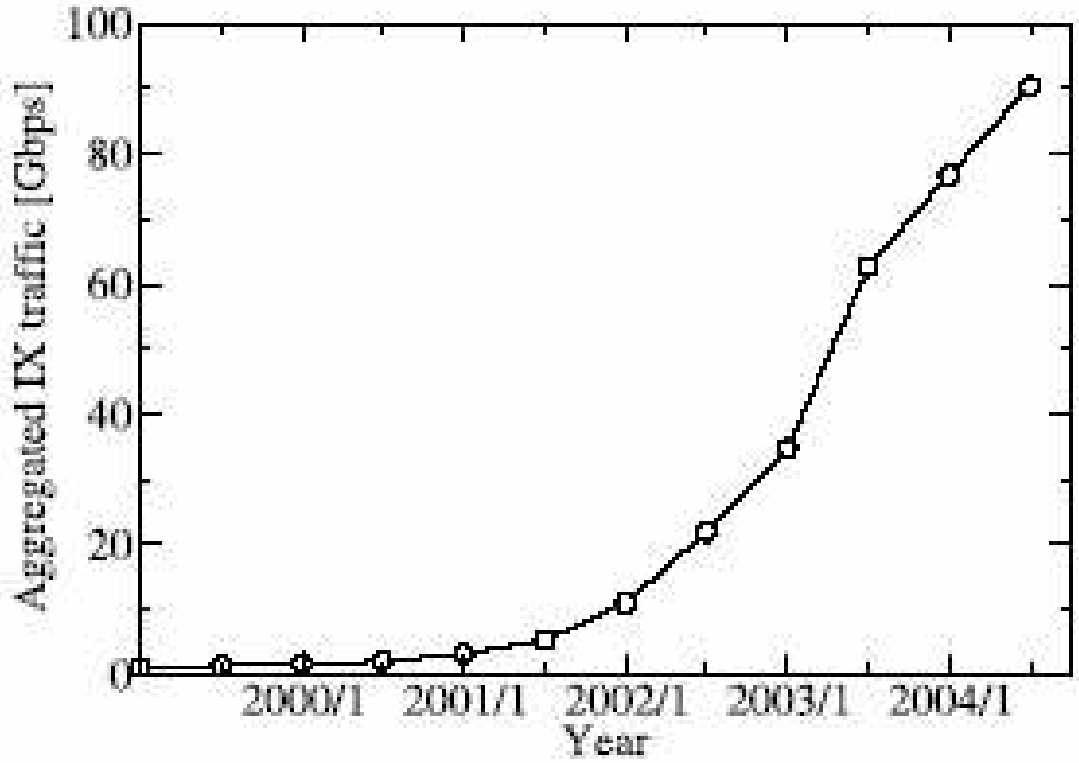
社会インフラとしてのインターネット

- 信用できない世界での運用
  - **Operation in an untrustworthy world**
- 一層要求条件の高いアプリケーション
  - **More demanding applications**
- ISPサーヴィスの差別化
  - **ISP service differentiation**
- 第三者の介入
  - **the rise of third-party involvement**
- 洗練されていないユーザ
  - **less sophisticated users**

r.f. Rethinking the design of the Internet: The end to end arguments vs. the brave new world – D.Clark et al.

The Internet - 最近の状況

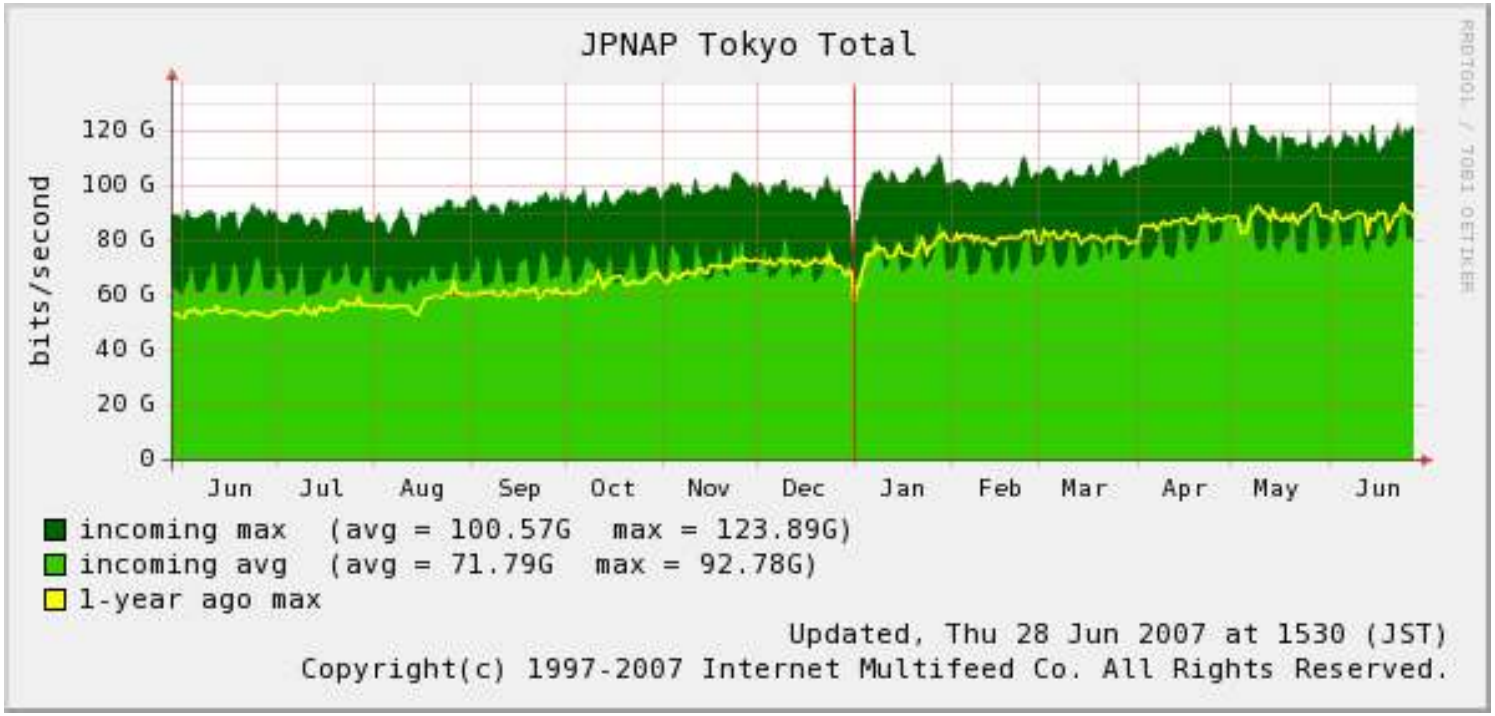
- **Trafficの急激な増加**
  - 前年比 > 30%増



http://www.iepg.org/march2005/kjc-iepg200503.pdf

The Internet - 最近の状況

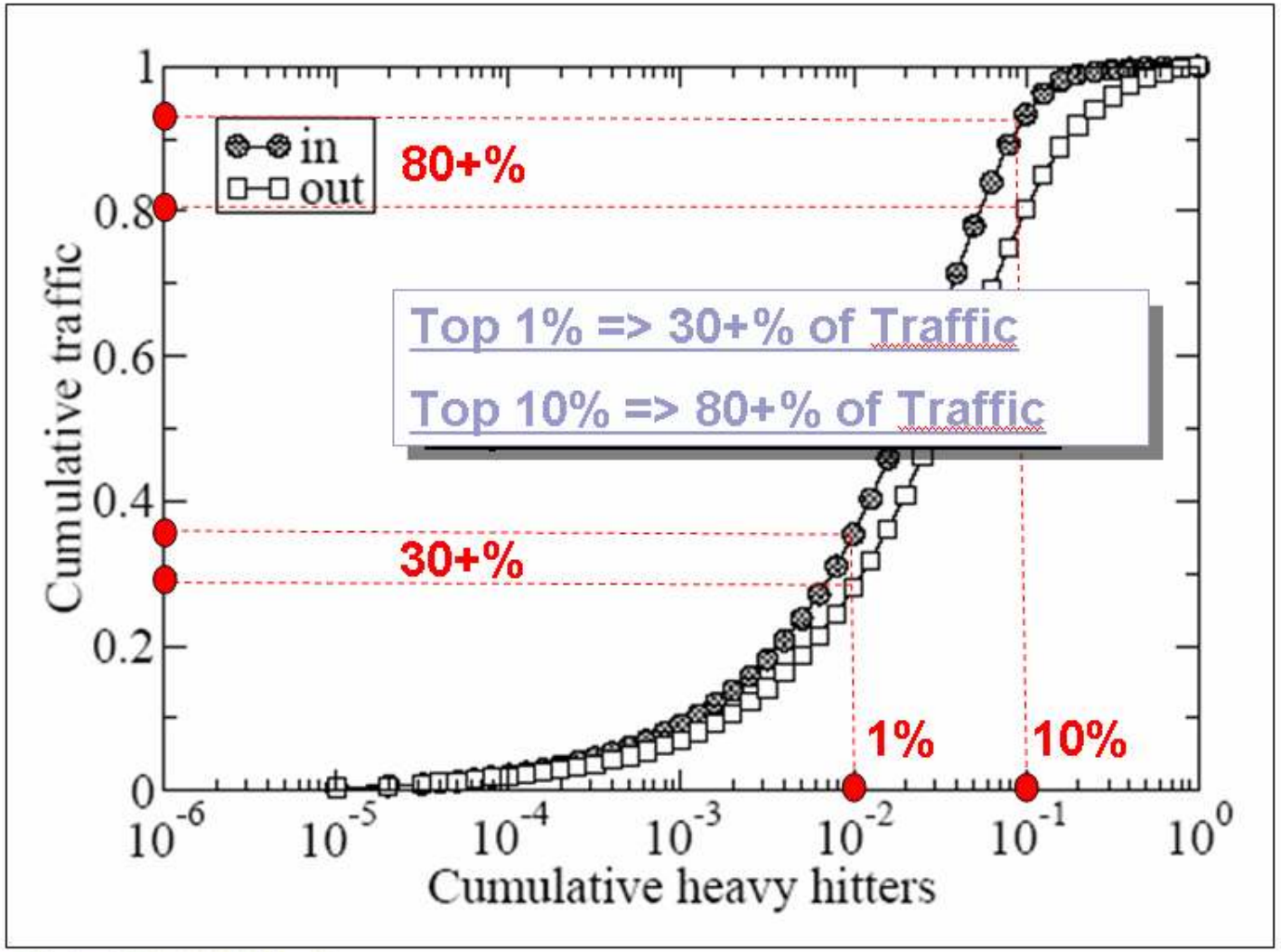
- **Trafficの急激な増加**
  - >120G！



http://www.mfeed.ad.jp/jpnap/traffic.html

The Internet - 最近の状況

トラフィック量の分布



C http://www.iijlab.net/~kjc/tmp/rbb-20060211.pdf

IPインフラに求められること

- ScalabilityとStability + α ？
  - Subscriber Control
  - Policy Control
  - Forwarding Control
  - Security
  - QoS assurance
  - High Availability

…

- …って、これってNGN？
- しかし、NGNはあくまでも電話(+映像?)が主。インターネットは従。
  - 最善(Best effort)から次善へ
  - (品質制御可能な)SNIは今のところSIP前提

IPインフラの価値とは

- 誰が提供しても同じ？
  - 単なる、A地点からB地点へのコネクティヴィティ
  - コモディティ
  - 価格勝負

- 本当にそうか ???
  - 下位レイヤは、上位から見れば抽象化されるのは当然
  - 実装、設計・運用技術が成熟したとは言えない
  - 複雑系、フラクタルな世界

IP Transportをコモディティ化させたくない…。

モジュールとしてのIP Transport

- **Transportを「部品」として切り出す**
- **特性、性能指標、品質指標を明示化する**

ことにより、

- Transport自体の差別化、高付加価値化
- さまざまなサービスを実現するenablerとしての形態

が可能にならないだろうか。

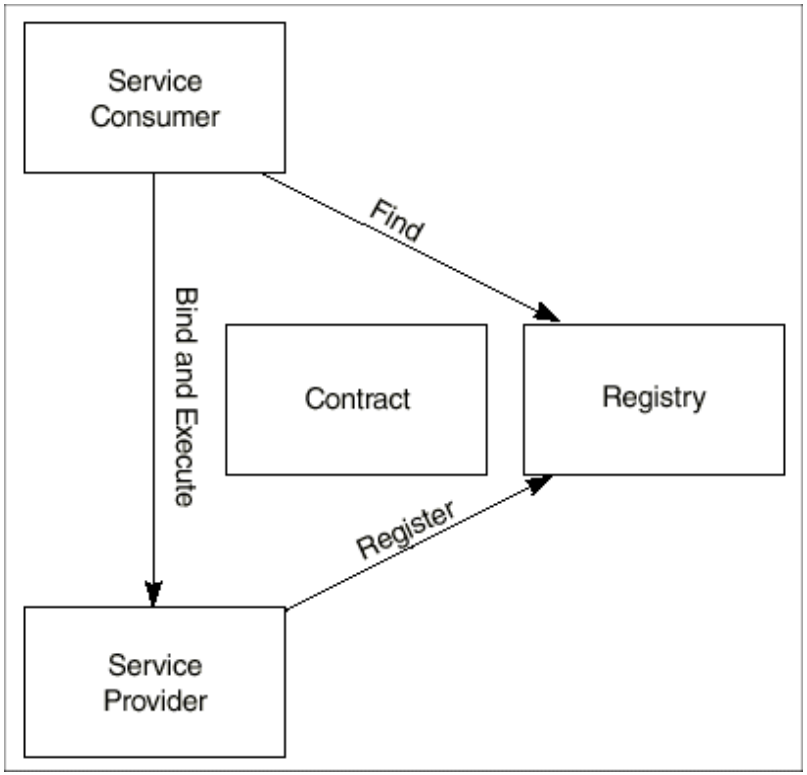
モジュールとしてのIP Transport

http://en.wikipedia.org/wiki/First-class_object

- **being expressible as an anonymous literal value**
- **being storable in variables**
- **being storable in data structures**
- **having an intrinsic identity (independent of any given name)**
- **being comparable for equality with other entities**
- **being passable as a parameter to a procedure/function**
- **being returnable as the result of a procedure/function**
- **being constructable at runtime**
- **being printable**
- **being readable**
- **being transmissible among distributed processes**
- **being storable outside running processes**

モジュールとしてのIP Transport

EmbeddedであったTransportをモジュールとして捉える



- 再利用性
- 可視化
- 高付加価値化

Service Oriented Architecture
http://www.developer.com/java/web/article.php/10935_2207371_1

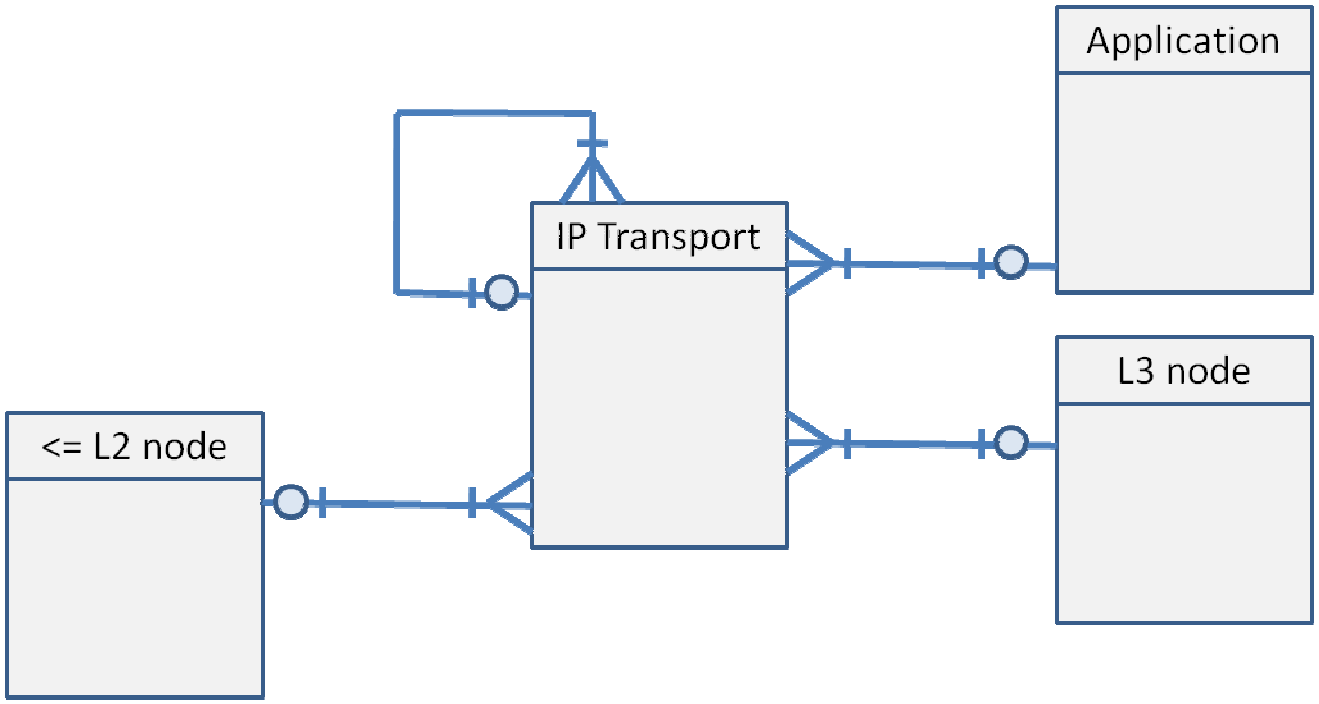
IP Transportとは何か

- ここではTransportを次のように定義してみる。
  - IPシステムを形成する各エンティティ(コンピュータ、通信ノード)間をつなぐ通信路
  - Transportも、またひとつのエンティティ

!= OSI参照モデルのTransport Layer

Transportを明示化、定義する

Transportの概念データモデル (Entity Relationship Diagram)

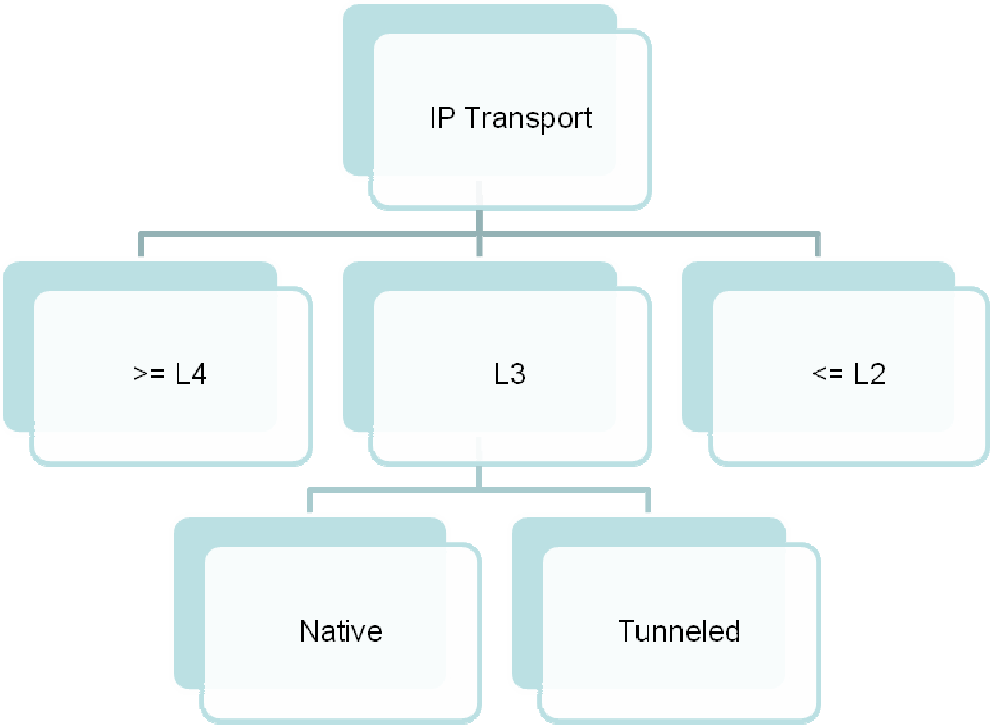


http://www.utexas.edu/its/windows/database/datamodeling/dm/erintro.html

Transportを明示化、定義する

- **IPシステムは、次のエンティティにより構成さされる。**
  - IP Transport
  - Application
  - L3 node
  - <= L2 node
- **IP Transportは、アプリケーション、または通信ノード(L3 または≦L2)により使用される。**
- **IP Transportはそれ自体が階層性を持つことがあり、再帰的に使用　される。**

Transportのサブクラス



>= L4		TCP, UDP, SCTP, RTP, TLS/SSL
L3	Native	Native IP Path
	Tunneled	MPLS LSP, IP in IP, GRE, IP sec
<= 2		PseudoWire, GMPLS path, PBT, T-MPLS

Transportの属性(例)

属性名	必須か	単一値か	ユニークか	備考
形態 （mp2p/p2p/p2mp/mp2mp, uni/bi-directional）	N	Y	N	
Source IPアドレス	N	Y	Y（L3サブクラス） N（L4以上サブクラス）	L3もしくはL4以上サブクラスのみ
Destination IPアドレス	Y	Y	Y（L3サブクラス） N（L4以上サブクラス）	L3もしくはL4以上サブクラスのみ
Source Port番号	N	Y	Y	L4以上サブクラスのみ
Destination Port番号	Y	Y	Y	L4以上サブクラスのみ
Transport Protocol ID	Y	Y	N	L4以上サブクラスのみ
Address Family （IPv4, IPv6）	Y	Y	N	L3もしくはL4以上サブクラスのみ
L2 ID （MAC, vlan, FR DLCI, ATM VPI/VCI..）	Y	Y	N	L2以下サブクラスのみ
Encapsulation (Native, IP in IP, GRE, MPLS, etc.)	N	Y	N	L3もしくはL2以下サブクラスのみ
経由Path	N	N	N	
Encryption	N	N	N	
帯域 – bps （最低、最大、平均）	N	N	N	
スループット – pps（最低、最大、平均）	N	N	N	
遅延	N	N	N	
遅延揺らぎ	N	N	N	
MTBF	N	N	N	
MTTR	N	N	N	
パケットロス率	N	N	N	

Transportの属性

- 必須かつ単一値を取る属性が、そのトランスポートを定義する。

- ユニークな値を取る属性が、他のエンティティから参照される際のキーとなる。

  (例)

  - **アプリケーションは、Dst IP, Port番号, Protocol IDにより、≧L4トランスポートを特定する。(実際は、ここにURI等のマッピングサーヴィスが介在する。)**

  - **NGNにおけるRACF(*)は、Service Stratumから、5 tuple (Src/Dst IP, Src/Dst port, Protocol ID)により、Transport Stratumを制御する。**

    (*) Resource Admission Control Function

Transportの属性

- 性能やAvailabilityに関する属性は、該当トランスポートの性質や品質を表す。
- この属性値を参照することにより、各エンティティは、トランスポートを使い分けることができる可能性がある。
- 理論値（構成により演繹される値）、または実績値（モニタリングにより計測した値）を入れるかは、運用方針に委ねられる。

IP Transportをどう使うか

- XML/Netconf等による動的provisioning
- Trafficの選択的マッピング方法
  - Prefixでわける
  - ToS/DSCP/expでわける
  - PBR (Policy Based Routing) / FBF (Filter Based Forwarding)
    - Flow(5-tuple)
    - BGP community attribute

課題

- ApplicationがIP Transportを使い分ける？
  - URIを使う
    - http: vs. https:のように
    - その上で、PBR/FBFを用いてマッピングする
  - ApplicationがToS/DSCPを付ける
  - ApplicationがIP addressを使い分ける?!
  - 3つの課題
    1. Scalability
    2. Inter-domain coordination
    3. 統計情報、課金…

1. Scalability

- Prefix分けたら、さらなるRIB/FIB増加問題
- PBR/FBFのためのclassifier　(#access control list)

2. Inter-domain Coordination

- **Provider間での取り決めが必須**
  - ToS/DSCPの扱い
  - Policy Mapping (for PBR/FBF)
  - PrefixやBGP community等の扱い
  - Inter-AS TEの扱い
  - Policy Server連携

- **差別化と協調をどう両立させるか**
  - キャリア連合？Star Alliance/One Worldのような

3. 統計情報・課金…

- 付加価値を価格に反映させることは必要
- 但し、できるだけシンプルに

Thank you !