5F-1 WWWのクライアント特性に適応するサーバーのモデル

河辺 岳人†, 王 家宏‡, 宮崎 正俊‡

†株式会社ＳＲＡ東北　‡岩手県立大学ソフトウェア情報学部

1．はじめに

World Wide Web (WWW) では、標準に準拠したシステムを用意することにより、情報提供者・情報利用者とも特定のベンダに依存することなくシステムの構築・利用が可能となっている。このため、旧来の情報システムと異なり、利用者の使用している端末ソフトウェアや取得後の情報の利用方法も多彩になってきている。

情報を提供する側には，このような多彩なクライアントの特性に対応した情報提供や運用を行いたいという要求がある．本稿では，このようなサーバーシステムを構築するにあたって必要となる，クライアント特性を予測する関数を組み込んだサーバーのモデルについて提案する．

2．背景

本研究の動機となっているWWWの現状と，クライアント特性の判別の必要性について次に述べる．

2.1 ブラウザであるという前提

WWWの利用者は大きく情報の提供者と情報の利用者に分けることができる．情報提供者はサーバーと情報(コンテンツ)を用意して情報提供を行い，情報利用者はクライアントとして閲覧用ソフトウェア(ブラウザ)を用いて情報を取得する，という形式が最も一般的なWWWの使われ方である．

従って情報提供者側は，取得された情報は主にブラウザによってレンダリングされて人間の目にふれるという前提のもとで情報を整理・作成・提供していることが多い．

2.2 非ブラウザの存在

しかし，クライアントとして使われるのはブラウザばかりではない．こういった多様性はベンダ非依存でアプリケーション構築が自由に行なえるWWWならではの特性と言えよう．非ブラウザ型クライアントの例としては

- 検索エンジンのためのデータ収集ロボット
- 更新検出
- 電子メールアドレスだけを収集
- 内容監視・検閲

といった目的を持ったものが存在する[4].

ブラウザとは異なった目的をもって作られたクライアントでは，サーバーより取得した情報の利用の仕方も通常のブラウザと異なる．提供側が「こういう風に見てもらいたい」といった意図をもって情報を提供していても，常にそれが満たされているとは限らないことになる．非ブラウザによるアクセスは全体の４割を占める場合もある[2].

2.3 関連研究とその適用性

このような事情から，情報提供者からは，意図せざる方法での利用を防ぎたい，あるいは防がずとも情報利用の特性によってサーバーの応答を変化させたいという要求が出てくる．このためには，サーバーに要求を出してくるクライアントの素性や特性を知ることが必要になる．

これに関連した手法や研究としては，Web Usage Mining と 侵入検出システムがある．

Web Usage Mining[3]は，主にサーバーのログファイルを解析することにより利用者の傾向や情報の利便性に関する評価を引き出すことを目的としている．しかし，Web Usage Miningにおいては利用者はブラウザのみを用いて情報を取得していることを大前提としている．キャッシュサーバーによる誤差についてはどの研究でも考察されているが，非ブラウザによるアクセスの影響については言及程度であっても極めて例は少ない[5]．また，事後のログ解析によっているため，今来た要求に対しすぐに応答を返さなければならないというサーバーには適用しにくい．

侵入検出システム (Intrusion Detection System, IDS) [6]は，機器やネットワークにおける不正なアクセス，あるいは不正の前兆と思われる現象を探知するシステムである．しかし，まずWWWでは誰でも情報取得が可能であることが基本であるため，「不正アクセス」が基本的に存在しない．また，侵入検出システムは現象の探知までがその適用範囲であり，出力は観測対象とは独立して警報を発するところまでである．探知出力によってサーバー出力の応答を逐一変えるといった用途には適用が難しい．

このため，従来のWWWサーバーにWeb Usage Miningや侵入検出システムそのまま適用して目的を達成することはできない．

A Server Model Adapting to WWW Client Characteristics

Taketo Kabe, Jiahong Wang, Masatoshi Miyazaki
Iwate Prefectural University, Iwate, JAPAN

3．サーバーのモデル

情報提供者の目的は，サーバーに対する要求に対応する応答を管理方針等に沿って最も適したものにすることである．要求と応答は必ず対になっているため，仕様は応答を生成する手順の中に組み込まなければならない．このようなサーバーを構築する際のモデルを次に示し，クライアント特性の検出がモデル中でどのような位置付けになるのかを示す．

3.2 情報提供者の要求のモデル

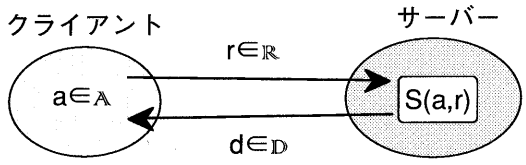

図１．設計段階でのサーバーモデル

𝔸: クライアントの特性の集合．

ℝ: クライアントの要求の集合．

𝔻: サーバーの応答の集合．

a: あるクライアントの特性．

r: ある要求に関してサーバーで観測される事象すべて．

d: その要求に対するサーバーの応答．

S(a,r): サーバー関数．情報提供者が要求するサーバーの応答仕様．

$$d=S(a,r). \quad (1)$$

情報提供者が要求する応答仕様は，実際には低レベルの事象そのものに対して記述されるわけではなく，事象を解釈した結果に対して行われる．したがって，サーバー関数S(a,r)はさらに以下のように分解される．

f(r): 事象rを解析・解釈する関数．

c: 事象に対する解釈．サーバーの応答仕様に現れる条件やデータを表す．

Qc(a, c): 要求としての応答仕様．

$$c = f(r). \quad (2)$$

$$S(a,r) = Qc(a, f(r)). \quad (3)$$

このモデルに基づくと，情報提供者の要求するサーバーは，aおよびrに対する応答dを出力する関数Qcを最適化することにより実現される．ただ，より適しているという評価基準が定量的であることは少ない．

◇ cの例：

・「事象(要求) rが表すURL」

・「自動検出用という特性を持つクライアント」

・「メールアドレス収集用のクライアント」

◇ Qc() の例：

・「URLが/pathであれば，*DocumentRoot*/pathを応答として返す」

・「自動更新検出であれば更新時刻のみを返す」

・「メールアドレス収集のアクセスは拒否する」

3.2. 実現可能なモデル

しかし，実際にWWW上の通信に使われているHTTP[1]ではクライアントの特性aをサーバー側から観測することは出来ない．したがって，実現可能なサーバー関数はaを引数に含んではならない．このため，aの代用としてaに相当するものを予測する関数gを導入する．

g(r): 事象rよりクライアントの特性を予測する予測関数．

a': 事象rから予想したクライアントの特性．

S'(a',r): 実現可能なサーバー関数．

$$a' = g(r). \quad (4)$$

$$S'(a', r) = Qc(g(r), f(r)). \quad (5)$$

実現されたサーバー関数S'の妥当性，つまり要求された応答仕様Sとの類似度は予測関数gの妥当性に依る．有意なgの存在にはaとrに相関があることが前提となるが，gの作成方法は自明ではない．作成のためには実験や統計的手法が必要となる．

4．まとめ

WWWの特性と実際の環境から，サーバーがクライアントの特性を判別する必要性について述べ，この機能を組み込むためのサーバーのモデルを提案した．今後はモデルを詳細化するとともに，予測関数の作り方や評価と実装について進めていく予定である．

参考文献

[1] Roy Fielding et al, "*Hypertext Transfer Protocol -- HTTP/1.1*", RFC2616, June 1999.

[2] 河辺岳人, 宮崎正俊, "WWWにおけるUser-Agent特定のためのアクセスログ解析手法", 情報処理学会第59回全国大会(平成11年後期) 3T-01.

[3] Robert Cooley, Bamshad Mobasher, Jardeep Srivastava, "*Data Preparation for Mining World Wide Web Browsing Patterns*", Journal of Knowledge and Information Systems, Vol.1,No.1,1999.

[4] Taketo Kabe, Masatoshi Miyazaki, "*Determining WWW User Agents from Server Access Log*", Proceedings of the Seventh International Conference on Parallel and Distributed Systems (ICPADS2000): Workshops pp173-178.

[5] Ian Marshall, Chris Roadknight, "*Linking cache performance to user behaviour*", 3rd International Caching Workshop, Computer Networks and ISDN Systems 30 (1998) 2123-2130.

[6] 武田 圭史, 溝江宏真, 武藤佳恭, "ネットワーク侵入検出手法の比較と脅威に応じた動的な検出", 情報処理学会第59回全国大会(平成11年後期) 4T-06.