

インターネットにおけるフィードバック型情報集約方式の構想

小泉 泰則 吉田 善幸 長岡 満夫
NTT ソフトウェア研究所

〒180 東京都武蔵野市緑町 3-9-11

あらまし

インターネットを利用した情報コンテンツへのリターンチャンネルと利用者への集約情報の提供は、視聴者参加型サービス（投票系、予約系、アンケート系等の従来電話系マスコリングサービス）の基盤として重要である。しかし、インターネットの特徴である不特定多数の利用、様々な形で増殖する情報コンテンツへの設備の効率化等を考慮すると、負荷と機能の最適な分散が必要となる。このため、リターンチャンネルで視聴者アクセスを適切に負荷分散し、かつ柔軟な情報コンテンツ配置による機能分散方式を提案する。さらに、本方式により負荷分散の評価を行い、方式の妥当性を検証する。

キーワード インターネット、負荷分散、フィードバック

A Conception of Method to Collect Feedback Type Information on the Internet

Yasunori Koizumi, Yoshiyuki Yoshida, Mitsuo Nagaoka
NTT Software Laboratories
3-9-11 Midori-cho Musashino-city Tokyo 180 Japan

abstract

It is important for the services which audience joins to give a return channel to information content and to send collected information to users on the internet. But as internet is used by many unspecified people and information contents increase in many types, distribution of user access and function is needed, considering efficiency of the equipment. This paper proposes the distribution of audience access in the return channel and distribution of function by placing information contents flexibly. And we estimate of the distribution method and verify the method.

Key words internet, distribution, feedback information

1. はじめに

インターネットを活用して、様々な情報コンテンツのリターンチャネルの提供と利用者への集約情報 [投票結果状況、予約状況等] の提供 (これをフィードバック型情報集約方式と呼ぶ) は、視聴者参加型サービス (投票系、予約系、アンケート系等の従来電話系マスコリングサービス) の基盤として重要である。しかし、インターネット利用者の増加に伴い、応答件数がサーバの能力を超え、サービスが十分に機能しないことが想定される。また、インターネットはTV系のマスメディアとの融合も進みつつあり、インターネットを利用した視聴者の応答が特定の時間に集中することが考えられるが、従来の個別の集中サービス方式では対処できないことになる。

本問題に対し、フィードバック型情報集約方式に適した情報サーバおよびネットワーク負荷の分散、機能の分散による解決策を提示し、さらに性能評価に基づく方式の妥当性を検証する。

2. 背景

2.1 情報発信需要の拡大と利用者情報の活用

放送型サービスのダウンサイジングが起こっており、インターネット放送という手段が実現できるようになってきた。また、個人や企業ミニコミュニティでの情報発信需要が拡大してきており、企業のイントラネットでの情報共有も進んでいる。このような状況で低コストかつ容易な情報発信技術が今後発展していくことが予想される。情報発信者は個人の意見をフィードバックし、プロモーション活動への反映 (的確な市場動向・ユーザ動向を把握する) とともに、利用者への魅力的情報コンテンツの提供維持が可能となる。

2.2 放送とインターネットの融合

放送とインターネットを融合することで、放送視聴者ユーザのインタラクティブな情報応答を放送系/情報プロバイダ/サービスプロバイダに効率的に反映することができる。このため、マスメ

ディアとの連動により、放送視聴者へのインターネットを介したリターンチャネルの提供、放送とインターネットのメリットをあわせた新たなサービスを得ることができる。

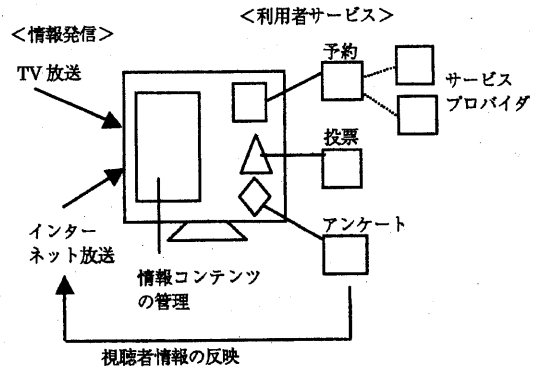


図1 視聴者参加型サービス形態

3. 視聴者参加型サービスの特徴

(1) 応答の集中

テレビ番組と連動した人気投票や議論に対する賛否の投票、クイズの回答などでは、テレビからの回答要求時に視聴者からの投票アクセスの集中が発生する。インターネット単独のサービスにおいても、利用者参加型のサービスでは参加する人数が多いほど情報量が増え人気のあるサイトとなるため参加者数の増加が必要だが、応答の数の多さが性能上のネックとなる可能性がある。

(2) 視聴者応答フィードバックの迅速性

視聴者参加型サービスでは、クイズの回答、アンケート調査など番組の中で収集した情報をもとに番組を進めることが多いので、視聴者から収集した情報をもとにした集計結果、回答正解者の提示などの視聴者からの応答を迅速に整理し番組に反映することが求められる。

(3) 参加者へのコンテンツ提供量の増加

インターネット単独のサービスでは、インターネット放送や参加者からの情報といったコンテンツを迅速に提供する必要があるが、参加者が多数のため提供するサーバやネットワークに大きな負担がかかる。

(4) 参加者からの収集情報の効率的利用

視聴者からの情報を迅速に番組に提供する形態に加工すること、チケット予約など管理しているチケット残数をもとに視聴者からの応募情報に回答すること、視聴者からの情報をもとに統計的に視聴者の嗜好等を分析し、今後の番組作りに反映することのために、収集する情報の構成および集計するデータベースの構成を効果的に作成する必要がある。

4. フィードバック型情報集約方式

上記の問題に対し、下記の方式により視聴者参加型サービスを提供するための基盤を構築する。

(1) コンテンツ配信および回答情報の負荷分散とフィードバック情報の分散集計

1台の企画サーバと複数のコンテンツサーバによりシステムを構成する。企画サーバは、アクセスするユーザに対し、適切なコンテンツサーバを指示する。コンテンツサーバはユーザとのインタフェース部となり、ユーザにコンテンツ情報を提供する。そして、コンテンツサーバはユーザからの回答を取得し、ローカルなデータベース上でアクセスしているユーザからの回答を管理する。企画サーバは定期的にコンテンツサーバのデータベースにアクセスし、全ユーザからの回答を集計し管理する。そして各コンテンツサーバに全体の集計結果を提供する。このように企画サーバおよびコンテンツサーバが、それぞれデータベースを持ち、それらを協調させることによりコンテンツ情報およびユーザ情報の効率的な流通を実現する。

(2) 仮想 URL による複数サーバの隠蔽とサービスの一元管理

企画サーバにおいてサービスごとの仮想の URL を管理し、ユーザが指定する仮想の URL をもとに企画サーバがサービス実施状況、サービスに対応するコンテンツサーバを検索し、適切なコンテンツサーバの実 URL に変換し、そのコンテンツサーバとユーザのコネクションを確立させる。

(3) ネットワーク輻輳への対応

コンテンツサーバを適切なユーザグループ単位に配置し、ローカルなデータベースによりグループ内でユーザ情報の一時集約を行いグループ外に対するネットワーク負荷を軽減する。一時集約したユーザ情報は企画サーバの要求に応じて、企画サーバとコンテンツサーバ間で送受信されるので、頻繁にデータが送受信されることがない。

(4) 流通する情報の分類

各コンテンツサーバのローカルデータベース、ローカルなデータベースの情報を集計する企画サーバ上のデータベース、フィードバック情報をユーザに提供するコンテンツサーバのデータベースを配置し、企画サーバの指示により集計、整合を行うことにより効率的なフィードバック情報の収集・提供を行う。

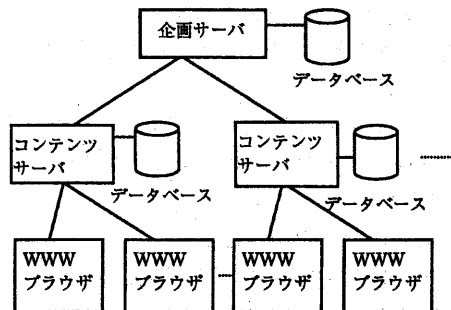


図2 システム構成

5. 方式の実装方法

5.1 URL の一元化

1台の企画サーバと複数のコンテンツサーバおよびサービスにアクセスするユーザがログインしているクライアントマシンからシステムが構成される。クライアントは、一度企画サーバに接続し、コンテンツサーバの URL を取得し、その URL でセカンダリサーバに接続を行う。

トラヒックの分散については、企画サーバが順次、接続コンテンツサーバの URL を変更すること

により、コンテンツサーバへの接続数を平均化する。

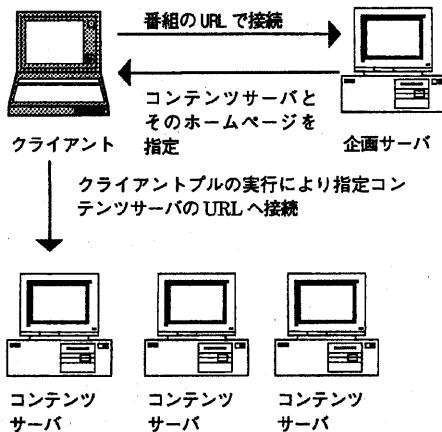


図3 URLの振り分け

5. 2 企画サーバ

企画サーバは、クライアントからの番組へのアクセス要求を事前に登録されたインターネットサーバフィルタを介して認識し、各番組固有のスク립トを実行する。

クライアントによる番組へのアクセスは、企画サーバ名および番組名を以下のように指定することで実行される。

http://企画サーバ名 /番組名

なお、このパス名は仮想のパス名であり、フィルタプログラムにより認識される。

各番組固有のスク립トは、番組の有効性（提供中か否か）およびクライアントが参照する実際の番組のホームページのコンテンツサーバ URL を作成する。

このスク립トがクライアントに送信されると、<META>タグによりクライアントの Client Pull が動作し、コンテンツサーバへのアクセスが開始される。

この各番組固有のスク립トは、番組名タイトルの記述と URL のパス名の記述を変更するのみで、各番組に対応することが可能である。

以下に、クライアントからの番組へのアクセスからスク립ト実行までのコンテンツサーバでの動作について説明する。

5. 2. 1 クライアントからの番組へのアクセス

企画サーバは、クライアントからのアクセス要求を受信するとフィルタを呼び出す。

フィルタは、要求によって取得したパス名をチェックし、パス名が「番組名」の場合、番組へのアクセスとみなし、受信したパス名を番組固有のスク립トを起動するように URL を変更する。次に変更した URL を使用して HTML スクリプトを起動する。以下に処理を示す。

5. 2. 2 スクリプトの動作

(1) 番組の有効性のチェック

番組の有効性は、現在時間がテーブルに記載されている開始時間と終了時間の範囲内にあるかをチェックし、範囲内にある場合はコンテンツサーバの抽出を行う。

現在時間がこの範囲内でない場合、当該番組が終了した旨を表示する html を出力し、終了する。

(2) コンテンツサーバの抽出

コンテンツサーバとして抽出した回数が最小であることを抽出条件としてサーバ名を抽出後、抽出したサーバの回数をアップデートする。

抽出したサーバ名は、<META>タグに番組のホームページのパス名と連結し、接続すべきコンテンツサーバの URL を埋め込む。

この<META>タグの CONTENT は0としクライアントに当該 HTML を出力後、クライアントが即座にコンテンツサーバにアクセスし、実際の番組ホームページへのアクセスが開始される。

5. 3 コンテンツサーバごとに分散したデータベースによる段階的集計

本システムでは、データベースをコンテンツサーバに分散配置し、投票結果は各コンテンツサーバごとに集計し、企画サーバが定期的（現在は10秒ごと）にコンテンツサーバの集計結果を収集

し、その合算集計結果をコンテンツサーバの表示用集計テーブルに反映する構成となっている。

(1) クライアントからの投票結果は、まず、接続しているコンテンツサーバにある投票カウンタ・テーブルを更新する。

フィールド名
コンテンツサーバ番号
投票項目番号
投票カウンタ
投票カウンタテーブル

(2) 企画サーバは投票カウンタテーブル内容を収集し、全投票結果を集計し、全投票結果集計テーブルを更新する。同時に、企画サーバは全投票結果を各コンテンツサーバに送信する。

フィールド名
投票項目番号
集計値
全投票結果

(3) 各コンテンツサーバは受け取った全投票結果を各クライアントに送信し、投票結果を更新する。

6. 性能評価

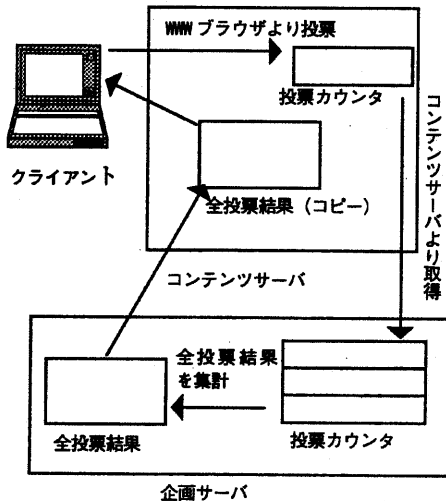


図4 フィードバック情報の流れ

本システムによる方式で、クライアントからの投票状況に応じて、どのような処理性能が予想され

るかについて、各処理時間をもとに評価する。各クライアントからの投票が行われる数を以下のよ

うに想定し評価を行う。

インターネット人口：500万人

投票率：2%

ピーク時集中率：3%

として、
ピーク時アクセス数を約 3000 人/分と想定する。
本値を以下の式では A 人/秒 (50 人/秒) と記述し評価する。また、各処理に要する時間を以下の値で評価する。

- ・企画サーバでの URL 払い出しに要する処理
…処理時間 T_1 秒 (約 0.01 秒)
 - ・投票に要する処理時間
…処理時間 T_2 秒 (約 0.2 秒)
 - ・企画サーバが各サーバより情報を収集する処理
…処理時間 T_3 秒 (約 0.2 秒)
- コンテンツサーバの数を N 台とする。

6. 1 コンテンツサーバへの投票時の負荷状況

各ユーザがクライアントから投票する間隔は、個々の投票者の意志であり、投票者間の連動はないため、ランダムであり、投票数の分布はポアソン分布とみなせる。また、各コンテンツサーバの投票処理時間は固定時間とみなせる。投票は複数のコンテンツサーバでそれぞれ行われるので、ポアソン到着・固定時間サービス待ち行列系に適合するので、分散系の M/D/1 型にあてはまる。

・ λ (到着率) : A/N
 ・ μ (サービス率) : $1/T_2$
 より、平衡条件

$$\rho = \lambda / \mu = A \cdot T_2 / N < 1$$

を満たすには、

$$N > A \cdot T_2$$

のため、少なくとも $A \cdot T_2$ 台以上のコンテンツサーバが必要である。

また、平均待ち時間は、

$$W_q = \rho / \{2\mu(1-\rho)\} = \frac{1}{2} \frac{AT_2^2}{N-AT_2}$$

となる。したがって、上記値を使用すると、平衡条件より最低11台以上のマシンは必要であり、かつ、平均待ち時間より、 $N > 20$ では平均待ち時間が0.1秒より小さくすることができるので、適切なサービスが可能となる。

6. 2 企画サーバにおける URL 払い出し時の負荷状況

企画サーバでは1つのクライアントがアクセスしてから実際に情報のやり取りをするコンテンツサーバの宛先を通知する処理を実行するが、このアクセス到着率も、個々に依存しないランダムなアクセスである。そして、本アクセスのサービス時間は固定時間なので、これをモデルに適用すると、M/D/1のモデルを適用できる。

λ (到着率) : A

μ (サービス率) : $1/T_1$

となる。これより

平衡条件は

$$\rho = \lambda / \mu = A \cdot T_1 = 0.5 < 1$$

を満たす。

平均待ち時間は、

$$W_q = \rho / \{2\mu(1-\rho)\} = \frac{1}{2} \frac{AT_1^2}{1-AT_1}$$

となるため、0.0001秒となり問題はない。

6. 3 企画サーバがコンテンツサーバの情報を収集するときの負荷

各コンテンツサーバが企画サーバにアクセスしデータベースを更新するときの負荷について考察する。このとき、アクセス頻度は定間隔であり、現在のデフォルトは10秒としている。コンテンツサーバの数をN台、企画サーバのデータベース更新処理時間を T_3 秒とすれば、アクセス頻度は $N \cdot T_3$ 秒以上にしなければならない。したがって、コンテンツサーバが数十台の条件であれば問題はない。ただし、コンテンツサーバの数を増やした場合、またはアクセス頻度を短くする場合には問題が生ずる可能性がある。コンテンツサーバを複数のグ

ループに分け、集計を段階的に行うことで集計負荷の分散を行うことを考慮する必要がある。

7. 終わりに

フィードバック型情報集約方式のためには、企画サーバとコンテンツサーバによる集計とユーザアクセスの両者の負荷分散が必要であることがわかった。しかし、コンテンツサーバからの集計については集計頻度の増加やコンテンツサーバの増加に対して、集計の段階的の実行による負荷分散が必要となる。机上評価では、現状想定したアクセスに対して適切なサービスが提供できる見通しを得た。

今後は、具体的な応用サービスへの適用実現を基に、実証実験での評価を行うとともに、適用性の検証を行う。本方式での実証実験結果は別途報告する。

参考文献

- [1] 本間：“待ち行列の理論”
- [2] 前川 守 (1992). ソフトウェア実行開発/環境, 岩波書店.