

ログ解析による Winny ネットワークの構成ピア数推定法の一検討

東京農工大学 *蜂須賀 大紀 HACHISUKA Daiki

東京農工大学 大坐畠 智 OHZAHATA Satoshi

0110437 東京農工大学 川島 幸之助 KAWASHIMA Konosuke

1 はじめに

近年、インターネットにおいて Peer-to-Peer (P2P) アプリケーションが急速に普及している [1]. そのなかで、国内で最も広く用いられている P2P アプリケーションのひとつが Winny (Version2) である. 特に Winny では大容量のマルチメディアファイルが扱われることが多く、帯域の多くを消費し、ネットワークに多大な影響を及ぼしている. このような利用に耐え得るネットワークを設計するには、Winny ピア数を把握することは重要である. Winny はネットワークを管理する中央サーバを必要とせず、全ピア数把握が困難なシステムである [2]. 本稿では Winny ピアの規模を推定する手法を提案し、実測定によって得られた Winny のデータをもとにコンピュータシミュレーション実験を行い、Winny ネットワークを構成するピア規模を推定する.

2 Winny ピアの規模推定法

まず、Winny ピアの IP アドレスを取得するマシンを用意する. Winny ピア数は時間の経過とともに増加し、重複のない IP アドレスを数えていく. はじめに、重複のない累計 IP アドレスの集合を U_a 、累計後の 1 時間で得られる IP アドレスの集合を V_a 、そして U_a 、 V_a で重複することのない新規 IP アドレスの集合を W_a とすると、

$$W_a = V_a - (U_a \cap V_a)$$

と表すことができる. どの程度の割合のピアがこの時点で新規ピアかは下記で表せる.

$$|W_a| / |V_a|$$

これを利用し、この時点で探索した空間の割合が推測できることから、この時点での推定される重複のない累計 IP アドレスの集合を U'_a とすると、

$$|U'_a| = |U_a| / (|W_a| / |V_a|)$$

となる. 次に推定される新規 IP アドレスの集合を X_a とし、 U'_a の t 時間経過後を U'_b とすると、単位時間あたりの新規ピア数は

$$|X_a| / t = (|U'_b| - |U'_a|) / t$$

と求まる.

Winny ネットワークを構成するピア数の規模を推定するためには、今まで求めてきた値に加え、ピ

アの生存時間を測定する必要がある.

この生存時間と新規 IP アドレス数 X_a の時間推移を組み合わせ、コンピュータシミュレーション実験を行うことにより規模を推定することが可能となる.

3 測定実験

3.1 測定実験の概要

Winny のピアを取得するマシンでは、20 の Winny プロセスを走らせ実験を行った. ここで得られたデータから、Winny の新規ピアの割合の算出と生存時間の測定を行った.

3.2 測定実験結果 1

まず Winny 新規ピアの割合を算出した. データは 2004 年 6 月 13 日から 2004 年 10 月 23 日の間に得られたものを利用した. ただし、2004 年 8 月 9 日から 2004 年 8 月 13 日の間はマシンを移動させていなかったため、その期間のデータはないものとする. 測定では、1 週間単位でピア数をまとめ、その次の日のはじめの 1 時間に参加するユーザとの重複を調べることにより、Winny 新規ピアの割合を算出した. 図 1, 2, 4 で“推定値”は前節の推定を用いたもので“実測値”はそのままの値を用いたものである.

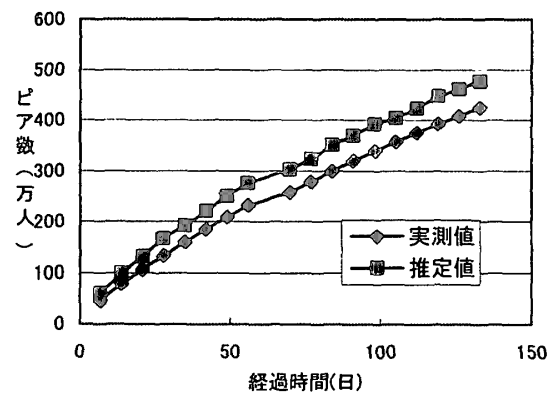


図 1 累計 Winny ピア数.

図 1 から、実測値も推定値も時間経過とともにピア数が増加し続けていることがわかる. 図 2 から、推定値が第 9 週辺りから増減を繰り返してい

ることがわかる。よって、値はこの付近で収束すると考えられる。

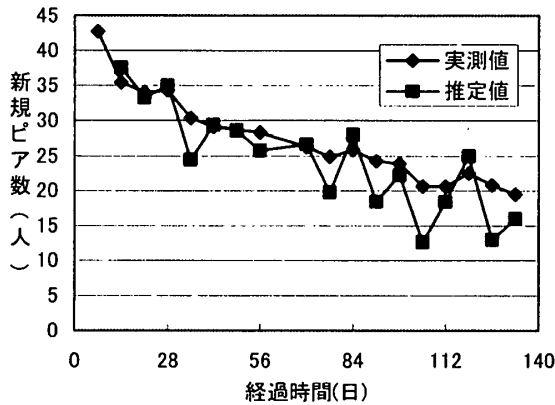


図 2 新規 Winny ピア数(1分当たり).

3.3 測定実験結果 2

Winny ピアの生存時間を測定した。ピアが生存しているかどうかは、SYN パケットに対して、SYN ACK が返されているかで判断した。しかし、前述したとおり実験で使用したマシンでは、ピアを取得できた場合 10 秒以内でその接続を切るよう設定している。このままの状態では生存時間を測定すると、生存時間が 10 秒以内のものが異常に多くなり、正確な生存時間を測定することができない。そこで、2 か所以上の IP ポートから接続があったピアの生存時間を測定することにした。このことにより、生存時間が短時間のピアを除くことができた。想定データは 2004 年 7 月 1 日から 2004 年 7 月 31 日の間に得られたものを用いた。

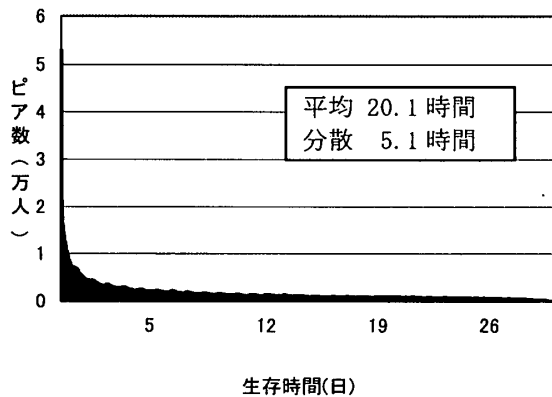


図 3 Winny ピアの生存時間.

図 3 から、Winny では短時間利用のピアが多いこととともに、長時間利用のピアも少ないながら存在することが確認できた。

4 シミュレーション実験

4.1 シミュレーションの概要

ピアの生存時間と新規 IP アドレス数の時間推移から、Winny ネットワークを構成するピア数をシミュレーションした。新規ピア数を 4 週間分まとめ、ピアの生存時間と組み合わせ推定数を算出する。また、新規ピア数の割合を時間経過にともない、1 週間分ずつ更新していく。測定データは、ピアの生存時間に関しては 2004 年 7 月 1 日から 2004 年 7 月 31 日まで、新規 Winny ピア数は 2004 年 6 月 13 日から 2004 年 10 月 23 日までに得られたものを利用した。ただし、2004 年 8 月 9 日から 2004 年 8 月 13 日の間のデータはないものとする。

4.2 シミュレーション結果

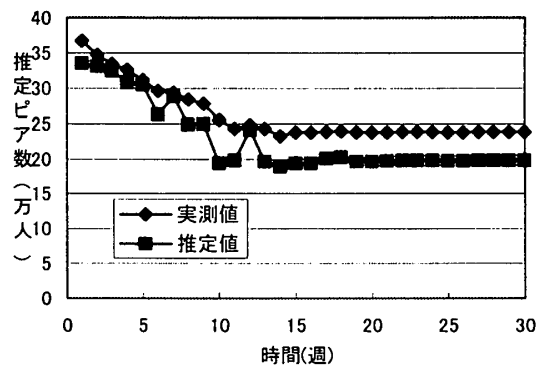


図 4 推定 Winny ユーザ数.

図 4 から、推定値が第 15 週辺りから増減を繰り返す、値はこの付近で収束すると考えられる。この区間の平均を算出すると、推定 Winny ピア数は約 20 万との結果となった。

5 おわりに

本稿では、全体測定が困難である P2P ネットワークにおける全体推定法の提案した。提案方式を用いて実際の大規模 P2P ネットワークである Winny のネットワークを構成するピア数の推定を行った。今後は日々変化する Winny ピアを正確に推定するために、常に新しいデータを用いて測定することが必要である。

参考文献

- [1] 亀井 聡, 内田 真人, “P2P ネットワークの規模推定法,” 電子情報通信学会技術研究報告, IN2003-204, pp. 41-45, 2003.
- [2] Steve Lawrence, C. Lee Giles, “Searching the World Wide Web,” Science, vol. 280, pp. 98-100, 1998.