

次世代インターネットにおける 研究の方向性

村田正幸


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

E-mail: murata@cmc.osaka-u.ac.jp

<http://www.anarg.jp/>

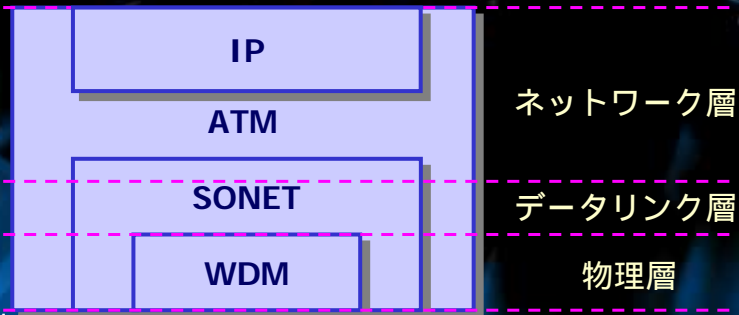


M. Murata

 Advanced Network Architecture Research

フォトニックインターネットに 対するいくつかのビュー

- IP over ATM over SONET over WDM
- IP over SONET over WDM
- IP over (PPP or HDLC over) WDM



The diagram shows a protocol stack with four layers: IP, ATM, SONET, and WDM. The IP and ATM layers are grouped under the label "ネットワーク層" (Network Layer). The SONET layer is grouped under "データリンク層" (Data Link Layer). The WDM layer is grouped under "物理層" (Physical Layer). Dashed horizontal lines separate the layers and group labels.

M. Murata 2

Advanced Network Architecture Research

マルチレイヤプロトコルスタックの問題点

- 機能の重複
 - 屋上屋を重ねる危険性
 - 複数レイヤにまたがった機能の最適化は容易ではない
 - ただし、機能分担の可能性はある
 - 経路制御、信頼性制御
- 非効率性
 - IP over ATM over SONET over WDM network
40バイトIPパケット / 2セル (106バイト)

M. Murata 3

Advanced Network Architecture Research

フォトリックインターネットアーキテクチャ

- 4つのアーキテクチャ
 1. WDM Link Network
 - 隣接ルータ間リンクをWDMで接続
 2. WDM Lightpath Network
 - エッジノード間を波長で直接接続
 3. Optical Burst Switching Network
 - パーストをエッジノード間を波長で接続して転送
 - Tell-and-Wait (TAW)、Tell-and-Go (TAG)
 4. Optical Packet Switching Network
 - パケット単位でスイッチング
- パケット交換 over (GMPLS-based)回線交換
 - パケット交換でも論理的回線は必要
 - 結局は時間粒度、PDU粒度 (プロビジョニングレベルの回線、コネクション、パケット)の問題

M. Murata 4

Advanced Network Architecture Research

フォトニックインターネットへの ロードマップ

Cross-Connect, Switch or Router?

payload header

ルーティング

フォワーディング

スイッチング

Queue Management

バッファリング

クロスコネクト + GMPLS
光バーストスイッチ + GMPLS
光パケットスイッチ + GMPLS
フォトニックIPルータ

M. Murata

5

Advanced Network Architecture Research

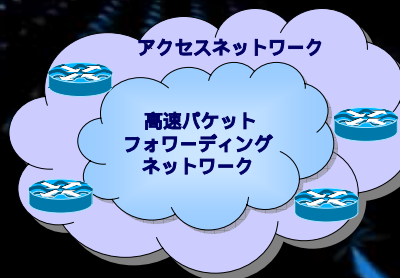
パケット交換 vs. 回線交換

機能	回線交換 (光クロスコネクトノード)	パケット交換 (電気ルータ)
回線効率	決して悪くない(回線の利用効率ではなく、回線数の利用効率) 波長数の増大が重要	一般に良いとされているが、遅延を小さくするためにはoverprovisioningが必要
エンド間パス可用性	コストをかけることにより維持	経路制御により維持
ノード可用性	機能が低い分高い	低い
ノードコスト	機能が低い分安い(半分から1/10)	高速化すればするほど多機能実現のためにコスト高
サービス機能の多様性	低い	高い



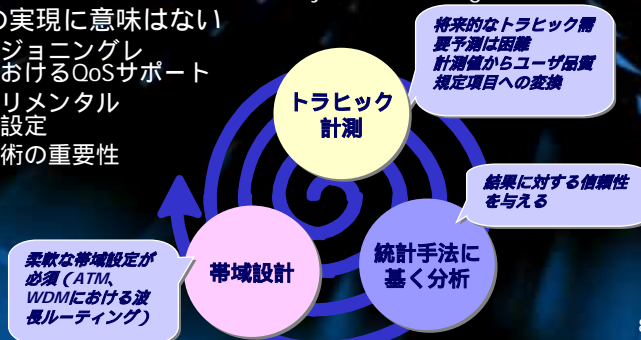
フォトリック技術の利用方法

- パケット・回線交換の融合？
 - アクセス系：パケット交換
 - バックボーン：WDM回線交換 (+GMPLS)：光パスネットワーク
 - スケーラビリティ確保のために、波長あたりの容量を増やすより、波長数を増やすことが重要
- 光パケットスイッチ + GMPLS
 - Deploymentに難あり
- しかし
 - 回線交換は必須ではない
 - 今のインターネットアプリケーションだけを考えればパケット交換で十分



データ系に適したQoS制御 スパイラルアプローチ

- データ転送の3原則
 1. Data applications inherently try to use the bandwidth as much as possible.
 2. Neither bandwidth nor delay guarantees should be expected.
 3. Competed bandwidth should be fairly shared among active users.
- QoS保証の実現に意味はない
 - プロビジョニングレベルにおけるQoSサポート
 - インクリメンタルなパス設定
 - 計測技術の重要性





経路選択・波長割当(RWA)問題

- WA : 経路は予め決めておいて「最適な」波長を選択
 - Random, First-Fit
- RWA : Multi-path Routing
 - 波長とともに、複数の経路から「最適な」経路を定める
 - Most-Used (同じ波長から埋めていく) : 集中化前提
- 中間解
 - 従来の経路選択方式とオンデマンド型波長割当方式の組み合わせ

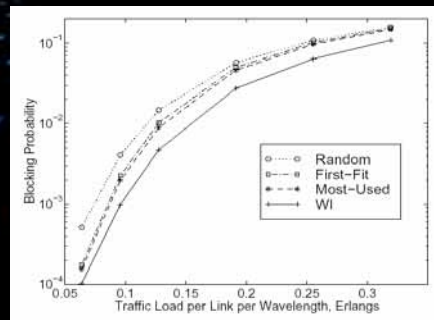


Figure 4 in E. Karasan, E. Ayanoglu. *Effects of Wavelength Routing and Selection Algorithms on Wavelength Conversion Gain in WM Optical Network*, ACM/IEEE Transactions on Networking, April 1998.



光パス設定の分散化

- どの経路情報を配布するか？
 1. Connectivity
 2. ホップカウント
 3. 負荷状況 (何本の波長が使われているか?)
 4. どの波長が使われているか?
- ルーティングプロトコル(1~3)は例えば OSPFで、波長割当(4)はオンデマンドで
- 制御プレーンの光化の重要性
 - 波長割当の光処理で実現可能



新しいアプリケーションの登場

- 大量のデータ転送
 - 広域分散SAN (Storage Area Network)
 - 大量データのバックアップ
 - グリッドネットワーク (特にデータグリッド)
 - QoS要求「瞬時に損失なくペタバイト級のデータを送ることができる」
 - CDN
- 結局、QoSは帯域と回線容量との比 (多重度) で決まる
 - 多重度が十分にあれば、パケット交換 (+ 輻輳制御) で十分
 - 個々のエンドユーザのQoSを保証したければ、回線を渡してしまう (回線を使い切るだけの能力がなければムダ)
- アプローチ?
 - データ転送プロトコル (TCP) の高速化
 - エンドユーザ間回線の提供



データ転送の高速化(1)

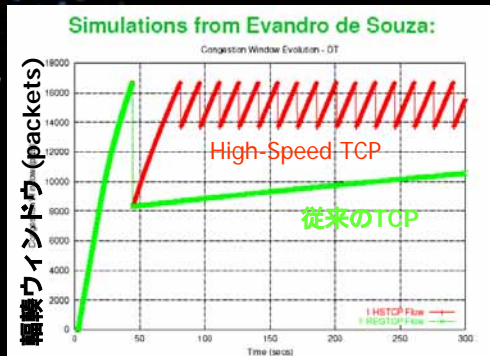
- プロトコルの高速化 : TCP
 - 最大ウィンドウサイズ > Bandwidth-Delay Product
 - Window Scale Option [RFC1323]: $65,535 \times 2^{14}$
= 1,073,725,440
 - スロースタート ファーストスタート
 - Large Initial Window [RFC2414]
 - ムダなパケットを再送しない
 - Selective Ack [RFC2018]
 - 統計的に見て意味のないパケットロスへの対処: たまたまパケットが落ちたとしても...
 - Fast Retransmit: 再送タイムアウトを待たずに再送パケットを転送する (TCP Tahoe)
 - Fast Recovery: ウィンドウサイズを半分にしかしない (TCP Reno)



データ転送の高速化(2)

■ High Speed TCP

- Sally Floyd らによる Internet Draft 化
- 輻輳制御の機能は残す
 - cwnd (輻輳ウィンドウ) が小さい時: 現在のTCPと同じ振舞い
 - cwnd が大きい時: よりアグレッシブに



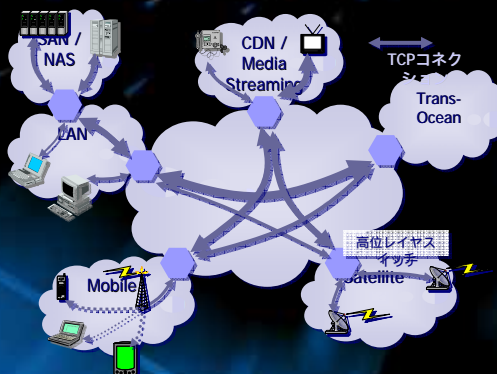
Source: <http://www.icir.org/floyd/talks/hstcp-Mar03.pdf>



TCP Overlay Network による高性能化

■ TCPコネクションの分割

- 帯域遅延積相当のウィンドウサイズの確保
- パケットロスへの早急な対処





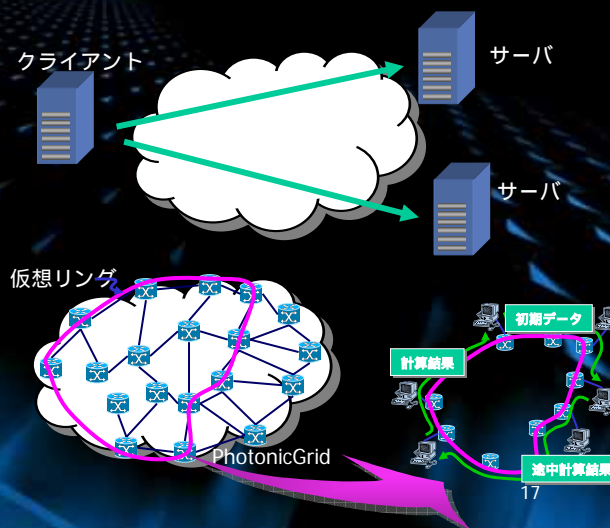
PhotonicGrid

- 目的
 - エンドユーザ間に波長を張り巡らすことによって、高速・高品質な通信パイプをユーザに直接提供
 - ユーザがカスタマイズ可能なネットワーキング基盤技術の提供
- 従来のアプローチ
 - キャリアがネットワークを管理し、帯域を切り売り（VPN）
 - IPパケットのサポート
- 技術課題
 - 波長数の増大（1000波～）；波長が豊富にあることが前提
 - 分散型ユーザ志向ネットワーク制御
 - 応用技術のためのミドルウェア（複数拠点ホスト間の共有メモリ、共有ディスク）
- 特徴
 - End-to-End Principleの精神は残す
 - 独自のプロトコル展開も可能



PhotonicGridの応用イメージ

- SAN、データグリッド
 - オンデマンド型回線提供
- グリッド計算





End-to-End Principle

- J. H. Saltzer, D. P. Reed, D. D. Clark, "End-To-End Arguments In System Design," ACM Transactions on Computer Systems, 1984.
- R. Bush and D. Meyer, "Some Internet Architectural Guidelines and Philosophy," RFC 3439, December 2002.
- "KISS: Keep it Simple, Stupid"
 - (1) ネットワークは特定のアプリケーションに基づいて、あるいは、特定のアプリケーションのサポートを目的として構築してはならない
 - (2) エンドノードで実現できる機能はそのノードに任せ、関係する状態情報はそのノードにおいてのみ維持すべきである
- 通信機能はできるだけエンドノードにおいて実現、ネットワークはビットを運ぶことに徹する



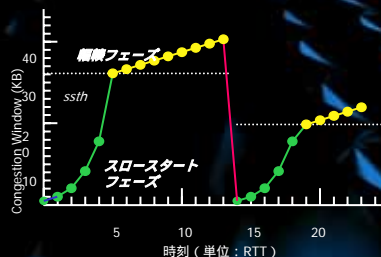
TCPの特徴：成功の要因

- ネットワーク層以下に特定の技術を仮定しない
- Self-adaptability
 - エンドホストはACKを受け取ると、ウインドウサイズを上げる
 - パケットロスがあると、ウインドウサイズを下げる
- タイムアウト制御
 - ラウンドトリップ時間 (RTT) の計測

$$RTT = RTT + (1 - \alpha) M$$
 M: 計測時間、 α : 重み(7/8)
 - 「ばらつき」の計測

$$D = D + (1 - \alpha) |RTT - M|$$
 - タイムアウト時間の決定

$$Timeout = RTT + 4 * D$$
- ネットワーク内輻輳制御を請け負う
 - ウインドウサイズ可変型フロー制御
 - 単純な制御
 - ACKを受け取る ネットワークは空いている 転送速度を上げる
 - パケットロス ネットワークが混んでいる 転送速度を下げる





ネットワーク制御に 求められる3要素

- 拡張性 (スケーラビリティ)
 - ルータ数やエンドホスト数、ユーザ数、情報機器端末数の増大への対応
- 多様性
 - 情報機器デバイスの多様性、ネットワーク技術の多様性、ネットワークサービスの多様性、トラフィックの多様性への対応
 - 単一のネットワークアーキテクチャによる統合ネットワークは存在しない
- 移動性
 - 利用者自身、ネットワーク資源 (ルータ、回線、サーバ) の移動
 - それらの生成・消滅が頻繁に発生



適応性

- 基本原則
 - エンドホストの適応性 (adaptability) を根幹とし、ネットワークはそのような適応性をサポートするための機構を提供する
 - インターネットの分散処理志向をさらに推し進め、それによって損なわれる資源利用の効率性については、エンドホストの適応性によって補償する
 - 今後も開発されていく多様な通信技術に対応しながら、スケーラブルでかつ耐故障性に富んだネットワークを構築しつつ、ユーザの多様な要求に対するサービスを提供する
- エンドホストの自律性がますます要求されるようになり、それを前提として、ネットワーク全体の調和的な秩序を保つ



複雑系としてのインターネット

Metcalfe's law

- "The value of a network increases exponentially with the number of nodes."
ネットワークの価値はノード数（あるいは、ユーザ数）に対して指数的に増加する
- ユーザ数 N 、ネットワークの価値 $V(N)$

$$V(N) \approx N^2$$

- Webシステムのクライアント/サーバモデルにより崩れつつあったMetcalfe's LawはP2Pネットワークの登場により、復活しつつある
- P2Pネットワーク：パワー則の観察
 - Preferential Attachment



Power Law Networkとしてのインターネット

Power Law分布

- 事象 X の確率密度
 $P[X=x] = x^{-k}$

自己組織的なネットワークに多くの事例

- 人のネットワーク (Small World)
- 文献引用ネットワーク
- インターネットのASレベルの接続リンク数
- HTMLページのリンク数

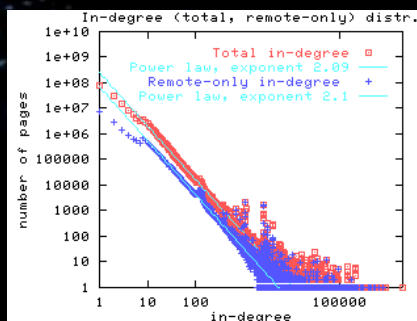


Figure 1 of "Graph structure in the web," authored by Andrei Broder et al., available at <http://www9.org/w9cdrom/160/160.html>

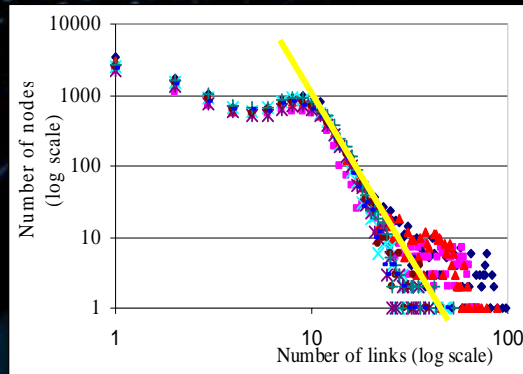


パワー則に従うことの意味 —ノードあたりのリンク数—

■ 特徴

- ランダムなノード故障には強い
- リンクの集中するノードの故障（アタック）に弱い

■ ランダム性を高める論理網（オーバーレイネットワーク）の構築



From "P2P Architecture Case Study: Gnutella Network," authored by Matei Ripeanu, available at <http://www.computer.org/proceedings/p2p/1503/15030099.pdf>

M. Murata

29



生物界に学ぶネットワーク制御

■ 背景

- 生物を複雑系として見た場合の、頑強性、安定性は次第に明らかになりつつある
 - 外乱に対する適応能力は高い
 - ただし、その適応速度は遅い
- 生物の自己組織的、自立的な制御をネットワークに持ち込み、その適応性 (adaptability)、頑強性 (robustness)、安定性 (self-stability)、耐故障性 (resiliency) を利用できないか？

■ 過去の例

- GA (Genetic Algorithm) : 遺伝子をモデル化し、それを最適問題の解法に適用
- ACO (Ant Colony Optimization) : ありの生態を模した最適問題の解法

■ 大阪大学21世紀COEプログラム「ネットワーク共生環境を築く情報技術の創出」

M. Murata 生物界の共生関係に学ぶ情報技術（ネットワーク制御）の創出 31



Biological Internet

- 複雑系としてのインターネットにおける、頑強性、安定性の確保
 - (フィードバック)制御: TCPそのもの
 - 冗長性: フォトニックネットワークのSelf-Healing
 - モジュール化: プロトコルの階層化
 - 構造安定: AS単位の階層化
- 研究課題
 - 現状の技術でインターネットはほんとうに安定しているのか、頑強なのか
 - 頑強な、また、安定な制御、ネットワーク構造?
 - 冗長性はどの程度、必要なのか?
- インターネットは設計を変更できる: 巨大な実験場!



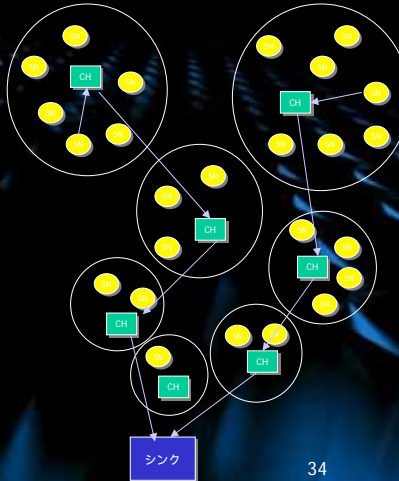
メタ情報環境を実現する ネットワーク基盤

- メタ環境 (仮想環境 + 実環境) を実現するネットワーク基盤の構築
 - 生活・社会・産業における神経系の創出
- 応用例
 - 現状では、遠隔監視
 - 流通管理
 - 環境情報の取得
 - 渋滞調査 (ITS)、極限地帯の気候・生態調査、地震情報の取得
 - パーソナルエリアネットワーク
 - 人の行動追跡 & 情報ナビゲーション
- ICタグ (RFタグ) との差異
 - ICタグは情報流通のチェックポイントを提供するのみ
 - メモリ容量、情報処理能力の限界
- インフラは誰が準備するのか?



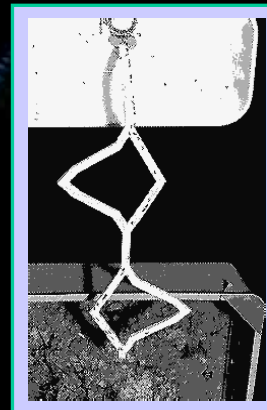
センサーネットワークの要素技術

- 省電力化のためのデータ伝送技術
- 位置検出
 - アドホック化、モビリティへの対応
- アドホックネットワークング技術
 - モビリティ、省電力を考慮した経路制御
- スケーラビリティの確保
 - 多数のセンサーノードの収容
 - 省電力化
- データセントリックネットワーク化
 - データ集約、途中ノードでのデータ処理
- 自己組織的、自律的適応型制御の実現



Ant Routingとは

- ありの餌採行動
 - フェロモンを介した相互作用により、全体の制御が実現される
 - フェロモンを道に残していく
 - フェロモンを追跡する
 - フェロモンは一定の割合で消滅する
- Ant Colony Optimizationの一種
 - Stigmergy
 - 間接的なインタラクションによって、全体の制御を実現する機構の一種
 - 環境を介した通信によって全体の制御を実現する 自己組織化
 - Complex System vs. Complicated System
 - 要素の寄せ集めではなく、パーツの集合体以上の振る舞いを期待





Ant Routingの現状

- (モバイルアドホック) ネットワーク経路制御への適用
 - 適応性、耐故障性の確保
- 必ずしも成功しているわけではない
 - Distance Vector Routingと同じ問題
 - blocking problem ; リンクが切れたとき、どうするか？
 - short-cut problem ; 新しいリンクができたときどうするか？
 - 解決策はあるが、アドホックな解
 - 制御パラメータ (ありの速度、フェロモンの消える速度、他の道を選ぶ確率) をどう選ぶか？
 - Trial and Errorが必要 : NN、GAと同じ問題
- 学べるのはPerturbation (故障、変動) に対する Tolerance、Resilience

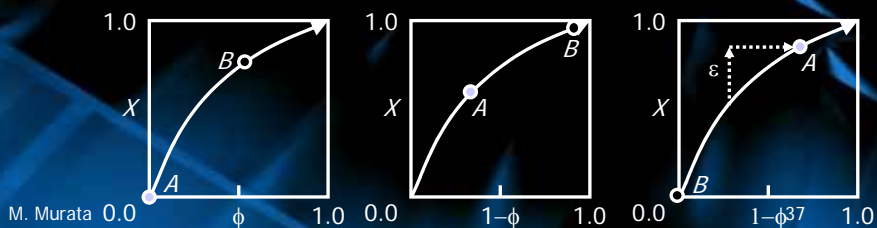


蛍の同期の例

- 「東南アジアのある蛍 (Pteroptyx Malacae, Pteroptyx Cribellata) は、周りの蛍と同期して光を点滅する」
- Integrate and Fire Model (Pulse-Coupled Oscillator)
 - リーダなしに同期可能
- センサー同期への適用 (電源節約)

$$\frac{dx_i}{dt} = S_0 - \gamma x_i, \quad 0 \leq x_i \leq 1$$

$$x_j(t) = 1 \Rightarrow x_i(t^+) = \min(1, x_i(t) + \varepsilon) \quad \forall j \neq i$$





研究の方向性に対する 現状認識

- 新しいアプリケーションの登場
 - グリッド、SAN、CDN、
 - 高速大容量データ転送
 - 再び、「パケット交換 vs. 回線交換」の議論
 - 既存トラフィックに加えて、新しいトラフィックが発生
 - オーバーレイネットワークの発展
 - 論理ネットワークと実ネットワークのインタラクション
 - 多重構造ネットワーク
 - システムの大規模化
 - ユーザ数、ホスト数、ルータ数
 - 自律分散制御の必要性
 - Robustness, Resiliencyの確保
 - 適応複雑系
 - それぞれのエンティティはシンプルなルールで動作、その結合体としてシステムが動作する
- M. Murata