信頼性の高いIP over WDM ネットワークの構築手法

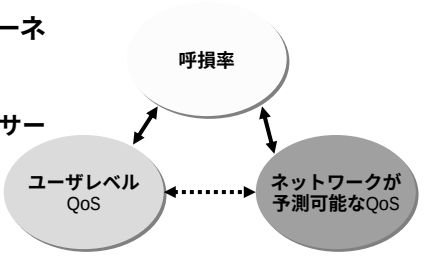
大阪大学サイバーメディアセンター
先端ネットワーク環境研究部門
村田正幸
e-mail: murata@cmc.osaka-u.ac.jp
http://www.anarg.jp/

Advanced Network Architecture Research

Advanced Network Architecture Research

電気通信網における ネットワーク設計

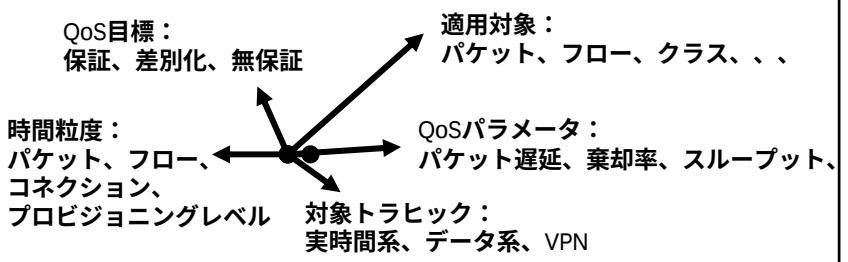
1. 過去の統計量に対する蓄積
   トラヒック特性
2. （古くは）単一キャリア、単一ネットワーク
3. アーラン呼損式
   ローバスト（ポアソン到着、一般サービス時間分布）
4. QoS測定　＝　呼損率
   キャリアが測定可能
5. 実時間転送；音声、動画像
   帯域保証のみで十分
   エンド間保証が前提



Osaka University

Advanced Network Architecture Research

ネットワークQoSのための要素



- 例
  - これまでのインターネット：全トラヒックを対象、無保証
  - IntServ：フローを単位とした遅延保証（実体はスループット保証）
  - DiffServ：クラスを単位とした遅延・スループットの差別化

Advanced Network Architecture Research

データ系アプリケーションにおけるQoSとは？

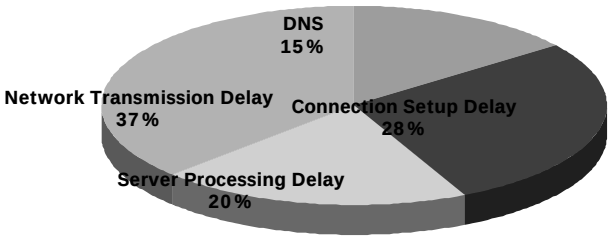
- データ系は帯域を食い尽くすアプリケーション
  - アクセス回線、エンドホストの高速化
  - TCP（＝エンドホスト）が輻輳制御を行う
  - バックボーンの高速化は解決策にならない
    - これまではアクセス回線がボトルネックになっていただけ
- データ系に適したQoS機構？
  - IntServによるQoS保証
    - パケット棄却率、パケット遅延を「保証」できるか？
    - トラヒック契約の考え方（QoSパラメータ、トラヒック特性を事前に申告）とマッチしない
    - RSVPのスケーラビリティに対する限界
  - DiffServによるクラスに対するQoS差別化
    - 実現はHOL優先権制御で十分（AFクラス）
    - 相対的なQoSはユーザのQoS要求とマッチするか？
  - QoS「保証」「差別化」なし
    - ただし、ネットワークプロビジョニングレベルでのQoS監視は重要
    - 帯域切り出し（VPN）は意味がある

データ系アプリケーションQoSの３原則

1. Data applications try to use the bandwidth as much as possible.
2. Neither bandwidth nor delay guarantees should be expected. Only network provisioning can satisfy user's QoS requests.
3. Competed bandwidth should be fairly shared among active users.

Webドキュメントダウンロードにおける遅延配分

- *Webのドキュメントダウンロード時間*
  - 回線容量を増やすだけでは性能向上に限界がある
  - エンドシステムの重要性
  - バランスのとれた資源配分が重要

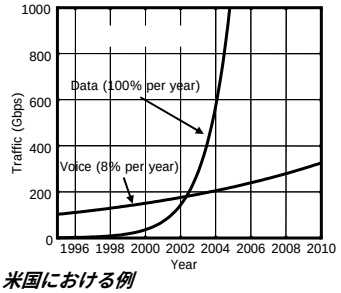


Produced from ftp://www.telcordia.com/pub/huitema/stats

Advanced Network Architecture Research

ネットワークディメンジョニングにおける課題

- 少なくともプロビジョニングレベルにおけるQoS予測が必要
- 電気通信網ではなかった新たな問題
  - データ系QoSとは何か？
  - QoSをどのように計測するべきか？
  - 「サービス」に対してどのように課金するべきか？
  - トラヒック特性を予測できるか？
  - エンド間性能はユーザしかわからない



米国における例

電話; 8%／**年、データ**; 100%／**年、3年で1桁上昇**

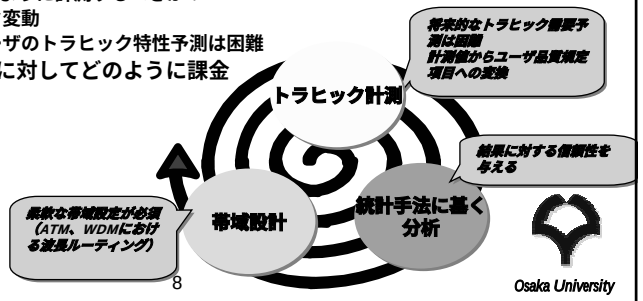
K.G. Coffman and A.M. Odlyzko, "The size and growth rate of the Internet," http://www.research.att.com/~amo

予測は不可能！

Advanced Network Architecture Research

データ系アプリケーションに適したQoS制御：スパイラルアプローチ

- 少なくともプロビジョニングレベルにおけるQoS予測が必要
- 電気通信網ではなかった新たな問題
  - データ系QoSとは何か？
    - パケット遅延、棄却率はエンドユーザレベルの性能指標ではない
    - エンド間QoSはユーザしかわからない（エッジルータの可能性はありうる）
    - QoSをどのように計測するべきか？
    - トラヒック変動
    - エンドユーザのトラヒック特性予測は困難
  - 「サービス」に対してどのように課金するべきか？



Advanced Network Architecture Research

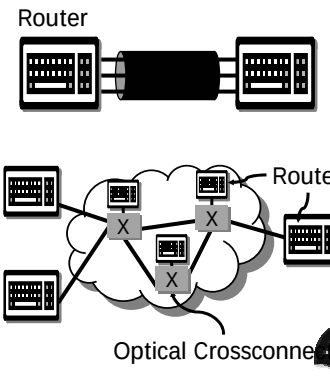
トラヒック計測の２つのアプローチ パッシブ／アクティブ

- パッシブな計測
  - OC3MON, OC12MON, ...
  - 点観測
    - 経路制御による経路の不安定性
    - TCPの誤り制御によるセグメント再送
    - 例：利用率が低いのは輻輳制御のため？　エンドユーザのアクセス回線が細いため？　エンドホストのパワー不足？
    - ストリーミングメディアのレート制御
  - ユーザQoSは不明
- アクティブな計測
  - Pchar, Netperf, bprobe, ...
  - エンド間ユーザQoSの計測
  - ある特定のユーザのQoSがわかったとしてもネットワーク設計ができるわけではない
  - ネットワークトラヒックの変動への対処

Advanced Network Architecture Research

フォトニックインターネットアーキテクチャ

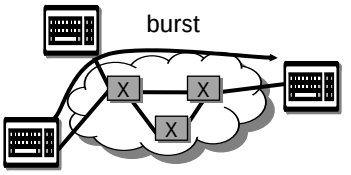
- ３つのアーキテクチャ
  1. WDM link network
     - 隣接ルータ間を複数波長で接続
     - 複数リンクが提供される
  2. WDM path network
     - 波長ルーティングに基き、論理トポロジーを形成
     - オプション：ネットワーク内部のルーティング機能
     - RWA (Routing and Wavelength Assignment) 問題



Advanced Network Architecture Research

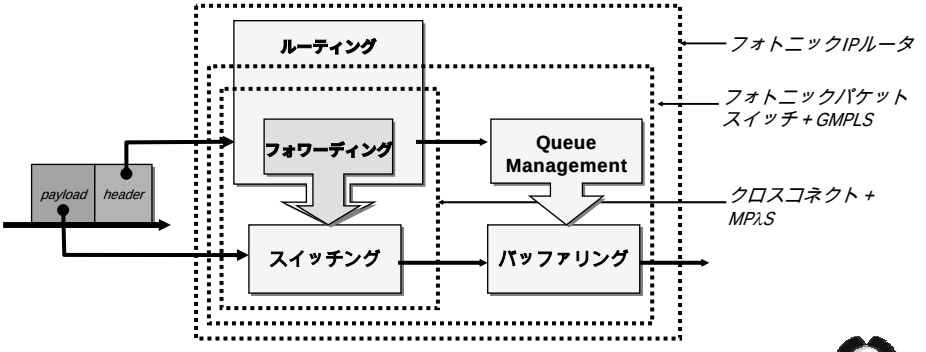
フォトニックインターネットアーキテクチャ（続）

3. WDM Packet-switched Network
   - バースト到着時に波長（＋経路）を定める光バーストスイッチング
   - 光パケットスイッチング



Advanced Network Architecture Research

IP over フォトニックのロードマップ
Cross-Connect, Switch or Router?

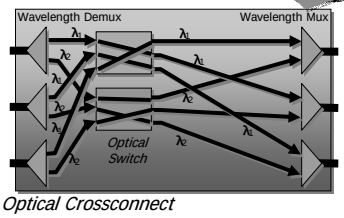


Advanced Network Architecture Research

MPLSと光MPLS（GMPLS)の対応

LSR (Label Switching Router);
光クロスコネクトが直接入力信号と出力信号を接続する

LSP (Label-Switched Path);
波長パス（ライトパス）



Optical Crossconnect

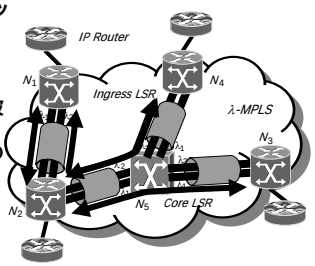
Ingress LSR;
IP**アドレスから波長への変換**

LDP (Label Distribution Protocol);
WRA**による波長による経路の設定**

Advanced Network Architecture Research

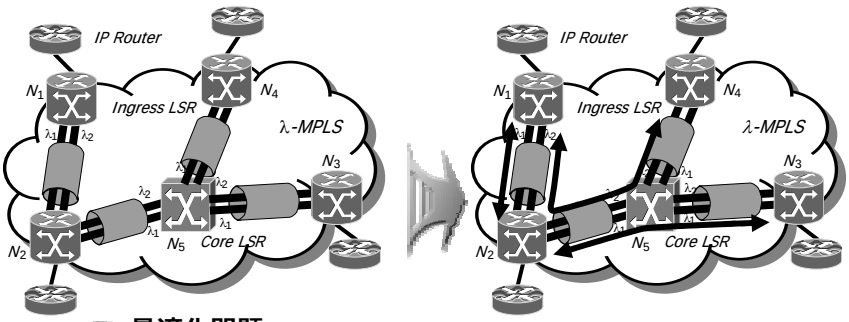
IP over WDMの実現に向けた課題

- **機能分割**
  - IP**ルーティング**とWDM**による波長ルーティングのマッチング**
  - Quality of Protection**をどう実現するか？**
- **論理パストポロジー設計問題**
  - **トラヒックエンジニアリングの観点に基づく段階的波長パス設定**
  - **ただし、これまでは、トラヒック量既知として全体のトポロジーを最適化問題として求めていた**
- **光パス設定の高速化**
  - **光バースト交換技術の応用**
- Ingress LSR**におけるボトルネック**
  - **ただし、**WDM Ring**などによる処理分散は可能**
- Label**の粒度が大きい；波長**
  - Label Merging/Splitting**は困難**
  - **４層スイッチングが困難**
- **バックボーンからフラットなネットワークへ：**
  PhotonicGrid



Advanced Network Architecture Research

WDM技術を用いた論理トポロジーの生成



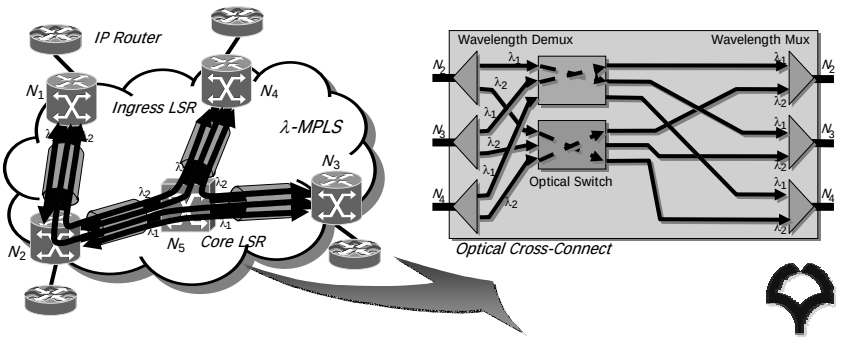
- 最適化問題
  - 例：トラヒックデマンドに基いて、各波長上のトラヒックを最大化するのに必要な最小波長数を求める
- 波長による直接パスをエンド間に設定することにより、ルータボトルネックを解消

Osaka University

Advanced Network Architecture Research

必要波長数の増大

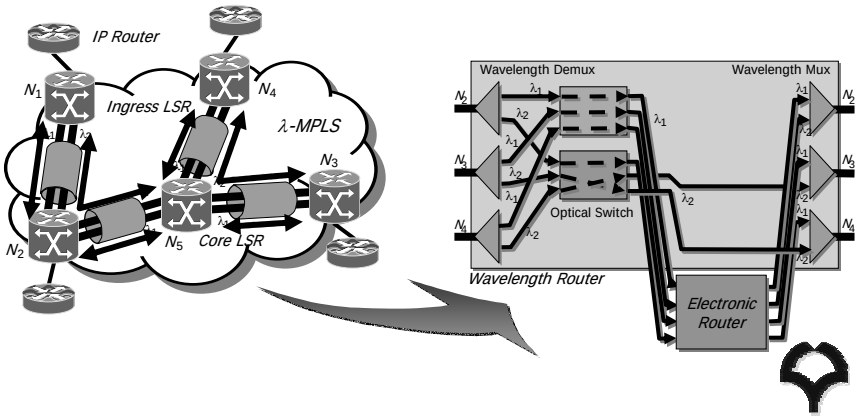
- すべてのノード間でAll-to-All Connectivityを保証するには多くの波長が必要



Osaka University

Advanced Network Architecture Research

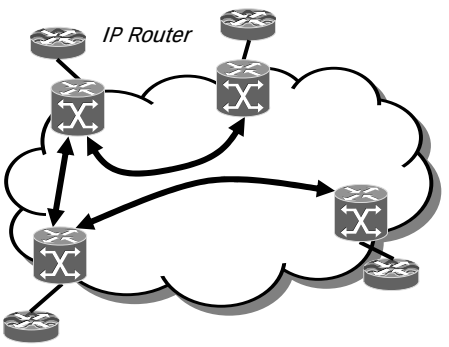
波長パスの積極的なカットによる必要波長数増大の抑制



Advanced Network Architecture Research

IPに提供される論理ネットワーク

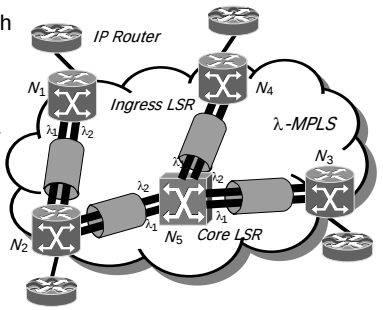
- 度数の増加による冗長性の高いネットワークの低居
- エンドノード間のホップ数の減少
- ルータにおけるパケット処理量の減少
- ルータボトルネックの解消



Advanced Network Architecture Research

IP over WDMへの適用における課題

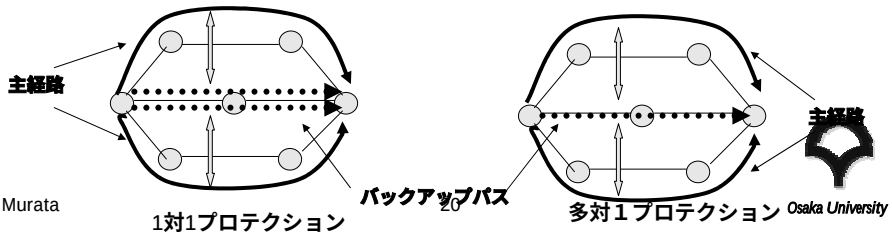
- **経路／波長割り当て問題** (Routing and Wavelength Assignment: RWA)**の例**
  - **与条件：**
    トラヒック量既知
  - **目的関数：**
    利用可能波長を使い切って各波長ごとのトラヒック量を最小化
- **ライトパスに流れるトラヒック量最小化**
  - **IPのメトリックス（Hop数、遅延時間）を考慮した場合、IPの経路が振動してしまう恐れがある**
- **ノードにおける負荷最小化**
  - **ルータボトルネックの回避**
- **エンドノード間遅延の最大時間の最小化**
  - **「物理ホップ数が大きければ、遅延時間が大きくなるのはしかたない」ことか？**
- **データ系の**QoS**？**
  - **性能指標はドキュメント転送遅延であるべき**



Advanced Network Architecture Research

WDMプロテクション

- **プロテクション技術**
  - **障害時にバックアップパスへ高速に切り替える**
    ~ 50ms
- **1対1プロテクションと多対1プロテクション**
  - **1対1プロテクションは複数の障害に対応可能**
  - **1対1プロテクションはより多くの波長を必要とする**



Advanced Network Architecture Research

IP over WDMネットワークのための プロテクション技術

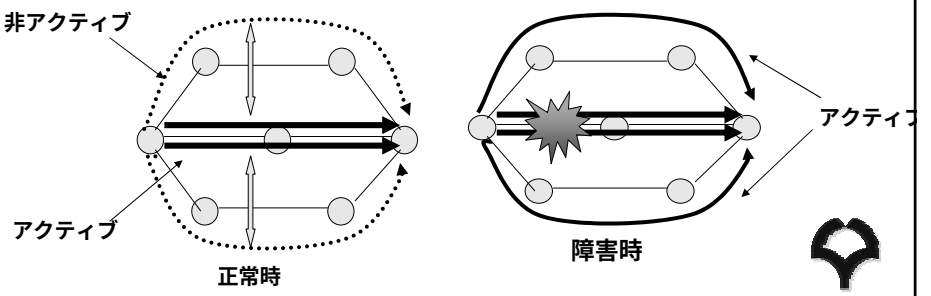
- １対１プロテクションにおいてはより多くの波長が必要
- IPネットワークにおける経路制御機構
- 迂回経路の設定による耐障害性
- 波長の有効利用

- 多対１プロテクション

Advanced Network Architecture Research

リンクプロテクション方式

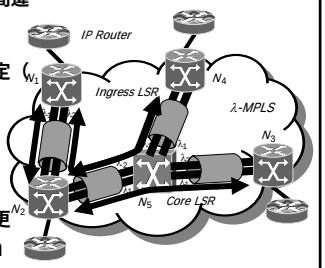
- リンク障害に対応可能であるプロテクション方式
- ファイバを通る全てのライトパスに対してプロテクション



Advanced Network Architecture Research

スパイラルアプローチの実現

- 段階的ネットワーク設計
  - 初期ステップ
    - 与えられたトラヒック量に基いたトポロジー設計；従来の設計手法が適用可能、ただし、トラヒック予測が間違っていたとしても、追加ステップで修正可能
  - 追加ステップ
    - トラヒック測定（パッシブ）、エンドユーザ品質測定（アクティブ）に基いた波長設定
    - 波長の追加、削減のみ
    - バックアップパスの有効利用
  - 調整ステップ
    - 全体の波長有効利用を考慮したトポロジー再設計
    - サービスの継続性を考慮した１波長ルートごとの変更
- WDM技術による高信頼化：IP＆WDM Integration
  - 共有プロテクション方式
    - 複数の障害には対応不可、必要波長数小
  - 追加ステップにおけるバックアップパスの有効利用
  - QoP (Quality of Protection)



Advanced Network Architecture Research

論理トポロジー再構成時のパス設定手法

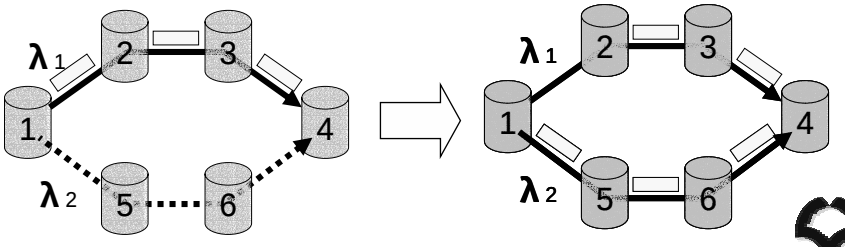
- 論理トポロジー再構成アルゴリズムで用いるパス設定手法
  - 新規プライマリ光パスの追加 (Append)
  - 現行プライマリ光パスの削除 (Delete)
  - 同じ送受信ノードをもつ現行プライマリ光パスと新規プライマリ光パスの切り換え (Exchange)
  - バックアップ光パスへのトラヒックの退避 (Switch)
  - バックアップ光パスの波長資源の解放 (Release)
- 用語
  - 現行光パス → 現在の論理トポロジー上の光パス
  - 新規光パス → 再構成後の論理トポロジー上の光パス

Append 操作，Delete 操作

- 新規プライマリ光パスの追加 (Append)
  - 必要な波長資源が確保できる新規プライマリ光パスを設定
  - 波長資源が確保できる条件
    - 波長資源が現行プライマリ光パスで使用されていない
    - 波長資源が現行バックアップ光パスで使用されていない
- 現行プライマリ光パスの削除 (Delete)
  - 現在設定されている光パスで使われている波長資源を解放
  - 削除した光パス上のトラヒックを損失

Exchange 操作

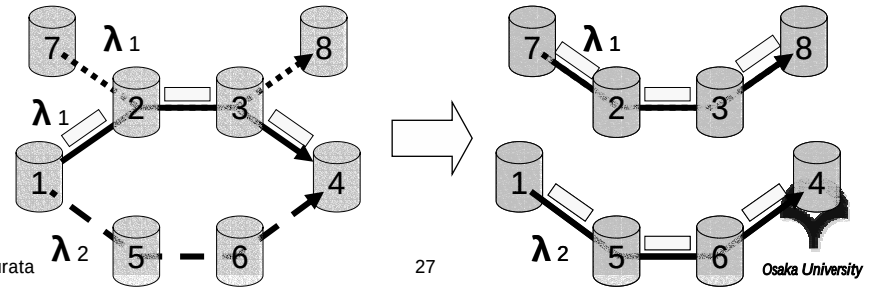
- 同じ送受信ノードをもつ現行プライマリ光パスと新規プライマリ光パスの切り換え
  - 現行プライマリ光パス上のトラヒックを保護
  - 現行プライマリ光パスとそのバックアップ光パスの波長資源を解放



Advanced Network Architecture Research

Switch 操作

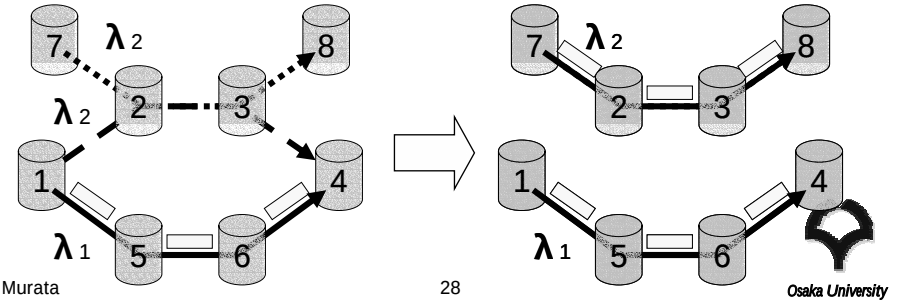
- バックアップ光パスへのトラヒックの退避
  - トラヒックをバックアップ光パスへ退避できる条件
    - ペアとなる新規プライマリ光パスがなく，Exchange 操作が不可能
    - 現行プライマリ光パスの資源が新規プライマリ光パスの設定に必要
    - バックアップ光パスが波長資源を専有 (共有バックアップ方式の場合)
  - 新規プライマリ光パスは Append または Exchange 操作で設定



Advanced Network Architecture Research

Release操作

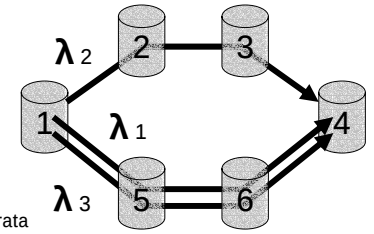
- バックアップ光パスの波長資源の解放
  - トラヒックが流れていないバックアップ光パスの波長資源を利用
  - 新規プライマリ光パスは Append または Exchange 操作で設定

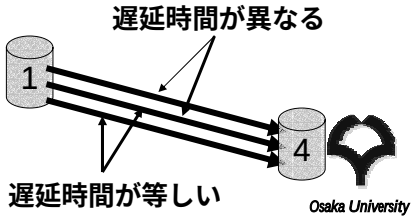


Advanced Network Architecture Research

割り当て波長の変更

- 割り当て波長
  - ある光パスの設定で用いるように，設計時に指定された波長
- 割り当て波長で光パスを設定できない場合，別の波長を用いて設定
  - ネットワーク上位層からみた性能は同等
  - 空き波長資源を有効に利用





Advanced Network Architecture Research

論理トポロジー再構成アルゴリズム

- 論理トポロジー再構成の手順を求める
  - Delete 操作回数の最小化が目標
- 再構成中は障害が発生しないと仮定

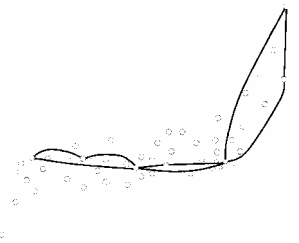
新規プライマリ光パスの設定手順を求める部分

Step1: P ← 0.
Step2: 設定可能な新規プライマリ光パスがあるならば Step2.1 へ．
　　そうでなければ Step3 へ．
　Step2.1: Exchange 操作で設定できれば P
　　そうでなければ Step2.2 へ．
　Step2.2: Append 操作で設定し，P ← P
Step3: 全ての新規プライマリ光パスが設定
　　そうでなければ Step4 へ．
Step4: 可能な限り Switch 操作を行う (操作ごとに P ← P+1)．Step5 へ．
Step5: P > 0 ならば P ← 0 として Step2 へ．そうでなければ Step5.1 へ．
　Step5.1: Release 操作が可能ならばそれを行い P ← 0 として Step2 へ．
　　そうでなければ Step5.2 へ．
　Step5.2: Delete 操作を行い P ← 0 として Step2 へ．

新規プライマリ光パスの設定に必要な波長資源を最も多く使用している現行プライマリ光パスから順に削除

Advanced Network Architecture Research

提案アルゴリズムの評価



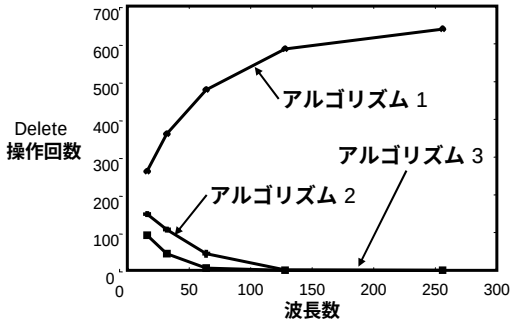
- 評価モデル
  - NTT 基幹ネットワーク
  - ノード数 49，リンク数 89
  - 波長数: 16，32，64，128，256
  - 最大ホップ数 4
- 評価方法
  - 31 個の論理トポロジーを，光パスをランダムに配置して生成
  - 30 回の再構成を行い Delete の操作回数の平均で評価
  - アルゴリズム 1，2，3 で比較

アルゴリズム 1	Append + Delete + Exchange + Release
アルゴリズム 2	アルゴリズム 1 + 波長の変更
アルゴリズム 3	提案アルゴリズム

Advanced Network Architecture Research

評価結果

- トラヒック損失の発生を大幅に少なくすることが可能
  - アルゴリズム 1: 光パス数が増えるにつれ，Delete 操作回数も増加
  - アルゴリズム 2: 波長数が増えるほど，割り当て波長の変更が有効
  - アルゴリズム 3: 波長数が多くないときに Switch 操作が比較的有効



プライマリ光パス数の平均

波長数	プライマリ光パス数
16	630
32	1080
64	1940
128	3353
256	5759

Advanced Network Architecture Research

再構成に要する操作回数

- 再構成に要する時間を操作回数で概算
  - 光パスの数が増えるに従い，操作回数が増加
  - IP の経路制御機能への影響？

波長数	16	32	64	128	256
アルゴリズム 1	1374	2217	3726	5920	9312
アルゴリズム 2	1741	2948	5095	8356	12284
アルゴリズム 3	1663	2798	4796	7645	11841

Advanced Network Architecture Research

今後？

- QoSに関して
  - バックボーンの高速化：フォトニックインターネット
    - GMPLSに基づくルーティング＋MPλSに基づく波長ルーティング
    - GMPLSに基づくルーティング＋フォトニックパケットスイッチに基づく波長ルーティング
    - フォトニックIPルータ
    - 高機能フォトニックIPルータ
  - エッジルータ、ゲートウェイにおける高品質化
    - プログラマブルルータの活用
  - エンドホストの高速化
    - ムーアの法則：CPUのコストパフォーマンスは18ヶ月で2倍に向上（10年で100倍）
    - ビルジョイの法則(?)：回線容量は9ヶ月～1年で2倍に向上（10年で1,000倍）

- インターネットが目の前にあったからこそ、それに適したWebというアプリケーションが生まれた
  - 背景：画像圧縮技術、GUI、画像表示能力
  - にわとりと卵（？）
  - napster, gnutella
- 波長の有効利用：PhotonicGrid
  - 波長をどれだけエンドユーザの近いところに持ってこれるか？
  - 多重波長数に依存

今後？

- エンド間QoSを保証、差別化することに意味があるか？
  - IntServ、DiffServの前提；回線固定、ノード固定、サーバ固定
  - P2P；サーバが突発的に現れる
  - モバイル環境；情報源が突発的に現れる
- ネットワーク資源の変動を前提とした、アダプティブなエンドホストによるQoS制御
  - 例：ストリーミングサービス vs. リアルタイム動画配信サービス
  - エンドシステムにとって利用可能な資源の実時間推定
  - ネットワークの資源管理は補助的な役割
- 次世代ネットワークのキーワード
  - Scalability
  - Adaptability
  - Mobility

オーバーレイネットワークの課題

- 論理網と物理網
  - データ転送は物理網をそのまま利用 (Gnutella)
- 効率的な論理網の構成手法
  - 論理網を構成する管理ノード（集中型、分散型）の設置
  - 物理網のQoS機能を利用する
    - IntServ、DiffServ
  - 論理ノードが物理網特性を自律的に把握
    - 計測（ホップ数、利用可能帯域、、、）