

JANOG13

～ 広がるP2Pサービスとインフラへの影響 ～

P2Pトラフィックとサービスの現状

2004年1月30日

NTTサービスインテグレーション基盤研究所

亀井聡

kamei.satoshi@lab.ntt.co.jp

P2Pとは？

Peer-to-Peer: 個々のノードが対称的な役割を持つ分散アプリケーションを構築するための技術 (IRTF P2PRG) .

P2P ファイル共有 違法

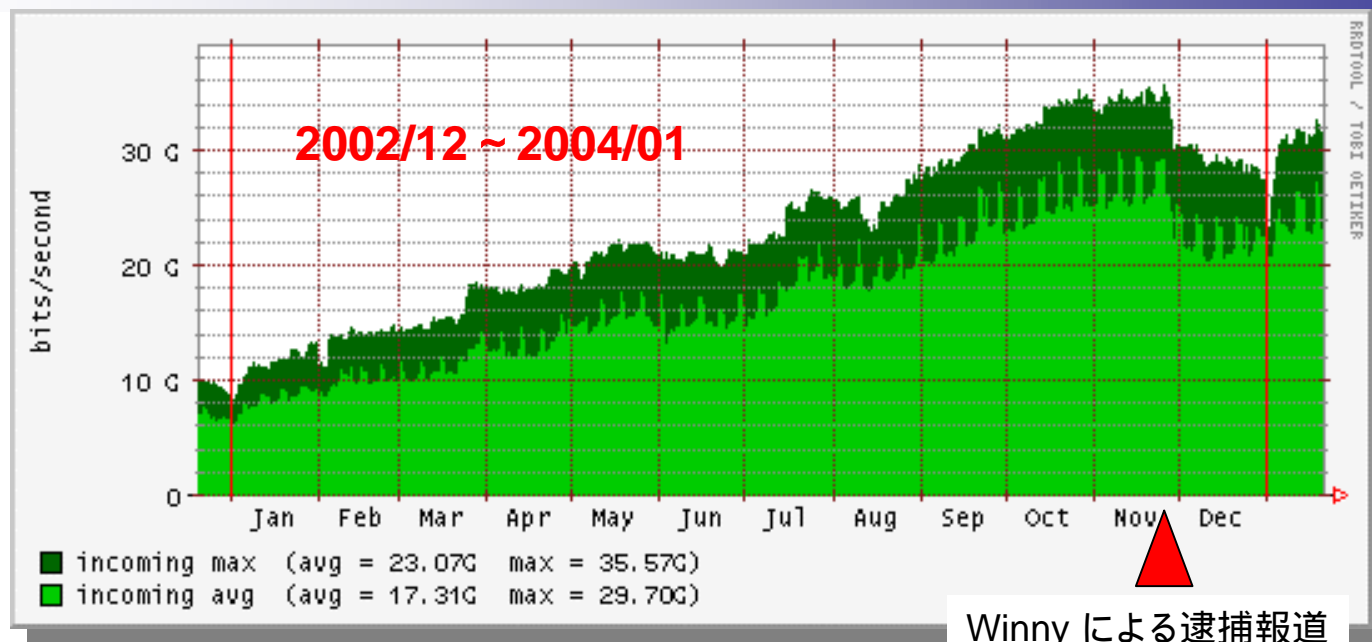
ではあるが...

ネットワークオペレータにとっての近々の課題は
P2Pファイル共有サービスによるトラフィック増 .

ネットワークを効率良く使える技術

ネットワークを効率良く喰い潰している現状

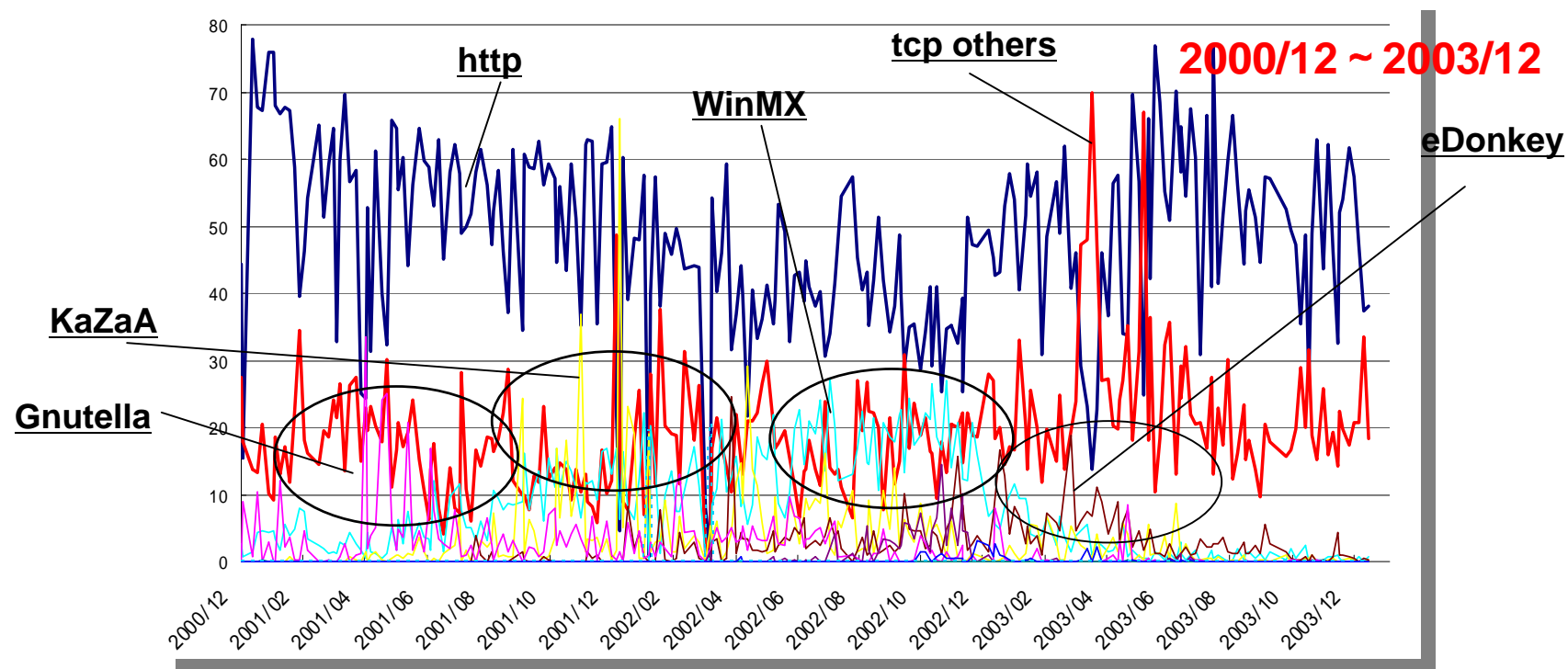
ネットワークの状況 ~ IX に見るP2Pトラフィック



mfeed, "JPNAP total traffic," <http://www.mfeed.ad.jp/jpnap/traffic.html>.

- 商用ISPのトラフィックを中心としたIXトラフィック。
- Winnyの逮捕報道後の急激なトラフィックの落ち込み。
- Winnyユーザの何割かのトラフィックが JPNAP の2割を占めていた。
- 既にトレンドは戻りつつあるようだが...

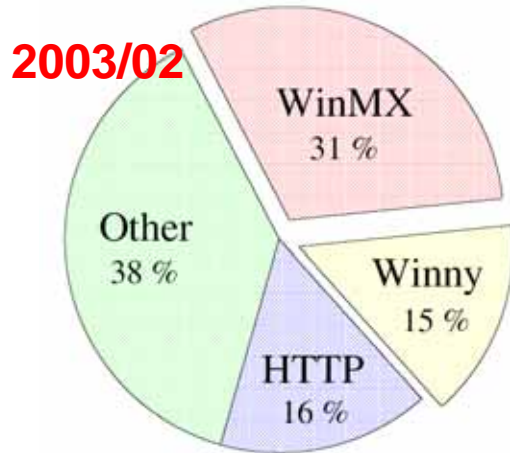
ネットワークの状況 ~ 国際線トラフィックから



WIDE Project, "MAWI Working Group Traffic Archive," <http://tracer.csl.sony.co.jp/mawi/>.

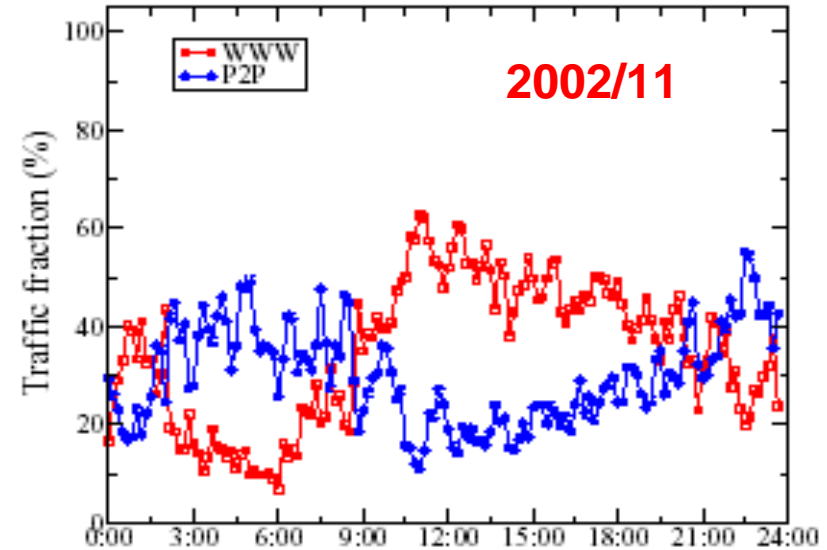
- 日本 ロサンゼルス間トラフィックの公開ダンプデータ(14:00-15:00)を元にポート番号により分析.
- 下流ネットワークはWIDE系学術機関が中心.
- 意外にP2Pトラフィックが伸びていない(落ち込みすら見られる)のは教育の成果?
- 主流アプリケーションの栄枯盛衰が激しい.
- tcp others の中にも見えない P2P トラフィックが含まれている可能性が高い.

ネットワークの状況 ~ エッジ側



閉域網 ISP間トラフィックのアプリケーション比率
2003/02 平日夜間

岡田, 川原, "IP網におけるトラフィック特性分析の一考察," 信学
技報NS-2003-5, 2003-4.

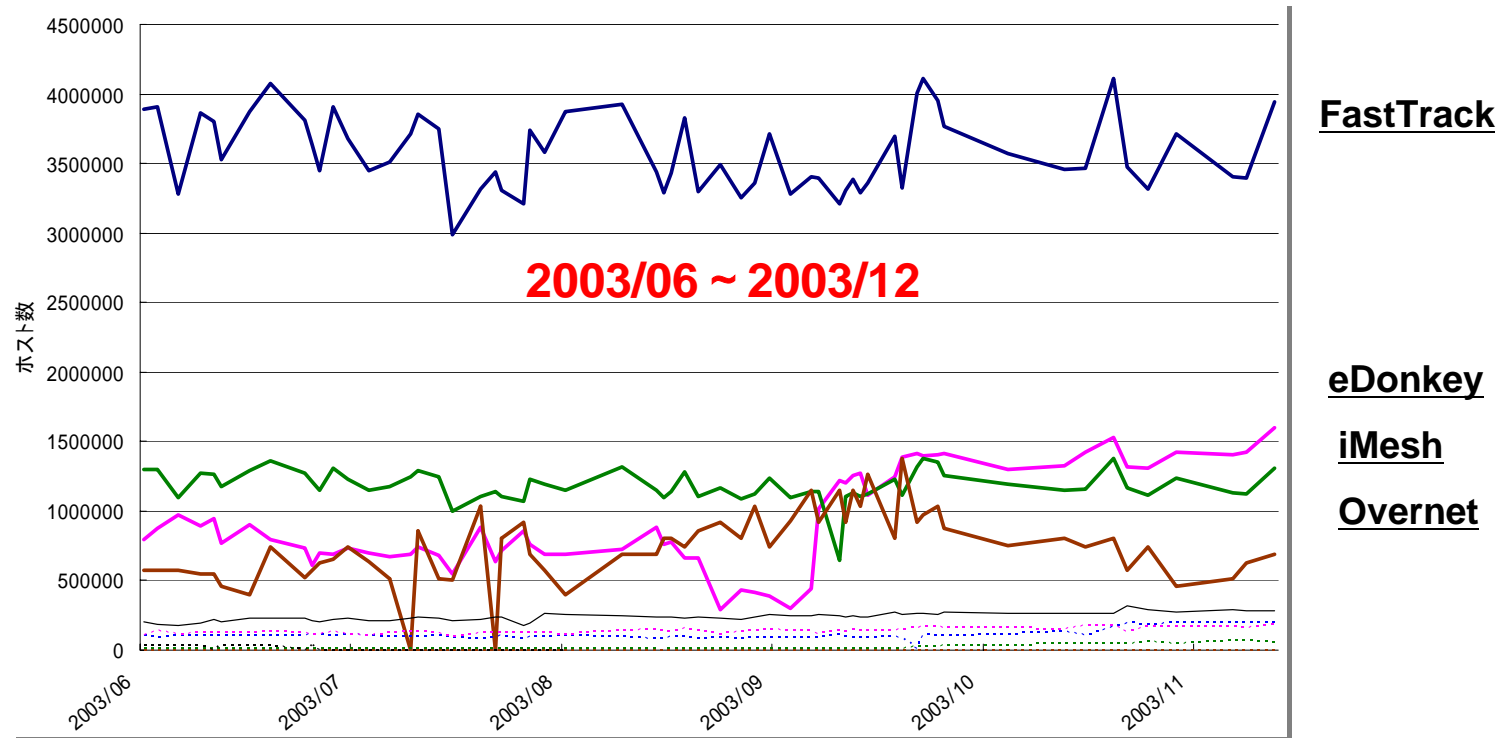


対外接続線のトラフィック占有率遷移(24時間)
2002/11 平日に測定

森, 亀井, 大井, "P2Pトラフィックの測定と特性評価," 信学ソ大, 2003-9.

- ISP内トラフィックとしてはHTTPを大幅に上回る状況も見られる.
- ポート番号による識別なので, Otherの中にも大量のP2Pトラフィックが含まれる可能性が.
- 利用者の生活パターンに依存しないため, Webと異なる挙動.

P2Pネットワークの規模とユーザーのふるまい ～ P2Pサービスの規模情報

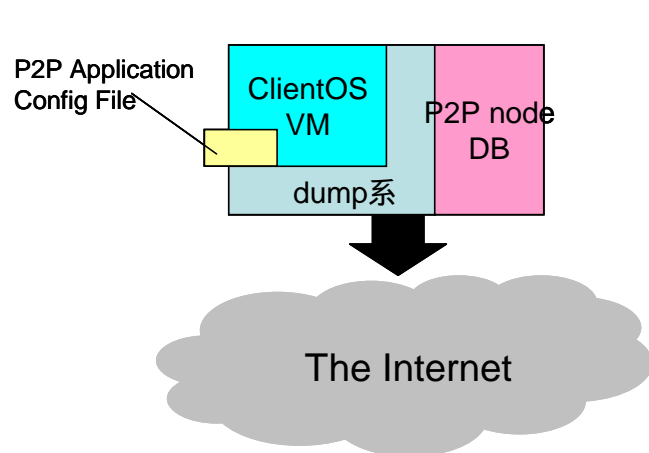


Slyck.com, "SLYCK – File Sharing News and Info," <http://www.slyck.com>.

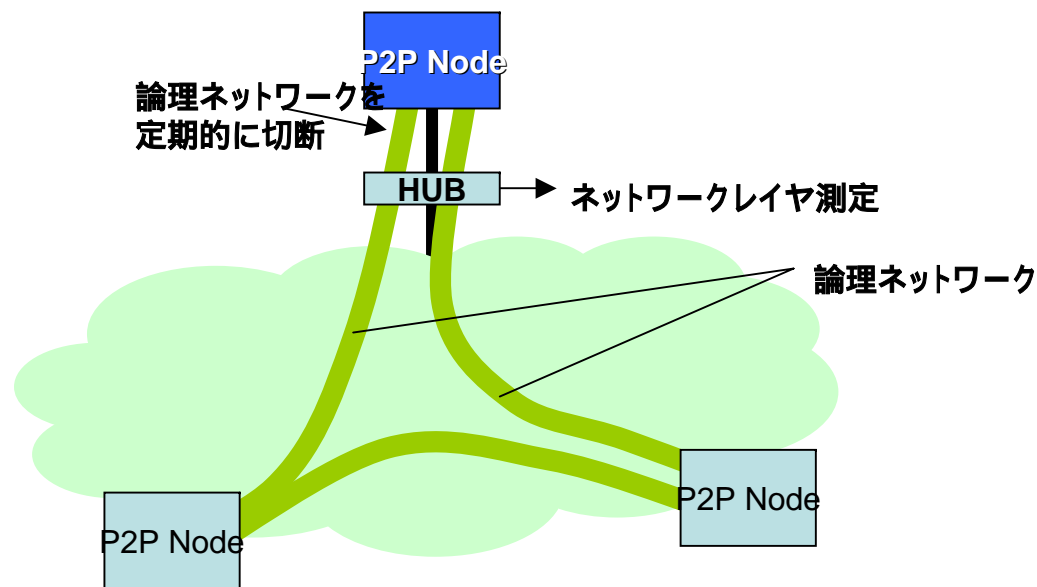
- サービス毎の参加ノード数を日々公開しているサイトからデータを取得して時系列プロット.
- 圧倒的なFastTrack(KaZaA)と,急上昇するeDonkey.
- 規模とトラフィック量は必ずしも比例しないが...
- RIAAの訴訟インパクトが現れてない. 測定方法が不明だが, 全数は取れていない?

論理網トラフィックの測定

- ユーザー権限でアプリケーションを起動し, P2P論理空間で隣接したノードとの通信をIPレイヤで測定する.
- (Pure)P2Pアプリケーションに対して汎用的に適用可能な測定方式
- P2Pネットワークを巨大な系(システム)として捉え, その内部挙動を推定するための基礎技術

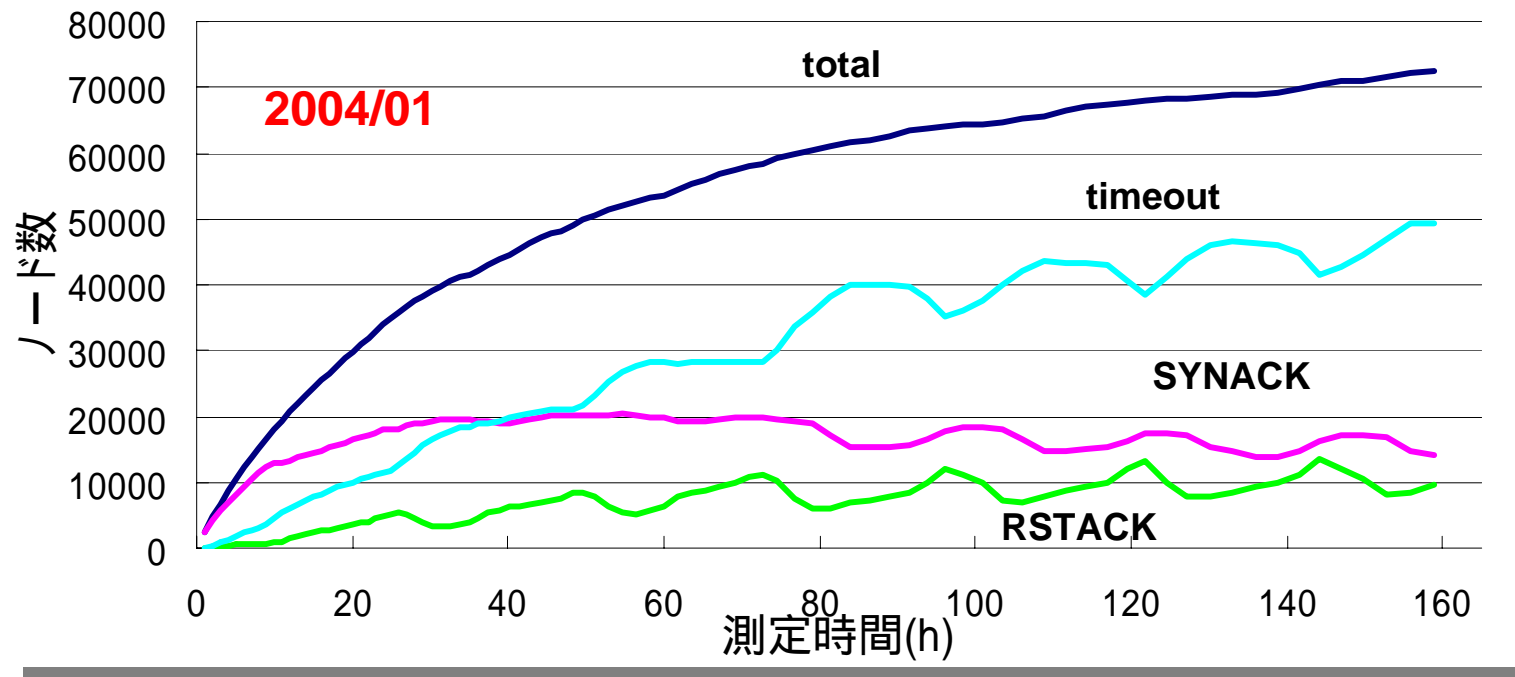


測定ノードのシステム構成



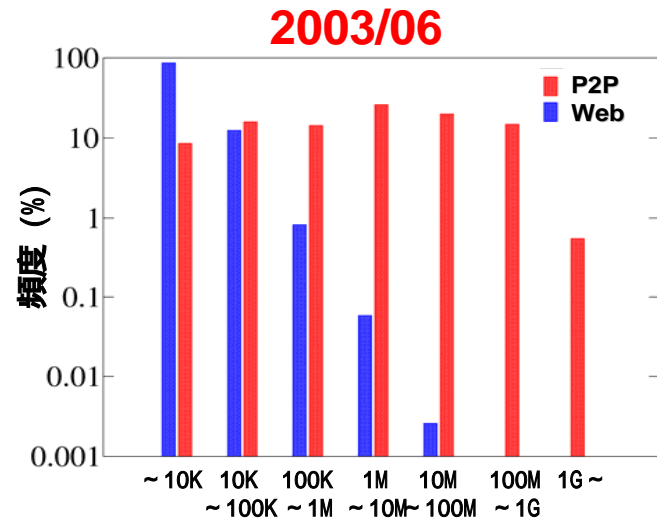
P2Pネットワークの規模とユーザーのふるまい

～ 論理網の測定の結果とアクティブノードの割合

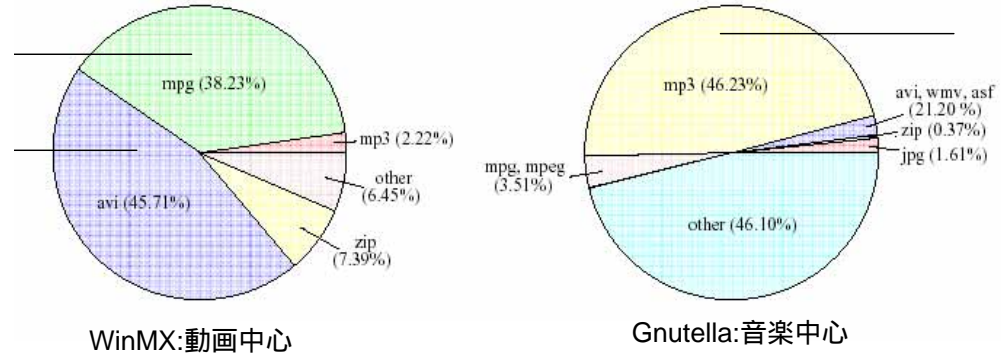


- 8並列で論理網経由でWinny2のノード情報を収集
- IP:port の組に対する hping でのアクティブ測定を実施
- 24時間での変動が見られるが、微妙な割合。ほとんどが常時ONか？
- 逮捕前にはグラフがサチることはなかったが...

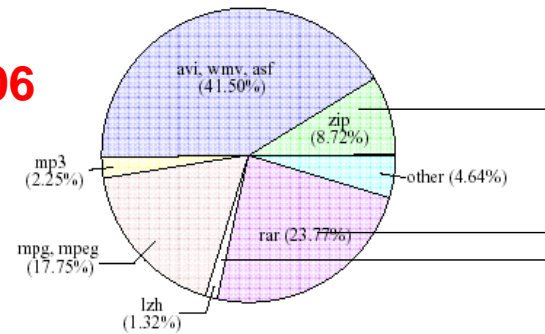
流通コンテンツの分析 ~ ファイルサイズとコンテンツ比率



コンテンツのファイルサイズ



2003/06

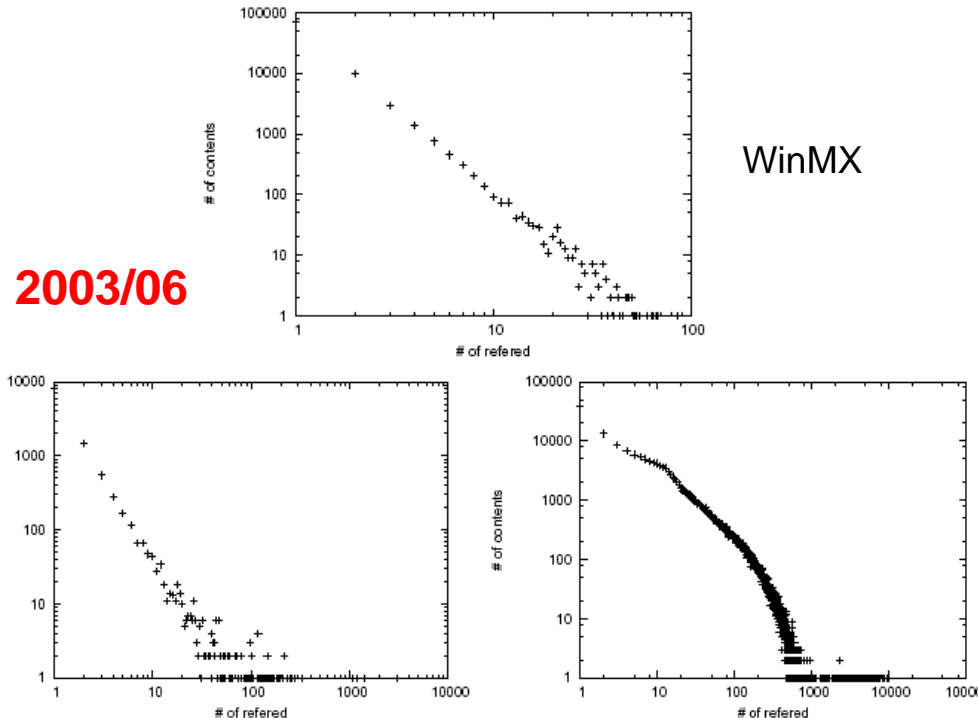


コンテンツの拡張子別コンテンツ分

- 亀井, 森, 大井, “P2Pファイル共有の実態と課題,” 信学技報 CQ-2003-40, 2003-7.
- WWWとは桁違い.

流通コンテンツの分析 ~ コンテンツの需給

➤ x/y : コンテンツ被共有数 n / n 人に共有されたコンテンツ数



➤ Lotokaの法則 (Zipfの第二法則) による直線性

- コンテンツが足りない場合に左上で曲線化 (マイナーコンテンツが不足) することが知られている
- ファイル数が万のオーダーにもかかわらず, 直線性を保っているのは驚異的

制御？規制？

➤ 規制を始めるISP

某ISPの場合(以後数社が追従, 情報公開していないISPも)

- 2002/12 規約改定(他の会員の統計的な平均利用方法と比較して大幅に上回る利用が継続して発生する場合に契約を解除する可能性)
- 2003/10 P2P(WinMX, Winny等)の規制開始を宣言

➤ 技術的には: ヒューリスティックな検知技術はいろいろ

- シグネチャ方式でウィルス/IDSと同様に検知
- 論理網測定で収集したデータを元に検知

亀井, 大井, “P2Pトラヒック分離法とその評価,” 信学ソ大, 2003-9.

➤ 今後は？

- Winny後継アプリケーションの開発(share2, mute etc...)
- 検知技術についてもいたちごっこになる可能性大(今のウィルスと同様?)
- 大規模ネットワークへの適用時には誤認識, 誤動作のリスクも.
- 緊急避難としては有効かもしれないが, 恒常的導入はコストに見合うか?

提言とお願い

➤ 測定と情報公開の重要性

- 公開の場での定量評価がきちんとされていない。
- 公開情報の積み重ねでもある程度の危機的状況は示せるが...
- 継続的な測定が重要。複数ISPでの測定を丸めて開示するとか？

➤ 世論形成が必要

- P2P ファイル共有 違法。
- ベストエフォート, 使い放題, 100Mbps っていったい？
- 「この状況だと規制もしかたない」でカスタマの同意を得るのは困難。

➤ 制御について

- 中長期には設計で対応, 短期的には制御で逃げるというのがあるべき姿。
- 緊急避難的な制御なのか, 恒常的制御なのか。
- リアルタイムアプリをきちんと守る, ベストエフォートの下にクラスを作る？
- P2P自体はネットワークを効率良く使える技術のハズ。
 - アプリケーションとネットワークの連携

参考文献

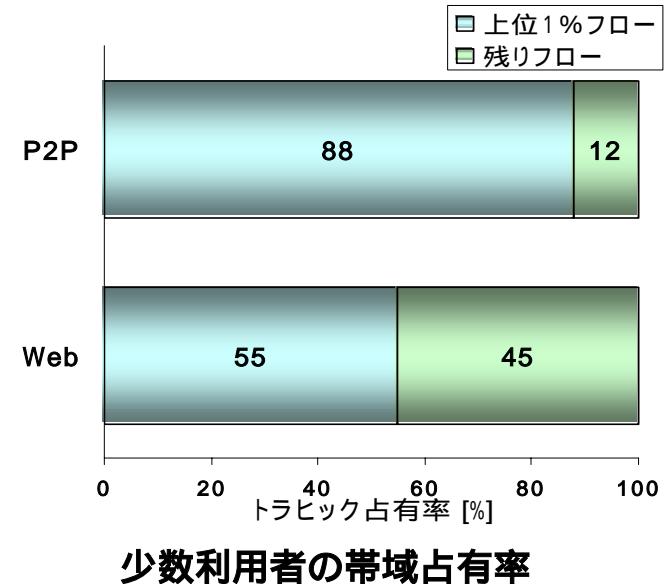
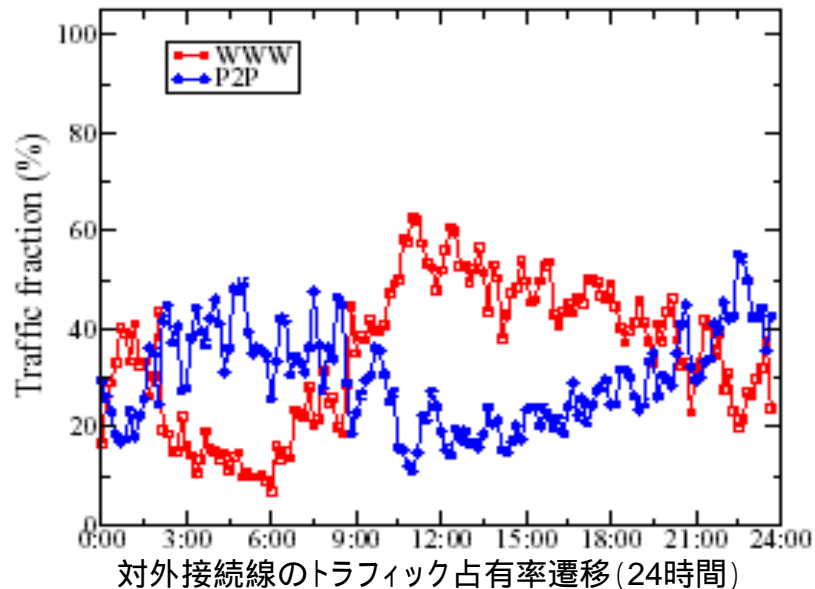
- mfeed, “JPNAP total traffic,” <http://www.mfeed.ad.jp/jpnap/traffic.html>
- 岡田, 川原, “IP網におけるトラフィック特性分析の一考察,” 信学技報NS-2003-5, 2003-4.
- 亀井, 森, 大井, “P2Pファイル共有の実態と課題,” 信学技報 CQ-2003-40, 2003-7.
- 森, 亀井, 大井, “P2Pトラフィックの測定と特性評価,” 信学ソ大, 2003-9.
- 亀井, 大井, “P2Pトラフィック分離法とその評価,” 信学ソ大, 2003-9.
- 亀井, 内田, “P2Pネットワークの規模推定法,” 信学技報 IN-2004-17, 2004-2.
- WIDE Project, “Measurement and analysis on the WIDE internet”, <http://www.wide.ad.jp/wg/mawi/>, 1988.
- Slyck.com, “SLYCK – File Sharing News and Info,” <http://www.slyck.com>.

➤ SFCでは2003/03に注意勧告

- JASRACよりの警告文書の存在？
- 研究目的での利用は届出制
- 対象アプリケーションは,
 - Gnutella 及びその互換ソフト (BearShare, LimeWire, Gnutella Light, FreeWire, NeoNapster, Qtraxmax, Gnucleus, Mactella など)
 - WinMX 及びその互換ソフト (iShare, Utatane など)
 - Winny
 - Bit Torrent (2003/08/13 追加)
 - KaZaA
 - eDonkey 及びその互換ソフト(OneMX など)

慶応義塾大学インフォメーションテクノロジーセンター, “現在, 利用申請が必要なP2Pファイル交換ソフトウェアのリスト,” <http://www.itc.keio.ac.jp/p2p/p2plist.htm>, March, 2003.

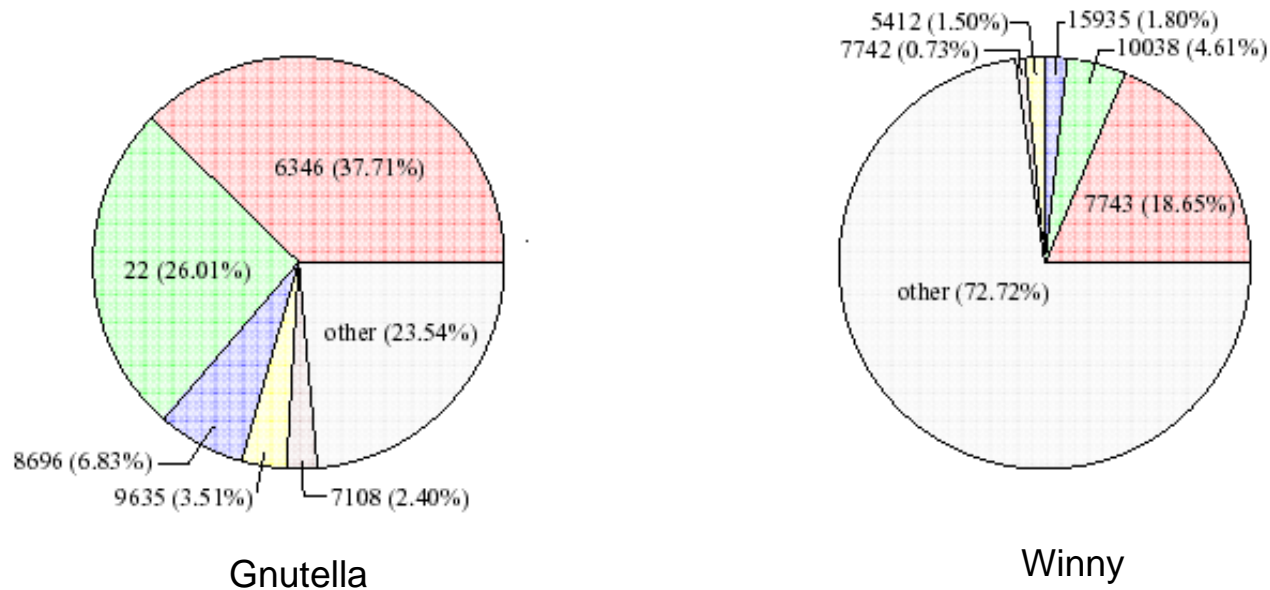
P2Pとそのほかのトラフィック共存制御に向けた分析



- 森, 亀井, 大井, “P2Pトラヒックの測定と特性評価,” 信学ソ大, 2003-9.
- P2Pでは少数利用者が帯域の大部分を占有している現状.

参考 アプリケーションレイヤ測定 ~ ポート番号分布

➤ デフォルトポートの使用率(パケット数)



- デフォルトポート利用はGnutellaで38%, Winnyで19%
- ポート番号による統計値の2~5倍の潜在P2Pトラフィックの存在

P2Pトラヒック分離技術 ~ 既存技術とその問題点

➤ 既存のトラヒック識別手法と問題点

a. ポート番号による識別

- TCPのポート番号を利用して識別
- ✓ ポート番号可変, well-known-portを使うアプリケーションの存在.

b. セッション追跡

- 中央サーバへの検索パケットをトリガーとして, 以降の通信を追跡
- ✓ 中央サーバが存在しないアプリケーションへの適用困難.
- ✓ プロトコルの解析が必要.

c. ペイロード監視通信, パターン認識

- ペイロード中の文字列やパケット長等の特徴を利用
- ✓ 誤認識の危険性
- ✓ プロトコルの解析, パターン分析がアプリケーション毎に必要.
- ✓ 大量のパケット監視, 通信内容の取得が必要, スケールしない.

d. アプリケーションレイヤ解析

- アプリケーションレイヤでプロトコルを解析
- ✓ パケット監視, 組立が必要でスケーラビリティが低い.
- ✓ プロトコルが既知である必要がある.



既存方式による識別法(ダンプデータ例)

市販装置:

c. P-Cube/Ellacoya/Packeteer/Allot

d. SandVine/CacheLogic/PeerCache

スケーラビリティ, アプリケーション依存の問題が大きい. また, 最近のP2Pアプリケーションは通信の暗号化や, ポート番号のランダム設定等, 隠蔽機能の高度化が進み, ウィルスパターン等と同様に個別の対応が必要となる.

参考

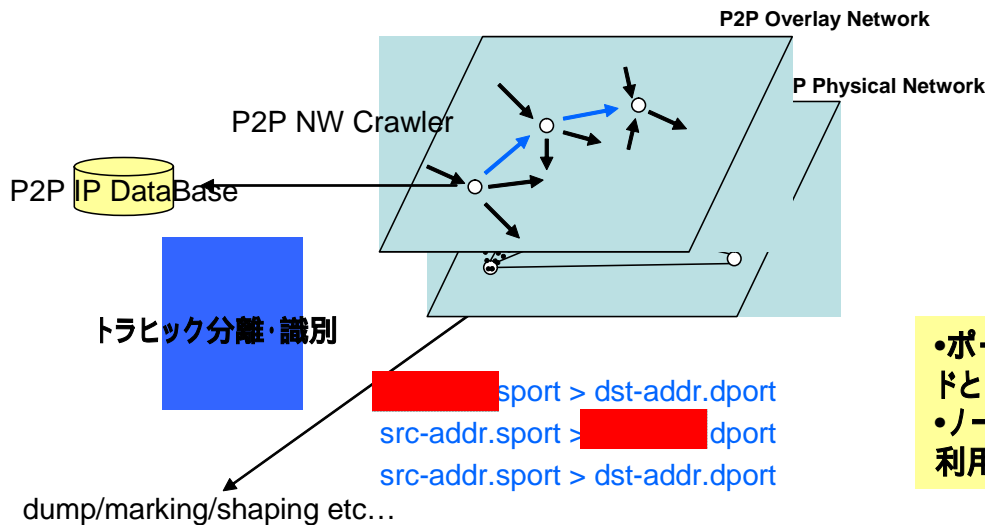
P2Pトラフィック分離技術 ~ 提案手法とその評価

提案アルゴリズム

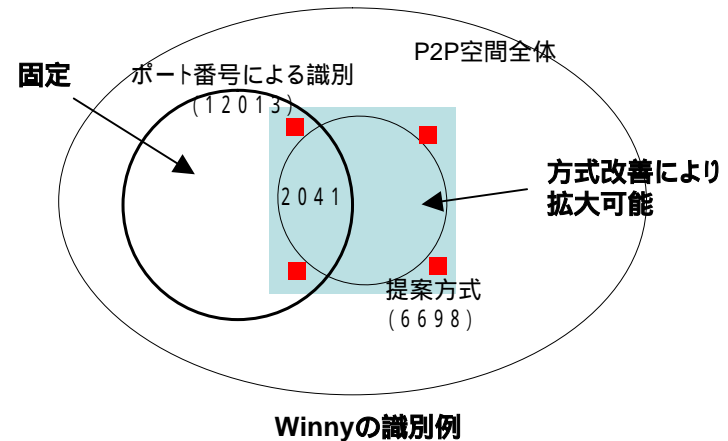
1. 任意の場所においてP2Pネットワークに接続したノードを設置, IPレイヤでの通信相手(P2Pレイヤでの隣接ノードのIPアドレス)を記録
2. 収集したアドレス情報に有効期限を設定し, P2P IPデータベースに登録
3. P2Pトラフィック識別対象の通信において送受信アドレスをデータベースに問合せ
4. 問い合わせ適合時にこれを P2P 通信と識別



ユーザ権限でソフトを起動できれば対応可能
アプリケーション非依存
アドレスとポート番号によるマッチング処理のみ
スケーラビリティに優れる



提案アルゴリズム



- ポート番号識別との併用で, 40%のノードを新たにP2P利用ノードとして識別.
- ノード情報収集方式の改善(P2P論理網トラフィック測定技術を利用)で, さらなる機能向上がみこまれる.