

読まれているニュースグループのみ自動的に購読する ニュースシステムの構築

渡辺健次† 小串英俊† 近藤弘樹†

†和歌山大学システム工学部 †佐賀大学理工学部

〒640 和歌山市栄谷 930

E-mail: watanabe@sys.wakayama-u.ac.jp

あらまし: インターネットの拡がりとともに、ネットニュースに投稿される記事の量が爆発的に増加したことで、(1) 記事の転送トラフィックの増加、(2) 記事を保存するスプール容量の不足、という問題点が表面化してきている。しかし実際には、送られてきた記事のほとんどが、読まれること無く消去されている。

我々は、ニュースサーバのアクセス情報から読まれているニュースグループを抽出して、読まれているニュースグループの記事のみを自動的に配送する仕組みを備えた、ニュースシステムを構築した。このシステムを利用することで、ニュース記事の量を 97% 節約することができた。

キーワード: ネットニュース, インターネット, システム管理

A Development of The DietNS

Kenzi Watanabe† Hidetoshi Ogushi† Hiroki Kondo†

†Wakayama University †Saga University

930, Sakaedani, Wakayama 640, Japan

E-mail: watanabe@sys.wakayama-u.ac.jp

Abstract: As the growth of the Internet, a lot of people join the NetNews community. Since the number of articles are increasing day by day, there are two serious problems as follows. (1) The traffic of article transmission increases rapidly. (2) The capacity of preserving storage for articles is not enough. However, most articles are never read in fact.

We have developed a new NetNews system called "DietNS". This system have a function that only read newsgroups subscribe automatically. This system can solve the above problems.

Keywords: NetNews, Internet, System Administration

1 はじめに

ネットニュースは、インターネットアプリケーションの中でも、最も利用されているものの一つである。歴史も長く、多くのニュースグループに世界中の人が参加して、活発な議論が行なわれている。

現在のネットニュースシステムでは、あるニュースグループに投稿された記事は、そのニュースグループを購読している全てのサイトに配送され、それぞれのサイトで保存される。購読しているニュースグループ数は組織によって違うが、例えば和歌山大学では平成8年6月6日現在で5,529のニュースグループを購読しており、多い日は1日に10万通以上の記事が送られてきている。

近年のインターネットの急速な拡がりとともに、ネットニュースの利用者も急速に増加している。これにもなって、流通するニュース記事の量も爆発的に増加してきている。このため、ニュースシステムの運用において、記事の配送トラフィックの増加、および記事を保存するスプール容量の不足という問題点が、表面化してきている。

ところが、5千以上あるニュースグループに対して、実際に読まれているニュースグループは、それほど多くない。ある日のデータでは、読まれているニュースグループは、全体のわずか9%にしか過ぎないことが示されている。それ以外のニュースグループについては、記事は届けられるものの、読まれることなく、保存期間が過ぎたら expire されている。単に消されるだけの記事を配送してもらい保存しておくのは、明らかに無駄である。

我々は、実際に読まれているニュースグループは非常に少なく、かつ限られていることに着目して、読まれているニュースグループの記事だけを配送することで、購読するニュースグループの数を減らすことなく、ニュース記事の配送量と保存量を節約できると考えた [1]。我々はこの考えに基づいた新しいニュースシステム “DietNS” を構築し、運用実験を行なっている。DietNS を利用することで、記事の量を 98% 節約することができている。

本稿では、まず第2章で、ネットニュースの利用統計に基づいて、無駄なニュースグループの実情を報告する。第3章では、我々が開発した DietNS について概要を述べ、第4章では、DinetNS の動作について述べる。そして第5章でシステムの評価を行ない、最後に第6章でまとめを行なう。

2 無駄なニュース記事

2.1 現在のニュースシステムの仕組みと問題点

現在のニュースシステムは、購読するニュースグループの全ての記事をバケツリレー方式で配送し、各サイトで保存する仕組みになっている。配送するニュースグループの設定は静的に行なう。

この方式は、UUCP 時代から行なわれてきた方式で、IP ネットワークで NNTP [2] が用いられるようになった現在でも、同じ方式で記事の配送が行なわれている。

近年のインターネットの急速な拡がりを受けて、ネットニュースの利用者も増加してきている。これにもなって、流通する記事の量も爆発的に増加しており、現在のニュースシステムの運用において、以下の2つの問題点が表面化してきている。

- (1) 記事の配送トラフィックの増加
- (2) 記事を保存するスプール容量の不足

これらの問題は、リーフサイトにおいては特に深刻である。(1)の問題については、例えば1日に150MBのニュース記事が送られてくるとすると、64Kbpsの回線では常に回線幅の20%を、ニュース記事の配送だけで消費することになる。また(2)の問題は、ニュースサーバの運用を難しくしている。例えばSunOS 4.Xでは最大2GBのパーティションしか扱えないため、1日に150MBの記事が届くと、わずか13日で一杯になってしまう。

これらの問題に対して、現在では(1)購読するニュースグループの数を減らす、(2)記事の保存期間を短くする、などの運用上の設定で対応しているのが現状である。しかしこれらの設定は、情報収集手段としてのネットニュースの価値を下げてしまうことになる。

2.2 読まれているニュースグループ

和歌山大学では、平成8年6月6日の時点で5,529のニュースグループを購読している。この日のニュースサーバのログによると、この日は1日に100,775通の記事が届けられている(この数からは、サーバが記事の受信を Refuse もしくは Reject した記事数は抜いてある)。この日はクライアントから合計692回のコネクションを受けており、のべ4,167ニュースグループから14,335通の記事が読まれている。

ところが、この 4,167 はアクセスされたのベニューズグループ数である。この数字から重複を省いた、この日 1 日で実際に読まれた (1 度でもニュースサーバに NNTP の読みだしリクエストがあった) ニュースグループはわずか 499 であり、全体の 9% にしか過ぎない。これは、読まれているニュースグループが少数に限られていることを示している。

この原因として、日本においては英語の問題があるため、fj などの日本語のニュースグループは集中的に読まれるが、comp や alt などの英語のニュースグループは、ほとんど読まれていないことが挙げられる。実際に、1 日に何度もアクセスされているニュースグループは、ほとんど全てが日本語のニュースグループである。

統計から、91% のニュースグループの記事は全く読まれていないことがわかる。これらの読まれていない記事のほとんどは、このまま読まれないまま保存され、保存期間が終了すると expire されている。これは明らかに無駄である。

3 DietNS の概要

3.1 着眼点

我々は、実際に読まれているニュースグループが非常に少なく、かつ限定されていることに着目して、読まれているニュースグループの記事だけを配送することで、ニュース記事の配送量と保存量を節約できると考えた。

すなわち、購読している 5,529 ニュースグループの全ての記事を配送するのではなく、読まれている 499 のニュースグループの記事だけを配送するようにすれば、記事の量を大幅に節約することができる。

さらに、記事を配送するニュースグループを、クライアントのアクセスに基づいて on demand に自動決定する仕組みを導入すれば、読者の状況を記事の転送に自動的に反映することができる。

我々は、この考えに基づいた新しいニュースシステム “DietNS” を構築し、運用実験を開始した。

3.2 システムの特徴

DietNS の特徴は、以下の通りである。

- 読まれているニュースグループの記事のみを配送する

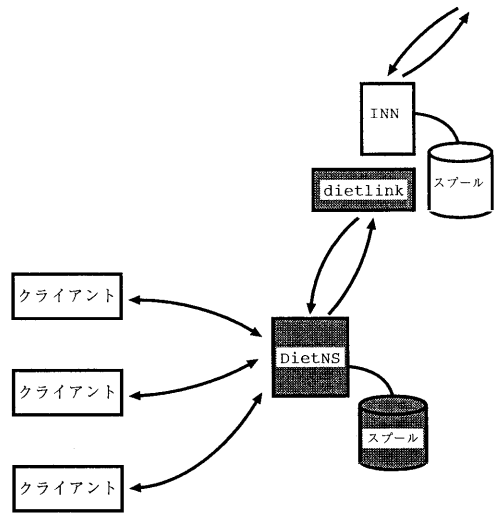


図 1: システム構成

- 記事を配送するニュースグループを on demand に自動決定
- 読まれなくなったニュースグループは自動的に記事の配送を中止する
- クライアントに対して快適なレスポンスの実現
- expire をニュースグループ毎に設定可能
- システムの状態を管理者にメールで通知する機能

3.3 システム構成

DietNS の構成を図 1 に示す。DietNS は、INN や Cnews といった通常のシステムと同様に、ニュースサーバとして機能する。クライアントは、単に NNTPSERVER を DietNS が動作しているマシンに設定するだけで、利用できるようになる。

記事の配送は、1 つ上流のサイトで動作するプログラム “dietlink” と DietNS が協調して行なう。dietlink は nntplink に相当するプログラムであり、daemon として動作する。

DietNS は、記事の情報を、active と history で管理している。また、DietNS は、通常のニュースシステムと同様に、記事を保存するスプールを持っている。

なお開発言語は、DietNS は Perl、dietlink は C 言語である。

```
saga-u.cc.request:1:1:14
saga-u.general:9:4:15
saga-u.misc:4:4:15
saga-u.q-and-a:8:6:15
saga-u.test:25:12:16
```

表 1: 読みだし要求の統計

4 システムの動作

4.1 記事を配送するニュースグループの決定

DietNS は、ニュースリーダ (クライアント) のアクセス状況から、「読まれているニュースグループ」を自動的に決定する。DietNS は、表 1 に示す、クライアントの読みだし要求の統計を取っている。これは、左からニュースグループ名、要求数、要求日数、最初に要求された日からの経過日数、を表している。この情報を基に、「読まれているニュースグループ」を決定すれば良い。

しかし現在は、1 回でも読みだし要求があったニュースグループは、全て「読まれているニュースグループ」として、記事の配送を開始するようになっている。既に述べたように、アクセス回数が 1 回以上のニュースグループを全て合わせても全体の 9% にしか過ぎないのが現状であり、当面は問題無いと考える。

記事を配送していないニュースグループに新たに読みだし要求があった場合は、そのニュースグループの記事を配送するニュースグループに自動的に追加する。DietNS は、記事を配送するニュースグループ情報の更新を 1 日 1 回行なう。

4.2 記事の配送と保存

上流の INN に新しく記事が届いた時、dietlink はまず DietNS に対して新しい記事が届いた旨を知らせる。通知は以下のフォーマットで行なわれる。

```
IHAVE <Message-ID> newsgroup
```

これは、「上流のサーバに、ニュースグループ newsgroup に、メッセージ ID が Message-ID である記事が届きました」、ということの意味する。

DietNS はこのメッセージを受け取ると、記事を配送するニュースグループの記事であれば、記事情報の history (次節参照) を参照して記事を既に持って

いるかどうかをチェックし、記事を持っていない場合は dietlink に記事の配送を要求して記事を取り込み、記事情報を更新して記事をスプールに保存する。記事を配送しないニュースグループであれば、記事情報の更新のみを行ない、記事の配送は要求しない。

DietNS は購読している全てのニュースグループの記事情報を持っているが、記事自体は読まれているニュースグループの記事のみ持っている。全てのニュースグループの記事情報を持っているのは、記事情報だけで答えられるクライアントからのリクエストに答えるためである。

4.3 記事情報の管理

DietNS は、受け取った記事の情報を、active と history で管理している。これらは頻繁に参照されるものであるため、NDBM を用いて実装している。

- active

購読可能なニュースグループの記事の情報。各ニュースグループで読むことができる記事番号を記録している。クライアントのアクセス時に参照される。

- history

受け取った記事の情報を記録する。記事のメッセージ ID、スプールでの保管場所、記事を受け取った時間を記録する。記事の expire 時に参照される。

4.4 クライアントへの応答

クライアントが発行する NNTP リクエストに対する応答は、記事情報だけで答えられる LIST や GROUP については DietNS が答える。記事自体を要求する ARTICLE や HEAD については、記事を持っているニュースグループについては DietNS が答え、持っていないニュースグループについてはリクエストを 1 つ上流のニュースサーバに転送し、そのレスポンスをクライアントに返している。

4.5 抜けている記事を取り寄せる

4.1 節で述べたように、記事を配送していないニュースグループに対してクライアントが読み出しのアクセスをした場合は、そのニュースグループの記事を配送するニュースグループに自動的に追加する。ただし、記事を配送するニュースグループ情報の更新は 1 日

に1回だけ行なわれるようになっており、更新後からDietNSはそのニュースグループの記事も受け取るようになる。

DietNSはアクセスがあった時点で上流のサーバからそのニュースグループの記事を取り寄せてクライアントのリクエストに対応する。しかし、クライアントの要求があった時点から更新が行なわれるまでの間、記事そのものは送られてこないため、その間に上流に届いた記事はDietNSでは抜けてしまうことになる。

このためDietNSは、配送開始後に抜けている記事を調べて、抜けている記事があれば上流のサーバから記事を取得して埋める。

4.6 記事の消去

記事の量が節約できるとはいえ、記事のスパールディレクトリには物理的な限界が存在することにはかわりはない。このため、INN同様に古い記事を消去するexpireの仕組みを備えている。

DietNSでは、ニュースグループ毎にexpire期間を設定できる他、expireの設定を行なうexpire.ct1にワイルドカードが使用できるようになっている。また、neverが指定されたニュースグループについては、永遠に記事を消去しない。

expireが実行されると、スパール中の記事が消去され、記事情報が更新される。

4.7 記事の配送の停止

全ニュースグループ数5,529のうち、読まれているニュースグループは499であるが、499のうち308(62%)が、1度しか読まれていないニュースグループである。時間がある時に普段は読まない沢山のニュースグループを斜め読みする人が多く見られるが、このような人の場合は、眺めたニュースグループをもう一度読むことは少ないと思われる。

現在のDietNSは、1度以上読まれたグループの記事を「読まれているニュースグループ」としているため、このような人が眺めたニュースグループの記事も配送することになるが、そのようなニュースグループの記事を永遠に配送し続けることは、明らかに無駄である。

したがって、DietNSでは一定期間読まれていないニュースグループについては、自動的に記事の配送を停止する機能を実装しており、defaultではこの期間を2週間に設定している。

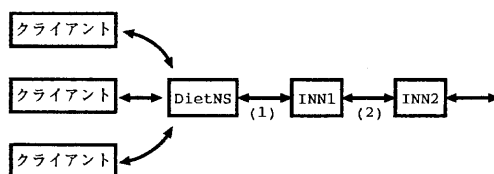


図2: 実験環境

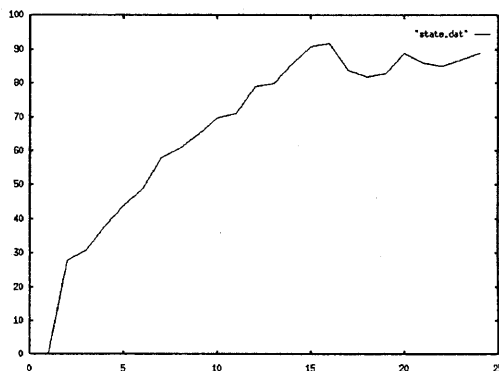


図3: 記事を配送するニュースグループ数の推移

5 システムの評価

5.1 実験環境

現在DietNSは、佐賀大学工学部情報科学科で、運用実験を行なっている。実験は図2に示す構成で行なっている。本章では、運用を始めてからの統計を元に、DietNSの評価を行なう。

5.2 配送しているニュースグループ数

平成8年6月6日の時点で、INN1が購読しているニュースグループ数が2,418であるのに対して、DietNSでは89ニュースグループの記事を配送している。これは全体のわずか4%である。

図3に、DietNSの運用開始から現在までの、記事を配送しているニュースグループ数の推移を示す。グラフからわかる通り、運用開始後しばらくして、グラフがほぼ横ばいになっている。読者が同じであるならば、今後もこのまま推移すると思われる。

	INN1	DietNS
記事数	25,961	542
バイト数	59,452,204	1,033,831

表 2: ある日に受け取った記事量の比較

5.3 ニュース記事の転送量

平成 8 年 6 月 6 日について、図 2 の (1) と (2) を通過したニュース記事の記事数とバイト数の比較を、表 2 に示す。表からもわかる通り、DietNS が受け取った記事数は INN の 2%、バイト数は 1.7% である。

5.4 ニュース記事の保存量

平成 8 年 6 月 6 日の時点で、INN1 のスプールディレクトリの容量は 140MB である。但し saga-u 以外のニュースグループについては、expire を 1 日に設定している。

同じ日の DietNS のスプールディレクトリの容量は、29MB である。但し expire は fj が 360 日、saga-u などのローカルニュースグループは永久保存、alt が 10 日、その他は 1 日になっている。

DietNS では、スプールディレクトリの容量が少なく済むため、expire の期間を長期間にすることが可能となっている。

6 まとめ

我々は、実際に読まれているニュースグループは非常に少ないということに着目して、読まれているニュースグループの記事だけを配送することで、ニュース記事の配送量と保存量を大幅に節約する新しいニュースシステム “DietNS” を構築した。運用実験では、DietNS を利用することで、記事の量を 98% 節約することができている。

on demand に記事を取り寄せるニュースシステムとしては、名古屋大学で開発された DNAS [3] がある。DNAS はニュースリーダーのフロントエンドとして機能し、クライアントが要求した記事をその時点でニュースサーバから取り寄せ、クライアントに送るとともにキャッシュを行なう。キャッシュにより、2 度目以降はサーバにアクセスする必要はない。

しかし DNAS は要求があった時点で記事を取り寄せるため、素早くクライアントに記事を送るためには、DNAS とニュースサーバの間の通信回線が高速であ

る必要がある。これに対して DietNS は、読まれているニュースグループについては常に記事を配送しているので、サーバにアクセスする必要はない。したがって、クライアントに対して高速にレスポンスを行なうことができる。記事を配送していないニュースグループについては、最初に要求があった時はサーバから記事を受け取る必要があるが、その後は自動的に記事を配送するようになるので、2 度目以降はレスポンスの問題はない。

DNAS はニュース記事を単位とした on demand なシステムであるのに対し、DietNS はニュースグループを単位とした on demand なシステムであると言える。

全てのニュースサーバ間で、バケツリレー式に全ての記事を配送する現在のニュース記事配送の仕組みは、コストパフォーマンスが非常に悪いと言わざるを得ない。我々は、DietNS はリーフサイトにおけるネットニュースの購読に、特に有効であると考えている。DietNS や DNAS の on demand に行なうキャッシュ機能を活用することで、ニュースシステムの記事の配送を最適化することが期待できる。

謝辞

DietNS の運用実験に御協力いただいている、佐賀大学理工学部情報科学科知能情報学講座の皆様へ感謝する。

参考文献

- [1] 小串英俊, 渡辺健次, 近藤弘樹: “インターネットニュースが抱える問題点と、その解決に関する考察”, 平成 7 年度電気関係学会九州支部連合大会講演論文集, p. 639 (1995).
- [2] Brian Kantor and Phil Lapsley: “Network News Transfer Protocol”, RFC977 (1986).
- [3] 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, pp. 39-47 (1992).