

P2P 方式による新規画像通信インフラ技術の研究

浅井 有人*, 湊川 宏*, 羽田 典久*

Peer-to-peer-based Image Networking for Adaptive Infrastructure

Arito ASAI*, Hiroshi MINATOGAWA*, and Norihisa HANEDA*

Abstract

While digital cameras and broadband connection to home are becoming common, data storage and network cost are growing heavy for service provides, especially for image-dealing service providers. It's important to distribute and lower these concentrations of cost.

As a result of this research, a P2P networking technology has been developed that enables services using low quality connection such as ADSL and noticeably reduces the cost for net printing service.

1. はじめに

近年、デジタルカメラの高解像度化による画像データサイズの巨大化、および大容量コンテンツを快適に通信できるブロードバンド回線の一般家庭への普及が進んだことにより、大容量静止画データをネットワーク上で流通させることによるビジネス環境は大きく変化している。たとえば、ユーザは高解像度デジタルカメラで撮影した大量の巨大画像データを、ネット上のプリントサービスに、安価な常時接続のブロードバンド回線を通じて気軽にアップロードすることができる環境が整備された。

サービスを提供する側では、サービス提供機会の増加の一方、高速回線を通じてアップロードされる巨大な画像データを受信するサーバ側の回線容量、価格は依然として高価であり、従来のサーバ・クライアントモデルでは、顧客数が増えるほど単一のサーバに負荷が集中し、サービス不能状態（パンク状態）に陥るなど、サーバがボトルネックになる状況下にある。

これらの状況を解決するためには、ボトルネックとなる集中サーバを排除し、効率よく巨大な画像データをネットワーク上で取りまわす新しいネットワーク方式が必須である（Fig. 1）。本報告では、ピア・ツー・ピア方式を利用した画像通信技術の開発を行い、実際にネットプリント実験サービス（Fig. 2）を行った結果と、その効果について報告する。

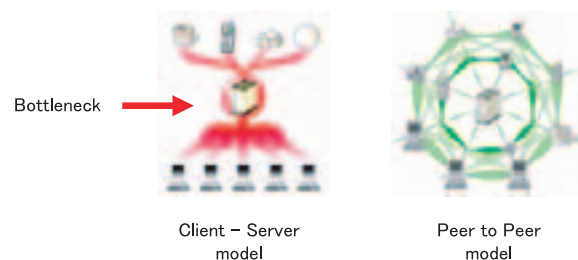


Fig. 1 Two models of network topology.

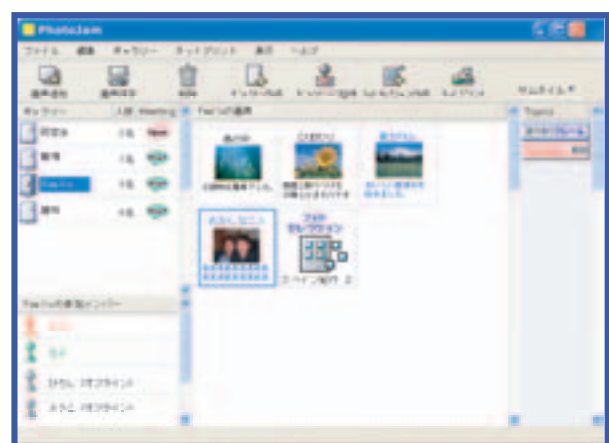


Fig. 2 Image viewer and net-print application with P2P.

本誌投稿論文（受理2004年11月22日）

*富士写真フイルム（株）新規事業開発本部

DIソフト事業推進部

〒351-8585 埼玉県朝霞市泉水3-11-46

* Digital Imaging Software

Business Development Division

Fuji Photo Film Co., Ltd.

Senzui, Asaka, Saitama 351-8585, Japan

2. 実験の概要

本研究の実証実験は、ピア・ツー・ピア方式の画像通信方式について研究を行った後、本技術を実際の市場にてネットプリント機能つき画像ビューアソフトとして提供することによって、実際のネットワーク上での通信負荷の分散効果の測定、およびピア・ツー・ピア方式に起因するさまざまな課題について検証を行うことを目的として実施された。

実際に、首都圏の20店舗のチェーンラボと、郵送発送可能な郵送ラボの合計21店舗を用いてネットプリントサービスを実施し、ピア・ツー・ピア技術を搭載した画像ビューア兼ネットプリントソフトを店頭、およびネットで配布・告知を行った (Fig. 3, Table 1)。

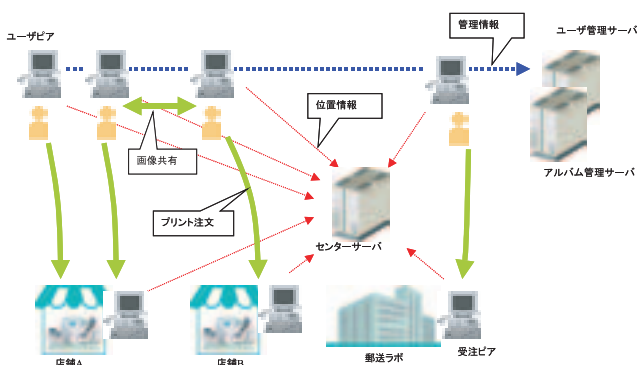


Fig. 3 System structure for the experiment.

Table 1 Experimental Conditions.

実験期間	2004.2.1～2004.3.31
実施店舗数	20店舗+郵送1店舗
参加ユーザ数	1181人
プリント注文枚数	31670枚
プリント注文件数	657件

3. クライアント・サーバ方式のコスト限界

ネットプリントサービスにおける、クライアント・サーバ方式の理論的限界について試算する。サーバのインターネット回線の下り通信速度を1Mbpsと仮定し、サービス停止とならないよう、ピーク時に回線容量限界のトラフィックが流入するものとして計算する。下記計算式中の「実測値」はそれぞれ、前章で述べた実証実験から測定された実際の値である。

(1) 1時間あたりの最大受信データ量

$$1\text{Mbps} / 8 * 60 * 60 = 450\text{MB} / \text{h} \quad \dots (a)$$

(2) 1時間あたりの最大受注枚数

ネットプリントにおける1枚の画像平均データサイズ (実測値) = 850KB / 枚

$$450 (\text{MB} / \text{h}) / 850 (\text{KB} / \text{枚}) = 530 \text{枚} / \text{h} \quad \dots (b)$$

(3) 1日あたりの最大受注枚数

実証実験における最頻時間係数を求めると、1日のプリント注文トラフィックの10%が1時間に集中することが判る (Fig. 4)。

$$\text{ネットプリントの最頻時間係数 (実測値)} = 10\% \\ 530 (\text{枚} / \text{h}) / 10\% = 5,300 \text{枚} / \text{day} \quad \dots (c)$$

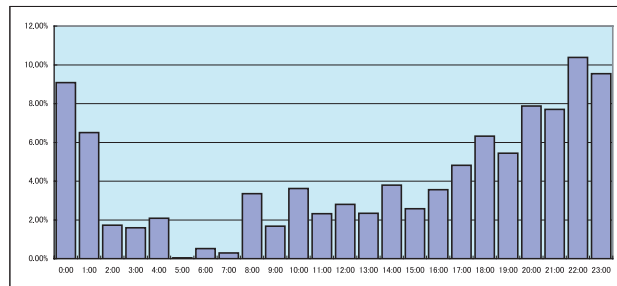


Fig. 4 Hourly traffic.

(4) 1週間あたりの最大受注枚数

同様に、1週間のプリント注文の20%が日曜日に発生する (Fig. 5)。

$$\text{ネットプリントの曜日分布係数 (実測値)} = 20\% / \text{day} \\ 5,300 (\text{枚} / \text{day}) / 20\% = 26,500 \text{枚} / \text{week} \quad \dots (d)$$

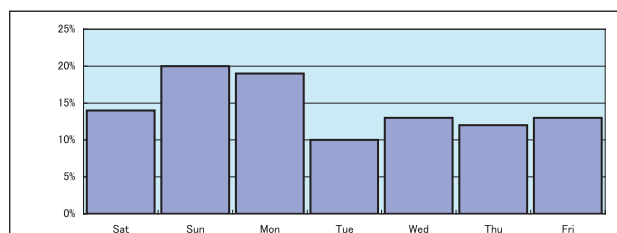


Fig. 5 Daily summary of orders.

(5) 1ヶ月あたりの最大受注枚数

$$1 \text{ヶ月は平均} 4.37 (= 30.6 / 7) \text{ 週である。} \\ 26,500 (\text{枚} / \text{week}) * 4.37 = 116,000 \text{枚} / \text{月} \quad \dots (e)$$

以上の計算により、1Mbpsの回線で運用するサーバ型ネットプリントサービスは最大で月間11,600枚の注文を受注することができる。1Mbpsの回線コストは15万円/月となる (O社IDC例 2004年5月調べ)。従って、ネットプリント1枚あたりの最低回線コストは以下のようになる。

$$1 \text{枚あたりの回線コスト} \\ = 150,000 (\text{円} / \text{月}) / 116,000 (\text{枚} / \text{月}) \\ = 1.29 \text{円} / \text{枚}$$

4. ピア・ツー・ピア方式のネットワーク負荷

実証実験では、ネットワークの負荷耐性について調査するために、故意にトラフィックの集中を発生させ

るトラフィック集中実験を行った。1,039人のユーザに対してメールによる告知を行い、注文数が最大になる日曜日の最頻時間帯の2時間限定で無料プリントキャンペーンを実施した (Table 2)。

Table 2 Conditions of the heavy traffic experiment.

注文件数	122件
注文枚数	5133枚
総データ転送量	4.33GB

この2時間のセンターサーバ、各店舗、郵送センターへの注文トラフィック流入量の集計結果はTable 3の通りである。特に、センターサーバへのトラフィックの推移はFig. 6の通り、キャンペーン期間中の集中的なトラフィックの増加が観測されている。

Table 3 Traffic Analysis.

	(単位 Kbps)	
	ピーク時	平均
プリント注文全体	7,998.5	4,115.1
郵送センター	3,625.0	2,339.1
各店舗	N/A	98.7
センターサーバ	2.6	1.5

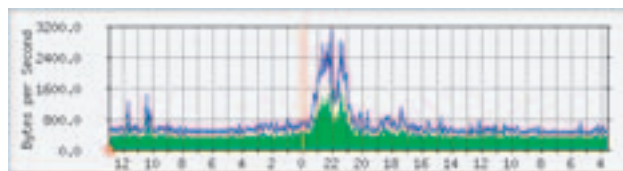


Fig. 6 Traffic into the center server.

センターサーバの通信量は、ピーク時でも2.6Kbps程度であり、通信量の負荷集中のスケールはきわめて小さい。プリント注文全体で平均して4.1Mbpsの帯域を消費したが、各店舗へのトラフィックは分散され、それぞれ平均100Kbps程度と、通常家庭用ADSL回線で十分に余裕をもって処理することができる。実験時の参加店舗が首都圏に偏っていたことも影響し、郵送センターへの通信トラフィックがやや集中したため、ピーク時3.6Mbpsと実際のADSL回線の限界に近い値を示している。

センターサーバへの流入トラフィックは、店舗管理のための定常的なトラフィックに加え、プリント注文時の店舗検索によるリクエストによって構成される。前者は店舗数に比例し、後者はプリント注文件数に比例する。プリント枚数については、実際の画像データがセンターサーバに直接流入しないので影響を受けない。

(1)1店舗あたりのセンターサーバへの負荷要求量

定常状態において、センターサーバへのトラフィックは平均0.7Kbpsであった。これは、郵送を含めた21店舗による管理トラフィックと考えると、1店舗あたりの負荷要求量は以下の通りとなる。

$$0.7\text{Kbps} / 21\text{店舗} = 0.033\text{Kbps} / \text{店舗} \quad \dots (a)$$

(2)1注文あたりのセンターサーバへの負荷要求量

集中実験において、平均1.5Kbpsのトラフィックで時間あたり61件の注文を処理している。上記管理トラフィックを差し引いて計算すると、以下の通りである。実際に注文行動を開始して注文完了にいたらないユーザも存在するので、これらの計算には実測値を用いる。

$$(1.5\text{Kbps} - 0.7\text{Kbps}) / 61\text{件} = 0.013\text{Kbps} / \text{件} \quad \dots (b)$$

(3)1店舗における注文込みのセンターサーバへの最大負荷要求量

店舗ADSL下り回線速度を平均2Mbpsと想定すると、1時間あたりの受注最大枚数は以下の通りである。

$$530\text{ (枚/h} \cdot \text{Mbps)} \times 2\text{Mbps} = 1060\text{ 枚/h} \quad \dots (c)$$

注文1件あたりの注文枚数は平均26枚である (当社調べ)。従って、1時間あたりの最大注文件数は以下の通りである。

$$1060\text{ (枚/h)} / 26\text{ (枚/件)} = 41\text{ 件/h} \quad \dots (d)$$

従って、注文トラフィック最大時の1店舗のセンターサーバへの負荷要求量は以下の通りである。

$$0.033\text{Kbps} + 0.013\text{Kbps} * 41 = 0.57\text{Kbps} / \text{店舗} \quad \dots (e)$$

(4)1Mbpsで接続するセンターサーバの収容できる店舗数

$$1\text{Mbps} / 0.57\text{Kbps} = 1750\text{ 店舗} \quad \dots (f)$$

(5)ネットワーク全体での1時間あたり最受注枚数

$$1750\text{ (店舗数)} \times 1060\text{ (枚/h} \cdot \text{店舗)} = 1,855,000\text{ 枚/h}$$

月間で計算すると、前章で述べた通りの計算方法で406,000,000枚の受注能力があることがわかる。

このとき、1枚あたりのセンターサーバ回線コストの最低値は以下の通りである。

$$\begin{aligned} &1\text{枚あたりのセンターサーバ回線コスト} \\ &= 150,000\text{ (円/月)} / 406,000,000\text{ (枚)} \\ &= 0.00037\text{ 円/枚} \end{aligned}$$

上記の試算は、センターサーバの回線に対する限界負荷を想定した試算であり、コスト的には理想的最低値であることに注意を要する。また、各店舗におけるインターネット接続のためのコストが必要であるが、これはセンターサーバ方式においても、センターサーバと店舗間を接続するコストと同じであると考えることができる。

5. ピア・ツー・ピア方式の通信技術

一般コンシューマー向けADSL回線を用いたピア・ツー・ピア型のネットワークにおいて、安定したサービスを提供するために解決すべき技術課題と、本研究における解決方法について述べる。

5.1 DHCPによるIPアドレスの動的割当

コンシューマー向けのインターネット回線のほとんどは、端末に対して固定的なIPアドレスを付与せず、DHCPによって動的にIPアドレスを割り当てる。受注店舗のIPアドレスが頻繁に変化する場合、既存のDNSによるIPアドレスの解決を行うことができない。ダイナミックDNSによる解決は、伝播タイムラグの存在やホスト停止時の検出に時間がかかるため最適ではない。

解決しなければならない問題は、(1)店舗のIPアドレスが変更されたときに、すみやかに顧客が変更後のアドレスを解決して接続することができること、(2)店舗ホストが万一障害などで停止している際に、誤った(古い)IPアドレスを解決して、無関係なホストへの接続試行あるいは接続失敗が発生しないこと、の2点である。

上記問題を解決するために、ホストのIPアドレスを動的に管理するセンターサーバが必要になる。店舗ホスト(ピア)は、センターサーバに接続してその存在とIPアドレスをセンターサーバに登録しなければならない。しかし、ピアは必ずしも自身のIPアドレスを知らないかもしれない(下記NATルータ参照)ので、センターサーバは接続されたTCPのソースアドレスをピアのIPアドレスと仮定し、逆に、センターサーバ側から当該IPアドレス宛てに別のTCPセッションを張り、正しく当該ピアが応答することを確認した後、そのIPアドレスを当該ピアのものとして登録しなければならない。セキュリティ上の店舗なりすましを防止するために、センターサーバは当該店舗の認証を同時に行う。

さらに、TCP/IP上では通知のない接続の切断を検知することができないので、ピアは十分短い時間ごとに定期的にセンターサーバにその生存を報告する義務を負う。センターサーバは一定期間生存報告のないピアを切断状態と判断し、登録から抹消しなければならない。

5.2 PPPoE接続の切断

コンシューマー向けのブロードバンド回線の多くは、回線の確立にPPPoEを用いる。PPPoEはアプリケーションで用いるTCP/IPの下部レイヤとなるため、上位TCP/IP層ではPPPoEの切断を検知することができない。PPPoEの切断は、局側の工事、不用意な電話線の抜線、落雷、回線上の雑音などの理由で発生する。PPPoEをブロードバンドルータに処理させる場合、LAN側の端末からPPPoEの再接続を指示することは一般にはできな

い。従って、ブロードバンドルータ自身にPPPoEの自動再接続を行う機能がついていることが必須である。

5.3 NATルータ

通常のブロードバンドルータは、ISPから付与される唯一のグローバルIPアドレスに対して複数台のLAN側の機器を接続するため、NATルーティングを行う。NATの存在はいくつかの問題を抱えている。(1)外部からNATルータを超えたLAN側の機器への接続到達性の問題、(2)長時間にわたるTCPセッションにおいて、無通信時のNAT変換テーブルのタイムアウト問題。

外部からの接続をLAN側機器にて受信する方法はいくつかの方法が存在する。

第一に、店舗側で注文を受けようとする端末については、NATルータの静的NATエントリに登録することで、特定ポート宛てのパケットを特定LAN側IPアドレスにマッピングすることができる。この方法が現在のところ最も安定である。ただし、ルータの設定変更を伴うため多少高度な知識を要求する。

第二には、uPNP(Universal Plug and Play)を用いて、LAN側端末上のアプリケーションからNATルータを検出し、NATルータに対して動的にマッピングを登録する方法である。uPNPに対応したNATルータであれば、おおむね正しく動作する。しかし、多くのNATルータのuPNP機能はNetMeetingアプリのみに対応したものであり、動作するものの細かな非互換性を逐次検証していかなければならない。また、ルータ機種・設定によって、uPNPによるNATエントリにはタイムアウトがあり、一定期間内に更新をしなければならない。

第三の方法として、第三者ピアに対してあらかじめTCPコネクションを確立しておき、この第三者ピアを経由してパケットを受信する方法がある。この方法は、ピア・ツー・ピアネットワーク枠におけるパケット中継技術であり、重要なキー技術である。

ユーザ間での画像共有、情報交換などの通信においては、ユーザ端末間でサーバを介さないコネクションの確立がピア・ツー・ピアの本質である。このとき、ユーザ端末のどちらかがNATルータの背後にある場合(以降ダーティピアと呼ぶ。同様に、NATを用いず直接接続されている端末をクリーンピアと呼ぶ)、双方向の単純なコネクションは拒否されるため、ダーティピアにパケットを中継する中継ピアの存在が必要である。

センターサーバは、各ピア接続時の逆接続診断の結果により当該ピアが外部から接続できないことを検知した場合、他の中継可能なピアを指示する。ダーティピアは、指示された中継ピアに中継依頼を要求し、中継コネクションの確立後、あらためてセンターサーバに登録要求を行う。ある一定の割合のクリーンなピアがネットワーク上に存在すれば、センターサーバで適切に複数の中継ピアを用いて経路を分散させることが可能である。

Fig. 7は、ピアAからダーティピアCに中継ピアBを介して接続するシーケンスを表したものである。

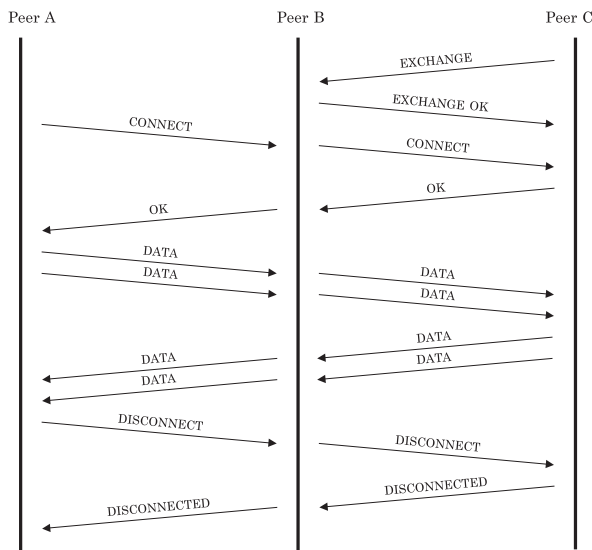


Fig. 7 Transport sequence via an exchange peer.

さて、ピアCから開始されたピアB-C間のTCP接続は永続的に保持され、ピアBを介して複数の接続を単一のTCPセッション上に展開しなければならない。このため、TCP上にピア・ツー・ピアの新たなトランスポートレイヤを用意することになる。

さらに、ピアB-C間のTCP接続は、ダーティなピアCが外部からのコネクションを受け入れるために長期間に渡り保持するものであるため、次に述べる二つの理由により、このコネクションの維持には新たな診断プロトコルが必要になる。第一の理由は、中継ピアBが通知無しに切断される可能性である。この場合、TCP上でデータを流さない限りピアCは切断を検知できず、かつピアCは受信待ちでブロックしているため、ピアBの切断に対して永久に外部からの転送を受けることができない。第二の理由は、NATルータの動的NATテーブルのタイムアウトである。ピアCのNATルータは、ピアCからのTCPセッション開始時にポート・IPアドレス組の変換テーブルを登録するが、中継するルータの位置ではこのTCPセッションの寿命を検知することができないため、一定時間の無通信をもって当該NATテーブルを破棄する。破棄された後、ピアBからの転送データが到着した場合、ルータはこのデータを破棄するし、ピアCはデータを受信できないし、NATテーブルが破棄されたことを検出できず、ESTABLISHEDのままの状態を保持してしまう。

中継コネクションを維持するためには、ピアB-C間で定期的に接続確認のパケットを流してタイマのリセット、および切断検知時のリカバリを行わなければならない。この機能の実装のために、中継TCPコネクションの上位レイヤには、アプリケーション独自のメッセージを流すための複数のトランスポートの他に、接続維持のための診断プロトコルを定義する必要がある (Fig. 8)。

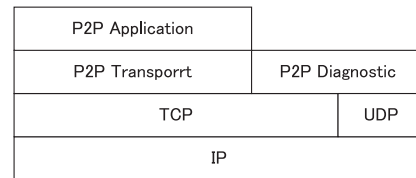


Fig. 8 P2P protocol hierarchy.

5.4 セキュリティー

通信におけるセキュリティーとは、(1) 盗聴、(2) 改ざん、(3) なりすまし、の三点の攻撃から安全に情報を保護することにある。特に、ネットプリントサービスにおいては、個人情報、課金情報をユーザ端末と店舗端末の間で安全に通信できなければならない。盗聴および改ざんについては、情報を適切に暗号化することでこれを実現できる。上記で述べたように、中継ピアの存在を常に仮定しなければならないので、TCPレイヤは、中継ピアで上位層を中継する必要から完全に解釈されなければならないため、P2Pのトランスポートレイヤにおいて暗号化を実施することが必須である。通信プロトコルとしては、複数の暗号方式をネゴシエートできる必要がある。本研究の実証実験においては、末端ピア間でのRSA公開鍵暗号によるRC4秘密鍵の交換を行う方式を採用した。なりすましについては、センターサーバが認証局となり、それぞれの店舗端末上のピアを認証することができる。

6. 今後の課題

常時接続のブロードバンドが一般に普及したとはいえ、各家庭のPCが常時起動していることはいまだ期待することはできない。本研究において、通信回線の分散化には一定の成果を納めたといえるが、ストレージの分散を行うためには、P2Pで接続された家庭のPCが搭載しているHDDが、P2Pネットワークで利用可能であるとはいえない。本研究の実証実験では、P2P方式による画像共有機能を同時に提供したが、規模・期間の制約があるとはいえ、その中で画像を共有するコミュニティの形成にはいたらなかった。同一コンテンツに対して多数のユーザが興味のあるような場合であれば、広くP2Pのピア間でコンテンツキャッシュを分散する方法で解決することができるが、写真コンテンツでの分散キャッシュの導入はそのコンテンツの利用機会の少なさから、有効であるかどうかはさらなる技術課題として研究を要する。

また、ネット上の複数ユーザによる共同作業としてアルバム編集機能を提供したが、本機能はP2Pネットワーク上でのトランザクション管理の技術が必須である。本研究中の実験においては、完全にP2Pでのトランザクション管理を達成できず、専用のトランザクション管理サーバを設置せざるを得なかった。

7. まとめ

デジタルカメラ、およびブロードバンド環境の普及に伴って拡大する画像データのネットトラフィックに対して、画像サービスを実施する立場のネットワークコストの大幅なコストダウンの可能性を検証するために、P2P技術の研究を行った。

結論として、ネットプリントサービスのように画像データの転送が主たる負荷であるようなサービスであり、サービス提供側において常時起動状態を維持することが比較的容易なサービスにおいては、現実サービスとして本技術を適用可能であり、規模を増すごとにそのコスト削減効果は大きくなる。サービスをこの点に特化すれば、本研究で用いた通信プロトコルをさらに単純なものにできるため、実サービスへの適用適合性は大である。

一方、ネットアルバムを代表とする、比較的長い時間その状態を保持しなければならないサービスにおいては、P2Pネットワークで状態維持のための技術が必須であり、アプリケーションの適用可能性はさらなる検討が必要である。

参考文献

- 1) Olaf Titz. "Why TCP Over TCP Is a Bad Idea".
(<http://sites.inka.de/sites/bigred/devel/tcp-tcp.html>), 2001.