NW設計の基本事項

吉田友哉
NTTコミュニケーションズ（株）

はじめに

- 大規模ネットワーク設計の基本
- 設計の際に考慮すべきポイント
- トポロジー県単位等、設計の肝となる要素の考え方
- アドレス設計や冗長設計
- その他、これは押さえておくべきというポイント

ISPネットワーク設計ポイント１

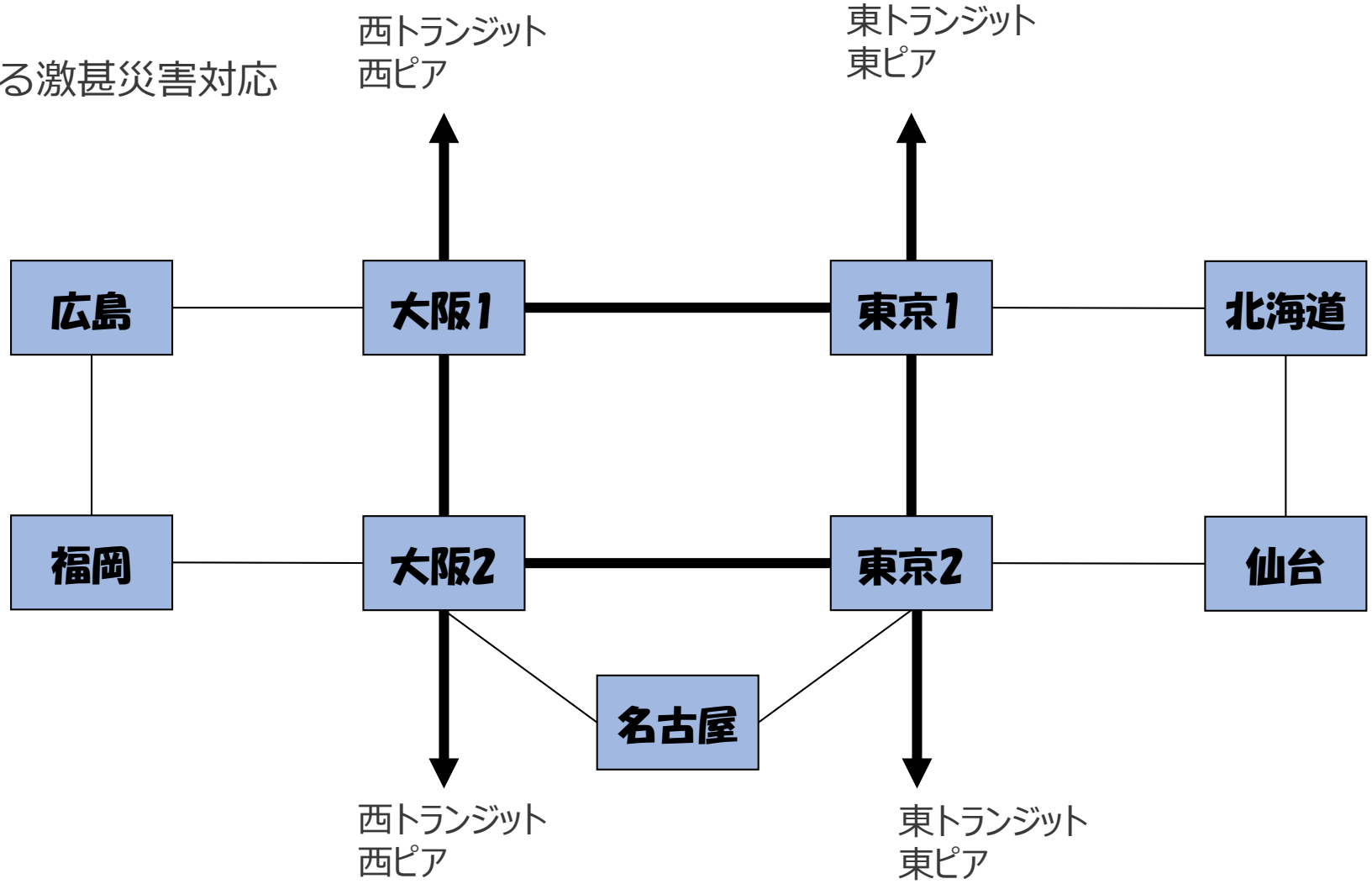
ネットワークを流れるトラフィックをどうさばくか
→ 必要帯域（ピーク時のトラフィック）の確保と論理設計

- 各POPのトラフィック
  - 地方POPのトラフィックは、東京や大阪のメインPOPに接続。あらかじめ設定してある迂回路にて救済
  - そもそもどこがPOPか？トラフィックの多い地域？定義はまちまち
- 国内ISPとのトラフィック交換
  - トラフィック容量の大きなISPとはPrivatePeerを活用、もしくはIXの積極利用
  - Transitで残りのトラフィックをさばく
- 海外トランジット
  - 国内Transitで賄うケースや複数の上流と接続しうまく使い分けるなど
  - コストの安い上流をメインとし、切れた場合には他に回す等
- 2重故障もある程度考慮にいれて設計するのが望ましい
  - 冗長をとっている接続先のトポロジーに依存

必要な接続性を確保し、トラフィック制御を実施する

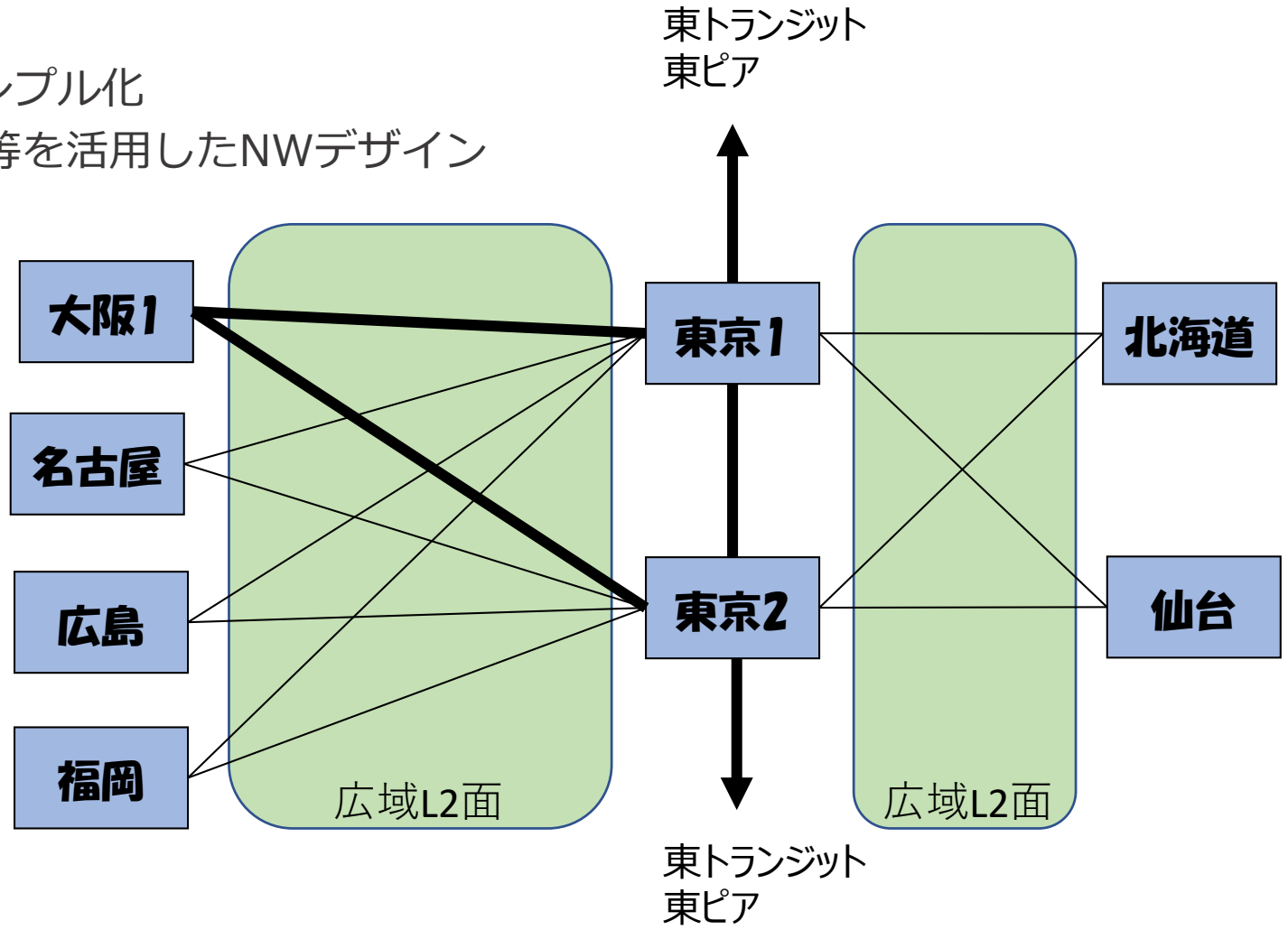
ネットワークトポロジー　例1

- 東西分散モデル
  - 東西分散による激甚災害対応



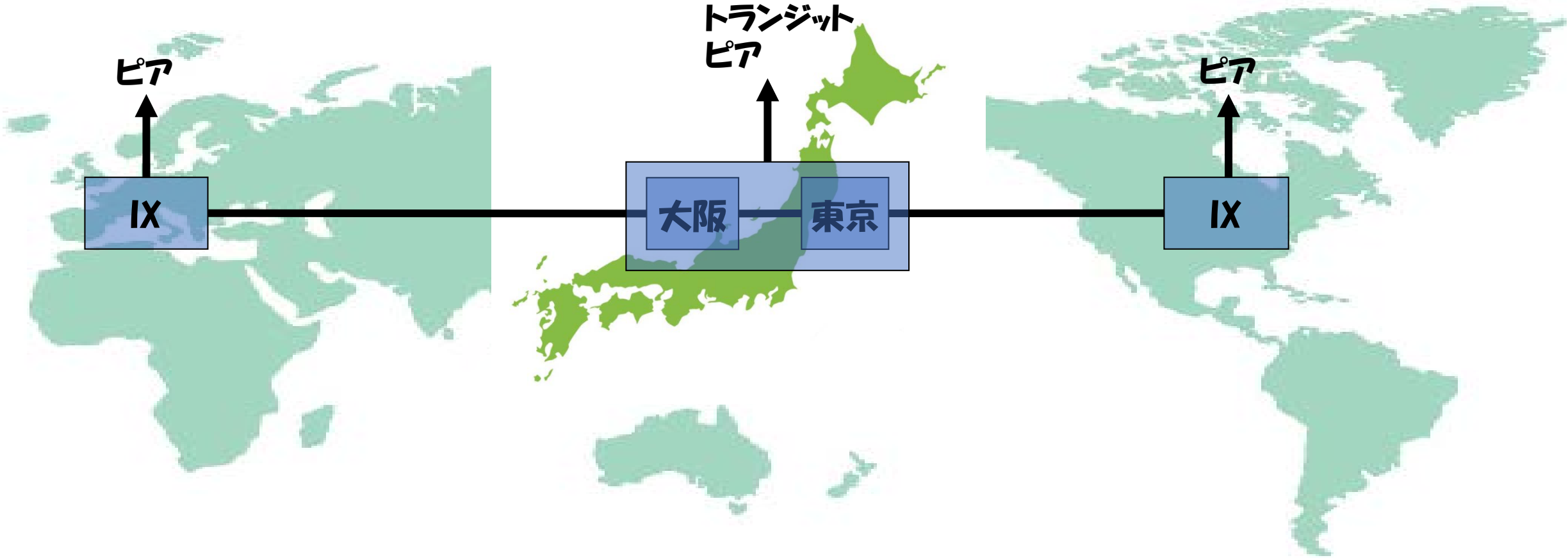
ネットワークトポロジー　例2

- 東京集約モデル
  - コスト追求、シンプル化
  - 広域L2サービス等を活用したNWデザイン



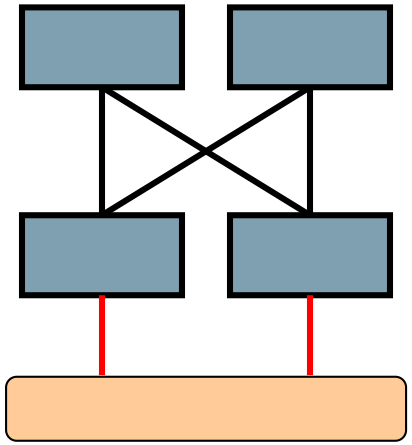
ネットワークトポロジー　例3

- グローバル展開モデル
  - トランジットの相対的な位置付けが低下
  - 海外IX等も活用してピア先を拡充



ネットワーク設計（基本）

- 信頼性（冗長性の確保）
  - 装置（ノード）、リンクレベルの冗長化、負荷分散
  - 同機能相当の装置はなるべく分散配備を念頭に
  - 電源系統の分散
  - ファイバー経路の異経路分散



- 品質
  - 装置単体、装置間における品質の確保（論理設計、最適なハードの選定）
  - 必要帯域の定義（コストと品質を勘案し最適な容量を確保）

- 運用性
  - シンプル性（網設計の一貫性、 HOP数の削減、パターン化）
  - トラブル対応時の解析性（NW/装置状態を監視、管理）

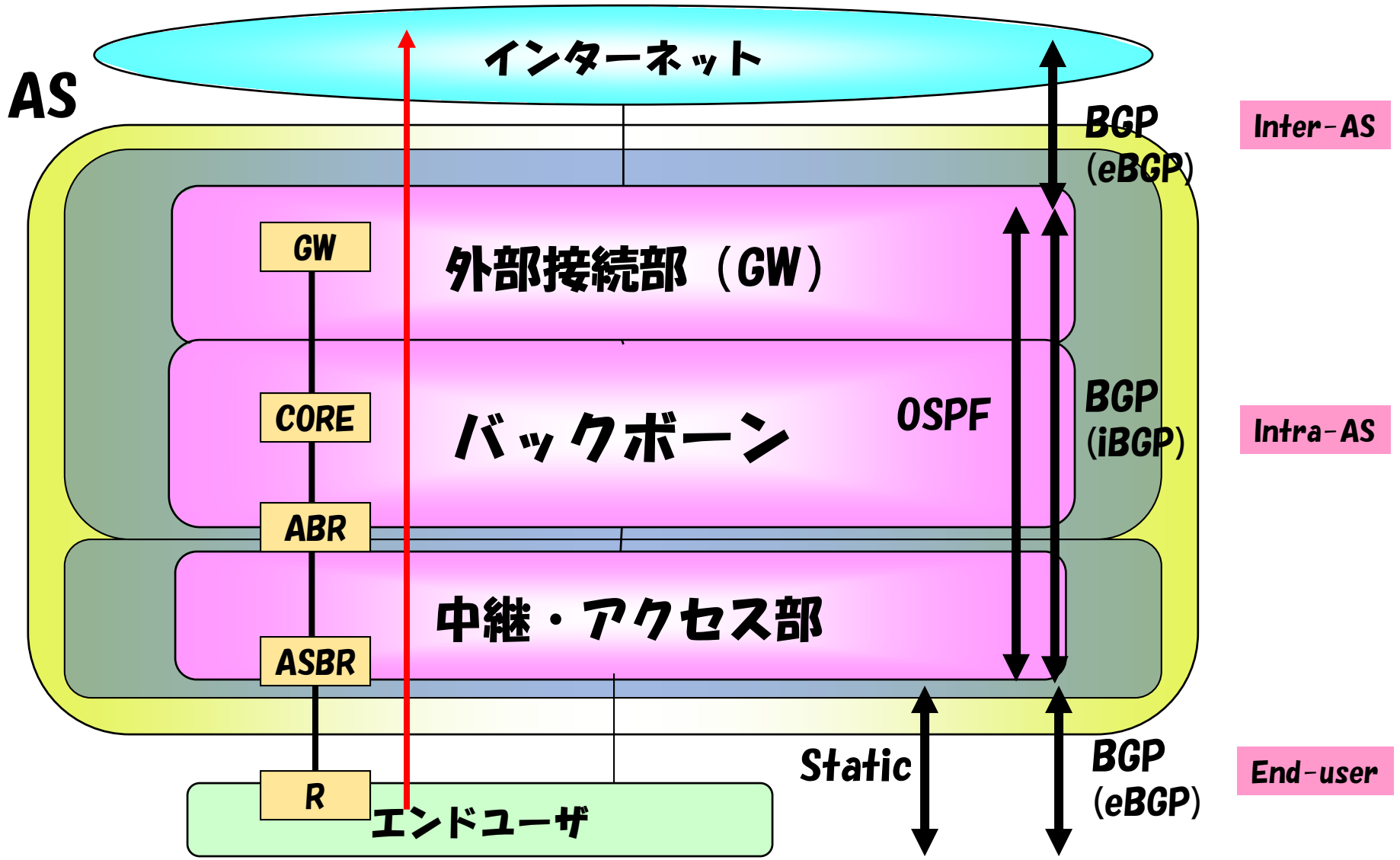
- 将来性・拡張性
  - 新たなサービスやネットワークの更改（End of Life）等に対応可能なネットワーク

ネットワークの規模・階層的構造

- 中規模・大規模なISPネットワーク
  - 物理ネットワーク
    - 外部から複数の上流経路を受信し、国内のピアも十数以上
    - GWは複数台、それぞれeBGPピア接続
    - 主な地域はPOPになっている
    - COREルータや境界ルータは基本は2重化構成
  - 論理ネットワーク
    - 内部のTopology管理はOSPF、経路情報の管理はBGP（OSPF）

- 階層的構造に沿ったルーティングの設計
  - AS間 [eBGP]　　inter-AS
  - AS内 [OSPF/iBGP]
    - 外部接続部（GW）　　intra-AS
    - バックボーン
    - 中継・アクセス部
  - エンドユーザ[static/eBGP]　　End-user

階層ルーティングネットワーク全体イメージ



トポロジー情報・経路情報

- トポロジー情報（ネットワークの地図情報）：OSPF
  - バックボーン全体のリンクのつながりを表す情報
  - OSPFのリンクステートデータベース（トポロジカルデータベース）に格納
    - 隣接とLSAを交換し、トポロジカルデータベースを作成

- 経路情報：BGP
  - ユーザの経路情報
    - PAアドレス、上流ISPからの経路情報（フルルート/トランジット経路）
  - 基本はBGPで交換
    - 最近は経路集成はあまり考えなくても大丈夫
  - 以下の場合にはOSPFも有効
    - ユーザ経路を簡単にロードバランスさせたい場合
      （protocol-nexthopの解決をOSPFで）

アドレス設計

- 基本方針
  - 使用目的別にアドレスを区分け（サービス毎、設備、お客様割り当て等）
  - 各アドレスの対外広報や到達性を考慮（セキュリティー観点）
  - なるべくなら経路集成可能な設計で（IPv4では難しかったがIPv6なら計画的にできる）

    - 例えば以下ように分類

      （１）バックボーンアドレス
      - ループバックアドレス
      - POP間アドレス、バックボーンスイッチセグメントブロック

      （２）ユーザアドレス
      - ユーザが実際に利用するブロック

      （３）外部接続アドレス
      - GWなどで外部と接続する部分のアドレス（実際は (2) に含める）

    - セキュリティーの観点
      - Telnet、SSHなどのリモートアクセス範囲の明確化
      - 経路広告の範囲の明確化（DOS対策など）

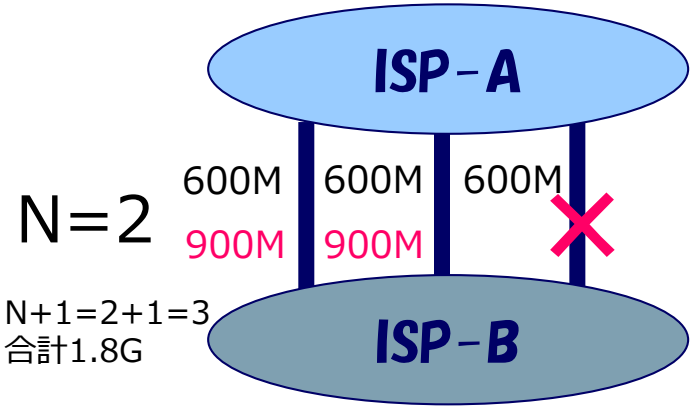
アドレス設計（IPv4の例）

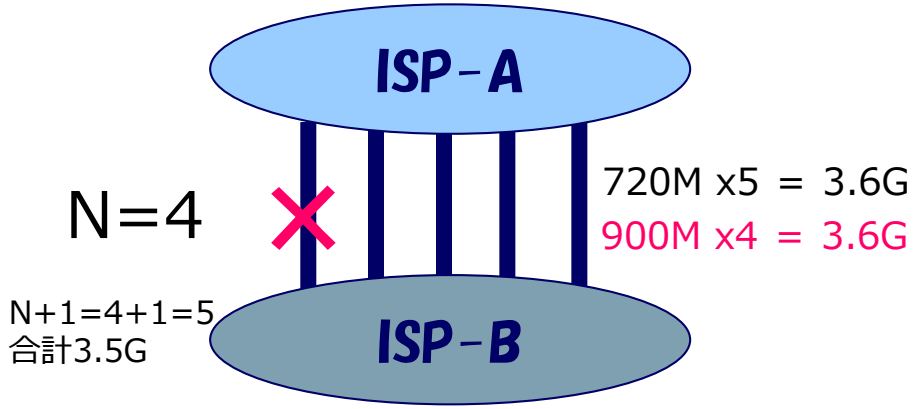
- PAアドレスを使用用途に応じて分類し明確化したアドレス運用、対外広報は必要最小限な経路に限る
- NW装置へのアクセスは、セキュリティーフィルタ設計を厳格に実施し、アクセス元を最小限とする

分類	用途	割当例	対外広報	自網装置へのアクセス許可
（1）バックボーンアドレス	ループバックアドレス スイッチセグメント POP間/POP内セグメント	/32 /26、/27等 /30等	不要（実際は広告）	許可 （更に特定の保守セグメントからのみアクセスを許容するべき）
（2）ユーザアドレス	お客様に払い出すアドレス（PAアドレス）	/24～/29等	必要（集約経路を広報）	拒否
	お客様が持ち込まれたアドレス（PIアドレス）	/16～/24等	必要	拒否
（3）外部接続アドレス	対外事業者との接続部で使用するアドレス	/30	不要（実際は広告）	拒否

N+1設計

- 実際に流れている利用帯域「N」に「＋1」の「N＋1」回線を用意し必要帯域確保
  - １Ｇ～２Ｇの場合　必要帯域　N＝２　⇒　２＋１＝３本で設計
  - ３Ｇ～４Ｇの場合　必要帯域　N＝４　⇒　４＋１＝５本で設計





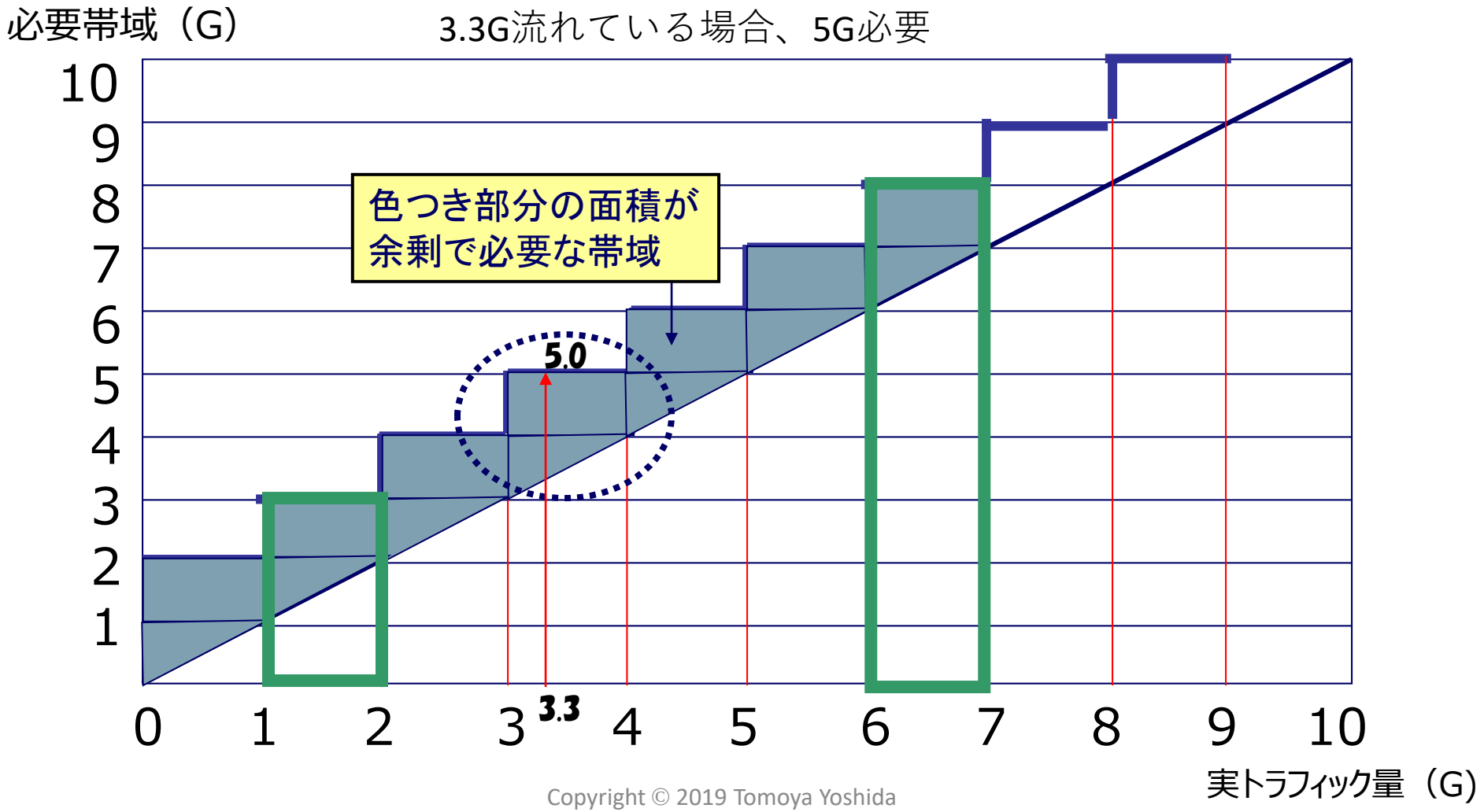
2GE相当のトラフィックに対して、3GEの容量を確保する必要がある
→ 3GEは、2GEの1.5倍の量に相当する

4GE相当のトラフィックに対して、5GEの容量を確保する必要がある
→ 5GEは、4GEの1.25倍の量に相当する

・トラフィック量の増加に伴い、束ねる回線数が増えれば有効利用が可能（バランシング注意）
・接続方法としては、eBGP multipath接続やLAG接続

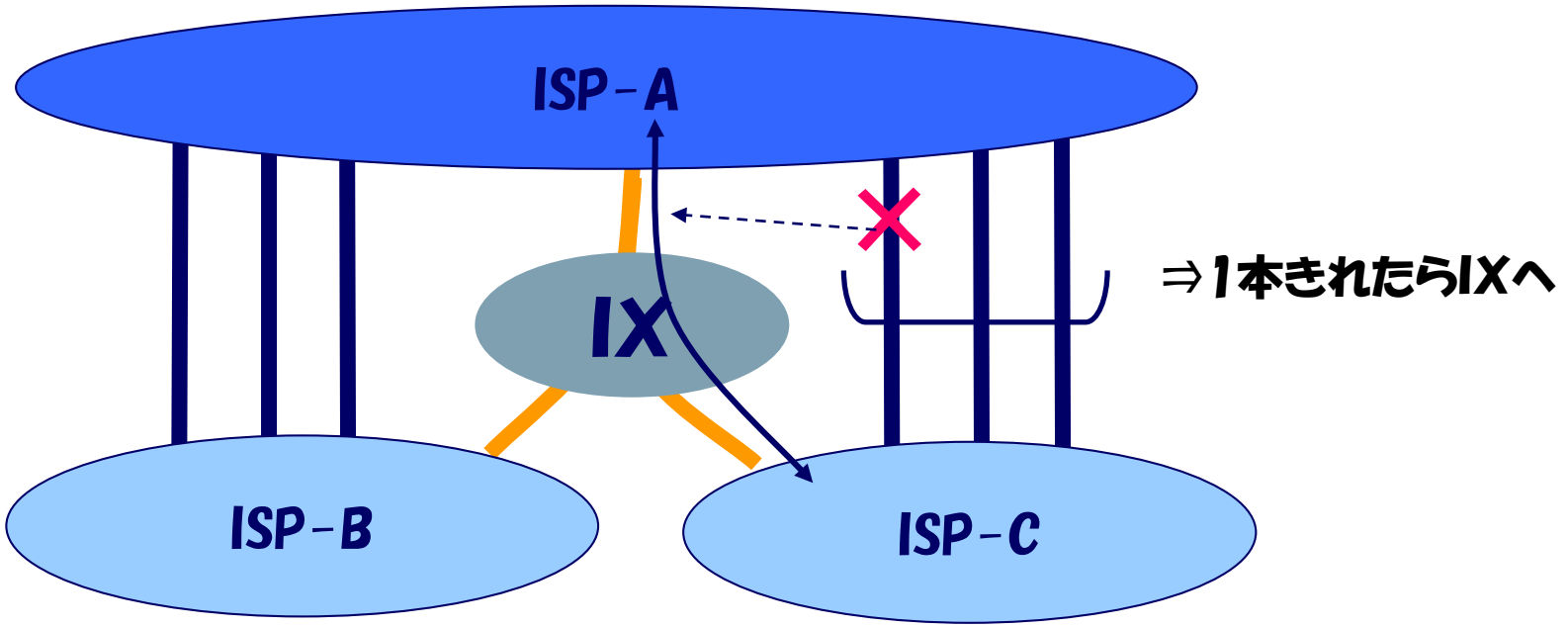
N+1設計 1Gの場合（10G/100Gでも考え方は同じ）

メリット：　実トラフィック量が増えるほど、効率的に回線が利用可
デメリット：増設ポイントが多い、トラフィック分散設計が大変



回線設計の応用

- ＩＸ（Internet Exchange）の回線等を利用し、メイン回線をフルに利用
  - ISP-Aが ISP-B, ISP-C と共にIXで接続していた場合



それぞれ＋１本用意する必要がないので回線の有効利用が見込める

需要予測と回線増設

- 過去から現在までのトラフィック量の伸びのデータをもとに、将来の需要を予測し、プロットした結果を線で結んでみる

- その上で、どの時期までにどのぐらいの帯域を必要とするかを判断
  - トレンド予測（バックボーン区間は特にトレンド予測が効果的）
  - エッジでは突発的な需要（大規模なお客様の新規収容等のイベント）を把握した設備確保が必要

- 実際に回線やファイバーを調達する時間を見込んで、最終的にいつまでに増設の判断をして行動に移さなければならないのか
  メディアの変更を考えるべきなのか（10GE x N本 ⇒ 100GE）の判断等

ハードウェアスペックについて

ルーティングエンジンの性能がNWの安定性、拡張性に直結
OSPFやBGPの経路数、ピア数を実網の負荷プラスαの（数年後を想定した）経路を処理できるか？
NW装置のスペックは机上検討に加えて、実機検証で評価するのが望ましい

- ルーティングエンジン（性能が高ければ、それに越したことはない）
  - 高性能であれば、大規模な経路フラップ、経路事故の発生下でもサービス提供を維持可能
  - ルーティングエンジンの冗長化とNSR（Non-Stop-Routing）：故障時のサービス無停止が可能
  - 高速な経路収束等もメリットあり：IGPの切り替え時間の差でピンポンになることもある
  - OSイメージやsyslog等を十分に保管できるストレージ （OS容量が多くて交換するケースも）
- フォワーディング
  - ラインレート転送能力の確認、自社のサービス仕様を満たせるスペックを選定
  - IF/SWカードは高額。 ポート数はNW規模と今後の需要を予測して適切なモデルを選定
    エッジで利用する場合は、アップグレード等も関連してくる
- 共通部：シャーシ、電源、FAN等
  - ハードウェアの交換作業が、通信に影響を与えない事は必須条件
  - 最適なスロット数のシャーシを選定（ラインカード分散は耐障害性を向上）

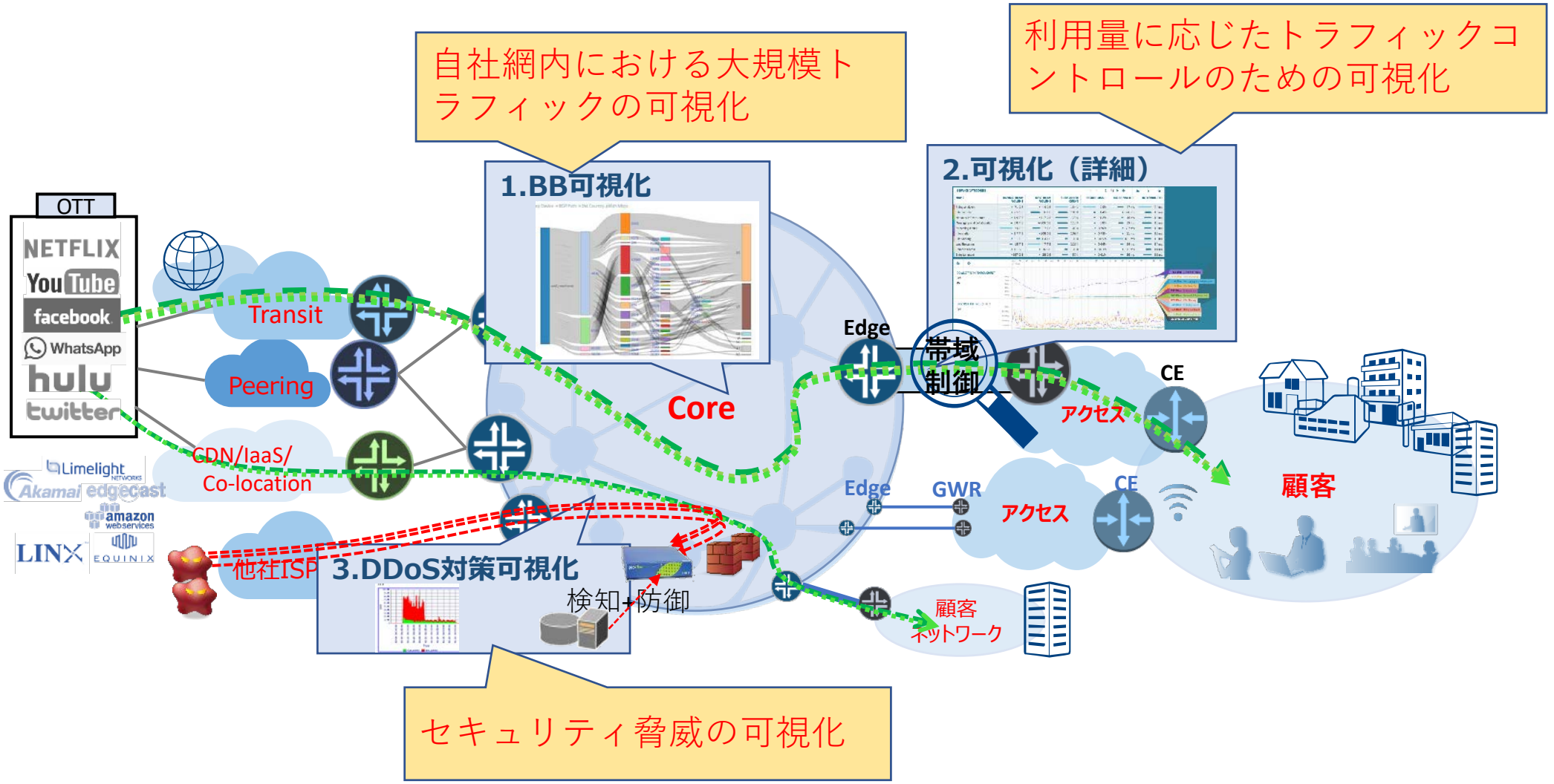
監視・可視化

- 目的
  - ヘルスチェック、故障検知、故障対応
  - 設備投資の判断、効率的なNW設計

- 使用される技術
  - 死活監視： ping、traceroute等
  - ログ監視： syslog等
  - トラフィック監視：SNMP（mrtg、cacti等）、flow
  - 経路監視（routeview）
  - 新しいトレンド： telemetry、複数装置（Kentik、Deepfield等）
  - それらの組み合わせ、複合監視

可視化のポイント

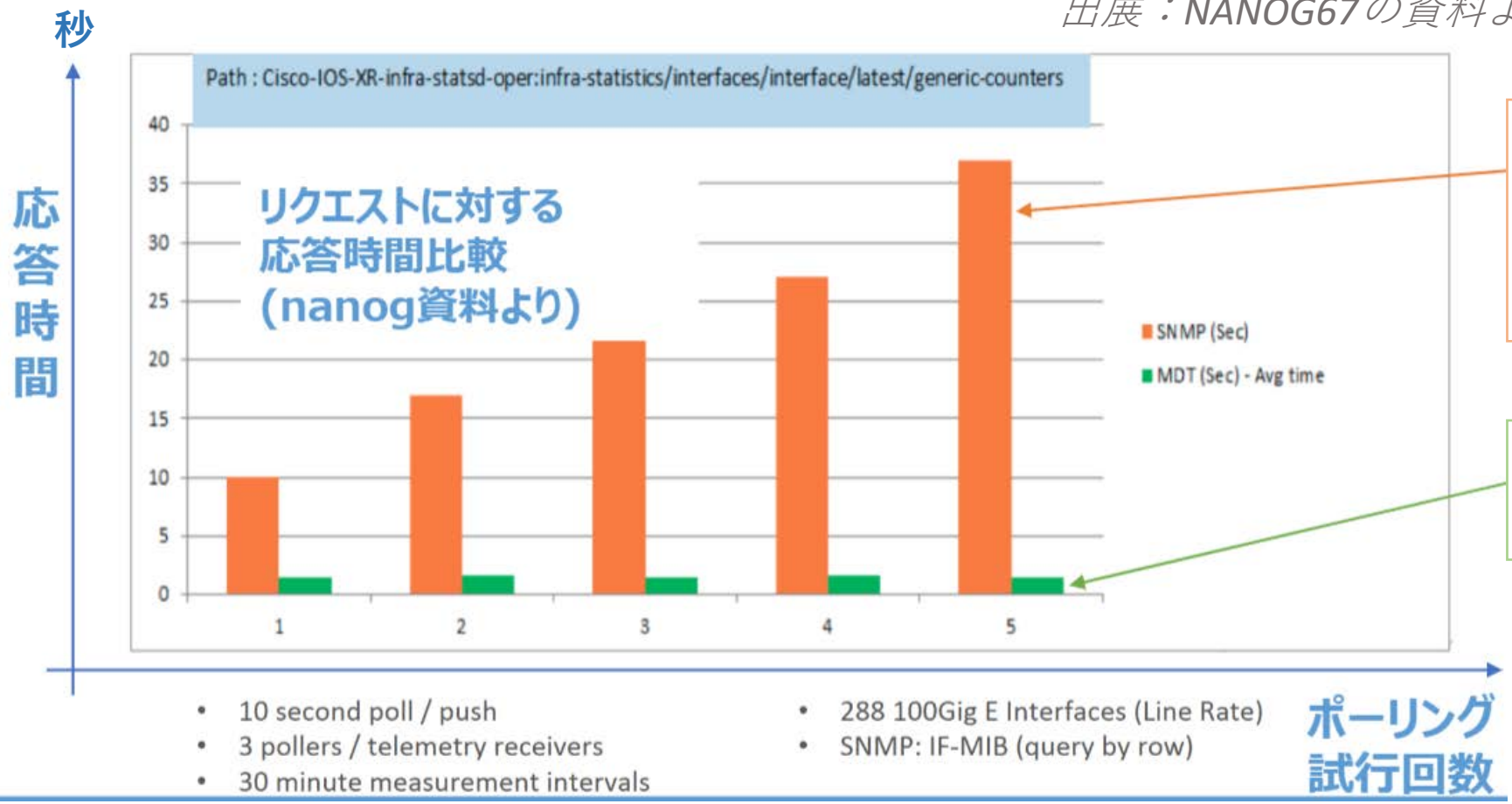
- 目的に合わせた様々な監視方法と関連技術の組み合わせ



SNMP and Telemetry

- SNMPのレガシーテクノロジーと次世代NWのTelemetry技術の併用が今後積極導入されていく想定

出展：NANOG67の資料より

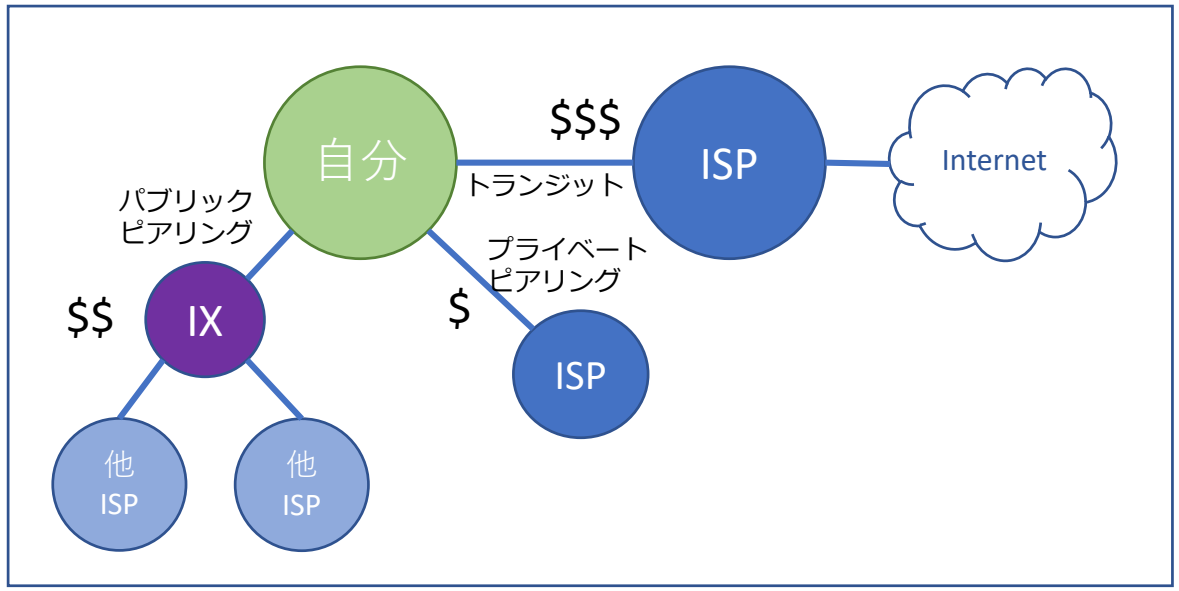


- レガシープロトコル
  - SNMPv1 1988年
  - SNMPv3 1999年
- 装置負荷が高い

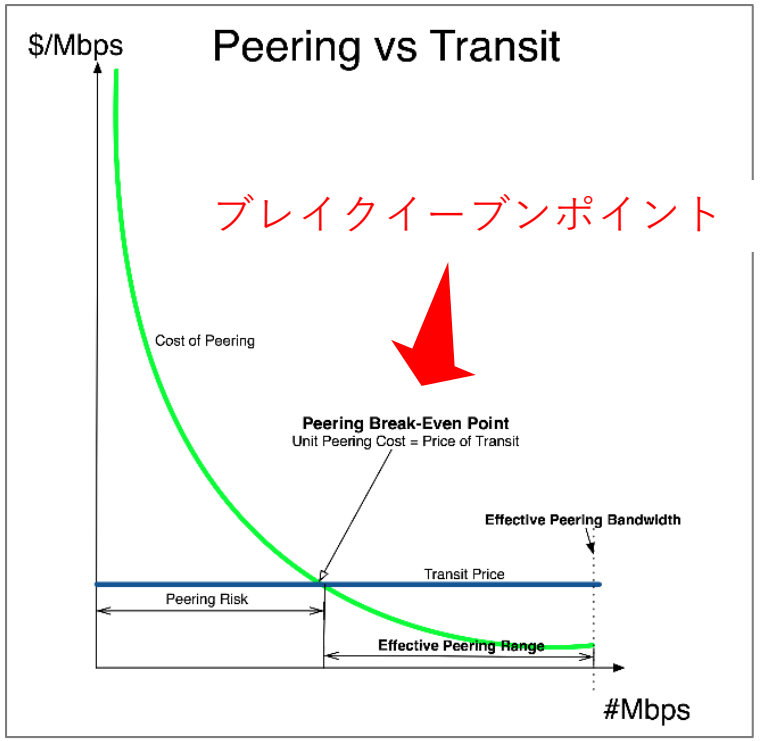
- リアルタイム性が高い
- 大規模にスケール

トランジット vs ピアリング

ピアリング & トランジット



ソース：*drpeering.net*

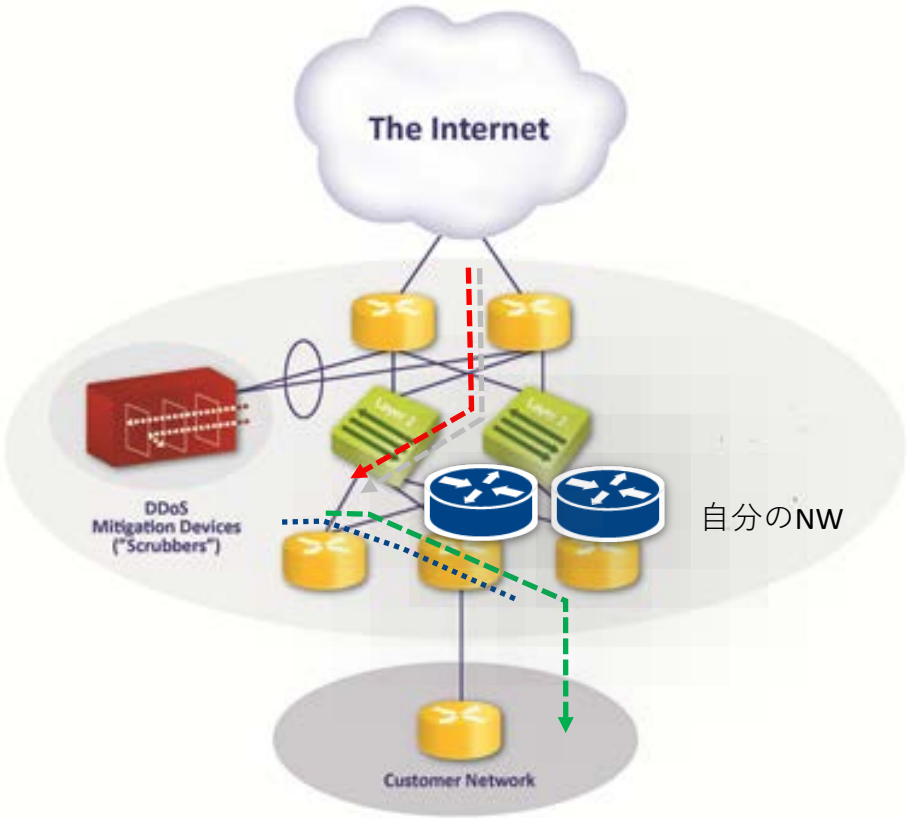


- 全Internet接続をTransitで賄う
- ピアリングで選択的な接続先を選定
- コストイメージは＄の数
- 運用コストも考慮する必要あり

→バランスが重要

セキュリティー

- セキュリティー
  - DDoSの攻撃ボリュームとパターンが増え続けている
  - 何を守る：
    - ・エンドユーザ
    - ・自分自身のNW
  - どうやって守る：
    - ・自分自身でDDoS緩和装置を導入
    - ・上流ISPからサービスを購入



- DDoSアタックからのISPバックボーン保護
  - ASボーダにおいて、典型的な攻撃は水際で止める（RTBHやフィルタの活用）
  - バックボーン区間のローバラ設計により帯域をフル活用でき、DDOS攻撃時の通信影響を極小化

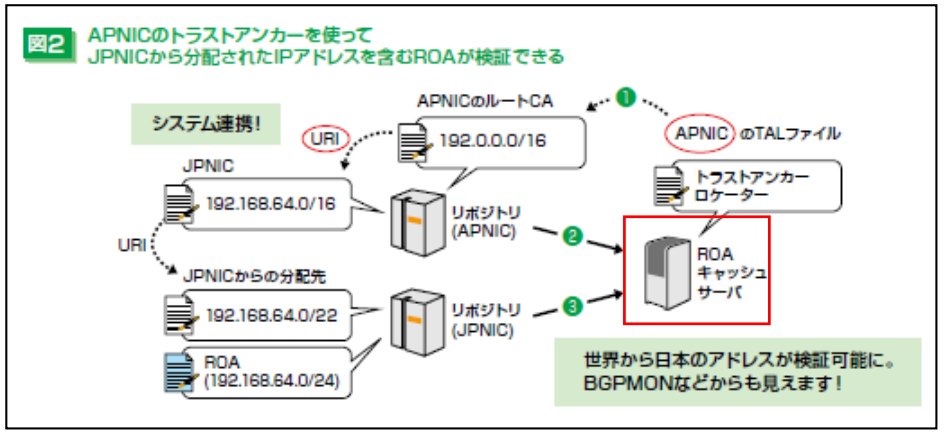
RPKI

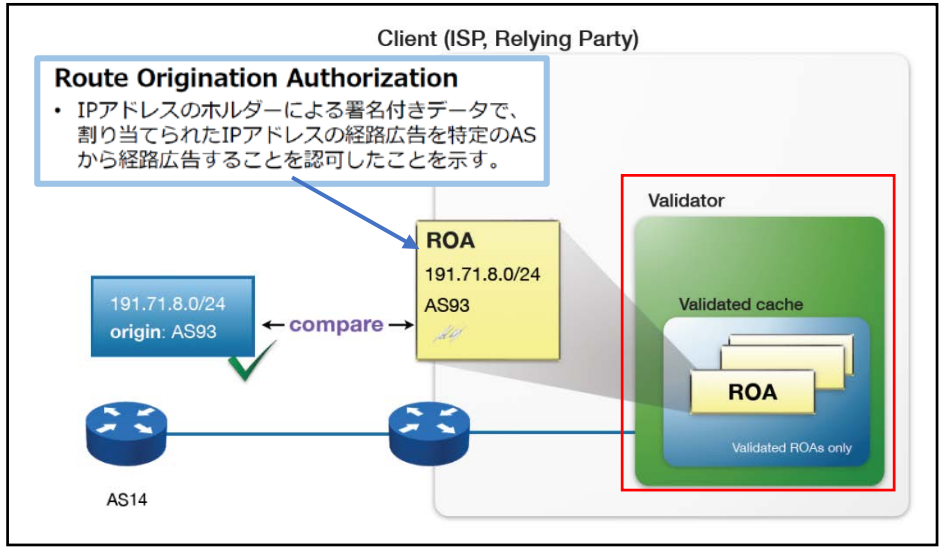
- 解決する課題
  - BGP経路ハイジャック（意図的）
  - BGPルートリーク（オペミス）

セキュアルーティングが必要

- Resource Public-Key Infrastructure
  - IPアドレスやAS番号といった番号資源（Number Resource）の割り振り／割り当てを証明するPKI
  - 1997年頃、Stephen Kent氏（BBN Technologies）によって提案され、IETFで仕様策定が行われている
  - https://www.mfeed.ad.jp/rpki/

ソース:JPNIC





その他トピック

- IRRの活用
  - BGPフィルタの自動化生成により、設定処理を省き、効率的な運用が可能
  - https://www.nic.ad.jp/ja/ip/irr/index.html
- キャッシュの導入
  - メリット： サービス品質向上、設備投資削減の期待
  - デメリット： CDN事業者への依存度、帯域設計の難しさ（故障時の余剰帯域設計等）
- IPアドレスの重複は危険
  - とりわけRouter-IDとなるループバックアドレスの管理は慎重に
  - アドレス管理データベースや、ping/tracerouteによる地道な確認で防止を
- MTU設計
  - L2サービスでJumbo Frame転送のためにMTU値を拡張するケースがあるが、
    L3サービス、対外接続(L3)ではデフォルトのMTU1500Byteとなっているかは要注意
    MTUミスマッチが起きるとマズイ（意図せずMTU値が異なってしまうケースあり、IP MTUで指定）
  - 自網内の場合、MTUミスマッチでOSPFネイバーがUPしないので気付きやすいが、
    対外接続部では気付き難く、お客様申告で表面化する場合あり

弊社エンジニア（なんでもお答えします）

Akio Ootaki（大滝亜紀夫）
a.ootaki[at]ntt.com

Huubach Nguyen（グエンホウバッ）
h.nguyen[at]ntt.com