

## NetNews における利用状況を反映する記事配送モデルの提案

知念 賢一, 山口 英, 山本 平一

奈良先端科学技術大学院大学

### 概要

現在、Internet では NetNews (Network News) における記事の配送による、サーバ間の膨大なトラフィックが問題となっている。これは NetNews が全ての記事を配送するモデルに従って動作することに起因する。本研究では NNTP 上の要求を観測し、配送される記事の大部分がクライアントから要求されないことを明らかにした。すなわち、ユーザは配送される記事の大部分は閲覧・印刷されることはない。さらに、利用される記事のみを配送する代理サーバを導入した新たな配送モデルを提案し、利用状況に応じた配送トラフィックを軽減した NetNews システムの評価を行なった。

キーワード: ネットニュース、代理サーバ

## A Proposal for Requested Article Delivery Model on NetNews

*Kenichi Chinen, Suguru Yamaguchi, Heiich Yamamoto*

**Nara Institute of Science and Technology**

### Abstract

In the current NetNews system, all the posted articles are delivered to the servers with a simple flooding model. This is one of the reasons, why the traffic of the NetNews articles has been exploded. However, we revealed that the most of the articles are not retrieved, through our analysis of the NetNews usage log data. In order to reduce the amount of NetNews traffic, we proposed a new NetNews delivery model. In this model, we promote to use the NetNews proxy server which get articles on clients demand maintains its local cache.

In this paper, we discuss the model and its implementation of the proxy server. Also, we provide a case studies to show how this model reduce the amount of NetNews traffic.

**Keywords:** *NetNews, proxy server*

## 1 はじめに

NetNews は記事 (article) と呼ばれる情報を提供するサービスであり、分散したコンピュータ間で配送と呼ばれる処理により記事を共有することが特徴である。記事の配送は全てのサーバへ全ての記事が転送されるように設計されている。このため、多数のサーバが接続され、多数の記事が配送されている現状の Internet では NetNews による膨大なトラフィックが生じている。そして、膨大な NetNews の配送トラフィックによる他のサービスの圧迫や、全てのサーバがトラフィックの増加に追従できない等の問題が発生している。

NetNews の記事の利用状況の調査として NNTP 上の記事の要求を解析した結果、配送されている膨大な記事の大部分は利用されていないことが判明した。利用された記事が利用されない記事に比べ非常に少ないことから、利用される記事のみを配送する配送システムが実現できれば、トラフィックの少ない配送モデルの構築が可能となる。

本研究では利用される記事を配送するシステムとして、要求された記事を転送する NetNews 代理サーバを設計し、従来のモデルに代理サーバを導入した配送モデルについて研究している。本稿では、NetNews の記事の利用状況の調査結果と代理サーバを導入した配送モデルの提案、そして代理サーバの設計について述べる。

## 2 NetNews のシステム

本節では NetNews の代表的なプロトコルである NNTP (Network News Transfer Protocol) を中心に、現在の NetNews のシステムについて述べる [1]。

NetNews はサーバ・クライアントモデルで構成されており、Internet 上には多くのサーバとクライアントが存在する。サーバは記事を蓄え、他のサーバへ新着記事の通知と転送を行なって記事の共有を図る。本稿ではこの新着記事の通知と転送を「配送」と呼ぶ (図 1)。

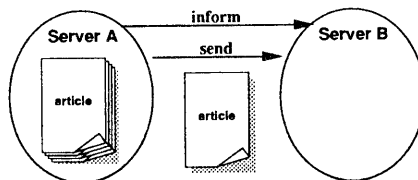


図 1: 配送-通知と転送

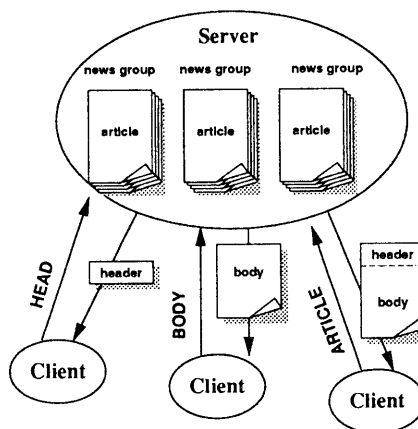


図 2: サーバクライアント間通信

現在の配送は全てのサーバへ全ての記事をバケツリレー式に配送するよう設計されており、ネットワーク上のサーバ間で配送を繰り返すことによって、遠く離れたサーバへ記事が配送される。記事は属性を記述したヘッダ (header) とボディ (body) で構成されており、ニュースグループ (news group) と呼ばれるカテゴリー毎にまとめられて、サーバに蓄えられる。クライアントはニュースグループとそのニュースグループ内の記事の番号を指定して、記事のヘッダ、ボディあるいは記事全体 (ヘッダとボディ) を要求することが可能である。サーバはこのクライアントからの要求に応じて記事を転送する (図 2)。

## 3 NetNews の問題点

本節は現在の NetNews で問題となっているトラフィックの増加と記事の利用度について述

ニュース グループ名	要求記事数	配送記事数	利用率	要求頻度		
				HEAD	BODY	ARTICLE
comp	17735	85202	20.8%	20803	20	3067
alt	9342	59062	15.8%	44646	7333	9769
fj	8491	10058	84.4%	62925	1080	18140
rec	3655	62108	5.9%	3634	26	965
soc	1837	39748	4.6%	2138	0	108
misc	1392	25984	5.4%	1352	0	42
tnn	377	755	49.9%	824	11	447
news	319	3421	9.3%	420	0	46
sci	195	13647	1.4%	193	0	49
bionet	184	1824	10.1%	655	0	133
other	1014	19745	5.1%	2408	137	893
total	44556	357790	12.5%	140048	8609	33706

表 1: ニュースグループ毎の記事の利用率と要求頻度

べる。

### 3.1 トラフィックの増加

Internet の拡大と NetNews の普及により、現在では膨大な数の記事が Internet 上で配送されている。具体的な例としてサーバ間のトラフィックが約 161MB/日 (1.91KB/秒)<sup>1</sup> に達することが ASTEC における観測によって報告されている (図 3) [2]。1.91KB/秒のトラフィックは転送容量が 64kbps のネットワークでは NetNews の配送トラフィックが全体の 24% を占めることになり、他のサービス(とくに遠隔ログイン等の会話的処理)を圧迫する。

そして、配送される記事数は増加し続けており、記事の配送によるトラフィックも増加し続けている。Internet はネットワークの集合体であり、サーバの存在する全てのネットワークの通信容量が一様に拡大されることは期待できない。このことから全てのサーバが配送トラフィックの増加に追従可能かどうか懸念されており、NetNews の配送トラフィックを軽減する手法が求められている。

### 3.2 記事の利用度

配送トラフィックはユーザに記事を提供するための処理によって生ずるものであり、ユー

<sup>1</sup> 95年4月の平均値

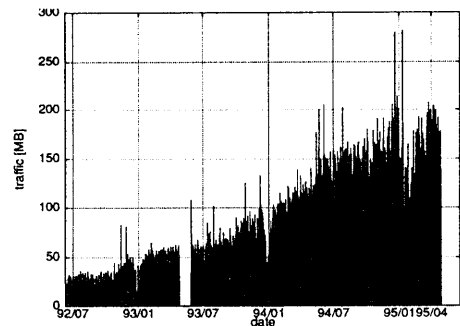


図 3: ASTEC におけるトラフィックの推移

ザの NetNews の利用が反映されていると思われる。本研究ではユーザの NetNews の利用をクライアントからサーバに対する記事の要求とみなして、奈良先端科学技術大学院大学の NetNews サーバにおいて NNTP による記事の要求 (HEAD, BODY, ARTICLE) を一週間観測した。

表 1 は一週間に配送された記事数、その記事のうちクライアントから要求された記事数、一週間に発生した要求の頻度をニュースグループ毎に集計した結果である。利用率は要求記事数と配送記事数の比率である。これより配送された膨大な記事の大部分が利用されていないこと、ボディの利用に比べてヘッダの利用が多いことが明らかになった。

表 2 は各記事への要求を HEAD, BODY,

HEAD	BODY	ARTICLE	組数	全体比
1	0	0	21363	47.9%
2	0	0	3308	7.4%
1	0	1	2489	5.6%
0	0	1	2161	4.9%
2	0	1	938	2.1%
3	0	0	908	2.0%
4	2	0	705	1.6%
2	1	0	531	1.2%
2	0	2	407	0.9%
3	0	1	397	0.9%
	その他		11349	25.5%
	計		44556	-

表 2: 記事要求の傾向

ARTICLE の発生頻度の組で集計した結果である。これより HEAD のみ実行されただけの記事が多いことが明らかになった。

これらの結果から、サーバは全ての記事を配送せずに NetNews のサービスを提供できることが明らかになった。要求されない記事の配送を省くことによって大幅なトラヒックの軽減が期待できる。

### 3.3 従来のトラヒック軽減策

配送トラヒックは配送される記事に応じて生ずることから、従来は管理者が配送するニュースグループを制限することにより配送トラヒックを軽減する工夫が行なわれていた。この手法はニュースグループの把握と、ユーザの利用状況の把握が必要である。しかし、現状では日々増減するニュースグループの把握と、ユーザの利用状況の把握は多大な管理コストを必要とする。

また、NetNews が多くのユーザが利用するサービスである点からみて、この手法はユーザの利用できるニュースグループを制限してしまうという問題がある。

## 4 新しい配送モデルの提案

記事の利用状況の観測から、記事を選択して配送するサーバを用いることによって配送

トラヒックの少ない NetNews サービスの実現の可能性が明らかになった。また、利用されない記事が利用される記事より多いことが判明し、転送を重視した現在の配送より要求を重視した配送が NetNews に適する。そこで、本節では要求を重視したサービスに適したシステムである代理サーバについて述べ、代理サーバを用いた新しい配送モデルを提案する。

### 4.1 代理サーバとその設計

代理サーバ(proxy server)とは特定のサービスを他のシステムに代わって行なうシステムである。NetNews における代理サーバはサーバの代わりにクライアントの要求を受け、クライアントの代わりにサーバに要求を行なうシステムである(図 4)。すなわち、代理サーバはクライアントから要求された記事のみを配送するシステムであり、従来の配送とは異なり通知と要求されない記事の配送を必要としない。

代理サーバは要求を受けた後にサーバからの転送をはじめめるため、サーバから直接転送をする場合に比べ長い待ち時間を必要とする。また、クライアントの要求のたびに代理サーバからサーバへ記事を要求すると、要求に応じて待ち時間と膨大なトラヒックが生ずるため、サーバと同様に記事を蓄えるキャッシュ(cache)を設けて同じ記事に対する再要求に備える必要がある。そして、前述のように HEAD のみ要求される記事が多いことから、記事のキャッシュへの格納の際にはヘッダとボディを別けて扱う必要がある。

代理サーバでは通知処理を省いたことから、新着記事の有無は代理サーバ側からサーバへ問い合わせる必要がある。

### 4.2 従来の配送モデルとの融合

代理サーバの登場によって、クライアントに記事を提供するシステムはサーバと代理サーバの 2 種類存在することになった。本節では従来の配送モデルの特徴である記事の共有を

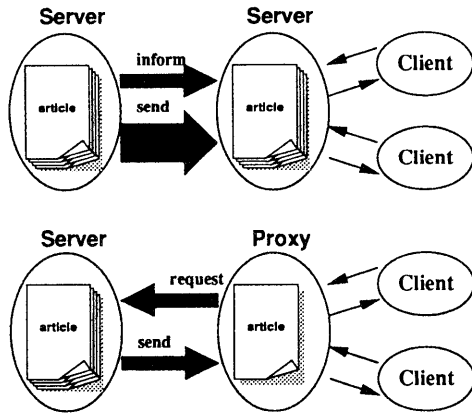


図 4: NetNews 代理サーバ

行ないつつ、代理サーバを用いた配送モデルについて述べる。

前述のようにサーバは全ての記事を格納するが稼働には大きなトラフィックを必要とし、代理サーバは要求された記事のみを格納するが少ないトラフィックで稼働が可能である。一方、サーバは複数サーバで協調して動作することが可能だが、代理サーバは協調できない。したがって、効果的な代理サーバの導入は、配送トラフィックが他のサービスを圧迫しているネットワーク上にあり、かつ複数のサーバと協調していないサーバを代理サーバへ交替することである。このとき、サーバの代理サーバへの交替はクライアントや他のサーバの変更を必要としないことから容易である。

### 4.3 新しい配送モデルの評価

本節では従来のサーバのトラフィックと代理サーバにおけるトラフィックを比較する。

サーバによる配送は全ての記事のヘッダとボディを配送している。H をヘッダの平均サイズ、B をボディの平均サイズ、Nh をヘッダの数、Nb をボディの数とすると、そのサーバのトラフィックは次式ようになる (図 5)。

$$T = H * Nh + B * Nb \quad (Nh = Nb)$$

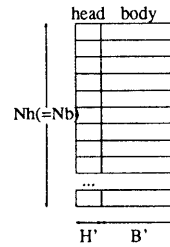


図 5: サーバにおける全記事配送

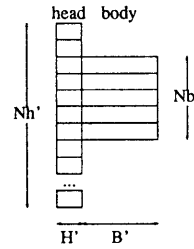


図 6: 代理サーバにおける選択配送

一方、代理サーバでは利用されたヘッダと利用されたボディのみ配送される。H' をユーザに利用されたヘッダの平均サイズ、B' をユーザに利用されたボディの平均サイズ、Nh' をユーザに利用されたヘッダの数、Nb' をユーザに利用されたボディの数とすると、代理サーバのトラフィックは次式ようになる (図 6)。

$$T' = H' * Nh' + B' * Nb'$$

以上のことから、サーバのトラフィックは全体の記事の数とサイズに依存するが、代理サーバは全体の記事に数とサイズには依存せず、利用される記事の数とサイズだけに依存する。すなわち、サーバ間の配送トラフィックが増加しても、代理サーバのトラフィックは影響を受ず、少ないトラフィックでサービスを提供できる。

また、前述の利用状況の調査より、一般的に  $Nh' < Nh$ ,  $Nb' \ll Nb$  であることが予想され、代理サーバトラフィックの軽減が期待できる。例えば、奈良先端科学技術大学院大学でサーバを代理サーバに交替した場合、表 1 から 80% 程度のトラフィックの軽減が見込める。

## 5 今後の課題

本稿では代理サーバで配送トラフィックを軽減する新しい配送モデルを提案したが、NetNewsはネットワーク上の全ての記事を共有することを目的とすることから、全ての記事を蓄えるサーバが必要であり、代理サーバで全てのサーバを交替することは不可能である。したがって、全てのサーバの配送トラフィックを軽減することは不可能であり、これ以上の配送トラフィックの削減は配送モデルのみではなく、NetNewsの目的である全ての記事の共有を再考する必要がある。

本稿で示した利用状況は、奈良先端科学技術大学院大学のサーバにおける一週間の観測結果であるが、利用状況はサーバ毎に異なることが予想される。したがって、より一般的な観測として複数のサーバでの長期の利用状況が必要である。

代理サーバは積極的に配送するサーバとは異なり、クライアントの要求を受けることによって転送を行なうので、サーバに比べ応答時間が長い。そこで、要求数あるいは利用率などから積極的に転送を行なう機構を検討する必要がある。

## 6 おわりに

現在 NetNews では膨大なトラフィックが生じ、他のサービスを圧迫する問題が生じており、通信容量の少ないネットワークでは大きな脅威となっている。記事の利用調査を実施することによって、配送されている記事の大部分は利用されていないことが判明した。したがって、利用される記事を選択して配送することによりトラフィックの軽減が可能である。

従来の配送重視のサーバにおいては利用される記事の選択が困難であり、配送するニュースグループを管理者の経験で選択する等の粒度の荒い選択しかできなかったが、要求重視のシステムである代理サーバを用いることによって、利用される記事の選択が可能となり、サーバ間で配送される記事が増加しても少な

いトラフィックの配送が可能となった。

そして、従来の配送モデルへ代理サーバを追加したトラフィックの少ない新しい配送モデルを提案した。この配送モデルの適用により配送トラフィックが脅威となっているサーバを代理サーバに交替し、脅威を排除することが可能となった。

## 謝辞

本研究においてトラフィックの解析は重要であった。トラフィックの資料を公開している ASTEC に感謝したい。また、奈良先端科学技術大学院大学内におけるトラフィックの調査を快諾して下さった、奈良先端科学技術大学院大学情報科学センターにも感謝したい。

## 参考文献

- [1] B. Kantor and P. Lapsley. Network News Transfer Protocol: A Proposed Standard for the Stream-Based Transmission of News, RFC977. February 1986.
- [2] news:/tnn.netnews.stats. ASTEC における日々の配送トラフィック.