

P2Pトラヒックの測定と特性評価

森達哉† 亀井聡† 大井恵太‡

† NTT サービスインテグレーション基盤研究所

‡ NTT 情報流通プラットフォーム研究所

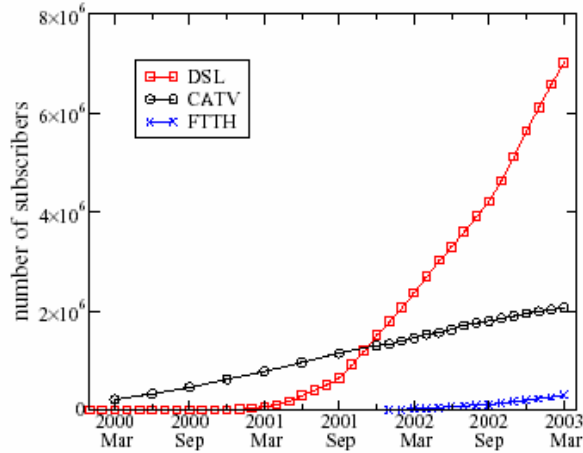
2003年9月24日

電子情報通信学会 ソサイエティ大会

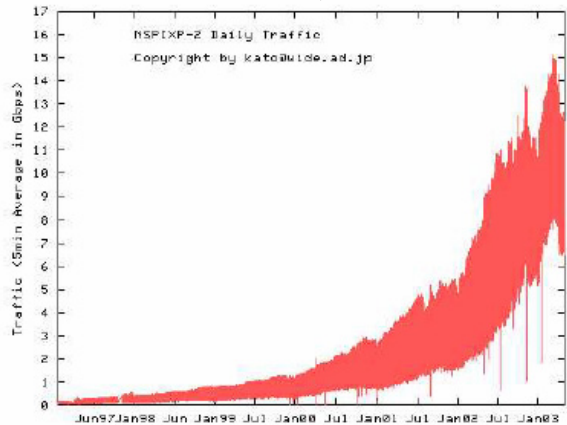
アウトライン

- 背景
 - インターネットトラヒックの変化
- 課題
 - P2Pトラヒックの測定と特性評価
- P2Pトラヒックの特性評価
 - フロー分析
- まとめ・今後の課題

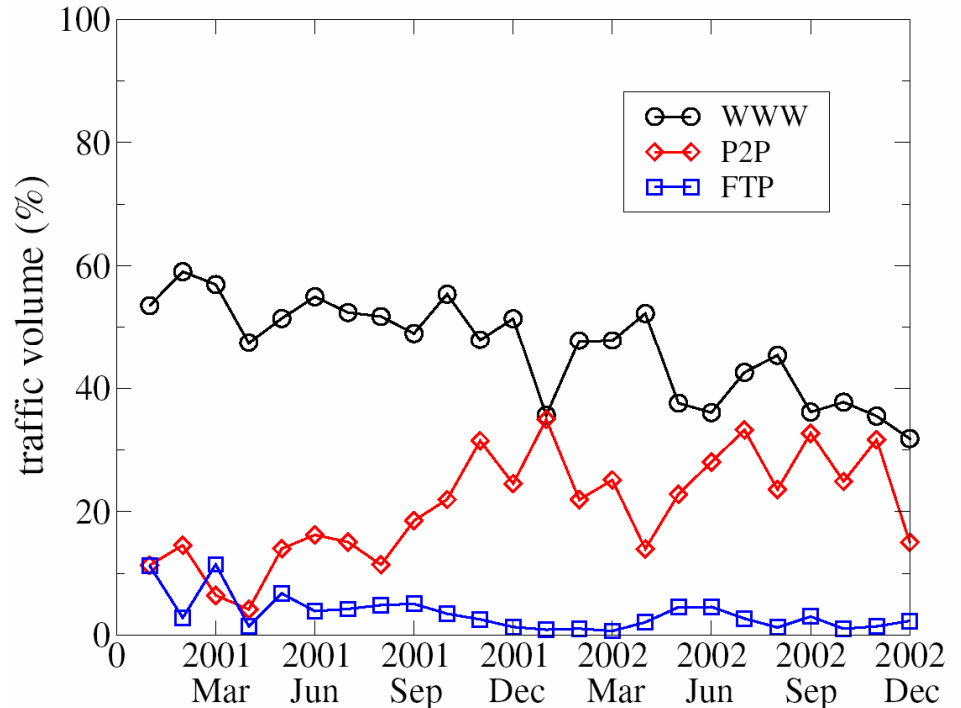
背景: インターネットトラフィックの変化



ブロードバンド回線の急激な普及
©総務省



バックボーン(NSPIXP2)トラフィックの増加
©kato@wide.ad.jp



トラフィックに占めるアプリケーション比率の変化
(WIDE 国際接続回線)

P2PがWWWに匹敵するまでに
※ ISPでは P2P > WWW の報告あり

背景: P2P トラヒックの性質

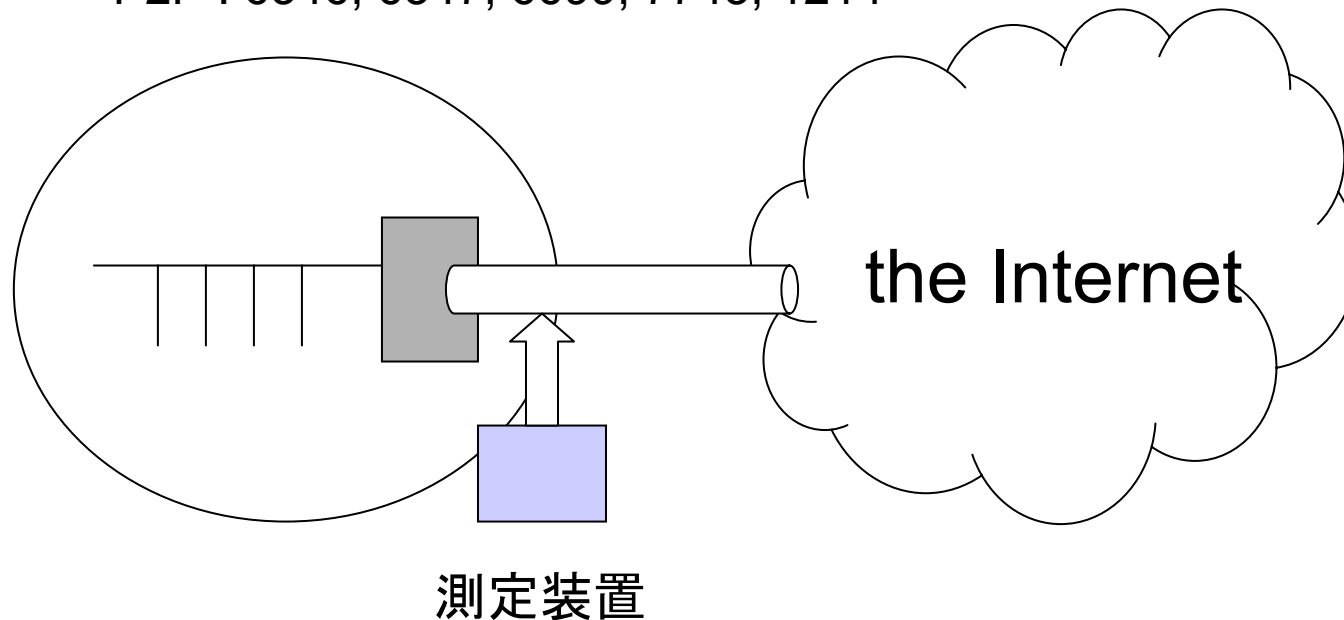
- 1セッションのサイズがWWWと比較して大きい (後述)
 - 流通ファイルサイズの特徴より (後述)
 - 現在ISPで大きな問題になっている
 - ごく一部のヘビー(P2P)ユーザが帯域を圧迫する
- Peer 同士の双方向通信
 - 物理的なトポロジーとは独立な通信
 - WWWのようなクライアントサーバモデルとは異なる
- 2000年代初頭より急速に普及
 - Post WWW
 - ブロードバンド回線の普及
 - クライアント(PC)の高性能化
 - CPU, ネットワークリソース, ハードディスク容量

課題: P2P トラヒックの測定と特性評価

- P2Pトラヒックの測定
 - P2Pトラヒックの特性評価のために
 - 実網の計測
- P2Pトラヒックの特性評価
 - トラヒックモデルを作るために
 - 特性の定量化
 - フロー分析
 - サイズ, レート, 持続時間, 到着間隔 等
 - 制御のために
 - 制御の対象は何か?
 - どう制御するか?

トラヒックの測定

- 対象回線: インターネットアップリンク (100Mbps)
- 測定期間: 2002年11月の平日夜間 (18:00-21:00)
- 測定方法: パケットキャプチャ
 - ポート番号によるアプリケーション識別:
 - WWW: 80, 8080, 443
 - P2P : 6346, 6347, 6699, 7743, 1214



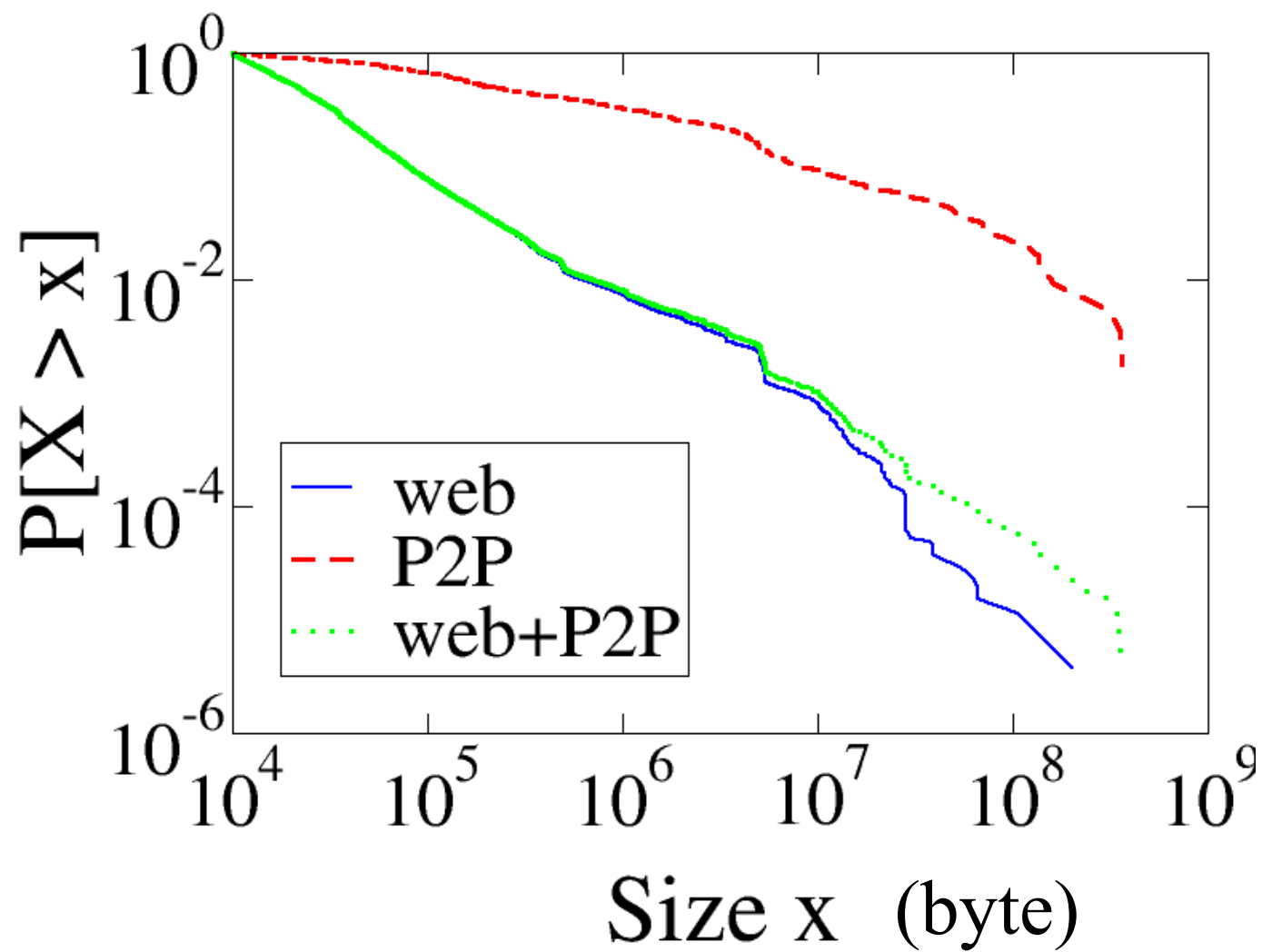
P2Pトラヒックの特性評価

- フロー分析を行う
 - フロー = 発着IPアドレス, 発着ポート番号, プロトコル
e.g., (192.168.1.12, 10.0.2.24, 53671, 6699, TCP)
 - ルータにおける制御、計測の単位
 - TCP の輻輳制御が実施される単位
- WWWとの比較
 - フロー数とトラヒック量
 - フローサイズ特性
 - フローレート特性
 - フローサイズとフローレートの関係
 - (フロー到着間隔, フロー持続時間)

フロー数とトラフィック量

	フロー数 (#/3時間)	トラフィック総量 (Gbyte/3時間)
全体	1223735 (100%)	37.72 (100%)
WWW	810418 (66.2%)	17.52 (46.4%)
P2P	12405 (1.0%)	10.25 (27.2%)

フローサイズ特性 (1/4)



フローサイズ特性 (2/4)

- フローサイズ分布はパレート分布
 $P[X > x] = k x^{-\alpha}$ で近似できる
 - P2P: $0 < \alpha < 1$ (平均, 分散が発散)
 - WWW: $1 < \alpha < 2$ (分散が発散)
- フローサイズ特性
 - 転送されるファイルサイズ特性に関係？

フローサイズ特性 (3/4)

- パレートの法則
 - 80/20 の法則, Elephant and Mice phenomenon

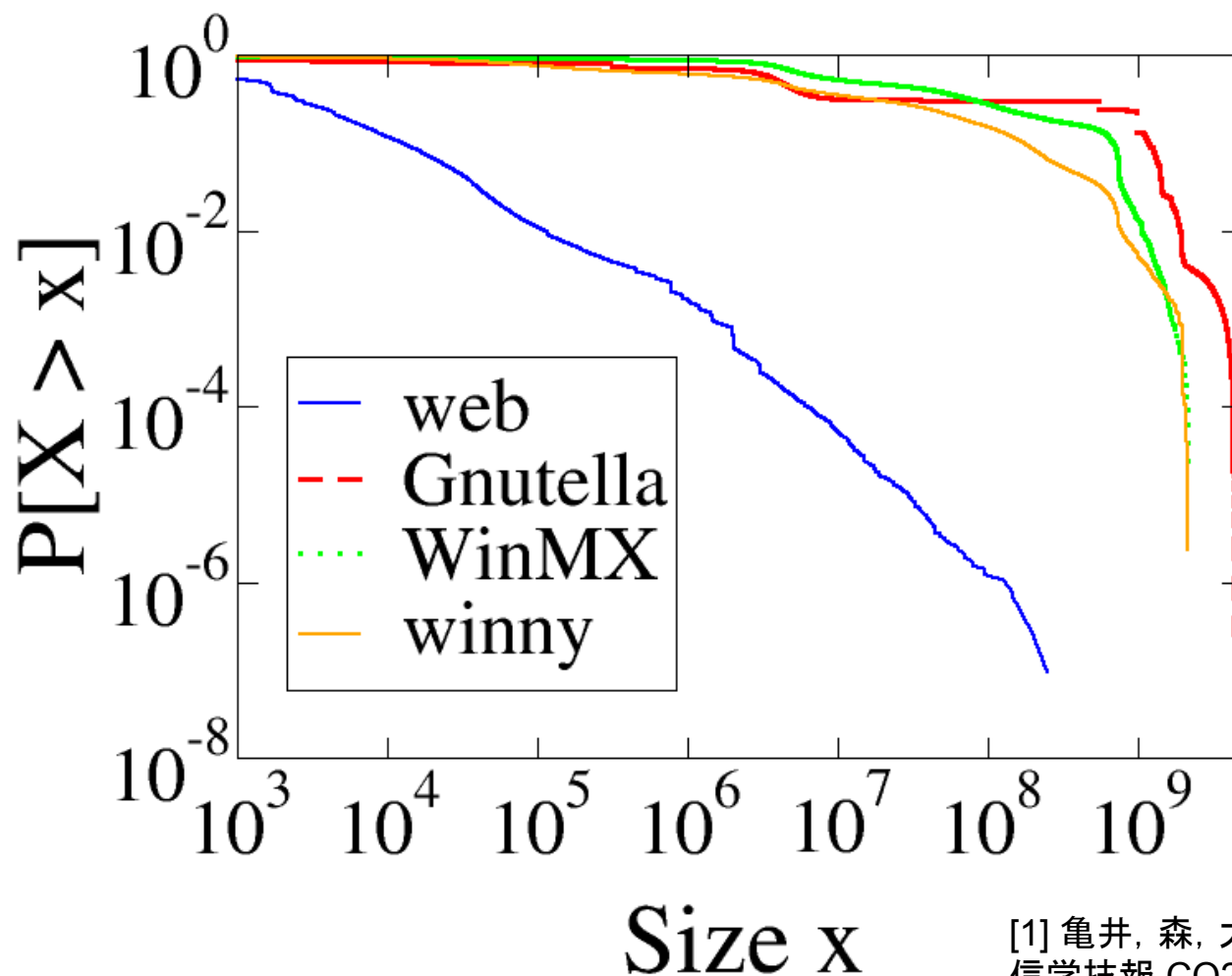
フローサイズトップ10のフローの統計

	全トラフィックに 占める割合
Web	2.78 %
P2P	12.39 %

P2Pはパレートの法則がより顕著である

フローサイズ特性 (4/4)

流通ファイルサイズ累積分布(補分布) [1]



分布関数
 $P[X > x] \sim x^{-\alpha}$
(パレート分布)

※右裾の方はカットオフあり

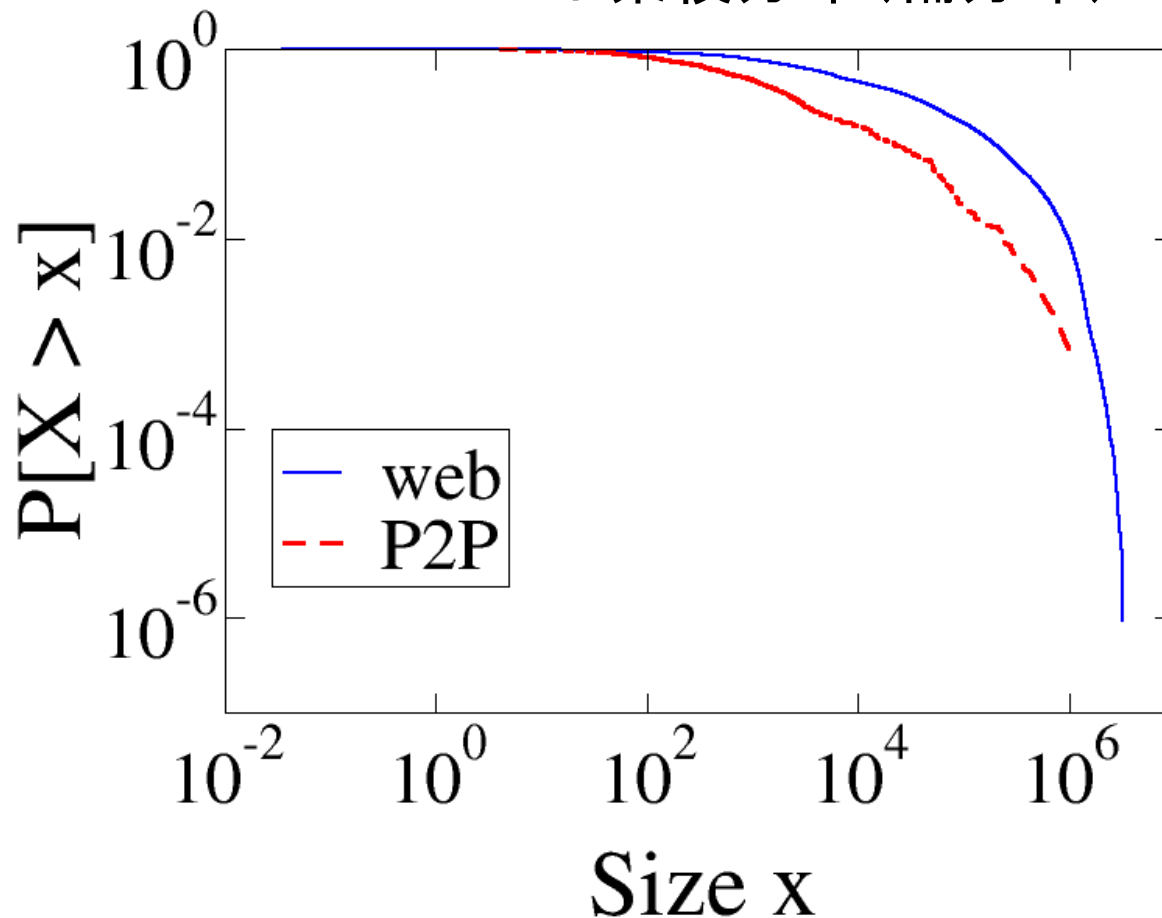
平均ファイルサイズ
WWW: 11.85 kB
G.: 326.48 MB
MX: 153.64 MB
ny: 62.82 MB

[1] 亀井, 森, 大井, P2Pファイル共有の実態と課題
信学技報 CQ2003-7, 2003

フローレート特性

フローレート = フローサイズ / フロー継続時間

フローレート累積分布(補分布)



分布関数
どちらも対数正規
分布で近似可能

平均フローレート
WWW: 72KB/s
P2P: 13KB/s

フローサイズとフローレートの関係

フローサイズS vs. フローレートR
(持続時間 > 5秒)

	S vs. Rの 相関係数
WWW	0.32
P2P	0.74

P2Pではサイズ大⇔レート大の傾向がある

その他の結果

- フロー到着間隔
 - WWW, P2P ともに指数分布で近似できる
 - WWW は時間相関あり。P2Pは時間相関なし。
- フロー持続時間
 - WWW, P2P ともに対数正規分布で近似できる
- P2P フローの増加の考察 (混合パレート分布)
 - ごく少量のP2Pフロー数増加→大幅なトラヒック増
 - 全体のフローサイズ分布はP2Pの分布に近づく

まとめ

- P2Pのフローサイズはきわめて裾野が厚い分布にしたがう（パレート分布の形状母数が1未満）
 - ファイルサイズ分布特性と関係
 - アプリケーションに固有な特性
- フローレートは $WWW > P2P$
 - 平均持続時間は $P2P > WWW$
 - 通信形態の違い
- 一定以上の通信を行うP2Pフローではフローサイズと平均フローレートに正の相関
 - レートが高い→通信量大きい傾向
 - ネットワークリソースが高い→通信量大きい傾向

今後の課題

- **トラヒックのモデル化**
 - シミュレーションにおける背景トラヒックモデル
 - P2P, WWW 混在環境を模擬したモデル
 - アプリケーション特性を考慮した現実的なモデル
 - 単純なパレート ON/OFF では不十分
- **トラヒックの制御・管理手法の提案**
 - フローサイズ特性より (パレートの法則)
 - ごく一部のフローが全体に占める割合が大きい
 - そのフローを制御した効果は大きい
 - そのフローをリアルタイムに検知する方法
 - レートとサイズの関係より
 - 高レート→サイズ大の傾向を利用する