2T-7 WWWサーバにおけるログの多次元階層化に基づく柔軟なトラフィック管理手法の検討

石塚 宏紀† 十川 基‡ 斉藤 裕樹† 戸辺義人†
東京電機大学 工学部 †情報メディア学科 ‡情報通信工学科
{isi, hajime, hsaito, yoshito}@unl.im.dendai.ac.jp

1. はじめに

World Wide Web (WWW) 技術は急速に社会に浸透し，多種のサービスが提供されている.しかし，トラフィックの増加による品質劣化が生じ，サーバの品質管理が望まれている．従来のサーバ・ログを利用したアクセス解析手法では，ログの形式が構造化されていないため，Webアプリケーション単位やアクセス元単位といった多様な観点での分析が困難である．また，性能面での記録が不足しているため，トラフィック計測には不十分である.

本稿では，ログをアプリケーション単位，アクセス元のドメイン名・IPアドレス，アクセス時間といった複数の観点から分類・構造化し，これらを組み合わせることで複雑なトラフィック解析を行う手法を検討する.

2.トラフィック管理技術

2.1 従来のトラフィック管理技術

- パケットトレースを用いた管理
  パケットトレースによる方法[1]では，解析やトレースに専用アプリケーションが必要であり，膨大なデータの分析に多くの計算資源を必要とする.
- アプリケーション・ログを用いた管理
  従来のサーバアプリケーションが記録するログを用いた手法は，主にアクセス傾向の解析を目的としたものであり，トラフィック解析の面では不十分である．また，ログのデータ形式が構造化されていないため，多観点でのログの解析は難しい.

上記の問題を解決するため，新たな手法を提案する.

2.2 サーバログの多次元階層化の提案

Webサーバのトラフィック解析を行うためには，アプリケーション単位や，アクセス元ごとや、時間帯ごと等の複数の観点からの解析が必要である．さらに，これらの観点を組み合わせ，「特定時間帯の特定アクセス元からのあるアプリケーションへのアクセス」というような複合条件での解析が必要である．よって，ログを構造的に分析する必要がある.

本研究では，ログのデータ構造としてtree構造を用い，複数条件下での解析を可能にするため，複数のtree構造を多次元化し解析を行うことでトラフィック解析を行う方法を提案する.

Traffic management based on multi-dimensional hierarchizing for WWW Server log, †Hiroki Ishizuka, ‡Hajime Sogawa, †Hiroki Saito, †Yoshito Tobe, Department of Information Systems and Multimedia Design, Tokyo Denki University(†), Department of Information and Communication Engineering, Tokyo Denki University (‡)

3．ログの構造化と多次元管理

3.1 ログ情報の構造化

本研究では，ログを「アプリケーション」,「アクセス元ドメイン名(srcFQDN)」,「IPアドレス(srcIP)」,「時間」の4つの観点から構造化し，トラフィックの管理を行う手法について検討する.

- サービス・アプリケーション単位の階層構造
  Webアプリケーションを構成するプログラムやファイル単位での解析、アプリケーション全体を単位とした解析、アプリケーション実行環境全体の評価といったように解析の対象範囲を変化させることができるように，プログラムやファイルの実行エンジンごとの分類や、所属するアプリケーションごとの分類を行う木構造を考える．木構造の例を図1に示す.
- アクセス元単位の階層構造
  アクセス元単位での管理を行うため，「srcIP」,「srcFQDN」を単位とした解析を行う．IPアドレス

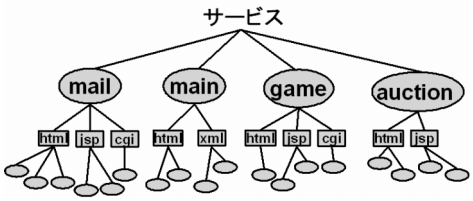

図1．サービス・アプリケーションの階層構造

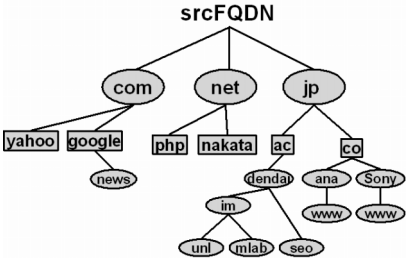

図2．srcFQDN単位の階層構造

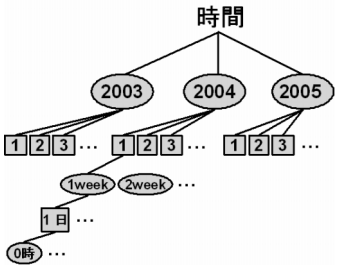

図3．時間単位の階層構造

単位の管理だけでなくネットワーク単位での解析を行うため，IPアドレスの階層化に則した構造化を行う方法[2]とする．またドメイン名の名前空間の階層化に則した構造化を行い，国や組織単位での解析を行う．例を図2に示す．

- 時間単位の階層構造

時間帯ごとのアクセスを管理するため，アクセス時刻と処理時間による解析を行う．管理する単位を「year」，「month」，「week」，「day」，「hour」とし，時間粒度の階層構造を利用する.さらに，より過去のログほど保持する粒度は荒く，最新のログほど細分化された粒度のノードを保持するようにする．例を図3に示す．

3.2 構造化したログ情報の多次元管理

ログ情報の構造化に加え，複数条件下でのアクセス解析を可能にするため，個々の木を一つのrootノードの元で一元管理する必要がある．

図4に示すように，各階層構造をそれぞれの次元と捉え，各次元のノードを全て関連付けた多次元階層構造化されたログ情報のクラスタ構造を考えることで，複雑なアクセス解析が実現できると考えられる．

4.クラスタへのアクセスモデル

4.1 問い合わせ言語の設計

実際にトラフィック解析を行う際には，多次元階層化されたログ情報クラスタに対して，解析の観点に応じた情報を抽出する必要がある．クラスタへのアクセスは，抽出条件を指定した問い合わせと応答の組で行われる．クラスタへの問い合わせ言語を設計し，クラスタへのアクセスを行う方法を検討する．階層化されたクラスタに対する問い合わせにおいては，クラスタは多次元構造であるため，不特定多数の条件を組み合わせ複合条件を指定する必要がある．そこで，次のようなXMLを用いた問い合わせ言語を設計する．

```
<getRequest>
  <query dimension=“content”>/game</query>
  <query dimension=“srcIPaddr”>/133/20/
  </query>
  <query dimension=“srcFQDN”>
  /jp/ac/dendai/im </query>
  <query dimension=“time”>/2005/1</query>
</getRequest>
```

4.2 クラスタからの能動的状態通知

クラスタへの問い合わせを行うpolling機能だけではなく，さらに，クラスタに何らかの障害が検知された場合にそれを能動的に通知するEventTrap機能が必要である．これを実現するためには，予めログ情報クラスタに条件を与えておき，これを満たすログ傾向があると，管理者側にクラスタから警告が成されるようにすることで実現できる．

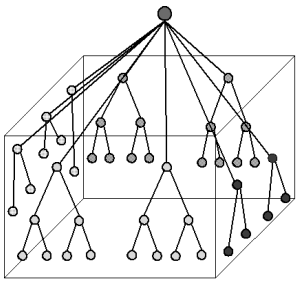

図4．多次元階層化構図

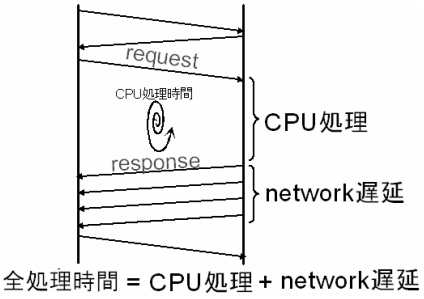

図5．トラフィック計測

5．システムの設計

我々は、これまで述べてきたようなトラフィック管理を行うシステムを設計中である。システムは次の部分から構成される．

- トラフィック計測部
  図5のように、通信フローのRequest TimeとResponse Timeを計測しフロー解析を行う部分．
- データ収集・整理部
  ログ情報を保持し，多次元化された木構造のデータ構造に基づく問い合わせ処理を行う部分．
- データ出力
  問い合わせ処理を行い、解析データの分析や可視化を行う部分．

6．おわりに

本稿では多次元階層化によるログ管理を用いたWebサーバのトラフィック管理手法を提案した．今後設計したシステムを実装し評価する予定である．

参考文献

[1] Y. Nakamura, K. Chinen, H. Sunahara, S. Yamaguchi, and Y. Oie: ENMA: The WWW server Performance Measurement System via Packet Monitoring, Proceedings of INET'99, Internet Society (1999).

[2] R.Kaizaki, K.Cho, :トラフィックプロファイラ AGURI の設計と実装，日本ソフトウェア科学会，Proceeding of WIT2001-G6-1 (2001).

[3] C.Estan, S.Savage, G.Varghese:Automatically Inferring Patterns of Resource Consumptionin Network Traffic, Proceeding ACM SIGCOMM2003 (2003) ,pp.137-148 (2003).