5E-04

情報セキュリティーポリシーを考慮した ネットワークトラフィック統計データ提供システムに関する検討

髙橋秋典† 高橋俊彦† 五十嵐隆治† 上田浩‡ 岩谷幸雄†† 木下哲男†‡

秋田大学大学院工学資源学研究科† 京都大学学術情報メディアセンター‡

東北学院大学工学部†† 東北大学大学院情報科学研究科†‡

1. はじめに

インターネットトラフィックデータをパケットレベルで解析することは、ネットワーク管理・運用やセキュリティー研究において重要な位置付けとなっている。しかし、教育研究機関においてパケットレベルのデータは、情報セキュリティーポリシーの関係で非公開情報と扱われることが多く、研究で解析等に用いる場合でも、守秘義務契約など繁雑な手続きを行わなければならない。また、正式に許可を得た学生に利用させる場合でも、情報漏洩や物理的な盗難について厳重な注意喚起を行わなければならない。さらに、データを用いた研究結果の公開化においても、プライバシー情報保護のため、公開可能情報の限定や統計処理等の加工を行うこと求められる。このように、研究用途であっても、共同研究者である学生に対しては十分な注意が必要とされる。厳密に情報セキュリティーポリシーを考慮した場合、ダンプファイル管理やデータの統計加工処理等は管理者が行うことが想定されるが、管理者が介在することで作業効率が悪くなるという問題点がある。

そこで本研究では、一般ユーザである学生に対して、管理者を介在させることなくパケットデータから統計データへ変換し、容易に処理後のデータを提供できるシステムを開発することを目的とする。試作として、キャンパスネットワークのトラフィックデータを各種統計データに変換して提供するシステムを構築した。

2. 情報セキュリティーポリシーを考慮したシステム機能要件

本研究目的を達成するため、システムに対する機能要件を以下のように設定した。

(1). 一般ユーザはキャプチャされたダンプファイルに直接アクセスできないようにする。
(2). ダンプファイルから一般的なネットワーク統計データのみならず、独自で開発した統計処理によるデータも変換可能とする。
(3). 解析対象とするパケットのフィルタリング設定は一般利用者が実行可能とする。
(4). 利便性を向上のため、OS に依存しない Web アプリケーションとして構築する。
(5). 変換後の統計データは、研究で再解析可能とするため、任意にダウンロード可能とする。

ダンプファイル複製による情報漏洩対策は、機能要件(1)によって満たされ、データの加工処理や研究解析に対する利便性配慮は、機能要件(2)～(5)で満たされると推測される。ここで、パケットレベルのデータファイルに対するプライバシー保護対策としては、tcpdpriv[1] のような IP アドレスやポート番号を他の値にして特定不能とさせるツールもあるが、機関ネットワークに対するトラフィック解析には不都合が生じるため、本研究では統計加工による保護対策を用いた。

3. 提案システム

3.1 システム構成

前章でのシステム機能要件に基づき検討したシステム構成を図 1 に示す。提案システムは、パケットキャプチャとダンプファイルを用いて統計解析を行うアプリケーションサーバと一般ユーザからの要求に対して対応する公開用 Web サーバから構成される。

Reverse proxy による公開用 Web サーバを前段に設置することで、アプリケーションサーバに保存されているダンプファイルに対するセキュリティーを高めている。

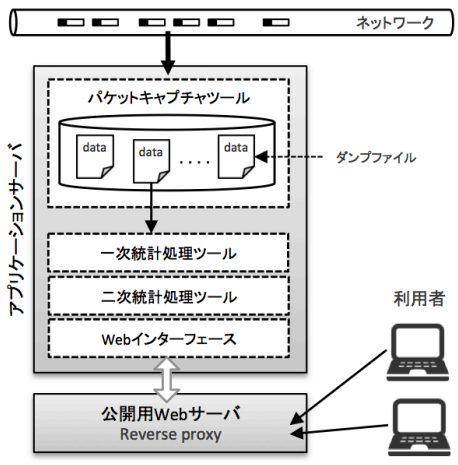

図 1 システム構成

Network Traffic Statistics Data Providing System Considering Information Security Policy
† Akinori Takahashi, Toshihiko Takahashi and Ryuji Igarashi, Akita University
‡ Hiroshi Ueda, Kyoto University
†† Yukio Iwaya, Tohoku Gakuin University
†‡ Tetsuo Kinoshita, Tohoku University

アプリケーションサーバは、パケットキャプチャから統計解析、およびデータ提供処理までを実現するため、パケットキャプチャツール、一次統計処理ツール、二次統計処理ツール、ならびに Web インターフェースにより構成されている。

パケットキャプチャツールは、長期間連続して観測を行うことが出来るように、dumpcap[2]コマンドを用いて 1 時間単位でダンプファイルをローテーションさせながら作業ディレクトリに保存する。次に、日付が変更された時点(00:00:00)で、作業ディレクトリのダンプファイルを管理ディレクトリに移動させる。その後の統計解析処理は管理ディレクトリに保存されたダンプファイルを用いるため、前日までのダンプファイルが解析対象となる。

一次統計処理ツールでは、各種統計解析を行う前処理として、pcap 形式のダンプファイルからパケット数、通信速度といった基本的なネットワーク統計量や、トラフィック量の時間的特徴解析に有用な時系列データへの変換を行う。本研究では、4 章で後述する pcap ライブラリを用いて開発した pcap2ts コマンドを使用した。

二次統計処理ツールおよび Web インターフェースは、Web ベースで容易に統計処理を行うことが出来る R 言語の Shiny server[3]を用いて構築した。R 言語を用いることで、標準的な統計解析処理はもちろん、独自アルゴリズムによる新たな統計解析も容易にプログラム可能となる。また、Shiny server により、一般ユーザは Web ブラウザを用いて容易にシステムにアクセスすることが可能となる。

3.2 システム処理手順

試作したシステムの Web アプリケーションを図 2 に示す。ユーザはインターネットブラウザを利用して、公開用 Web サーバに接続して、ダンプファイル選択のための日時、および時系列生成のための単位時間、フィルタリングオプションを指定する。これらのパラメータは、リバースプロキシを介して、アプリケーションサーバに送られる。アプリケーションサーバではパラメータに基づき、一次、二次統計処理を行い、解析結果を表示する。このとき、結果は R 言語のプロット描画に基づき、グラフとして表示される。さらに、解析結果データは Shiny server のダウンロードハンドラを用いて、ユーザがダウンロードすることができる。

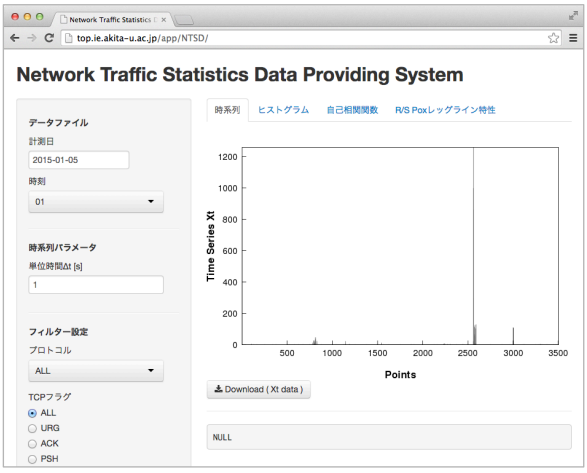

図 2　Web アプリケーション

4. 一次統計処理ツール pcap2ts

ダンプファイルから統計処理を行う標準的なツールとして tshark[4]があるが、時系列データへの変換を行う場合、ダンプファイルの先頭パケットのタイムスタンプが時系列の開始時刻、最終パケットのタイムスタンプが時系列の最終時刻として変換される。しかし、連続した実時間を 1 時間単位で分割保存されたダンプファイルでは、時間帯の正時(%H:00:00)から時系列データの単位時間 Δt を超えて最初のパケットがキャプチャされた場合や、最後のパケットがキャプチャされてから次の時間帯直前までの時間が単位時間 Δt を超えていた場合、パケットの存在しない時間での時系列データに欠損が生じてしまう。

そこで、この問題点を解消するため、ダンプファイル内の先頭パケットのタイムスタンプより当該時間帯の正時刻を求め、その正時刻からの各パケット到着時間を用いて時系列データに変換するツール pcap2ts を開発した。

キャンパスネットワーク、および WIDE Project MAWI ワーキンググループ国際線から収集されたダンプファイル[5]を用いて、pcap2ts の評価を行った。ダンプファイルに対して時系列データ変換処理を行ったところ、tshark ではファイルごとの時系列データサイズに変動が生じたが、pcap2ts では全てのファイルにおいて単位時間 Δt で決定されるデータサイズとなることが確認された。また、変換処理時間は、pcap2ts のほうが大幅に短縮されることが確認された。

5. あとがき

提案システムにより、一般ユーザに対して情報セキュリティーポリシーに配慮しつつ、管理者を介在させることなく容易な操作性により統計データを提供することができた。

謝辞　本研究の一部は東北大学電気通信研究所における共同プロジェクト研究 H25/A15 の助成を受けたものである。

参考文献

[1] WIDE Project 2000 年度研究報告書 : http://www.wide.ad.jp/project/document/reports/pdf2000/part04.pdf
[2] dumpcap: https://www.wireshark.org/docs/man-pages/dumpcap.html
[3] Shiny by RStudio: http://shiny.rstudio.com/
[4] tshark: https://www.wireshark.org/docs/man-pages/tshark.html
[5] MAWI Working Group Traffic Archive: http://mawi.wide.ad.jp/mawi/