

アプリケーション毎のトラフィックの統計情報を 収集するトラフィックアナライザの設計[†]

大岸 智彦 長谷川 亨 加藤 聡彦 中村 元

KDD 研究所

1. はじめに

インターネットの急速な発展により、バックボーン回線において IP トラフィックが急増し、帯域の増大が望まれている。これに対し、将来的な帯域の需要を予測するために、トラフィックの変動を長期間測定する試みが行われてきている。しかしながら、従来の IP レベルのトラフィック解析では、ユーザが実際に転送したデータ量や、ネットワークの過負荷のためユーザが転送を断念したデータ量など、必要な帯域の予測に有効な情報を推定することができないという問題があった。帯域の予測を行うには、アプリケーション毎のトラフィック量やその発生間隔などを把握することが重要となる。そこで筆者らは、アプリケーション毎のトラフィックの統計情報を収集するトラフィックアナライザを設計した。本アナライザは、特徴的なパラメータを抽出するとともに、抽出した結果に対し、統計処理を行うことを可能とする。本稿ではその詳細について述べる。

2. 要求仕様

本アナライザは以下の機能を持つこととする。

- ・tcpdump や snoop などを用いて、予め収集したトラフィ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。
- ・抽出するパラメータは、1 回の WWW サーバへのアクセスや 1 通のメール転送など、個々のアプリケーションにおける、まとまったトラフィック (アプリケーションフロー) を対象とする。
- ・抽出したパラメータをもとに、確率分布や時系列の解析を行うための統計データを作成する。
- ・OPNET などのシミュレータへ入力可能な形式で統計データを作成することにより、観測されたトラフィックを再生できるようにする。

3. 詳細設計

3.1. 全体構成

本アナライザの構成を図 1 に示す。本アナライザは、パラメータ抽出モジュールと統計処理モジュールから構成する。パラメータ抽出モジュール

は、TCP コネクションテーブルなどの内部状態を管理しながら、発生時刻と値からなるパラメータリスト、対応するパラメータの組からなるパラメータ相関リストを作成する。統計処理モジュールは、パラメータ/パラメータ相関リストに対し、時系列データへの変換や確率分布への当てはめなどの統計操作を行い、統計データを作成する。

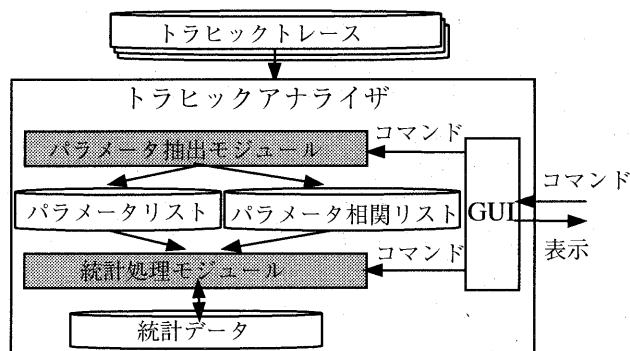


図 1 トラフィックアナライザの構成

3.2. パラメータ抽出モジュール

アプリケーション毎に抽出するパラメータ/パラメータ相関の一覧を表 1 に示す。パラメータ相関としては、Email におけるメール長と添付ファイル長のように、その発生数が等しくなるもののみを抽出する。

表 1 パラメータ/パラメータ相関の一覧

アプリケーション	パラメータ、または、パラメータ相関の系列 (パラメータ、...)
WWW	リクエスト系列 (リクエスト長、オブジェクト長)、セッション系列 (WWW セッション時間、閲覧時間、WWW セッション内の WWW ページ数、WWW セッション内の TCP コネクション数) 等
FTP	セッション系列 (セッション長、セッション間隔)、コネクション長、GET/PUT 比率等
Email	メール系列 (メール長、添付ファイル長)、バースト系列 (バースト数/長)、メール発信間隔等
Net News	記事長、バースト系列 (バースト数/長) 等
H.323 ビデオ会議	有音/無音時間、会議系列 (I/P フレーム長、ビデオ/音声符号化方式、会議持続時間) 等
Real Video/Audio	パケット到着間隔、シーン系列 (シーン持続時間、シーン発生間隔)、セッション持続時間等
IP 電話	有音/無音時間、セッション系列 (セッション持続時間、セッション発生間隔) 等

パラメータ抽出モジュールの設計にあたり、以下の方針を立てた。

(1) トラフィックトレースには、複数のアプリケーション

[†]: 本研究は、INTELSAT からの委託により実施している。

セッションフローが混在すると考えられる。そこで、アプリケーションフロー毎に独立な解析を行うため、TCP コネクションやアプリケーションフローにおける状態と内部変数、さらに、それらの対応づけを管理する^[1]。

(2) FTP では、制御用 TCP コネクション、データ転送用 TCP コネクション、1つの制御用 TCP コネクションで連続的に発生したコマンド/リプライを識別するセッションフローを管理する。

(3) Email では、個々のメールを転送するための TCP コネクション、ある SMTP サーバからの連続的なメール発信を識別するバーストフロー、発信者毎のメール発信を識別する発信者フローを管理する。

(4) H.323 ビデオ会議は、呼制御トラヒックと RTP によるデータトラヒックで構成されるが、パラメータ抽出においては後者のみを対象とする。本アプリケーションでは、1本の映像/音声チャンネルを識別するチャンネルフローと、それらの複合で構成される会議を識別する会議フローを管理する。

(5) Real Video/Audio では、RTSP による SETUP/PLAY 等の制御コマンドと、UDP/TCP によるデータストリームを対応づけて解析を行う。本アプリケーションでは、制御用 TCP コネクション、1回の再生における連続的な転送を識別するセッションフロー、個々のストリーム転送を識別するセッションフローを管理する。

3.3. 統計処理モジュール

統計処理モジュールが出力する統計データとその出力順序を図2に示す。本モジュールは、以下の方針で設計した。

(1) ヒストグラム/PDF/CDF の出力や確率分布への当てはめなどの統計処理は、アプリケーションフロー毎のトラヒック量の分布や閾値以上となる確率などを評価するために行う。

- ・ヒストグラム/PDF/CDF を出力する場合、異常値の除去などのため、値の範囲を指定可能とする。

- ・確率分布への当てはめにおいては、PDF/CDF に最も適合する確率分布と分布関数のパラメータ値を推定する。指数分布、正規分布では、それぞれ、平均、平均および分散の理論値と実測値を一致させることにより、ワイブル分布、パレート分布では、対数を用いた直線回帰を行うことにより、パラメータ値を推定する。また、確率分布への適合度を示すカイ2乗検定の結果を出力する。

(2) 時系列データ、自己相関関数、自己相似性などの統計処理は、アプリケーションフローの発生間隔の変動や、その周期性・長期依存性などを評価するために行う。

- ・時系列データは、ある時間間隔毎に、発生時刻

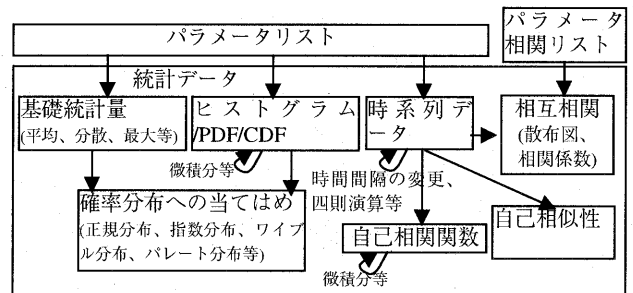
がその時間間隔に含まれるパラメータの値を合計して求める。時間間隔は任意に指定可能とする。

- ・自己相似性の計算においては、時系列データにおいて、時間間隔を変更した場合の分散の変動を出力するとともに、その傾きとなる Hurst パラメータを計算する。

(3) 相互相関の解析は、パラメータ間の依存性を検出するために行う。本解析は、パラメータ相関リストあるいは任意の二つの時系列データをもとに行い、散布図と相関係数を出力する。

(4) 特定の統計データに対しては、四則演算、微積分などの数学的操作を行うことを可能とする。

(5) 各種統計データは、テキスト形式で格納するとともに、グラフィカルに表示できるものとする。



PDF: Probability Density Function (確率密度関数)
CDF: Cumulative Distribution Function (累積分布関数)

図2 統計処理モジュールの構成

3.4. 画面表示例

本アナライザの設計により、トラヒックトレースをもとに、図3に示すような統計処理結果を出力することができる。左図は、WWWセッション内のWWWページ数とWWWセッション内のTCPコネクション数の相互相関を示す。右図は、WWWページ内のTCPコネクション数のPDFをパレート分布に当てはめた結果を示す。

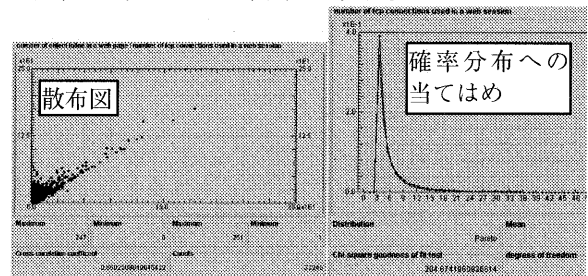


図3 トラヒックアナライザの表示例

4. まとめ

本稿では、アプリケーション毎のトラヒックの統計情報の収集と、その統計処理を行うトラヒックアナライザの設計について述べた。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] 大岸、長谷川、加藤、「WWWトラヒックにおけるアプリケーションレベルのパラメータ抽出方法に関する一検討」、第60回情報処大会、Mar. 2000.