5U-06　アプリケーショントラヒックアナライザを用いたトラヒック特性評価実験に関する一考察

大岸　智彦　　長谷川　亨　　中尾　康二　　中村　元

KDD 研究所

1. はじめに

インターネットのトラヒック解析については、IP パケット・バイト量などの IP レベルの解析が盛んに行われており、その結果として IP レベルにおいて長期依存性の特徴を持つことが明らかにされている[1]。本特徴は、アプリケーションレベルのプロトコルやユーザの振舞いに起因しているが、アプリケーションレベルの特性に関してはこれまで検討されていない。そこで筆者らは、収集済みのトラヒックトレースを基に、WWW アクセスを対象として、リクエスト長や WWW セション時間など、WWW トラヒックを特徴づけるパラメータに関し、その分布・相関とその時系列における長期依存性の観点から評価を行った。上記のパラメータの抽出に関しては、KDD 研究所で収集したトラヒックトレースを基に、開発済みのアプリケーショントラヒックアナライザ[2]を用いて行った。本稿では評価の詳細について述べる。

2. アプリケーショントラヒックアナライザの機能

実験に用いたアナライザは以下の機能を持つ。

・tcpdump や snoop などにより、予め収集したトラヒックトレースをもとに、アプリケーション毎のパラメータの発生時刻と値を抽出する。

・各パラメータは、トラヒックトレースより正確に抽出できるように定義されている。表 1 に WWW のパラメータの一覧を示す。

・抽出したパラメータをもとに、ヒストグラムや時系列などの統計データを作成する。

表 1　WWW のパラメータ名とその定義

パラメータ名	パラメータ定義
リクエスト長	1 個の HTTP リクエストを構成する TCP の DATA セグメント長の合計
オブジェクト長	1 個の HTTP レスポンスを構成する TCP の DATA セグメント長の合計
レスポンス待ち時間	HTTP リクエストの終了が検出されてから、HTTP レスポンスの先頭が検出されるまでの期間
TCP コネクション持続時間	SYN+ACK セグメントによるコネクション確立完了時刻から、双方向の FIN セグメントによるコネクション終了時刻までの期間
TCP コネクション確立待ち時間	SYN セグメントによるコネクション確立要求時刻から SYN+ACK セグメントによるコネクション確立完了時刻までの期間
リクエスト数	1 本の TCP コネクション内で転送されたリクエストの個数
WWW セション時間	1 つの WWW ページのアクセスに関連すると考えられる複数の TCP コネクションにおいて、最初に確立された TCP コネクションの開始時刻から最後に終了した TCP コネクションの終了時刻までの期間
オブジェクト数	1 つの WWW ページに含まれるオブジェクトの個数
TCP コネクション数	1 回の WWW セションで使用された TCP コネクションの個数

A study on the Experiment of Traffic Characteristic Evaluation Using an Application Traffic Analyzer
Tomohiko Ogishi, Toru Hasegawa, Koji Nakao and Hajime Nakamura
KDD R&D Laboratories

3. 実験方法

3.1. トラヒックトレースの測定環境

トラヒックトレースとしては、KDD 研究所の外向けのセグメント (10Mbps でインターネットに接続) において、約 3 日間収集したものを利用した。本セグメントでは、当研究所からの全ての WWW アクセスのトラヒックが観測可能である。

3.2. 実験手順

(1) アプリケーショントラヒックアナライザにより、3.1.節で得られたトラヒックトレースを基に、WWW のパラメータを抽出する。

(2) パラメータ値の特徴を調査するため、各パラメータにおいて、総数、平均、最大／最小などの基本情報とともに値の分布を解析する。また、同時刻に発生したパラメータ値間の依存性を表す相互相関を解析する。

(3) 各パラメータに対し、定常的にトラヒックが発生した期間 (平日 1 日の 10:00～18:00) を対象として時系列データを作成する。時系列データとは、単位時間毎に、その時間に発生したパラメータ値の合計を算出したデータを表す。さらに、時系列データより、長期依存性の程度を示すハースト値、および、自己相関を算出する。

4. 実験結果

4.1. 基本情報の解析結果

WWW のパラメータの基本情報を表 2 に示す。各パラメータは、その発生頻度に従い、リクエストグループ、TCP コネクショングループ、WWW セショングループに分類した。総数、平均、最大／最小は、全観測期間におけるものを表す。本表には、総数のうち 4.3.節における時系列解析で対象としたサンプル数を付加する。表 2 より以下のことが分かる。

・1 本の TCP コネクションには、平均 1.36 個程度のリクエストが含まれる。

・1 つの WWW ページにアクセスした場合、平均 3.5 本の TCP コネクションが確立される。

表2　WWW トラヒックのパラメータの基本情報

	総数	平均	最大／最小	時系列解析のためのサンプル数
リクエスト長（Byte）	288,938	365	8,150 / 13	124,632
オブジェクト長（Byte）	268,492	28,873	12,136,291 / 18	120,608
レスポンス待ち時間（秒）	268,492	0.278	5.144 / 0	120,609
TCP コネクション持続時間（秒）	211,333	2,721	1,693,544 / 0	74,137
TCP コネクション確立待ち時間（秒）	185,646	1.974	2,998 / 0	74,092
リクエスト数	211,333	1.367	228 / 0	74,137
WWW セション時間（秒）	52,921	3,636	3,393 / 0	22,596
オブジェクト数	49,634	5,409	1,673 / 1	21,604
TCP コネクション数	52,921	3,470	1770 / 1	22,596

（注）網掛けの色の薄い順にリクエストグループ、TCP コネクショングループ、WWW セショングループを表す。

4.2. パラメータ値の分布の解析

図 1 にパラメータ値の分布 (横軸は対数スケール)、図 2 に同一グループ内のパラメータの相互相関を示す。本解析により以下の結果が得られた。

・いずれのパラメータも、heavy-tailed (長いすそを引くよう) に分布している。

・パラメータ毎に最頻値が存在する。例えば、リクエスト長は、300～600 バイト、TCP コネクション持続時間は、3～8 秒の範囲の値を取りやすい。

・全体的に見て、WWW セショングループのパラメータ間のみに相互相関がある。

4.3. パラメータ時系列の解析

パラメータ値を合計した値が有意と考えられる、バイト数および個数を単位としたパラメータを対象として、時系列解析を行った。ハースト値の算出には、最小の単位時間 (⊿t 秒) において十分なデータを含むように、⊿t=60 とし、1⊿t～16⊿t の範囲のデータを用いた。⊿t 数を横軸、相関係数を縦軸とした自己相関関数のグラフは、いずれも⊿t 数=1 の値より下降する曲線を描いた。そこで、自己相関の解析として、⊿t 数=1 での相関係数と相関係数が最初に 0 以下になる⊿t 数をそれぞれ算出した。表 3 にその解析結果を示す。解析結果より、以下のことが導出できる。

・発生頻度別に分けたグループに関係なく、いずれのパラメータも、長期依存性を持つ。heavy-tailed に分布しているパラメータの方が、長期依存性がより大きい。

・いずれのパラメータも、60 分以内において、短期的な相関の傾向を持つ。

4.4. 考察

以上の結果より、以下のことが考察される。

・WWW トラヒックのパラメータの値は、いずれも heavy-tailed に分布していた。この特性が IP レベルのトラヒックにおける長期依存性の性質を形成する要因と考えられる。

・IP レベルのトラヒックのみならず、WWW トラヒックのパラメータに関しても長期依存性の傾向がある。

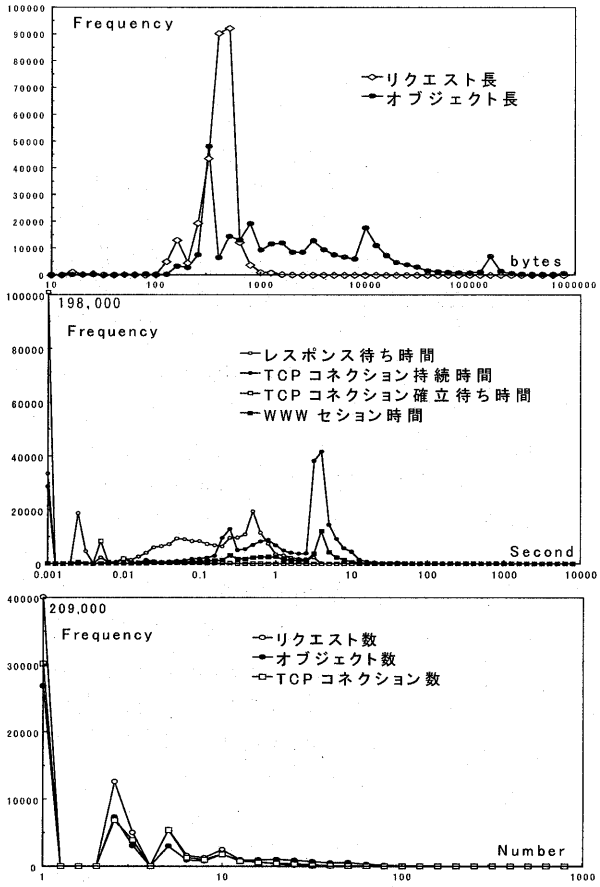

図1　パラメータ値の分布

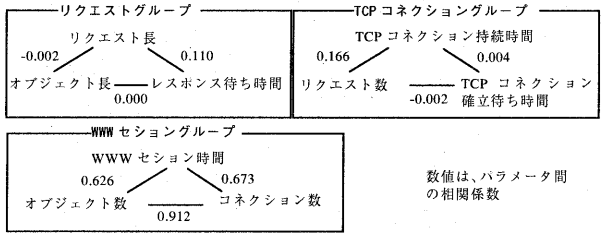

図2　パラメータ間の相互相関

表3　パラメータ時系列の解析結果

	ハースト値	自己相関
リクエスト長（リクエストグループ）	0.796	0.40, 64
オブジェクト長（リクエストグループ）	0.951	0.90, 66
リクエスト数（TCP コネクショングループ）	0.863	0.55, 77
オブジェクト数（WWW セショングループ）	0.827	0.36, 64
TCP コネクション数（WWW セショングループ）	0.888	0.48, 68

5. まとめ

本稿では、アプリケーションレベルのパラメータに基づいたトラヒック特性の評価実験とその結果について述べた。パラメータの抽出には、アプリケーショントラヒックアナライザを用いた。

最後に日頃ご指導頂く KDD 研究所秋葉所長に感謝する。

参考文献

[1] W.E. Leland, et. al, "On the self-similar nature of Ethernet traffic," SIGCOMM '93, 1993.

[2] 大岸他、「アプリケーション毎のトラヒックの統計情報を収集するトラヒックアナライザの設計」、第 61 回情処大会、Oct. 2000.