

ニュースグループごとに配送方式を自動変更する NetNews システムの評価

舟阪淳一 山本和彦 福田晃
奈良先端科学技術大学院大学

概要

インターネット利用者の増加に伴い、NetNews の記事の流量が増大してきている。しかしながら、サーバによっては配送された記事の数の約 10% しか読まれていない。利用されない不要な記事を大量に送りつけてしまうのは、従来の NetNews システムが利用状況を考慮していないことに起因する。そこで、利用状況を反映して配送方式を選択するシステムを提案する。このシステムを用いて、頻繁に利用されるニュースグループの記事を全て配送し、残りのニュースグループの記事を要求に応じて配送する実験を行った。その結果、記事の利用率が従来の約 10% から約 50% に増加し、配送記事量も軽減できた。本手法は全体の約 60% のサーバで適用可能と考えられるため、配送記事量の軽減に有効である。

Evaluation of NetNews system applying different feeding method for each newsgroup

Junichi Funasaka, Kazuhiko Yamamoto, Akira Fukuda
Nara Institute of Science and Technology

Abstract

As number of Internet users are increasing and NetNews traffic is also increasing. Only 10% of articles are, however, read on some servers. A lot of unwanted articles are delivered because usual NetNews systems do not take account of user's requirement. We propose a new system which can select feeding method reflecting user's requirement. With this system, an experiment was carried out. All articles are delivered for frequently required newsgroups and requested articles are delivered for other newsgroups. As a result, ratio of read articles increased from 10% to 50% and delivered articles decreased. This method can be applied on 60% of all news servers, thus it is very effective to decrease articles to be delivered.

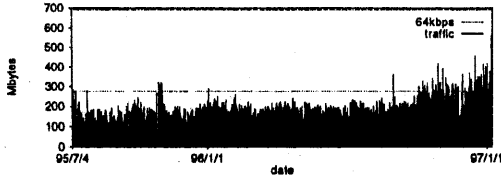


図 1: 奈良先端科学技術大学院大学のニュースサーバに一日で流入する記事の総量

1 はじめに

インターネット利用者の増加に伴い、NetNews に投稿される記事も多くなってきた。一方、広帯域の回線が登場し、より多くの情報の交換が可能になってきた。しかしながら、狭帯域の回線での情報交換を強いられる環境も依然として残っており、帯域を効率よく利用する方法が必要である。

図 1 は奈良先端科学技術大学院大学のニュースサーバに一日で流入する記事の量を示している。これを見るとわかるように、記事の量は増加する傾向にある。また 96 年末には一日約 300Mbytes に達している。

また図 1 の点線は 64Kbps の約 40% を示す。従って、現在の NetNews の記事量は 64Kbps の回線の約 40% を占める。狭帯域の回線しか用意できない環境では大きな負担になる。

しかしながら NetNews の記事の全てが利用されているとは限らない。表 1 は奈良先端科学技術大学院大学のニュースサーバで利用されている記事の割合を示している。表によると利用されている記事はせいぜい 1 割である。従って全て

表 1: 記事の利用率

	97.1.31-2.6
利用された記事数	60811
流入した記事数	996603
利用された割合	6(%)

表 2: ニュースグループの利用率

96.9.20 -26	ニュース グループ数	割合 (%)
利用された ニュース グループ	971	9
利用され なかった ニュース グループ	9873	91
合計	10844	100

の記事を配送するのは回線資源の浪費である。

NetNews の記事はニュースグループによって分類される。記事と同様、ニュースグループについても利用状況に偏りがある。表 2 は 96 年 9 月 20 日から一週間のニュースグループ利用状況を示している。ニュースグループについてもやはり約 10% しか利用されていない。

この問題点は、従来の配送方式が利用状況を考慮に入れていないことに起因する。RFC977[1] に記述されている NetNews の通信手順は、ニュースグルー

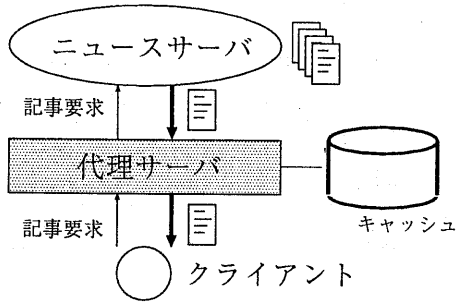


図 2: 代理サーバ方式

ブを考慮に入れた配送方法を規定していない。またサーバプログラムによる独自拡張手順[2]の中でも規定されていない。

考えられている解決策のひとつとして、要求に応じた配送方式[3]が挙げられる。この方式では、代理サーバをサーバとクライアントの間に置き、クライアントから要求のあった記事をサーバから受けとって保存する(図2)。また同じ記事の2回目以降の要求にはキャッシュに保存されている記事を返す。この方式の利点は必要な記事のみを受けとれることである。欠点は、初めて参照される記事の取得に時間がかかることである。頻繁に利用されることが予想できるニュースグループの記事であっても、要求があるまで記事を取得しない。

本研究の目的は、利用されない記事の配送を減少させ、利用される記事をあらかじめ配送することである。従って、配送記事量の減少とそのうち利用される記事の割合の向上が目標となる。また代理サーバ方式ではできなかった頻繁に利用されるニュースグループの先読みを可能にすることをめざす。

次章では、前述の目的を達成するた

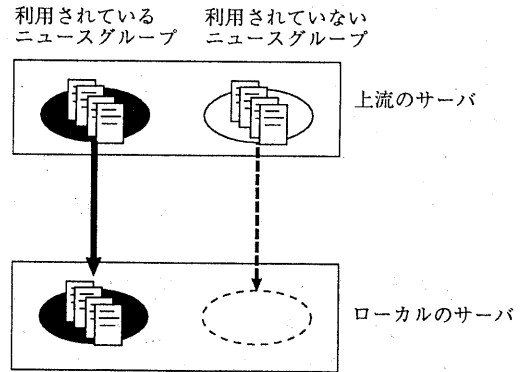


図 3: 本研究の配送方式

めに本研究で採用したニュースグループごとに配送方式を変更する手法について説明する。次に実装したシステムを用いた実験の結果を3章で示す。4章で本研究の手法を評価し、最後にまとめる。

2 本研究の手法

本研究ではニュースグループごとに配送方式をかえられる手法を考えた。この手法では、頻繁に利用されるニュースグループの記事は利用の有無にかかわらず全て配送され、それ以外のニュースグループの記事は要求に応じて配送される。本手法を用いることで、ほとんど利用されないニュースグループの記事を全て送りつけるような無駄な配送を避けられる。また、頻繁に利用されるニュースグループの記事を、要求を待たずにあらかじめ配送できる。(図3)

現在の実装では、ニュースグループの利用頻度によって配送方式を変更し

ている。例えば、週一回以上利用されるニュースグループは常に全記事を配送し、それ以外のニュースグループは要求のあったときに送るというぐあいに設定できる。利用頻度のしきい値を決めることで、自動的に配送方式が変更できる。

また利用状況の情報を交換することで、3つ以上のサーバにも適用できる。当然サーバごとに利用傾向が異なると考えられるため、取捨選択のポリシーが必要である。このポリシーは実際の利用状況から決める。

本手法を用いることで、従来管理者が静的に指定しなければならなかった配送ニュースグループを動的に選択できるようになる。配送ニュースグループの動的変更には、新規ユーザの急増による要求傾向の変化に対応できる利点がある。

ところで、本手法はどんなサーバにも適用可能というわけではない。本システムでは上流が下流の利用状況を把握し、上流と下流の間で配送グループの情報をやりとりしなければならない。従って適用できるサーバの条件は、循環して接続していないこと、および下流のサーバの数が把握できることである。例えば木構造を形成して接続するサーバに適用できる。

この適用条件を満たすサーバが全体に占める割合を調べた。NetNewsの記事には経由してきたサーバの名前が記録されているので、これを基にサーバ間接続図を作成した。fj.*の96年10月の記事を用いて国内のサーバを調査したところ、本手法は表3のように約60%のサーバに適用可能であることがわかった。

表 3: 適用条件を満たすサーバ

	サーバ数	割合 (%)
適用可	268	59.7
適用不可	181	40.3
total	449	100

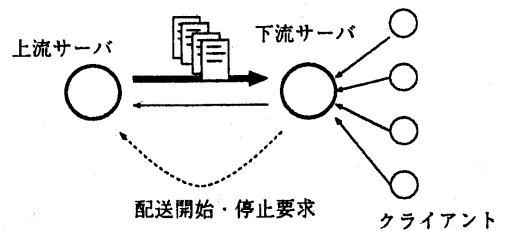


図 4: 実験環境

3 実験

本章では実験の概要について説明し、結果を紹介する。まず、本当に必要なニュースグループが配送の対象となっているかを検証するために、ニュースグループの再利用率の結果を示す。次に記事の利用率をみて、配送された記事が有効に利用されているか調べる。最後に、記事の配送量の変化について述べる。

3.1 実験環境

実験は図4に示す2つのニュースサーバで行った。この両サーバ間でニュースグループごとに配送方式を変更し、その効果を調べた。

実験に参加したクライアント数は図5に示す通りであり、平均4.3クライアントである。参加者は全て奈良先端科学

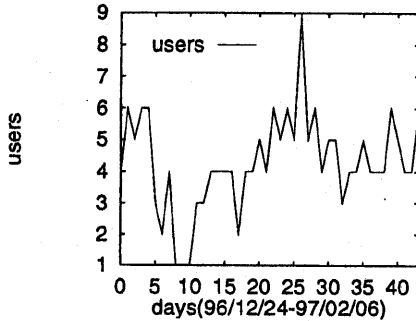


図 5: 実験への参加クライアント数

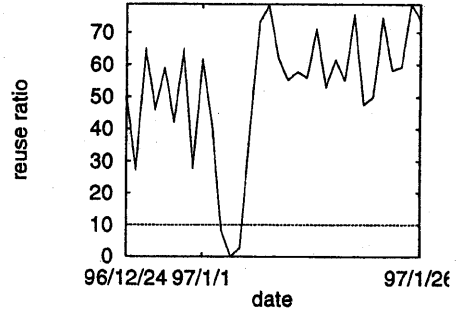


図 6: ニュースグループの再利用率

技術大学院大学の一員である。

図 4の上流のサーバは奈良先端科学技術大学院大学のニュースサーバを利用した。従って全記事を受けると1日約300Mbytesの量になる。また上流サーバの利用は1日平均100クライアントである。

また配送方式変更のためのしきい値は、一週間に一回の利用頻度に設定した。上流のサーバにおいて、保持された記事が消去される期限を平均一週間に設定しているので、妥当な頻度である。

3.2 ニュースグループの再利用率

前日配送されたニュースグループ数に対する、読まれたニュースグループ数の割合を図6に示す。中央の大きな減少は、年末年始のクライアントの減少による。この時期を除けば、利用者が少なく限定されていることを反映して、かなり高い再利用率(約60%)が得られている。図6ではさらに、上流のサーバで全グループを配送していた場合の利用率を点線で示している。この約10%に比

表 4: 記事の利用率

要求される 記事数	流入する 記事数	割合 (%)
4040	7671	53

べて約60%という値は十分大きく、本研究の手法が有効であることを示していると考えられる。

3.3 記事の利用率

配送される記事のうち、実際に利用された記事の数を調べた結果を表4に示す。記事の利用率は約50%である。これは全グループの記事を配送していた場合の約10%に比べて十分大きく、本研究の手法の効果を示している。

3.4 記事の配送量

実験期間中、97年1月31日からの一週間について配送記事量を調べると、一日平均3Mbytesであった。これは一日300Mbytesと比べて極めて小さい。

ただし実験では利用者数が1/20程度であることに注意しておく必要がある。配送記事量を1クライアントあたりで考えると、約1/5に軽減されている。

4 評価

本研究の手法により、全ニュースグループの記事を送っていた場合よりニュースグループの利用率、記事の利用率が向上した。また配送される記事の量が減少した。ただし今回の実験が限られた少人数の参加者の利用についての結果であることに注意しておく必要がある。また利用者が一大学に限定されているため、利用に偏りがある可能性がある。

一方、代理サーバ方式では、利用された記事しか送られないので記事の利用率は100%である。しかしながら、初めて利用される記事は要求があった時に送られるので、記事が得られるまでの所要時間が大きい。本システムでは頻繁に利用されるニュースグループの記事があらかじめ配送できるので、代理サーバ方式より所要時間を少なくできる。

本システムにより、記事の利用率を高め、配送量を軽減するという目的が達成された。また代理サーバ方式よりも所要時間を少なくできる。

5 おわりに

従来の配送方式では利用されない記事を大量に送ってしまうため、狭帯域の回線をもつ環境では大きな負担となっている。そこで、頻繁に利用されるニュー

スグループの記事のみを全部配送し、それ以外のニュースグループの記事は要求に応じて配送するという方式を提案した。

この方式により、従来の全記事を配送する場合に比べ、記事の利用率を高め配送量を軽減することができた。また全体の約60%のニュースサーバで適用可能であることがわかった。

ただし今回の結果は、利用者が少なく偏りが大きいことの影響を受けている可能性がある。より多くの利用者についての結果を調べ、有効性を検証する必要がある。また利用傾向の異なる組織間でも有効であるか、調べる必要がある。

参考文献

1. Brian Kantor and Phil Lapsley, Network News Transfer Protocol: A Proposed Standard for the Stream-Based Transmission of News, Request for Comments 977, 1986.
2. Stan Barber, Common NNTP Extensions, internet draft, work in progress.
3. 知念 賢一, 山口 英, 山本 平一, NetNewsにおける利用状況を反映する記事配送モデルの提案, マルチメディア通信と分散処理研究会研究報告 No.71, 1995.