

IP Multicast Deployment

KDDI(株) 加藤 利雄
ti-katou@kddi.com

Agenda

- IP Multicastで実現できるサービスは？
- IP Multicastインフラを構築するには？
- IP Multicastインフラの耐障害性は？
- 今後の課題

IP Multicastを使うサービスとは？

- 「チャンネル」という一種のデータプレーンでの利用形態に差
- PtoMP
 - 送信元が1箇所、複数の「特定」受信者に対して一斉配信
 - コンテンツ配信、プッシュ型メディアなどで利用
 - ブロードキャストではない
 - 「チャンネル」毎に別々の情報が流れ、受信側で選択
- MPtoP
 - 1箇所の受信元に対して、複数の送信者が存在
 - 送信元毎に、別々の「チャンネル」で情報を送信
 - 監視、データ収集などで利用
- MPtoMP
 - 1つの「チャンネル」内に、複数の送信者/受信者が存在
 - チャット、遠隔会議などで利用

IP Multicast技術で実現したサービス例

- “Triple-Play”サービス

- KDDI 光プラス

- FTTP (Fiber to The Premises), High-Speed VDSL, ...
- 放送 / 電話 / インターネットを同じインフラ上で提供
- 2003年10月よりサービス開始
- 放送サービスを提供するために、IP Multicast技術を採用

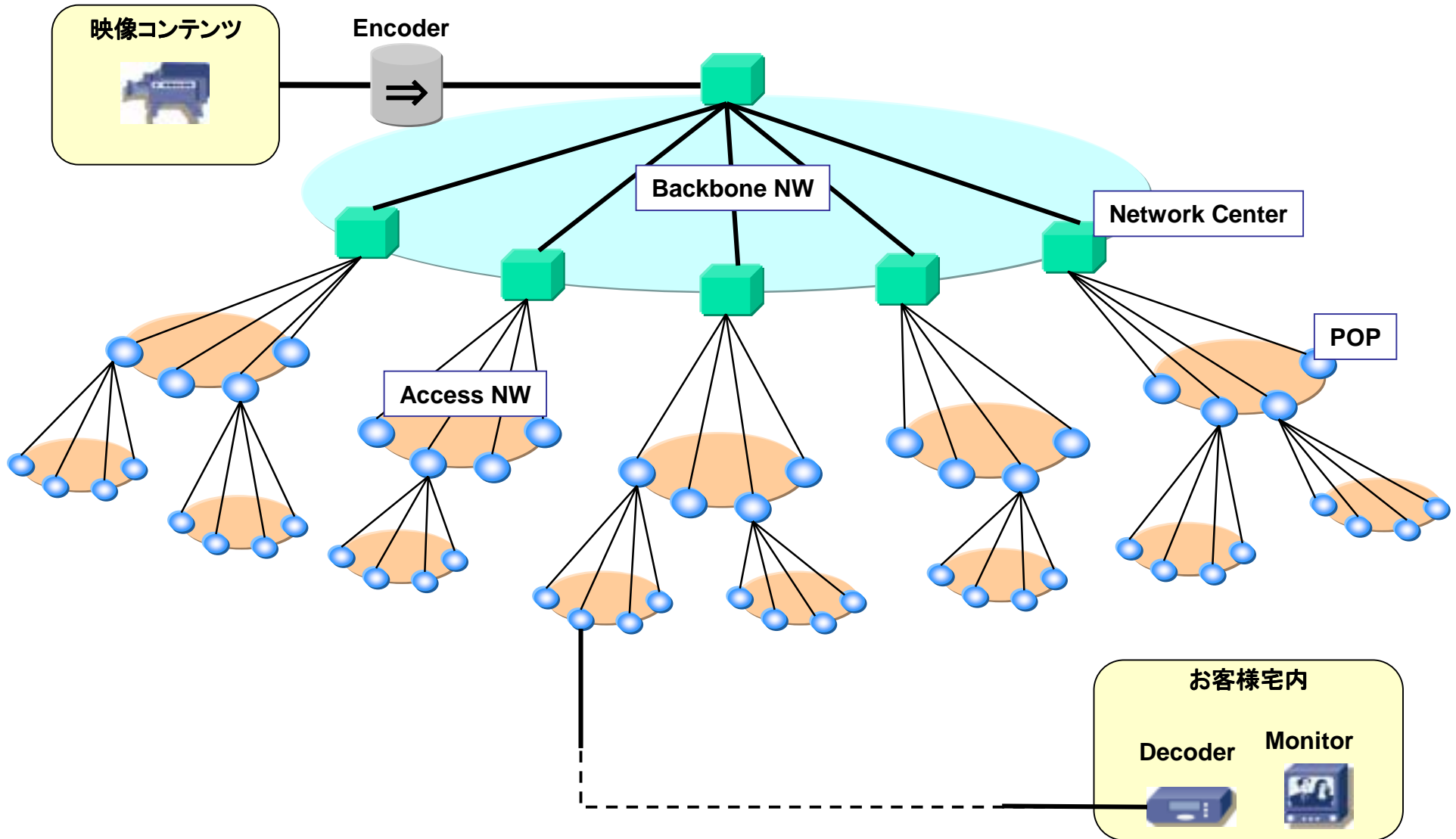


IP Multicast インフラを構築するには？

- 基本技術は既に存在
 - IPv4: PIM, IGMP, IGMP snooping, ...
 - IPv6: MLD, MLD snooping, ...
- ルータ自体の設定はシンプル
 - 機能を有効にすれば、とりあえず動かすことはできる
 - ex. PIM sparse mode, dense mode どちらを使用すべきか？
- インフラの基本設計
 - 「どこから」流したいのか？
 - 「どこに」流したいのか？
 - 「勝手に」流されないようにする
 - サービスインフラ外に流れないように注意

⇒ では、簡単に構築できるものなのか？

IP Multicast インフラ



全国規模のインフラを構築する難しさ

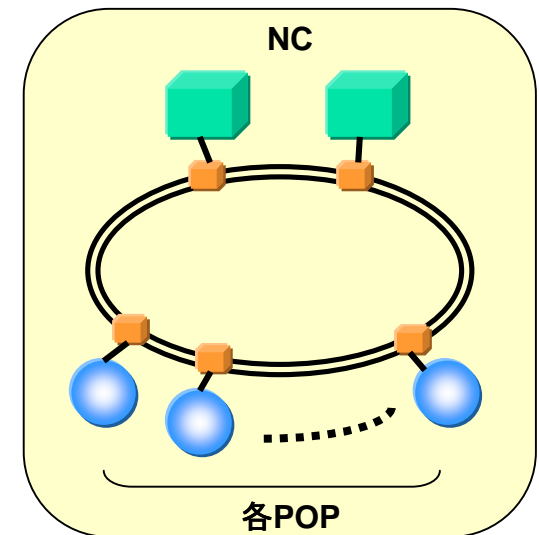
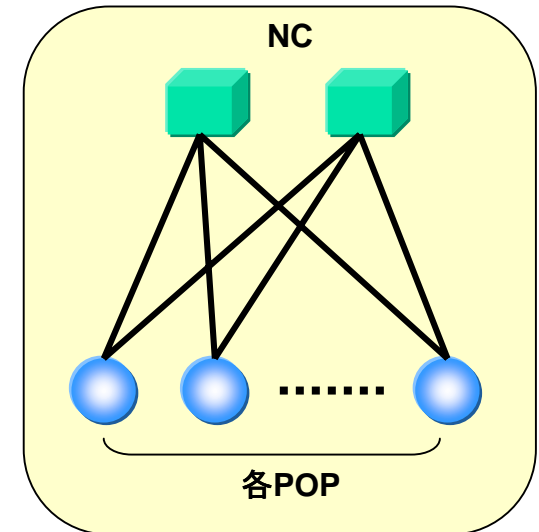
- ネットワークセンター(NC)間を結ぶバックボーン
 - 伝送ルート、伝送機器、伝送媒体も設計時点で選択可能
 - 異ルート化による信頼性向上
 - SONET/SDH オーバヘッドによる警報転送機能
 - Ethernet系I/Fでも警報転送機能のある機器を選択
 - トポロジーとしてPtoPで接続可能

⇒ バックボーン構築は、それ程ハードルは高くない？

- 国内で構築できるトポロジーは、地理上限定されてしまうのは止むを得ない
- どこに配信局を設置するか？
 - RPを起点にツリーが構築される
 - 共有木、SPTのルートをよく考える必要あり

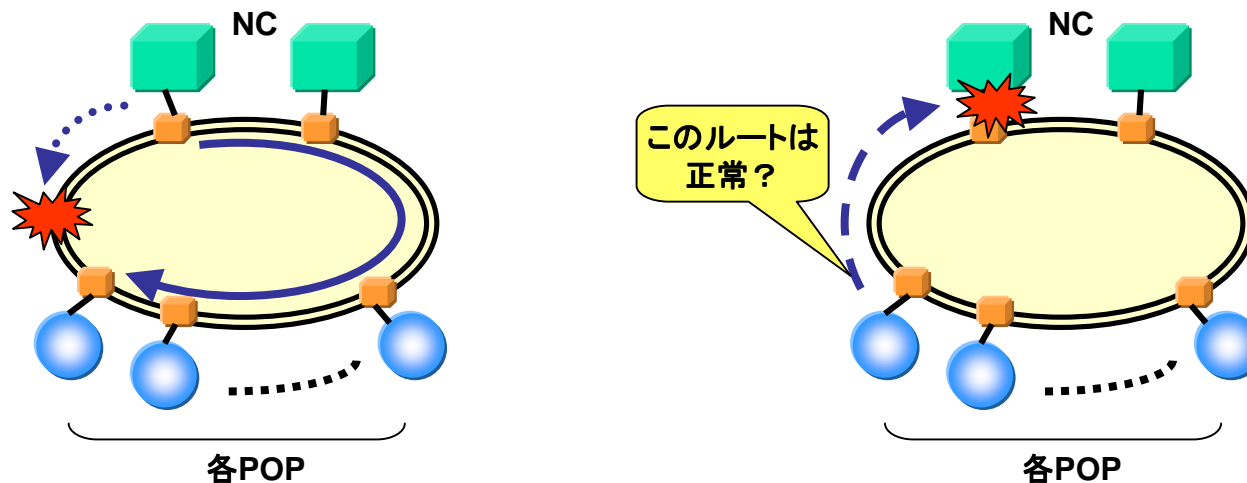
アクセスネットワークの構築の難しさ

- アクセスネットワーク
 - NCとユーザ収容ポイント(POP)を結ぶネットワーク
 - POPの数は、NC数と比べて格段に多い
 - PtoPでNC-POP間を全て接続することは困難
 - リングNWを構築するのが現実的 (構築せざるを得ない)
 - 設備面、コスト面、...
- バックボーンNWより、むしろ大規模に
 - 障害ポイントも飛躍的に増加
 - 対象POP数が増加すればするほど、深刻な問題に
 - 運用面での負荷 (監視対象数の増加、障害対応など)
- Layer2で構築する？ Layer3で構築する？
 - どこまでブロードキャストドメインを広げるか？



アクセスネットワークの構築の難しさ (cont.)

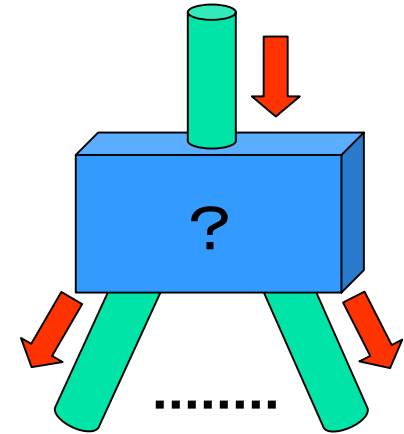
- リングネットワークの構築技術
 - ADMリング、RPRリング、その他...
 - トラフィックをPOP毎にコピーしなくて済む技術が必須
- リングネットワークでの問題点
 - ノード間のリンクプロテクションは得意 (50msec以下で復旧可)
 - リングノード⇔ルータ間の障害を、他ノードが知ることが出来ない
 - 上流のNW障害を、如何に早く下流側に伝えるかが重要
 - では、どうするか？ (ここが工夫するポイント)



ルータ選定の難しさ

- Multicast特有の問題

- 上流からのマルチキャストトラフィックは、ルータ内部で複数の下流I/Fに対して複製しなければならない
- 必要な複製能力
 - トラフィック量と複製先I/F数に比例



- ルータは、どこでパケットを「複製」するのか？

- あまり公開されない情報ですが...非常に重要です
- ルータ内部のアーキテクチャに大きく依存
 - ラインカード、ファブリック部分など
 - 内部バス、スイッチファブリック容量にも注意
- 複製ポイントが内部に複数あった方が効率的

IP Multicastインフラの耐障害性は？

- IP Multicastインフラでの障害発生ポイント
 - ノード単体障害 (冗長構成部)
 - ルータ自身のHA機能とMulticastの連携は十分か？
 - StatefulなCPU切替が出来れば、影響は最小限に抑えられる
 - リンク障害、ノード障害
 - 障害発生ポイントの周辺ルータが如何に早く、障害検出できるか
 - 伝送区間内でバックアップルートに切替えられれば、ルータ側は関与しない
 - 伝送区間で救済できなかった場合は、直ちに配信経路の再構築が必要
- 配信経路の決定方法
 - Unicastルーティング情報に依存する
 - コスト、アドレス配置などで、ある程度は調整可能
 - RPF (reverse path forwarding) check
 - RPF checkにより、「トラフィックを受け入れる」I/Fを決定

IP Multicast (PIM-SM)における経路切替方法

- RPF Neighbor変更によるDistribution Treeの再構築

(1) Unicast Routingに変更が発生

⇒ Unicast Routing収束

(2) Source/RPに関するRPF Neighborが変更

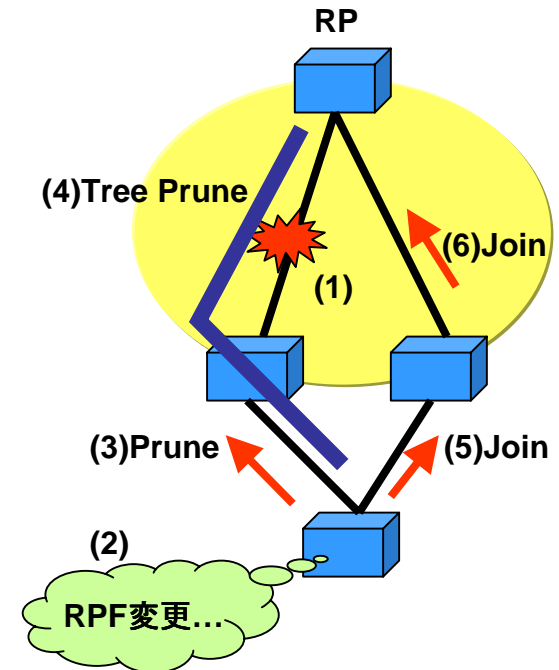
(3) 旧RPF Neighborに対してPrune送信

(4) 旧TreeがPruneされる

(5) 新RPF Neighborに対してJoin送信

(6) 上流へJoin送信

⇒ Distribution Tree再構築



- Multicastの収束時間の短縮化

= (1)のUnicastルーティング収束時間を如何に短縮するかがポイントに

- 何も対策しなければ、数秒～数10秒程度かかる

今後の課題

● IP Multicast Deployment

- 基本技術は(ある程度)完成されている
 - 如何にシンプルに動かせるか (結果的に耐障害性を向上)
 - PIM-SSM, Bi-directional PIMなどもあるが...適材適所を目指したい
- 性能向上させるためには、他の要素技術との連携が不可欠
 - 伝送機器との連携を十分に考慮してNWを設計
 - IGP fast convergence
- HA機能が、OSPF/BGPなどに比べて弱い
 - 同じタイミングで、機能実装して欲しい!
- MIBの充実

● MPLS技術との比較

- メリット、デメリットを十分に見極めたい
 - MPLS実装のために要求されるH/W, S/W仕様
 - N/Wトポロジー
 - 実現可能なスケーラビリティ
 - 「どの範囲まで」、MPLS化すべきか?