HOKKAIDO UNIVERSITY

Title	インターネット上の分散型共有情報探索問題
Author(s)	三谷, 和史; 斎藤, 文武; 吉本, 哲郎 他
Citation	北海道大學工學部研究報告, 167, 29-38
Issue Date	1994-01-14
Doc URL	https://hdl.handle.net/2115/42406
Type	departmental bulletin paper
File Information	167_29-38.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北海道大学工学部研究報告
第167号（平成6年）

Bulletin of the Faculty of Engineering
Hokkaido University, No. 167 (1994)

インターネット上の分散型共有情報探索問題

三谷　和史　　斎藤　文武
吉本　哲郎　　宮本　衛市

（平成5年8月31日）

Distributed Cooperative Information Retrive over the Internet

Kazufumi MITANI, Fumitake SAITO,
Tetsuro YOSHIMOTO, and Eiichi MIYAMOTO

(Received August31, 1993)

Abstract

There are many information resources which are shared among the people on the Internet. These resources vary both in quality and quantity. NetNews, a distributed BBS system, is one of the most common use in the Internet. In this paper, we propose a new method to manage each information in NetNews and introduce a search method to obtain the desired information over the Internet. Adding a uniq increasing densely number to each NetNews articles makes NetNews System more fault-tolerant. Simulation of 100 nodes network shows that a backet-relay search of NetNews articles with newly devised caching system is useful and tolerable on the Internet.

1．まえがき

身の回りの情報メディアを見渡してみると，電話，手紙，会話といった個人的なもの，テレビ，ラジオ，新聞，雑誌，単行本といったマスメディア等々沢山の種類があり，それらを適宜使い分けていることがわかる。これらは，情報を共有するコミュニティの範囲，同時性，情報が一過的か保存的か，一方的か双方向的か等によって分類される。

インターネットはTCP/IPプロトコルによって通信が行なえる全世界的なネットワークで，北大のHINESもその一翼を担っている。規模としては，概算でホスト数が200万，ユーザ数は600万を越えるといわれている。

インターネットワーク上で利用できる共有情報資源は多数ある。それらの情報は，特定のノードに接続して利用する集中型サービスと，共有情報を2次的にキャッシュしたコピーがインターネット上の各所にあり，それを利用する分散型サービスの2種類に大別される。

前者では，特定のデータベースサービスやライブラリサービス，例えば遺伝子の塩基配列のデータベースや，気象情報のデータベース，文献検索や合衆国議会の記録等，後者では，Anonymous ftpで公開されているソースコード，ドキュメント等と，NetNewsと呼ばれているBBSが有名で

情報工学科　言語情報工学講座

ある。

後者の分散型共有情報資源の特徴をまとめると，表1のようになる。

その中でもここでは共有情報資源，とりわけ NetNews で代表されるような分散型の共有情報資源について取り上げる。これは日常生活上から言えば，新聞，雑誌の類に近いものである。

本論文では，このような分散型共有情報資源を有効に活用するための情報の名前付けと探索方法についての新たな枠組について NetNews を例として述べる。

	frequency	granularity	quality	integrality	access method
source code or document	at most 10／day ?	medium——large	high	good	archie, wais, gopher, www, telnet, anonymous-ftp
NetNews	over 20,000／day	small——medium	low ?	bad	the current NEWS System

表1　分散型共有情報資源の特徴

2．NetNews の特徴

NetNews の起源は現在のようにインターネットが普及する以前に遡る。インターネットが普及し，NNTP[2] による記事の配送へ移行しても，配送は依然バケツリレー式であって，ユーザに提供する記事もユーザがサービスを受けるノードのスプール（ディスク装置）にあるものだけを想定している。本節では，現行の News Syetem（一般に Bnews，Cnews，INN が使用されている）とその問題点について簡単に説明を行なう。

2.1　現行の News Syetem

NetNews では記事(Article)を情報資源として，記事の配送と参照によって情報が利用されている。記事の形式については，[1,5,6]で定義および議論されている。図1に概略を示す。

```
Article  :: = Message
           +  NewsGrouup
           +  From
           +  Date
           +  Subject
           +  Other headers
           +  Article_body
```

図1　記事の形式

ここで，

- Message-ID は記事を唯一に特定する文字列
- NewsGroups は記事が属するニュースグループと呼ばれる分類のリスト
  (ex. fj.jokes, fj.news.policy, fj.news.group)
- From は記事を書いた Poster を特定するアドレス
- Date は記事が投稿された日時
- Subjects は記事の表題

- Other headers はそれ以外の記事のヘッダ部
- Article_body は記事の内容

となっている。

また，NetNews で扱われる記事は分散データベースで扱われているような情報とは違い，一旦書かれると，取消もしくは時間が経過して消去されることはあっても変更されることはなく，順序性に関してもそれほど厳密である必要はない。

2.2 現行の News Sysytem の問題点

一般に，配送を受けた記事は，各ノードのスプールに，ノード毎に異なる昇順の番号（記事番号と言う）を名前としたファイルとして格納・管理されている。記事を読み書きするための News Reader Program もこれを前提に作られており，記事の読み出し時に指定されたノード以外のスプールを参照する機構を持っていない。

DNAS[11]のような複数のスプールを参照する機構でも，特定の NewsGroup は特定のノードのスプールから参照するだけで，スプールを重ね合わせることはできない。

他のスプールの記事を参照するためには，各スプール毎に既読履歴を持たねばならない。また，ある既読履歴を他のスプールの既読履歴に反映させるためには個々の記事についての突き合わせを行なわなければならず，膨大な手間がかかる。ネットワーク単位の管理情報である Message-ID で既読履歴を管理すると複数のスプールを同時に扱えるようにはなるが，番号での管理に比べて大きなサイズの既読履歴を持つことになり，ディスク容量や Message-ID のマッチングコストから見てまだ現実的ではない。

一方，ネットワークや各ノードが全く誤りなく動作していれば，News System は正常に機能する。しかし，ネットワークの切断や，ノード上での誤り（ディスククラッシュや設定ファイルの間違い，ファイルの消去等）は必ず発生し，それを考慮に入れたシステム作りをする必要がある。現状の News System では，redundant ihave/sendme という手法を使って記事の配送確認を行なうことができるが，それによって失った記事が常に回復されるとは限らない。

個々のノードがポストされた記事全体の集合（これを {Article} と書く）を知ることができれば，失った記事を特定してそれを得るための手段を講じることができる。{Article} は個々の Article を唯一に特定できる {Message-ID}，つまりポストされた記事中の Message-ID の集合を使って定義できる。しかし，各ノードで知ることのできる {Article} は，記事の配送よって外部からもたらされたものと，自分のノードでポストされた記事からなるものだけである。即ち，現状の News System では {Article} を知ることは不可能であって，これが記事の配送をフォールトトレラントにすることを困難にしている。

3．Mnews System でのアプローチ

前述のような問題点を克服するために，我々が構築を進めている Mnews System[7,8,13,14,15]でのアプローチについて説明を行なう。Mnews System では，

- 記事番号の統一によるフォールトトレラント性の獲得
- ネットワーク上での記事の探索によるニュースプールの重ね合わせ

を主眼としている。

記事の形式は，図2のように変更する。

現在の News System との違いは，各々の記事に対して NewsGroup 毎にインターネット上で唯一の昇順の記事番号を与え，記事のヘッダ中にNewsGroup/Numbersとして埋め込むことである。各ノードではこの番号で記事をスプールに格納する。

記事番号がインターネットで統一される事により，{Article}は Message-ID の他に，記事を唯一に特定できる情報である NewsGroups/Numbers を使って，{NewsGroup/Numbers}とも定義できる。つまり，自分のノードにある記事が未着かどうかは記事番号を見ることで判断が可能となる。

記事番号を使わずに，記事中の Date:フィールドの時刻をもとに順序をつけるシステム[12]も考案されているが，{Article}を知る事ができないのはこれまでの News System と同様である。

```
Article    ::= Message-ID
             + NewsGroup/Numbers
             + From
             + Date
             + Subject
             + Other headers
             + Article_body
```

図2　Mnews System での記事の形式

次節以下で，記事番号サーバと記事の探索手法，それによって起こるネットワークトラフィックについて述べる。

3.1　記事番号サーバ

Message-ID と NewsGroup が与えられたとき，その記事に一意な番号を与えるサーバを記事番号サーバと呼び，[10]で提案されているものと基本構想は同一である。

我々が考えているサーバ[16]は，

- DNS[3,4]の Resource Records に NewsGroup から記事番号サーバのアドレスを引くためのレコードを追加する。
- 記事番号サーバは各 NewsGroup につき3つ置き，それらが協調して一連番号を発行し，発行の履歴を保存する。

というもので，[10]で発生し得る記事番号の飛びを排除している。

Mnews System では，記事をポストされたノードが記事番号サーバから記事番号を取得して，それをヘッダに挿入して記事を格納，転送する。記事を受けたノードではその番号で記事を格納する。記事番号サーバと News System の関係を図3に示す。

記事番号サーバは，Message-ID から NewsGroup/Numbers への変換をするためにも用いられるが，そのための専用の参照サーバを置くことも可能である。

NewsGroup の moderate/unmoderate や newgroup/rmgroup の問題[†] に関しては，DNS の管理に任せることで混乱は避けられる。moderate な NewsGroup では DNS に別の Resource Records を導入して moderator のアドレスを示せば良い。

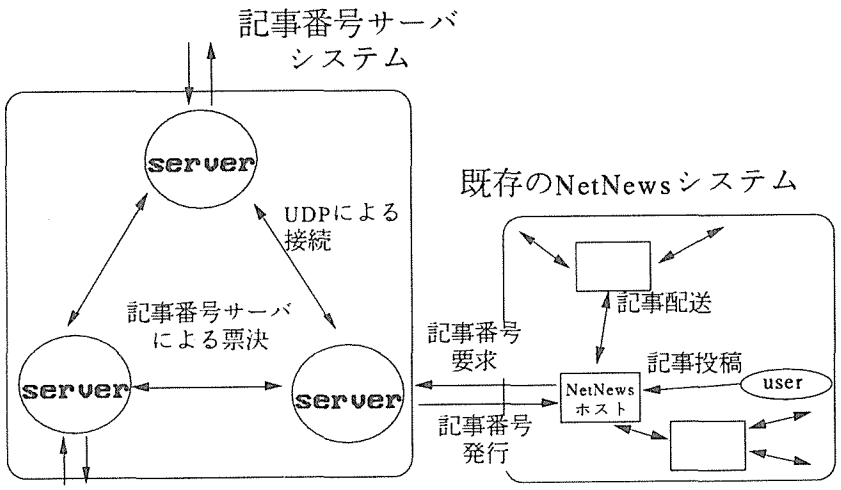

図3　記事番号サーバシステムの概念図

3.2　探索手法の比較

探索とは，NewsGroup と記事番号の探索範囲の上限，下限を与えて，その範囲の記事のうち取消されたものを除く記事を保持しているノードを探すことである。実際に記事を読む場合は，捜し当てたノードから記事を転送して貰って読むことになる。探索のための機構については，大略以下の3つが考えられる[17]。

3.2.1　集中管理方式

集中管理方式は，外部からの記事の参照を許す全てのノードでの記事の保存状態を保持するサーバを置き，そこに問い合わせる方式である(NewsGroup 毎のサーバやそれらを複数個置くというバリエーションはある)。この場合は，問い合わせの負荷の集中，全てのノードへ記事の保存状況を問い合わせる，あるいはノードから通知して貰うための通信負荷と手間がかかる。それ以上に探索の問い合わせに対する答をどうするか，例えば5,000ノードが必要とする記事を保持していることが判ったとして，それを全部問い合わせの答として返すのが妥当かどうかが問題となる。問い合わせたノードの状況に合わせて最適なものを返そうとすると最適なものの基準を決める必要があり，さらにネットワークの状況も考慮しなければならず，大変困難である。

3.2.2　バケツリレー方式

集中方式の対極としてバケツリレー方式の探索が挙げられる[13,15]。バケツリレー方式の探索では，NetNews に参加しているノード間の関係を nntp 等による記事の配送（これもバケツリレー方式

↑例えば fj.news.policy 等の今年初めの記事を参照

である）によって作られたグラフと見做す。探索は，あるノードがグラフ上の隣接ノードに探索要求を出し，そこが要求を満たす記事を保持している，もしくはどこかが保持しているという情報を持っている場合に答を探索元に返し，知らなければ探索元を除く隣接ノードに探索要求を再送する。探索が成功したなら，バケツリレーを逆に辿って探索元に答が返され，バケツリレーの中継をした各ノードでも探索結果を一定期間キャッシュして，さらなる問い合わせに備えるものである。隣接ノードが複数存在する場合，答が各々の側から返される可能性があるが，最初に到着した答を使用してあとはリレーしないこととする。

この場合，インターネット全体の状況を知ることは不可能である。局所的な記事の所在の情報をキャッシュして探索を減らしたり，探索に様々な制限をつけたり，キャッシュや探索手法に工夫を凝らすことで，探索のために生じるパケットの数をかなり減らせることが判ってきている。

3.2.3 geometric region の応用

これら2つの中間的なものとして，geometric region[9]を応用したものがある。これは，ネットワーク全体を適当な大きさの地域に分けて，その地域の中の情報は集中的に管理し，その地域中にあるノードは管理ノードに問い合わせを行なう方式である。管理ノード同士はバケツリレー方式と同様なグラフを形成しており，自分の管理地域で探索が失敗した場合はバケツリレー的に隣接の管理ノードに問い合わせを行なう。この方法は，管理可能な大きさにネットワークをうまく分割することにより，かなりうまく機能すると考えられるが，管理ノードにある程度負荷が集中することは避けられない。

これら3つの探索の伝わり方を図4に示す。黒点が入った丸がサーバーの役割を担う。

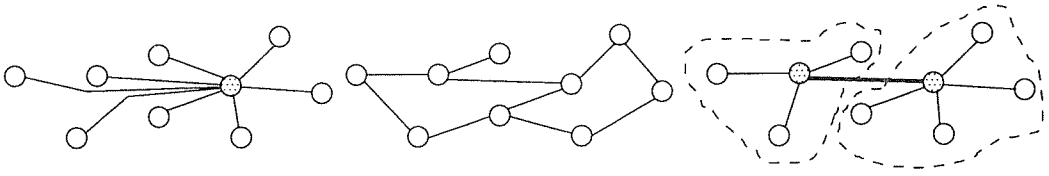

図4 探索手法によるグラフの違い

3.3 バケツリレー方式の探索に伴うネットワークトラフィックの予測

Mnews System では現在バケツリレー方式の探索を考えている。ここでは，新たに発生するトラフィックとして，探索要求のパケットによるネットワークトラフィックを考える。記事番号サーバとの通信は探索パケット数に比して少ないので無視する。また，記事本体を取得するためのトラフィックは，通常の nntp による記事の配送トラフィックに隠れると考え，これも無視する。

単純なバケツリレーではネットワーク負荷が過大であるため，探索負荷軽減のために以下の6種類のキャッシュを持たせ，どの程度キャッシュが有効かを調べる。キャッシュは上から順に試され，それでも探索に失敗した場合にバケツリレー方式の探索を行なう。

1．隣接ノードから配送されてスプールされている記事（pre cache と考える）。
2．探索の結果，他のノードから得た記事本体。
3．自分のノードからの探索中の記事のリスト。

4．他のノードからの探索要求と要求元のリスト。

5．他のノードが出した探索要求の成功時の探索要求と記事を保持しているノードのリスト。

6．探索の過程で発見した，大きいスプールをもっているらしいノードのリスト。

[14),15)]によって，以下の結果が得られている。なお，シミュレーションは，1991年末の国内のネットワークトポロジを模した100ノードからなるネットワークで，隣接ノードは物理的隣接ノードとした。

ユーザの記事へのアクセスが，縦軸をアクセス要求頻度，横軸を記事の古さにとった自由度3のポアソン分布に従うと仮定する。また，ポアソン分布の内で最近のものから80%分の記事を常に自分のノードのスプールに持っているとし，単位時間毎にネットワーク全体のうちの半数のノードで記事の探索要求が発生し，そのうちの20%が自分のノードで処理できないと仮定する。また，パケットが隣接ノードに届くまでの時間および隣接ノードでの処理時間にも単位時間かかると仮定する。

この仮定に従って，システムの動作を再現するシミュレータを作成し，計測を行った。ネットワーク全体に存在する探索パケット数に関しての結果を図5，6に示す。

シミュレーション結果から次のような傾向が確認された。

- ユーザは新しい記事から読む傾向があるため，pre cacheとしての現在のニュースのバケツリレーは意味がある。
- 記事の所在をNewsGroup/最小番号：最大番号でキャッシュすることがかなり有効である。
- 記事の所在をNewsGroup/番号という形でキャッシュした場合には，大きなスプールを持つノードを良く使う傾向であったが，最小と最大で押えるとこれはあまりヒットしない。
- 定常状態になると，よく参照される記事はほぼキャッシュ経由でアクセスされ，バケツリレー方式の探索は行なわれない。
- 記事の存在を否定する情報がないため，ネットワーク中にない記事を探索した場合にその頻度に比例してネットワーク中のパケットが増える。これをどう減らすかが今後の課題である。
- キャッシュで参照しているノードから記事が消去されたり，ディスクがクラッシュしたりしても，かなり短時間（100単位時間程度）で次の定常状態になる。そのノードが回復しても元の定常状態には戻らない。状態遷移の間のパケット数は定常状態の2倍程度である。
- トポロジー的に沢山のノードと繋がっているバックボーンノードに向けてパケットは集中する。最大部分でも，16パケット/単位時間程度であった。

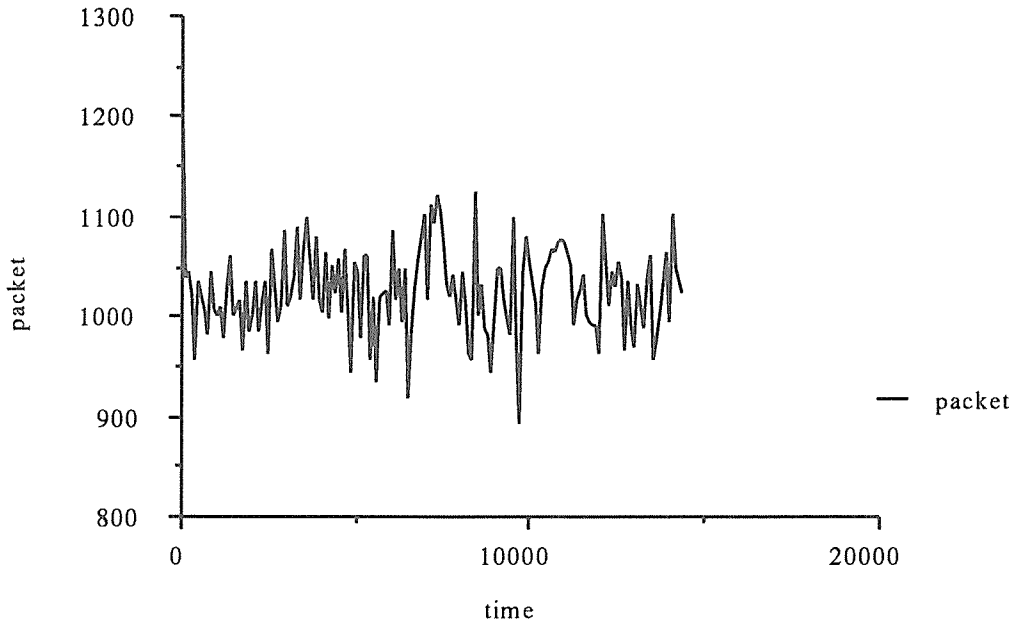

図5　パケット数の推移（キャッシュ等が無い場合）

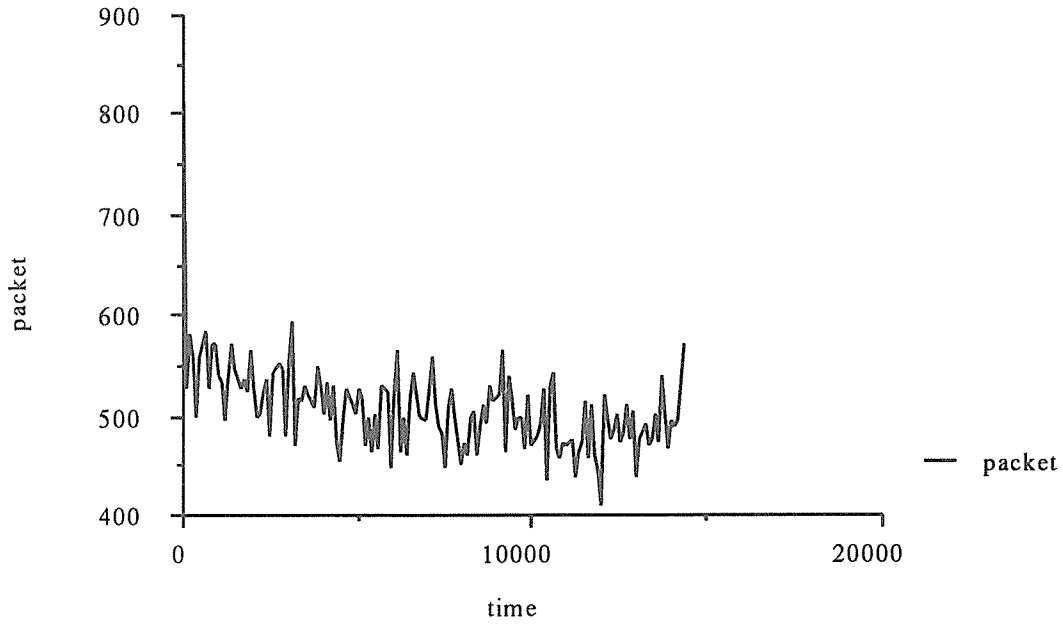

図6　パケット数の推移（キャッシュ等が有る場合）

4. おわりに

本論文では，分散型共有情報資源としてNetNewsを取り上げ，Mnews Systemで提唱する記事番号の統一がフォールトトレランシに優れ，記事の探索にも役立つことを述べ，記事の探索がバケツリレー方式でもネットワークトラフィック上からも実現可能であることを示した。

しかし，探索パケットによるトラフィックのシミュレーションの前提が，このようなシステムが運用された場合に変化する可能性が十分あり得るので，探索に関してさらに検討を加える必要がある。特に探索パケットにhop数や有効時間(ttl)を設けたり，コストを導入して一度に沢山の探索パケットを出せないようにするなどの方策も採り入れなければならないであろう。

また，記事の既存のヘッダ部を変更せずに，拡張ヘッダとして記事番号を埋め込めば，既存のNews Systemとの併用も可能であり，さらに一部のNewsGroupのみをこの手法で取り扱うことも可能である。

分散開放型の環境としては，特定のノードに情報が集中せず，ノードの近傍での情報の所在の知識をうまく活用してやることが望ましいと考える。しかし，Mnews Systemでは番号付けの部分は集中管理されてしまうが，これは情報の存在を知ることができなければ探索もできないので，最低限管理されるべきものであると考える。このための情報量や通信量もインターネットで十分賄い得るものである。

適当なネットワークの地域分けができた場合とバケツリレー方式では，ネットワークに与える負荷や，管理ノードのトラブルに対して，どちらが有利であるかをはっきりとさせることが今後の課題である。

さらに，情報がますます洪水のようになっていく現状に即して，情報のフィルタリングをどう行なうかについても検討を行ないたい。

参考文献

1) Mark R. Horton : Standard for Interchange of USENET Messages, RFC850, (1983)
2) BrianKantor, PhilLapsley : Network News Transfer Protocol, RFC977, (1986)
3) P. Mockapetris : Domain Names —— Concepts and Facilities, RFC1034, (1987)
4) P. Mockapetris : Domain Names —— Implementation and Specification, RFC1035, (1987)
5) Mark R. Horton : Standard for USENET Messages, RFC1036, (1987)
6) rfcerratta in Cnews doc directory.
7) 磯野京介，三谷和史，宮本衛市：LAN上の分散ニュースシステムについて，電気関係学会北海道支部連合大会，(1990) p. 366
8) 三谷和史，斎藤文武，宮本衛市：インターネットにおけるニュースのキャッシングについて，日本ソフトウェア科学会第8回大会，(1991)，pp. 469—472
9) 太田昌孝：国際電子メイルの配送とgeometric region，日本におけるアカデミック・ネットワークの相互接続の諸問題シンポジウム論文集，(1992)，pp. 68—73
10) 辻健二郎，太田昌孝：IP接続されたネットワーク上でのニュースシステムに関する研究，日本におけるアカデミック・ネットワークの相互接続の諸問題シンポジウム論文集，(1992)，pp. 74—83
11) 梶田将司，今井祐二：複数NNTPサーバへの統合的アクセスの実現，JAIN Symposium'92 論文集，(1992)，pp. 39-47
12) 相原玲二，栄本健二，山下雅史：ナンバリングサーバを用いないIPネットワーク上のニュースシステム，JAIN Symposium'92 論文集，(1992)，pp. 48-57
13) 三谷和史，情報システムとしてのNews System —— Mnews，JAIN Symposium'92論文集，(1992)，pp. 67—76

14) 斎藤文武, 三谷和史, 宮本衛市：ネットワーク分散環境における共有情報探索法, Info—Hokkaido '92, (1992), pp. 67--70
15) 三谷和史，斎藤文武，吉本哲郎，宮本衛市：ネットワーク分散環境における共有情報探索について，日本ソフトウェア科学会第 9 回大会，(1992), pp. 425—428
16) 吉本哲郎，三谷和史，宮本衛市：分散型共有情報環境における情報の名前付けについて，Info—Hokkaido'93, (1993), pp. 99—102
17) 三谷和史，宮本衛市：インターネット上の分散共有情報探索について —— NetNews を例として ——，電気関係学会北海道支部連合大会，(1993), pp.186-187