E-15

自律システムを用いたトラフィック可視化手法の提案

Traffic Visualization by Autonomous System

畑井 健司†　　川橋 裕‡
Kenji Hatai　　Yutaka Kawahashi

1. はじめに

インターネットは自律システム(以下 AS)と呼ばれる,単一ポリシーにおいて運用されているネットワークの集合体である.それぞれのネットワークには管理者が存在し,管轄内とそれにより他のネットワークに生じた障害に対応している.従来の考え方では,障害が起こらないシステムの構築並びに管理が理想とされてきた. しかし,インターネット自体の急速な拡大と内部ネットワークの肥大化により,障害を起こさない環境整備というのは極めて困難である. この問題に対して、障害を未然に伏せ g のではなく障害は怒るものであると考える「事故前提社会」[1]への対策が経済産業省より提言、推進されている。 事故前提社会ではインシデントの局所化・最小化,並びに早期完治と復旧が可能なシステムの構築が要求される. ネットワーク管理者(以下管理者)は上記に加え,障害の原因究明と再発防止も求められる.

ネットワーク障害の感知や原因究明において重要なのは,管理者が管轄するネットワーク(以後内部ネットワークと呼称)とそれ以外のネットワーク(以後外部ネットワークと呼称)の間で過去に行われた通信の記録(以下通信記録)の把握である. 通信記録とは「誰と誰が,いつ,どの程度の量の通信を行ったか」という情報である. 通信記録の解析により感知可能な障害は数多く存在する. 通信先ホスト数により P2P ファイル共有ソフトウェア利用検知,ウイルス等の感染経路の特定や観戦端末の把握,対外接続線の逼迫による通信障害の発生や,脆弱性により踏み台となった学内端末の特定などが挙げられる.

しかし,通信記録による問題の把握には問題が存在する. ネットワークの大きさに比例して管理者と記録を収集,保持,分析するシステムに対して,負担が大きくなるという点である. 端末数の増加に依る通信料の増加と,端末単位での通信料の増加とが複合し,通信量は増加の一方である.

管理者による障害の感知と原因究明には,現在のものだけではなく過去の通信記録が必要となる場面も多い. 和歌山大学の対外接続線を例にとれば,1 日で最も通信の多い時間帯には毎秒 1 万パケット(10kpps)を超える通信が発生する. この通信記録を用いて管理者がネットワークを監視することは非常に困難である. よって,管理者に対して通信記録のうち,障害の感知や原因究明に役立つ部分を提示することが必要となる.

本研究では,トラフィックを「質」と「量」の 2 つの観点から分析し,通信記録の活用を行う. 本論文は,トラフィックの質的情報を,内部ネットワークにある端末が外部ネットワークにある端末と通信するときの,個別の通信量,時刻,相手先などの情報として定義する. さらに,トラフィックの量的情報を,内部ネットワークと外部ネットワークとの間に発生する総トラフィック量として定義する.

トラフィックを「質」の観点から分析すると,通信っ記録が得られ,先の述べた通りウイルス感染や P2P ソフトウェア利用端末の特定などが可能となる. 一方,トラフィックを「量」の観点から分析すると,時間帯/曜日/季節による増減などの,トラフィック量の全体像を把握することが可能となる. 正常なトラフィック量の増減を把握することは,トラフィックの異常状態を感知することに繋がる. 量的情報から感知した異常を,トラフィックの質的情報を用いて分析することで,障害の原因究明を行うことが可能である.

本研究では,トラフィックの量的情報を,質的情報を用いて分析した結果と,質的情報の記録を管理者に提示し,トラフィックの正常/異常状態の把握,異常の原因究明を支援するシステムを提案する. 具体的には,内部ネットワークから外部ネットワーへの通信(以下 outgoing),外部ネットワークから内部ネットワークへの通信(以下 incoming)を通信記録として取得し,処理を加えて通信履歴として保持,通信履歴を用いてトラフィックの量的情報を分割し,通信履歴によって異常感知後の原因究明を支援するシステムを構築する.

2．技術概要

本章では,ネットワーク運用管理における本研究の位置づけを説明したのち,本論文において用いられる用語の説明をする. 加えて,既存技術・研究について述べる.

2.1　ネットワーク運用管理における本研究の位置づけ

ネットワーク運用管理[2]とは,大正のシステムを効率よく,円滑かつ安全に利用できるようにする業務である. 業務内容は,構成管理,障害管理,性能管理,設備管理,セキュリティ管理の 5 つの管理項目が挙げられる.

本研究では,障害管理,性能管理を対象とする. 障害管理とは,ネットワークで発生する障害の検出や対応,記録などを行う管理業務である.

本研究は,「トラフィック」を従来にない形で可視化することと,活用可能な形で通信履歴を保持することにより,障害管理,性能管理を補助する.

2.2　用語の解説

2.2.1 自律システム

自律システム(Autonomous System)(以下 AS)とは,インターネットに繋がっている単一のルーティングポリシーによって運営されたネットワークの事である. 各 AS は固有の番号を持ち,本学の AS 番号は「24248」である

2.2.2 Hyper Giants

本論文では,Google.Akamai,Youtube,[3]などの,インターネットの総通信量に占める割合が上位の組織の内一定数の事をさす. これら上位組織の占める割合は,年々増加し続けている.

2.2.3 通信記録

インターネットにおける最小単位の管理情報. 「誰と誰が,いつ,どのくらいの情報量をやりとりしたか」という情報の記録とする

†和歌山大学, Wakayama University

‡ 和歌山大学 システム情報学センター, Center for Information Science, Wakayama University

2.2.4 トラフィックの量的情報

トラフィックの量的情報とは,管理者が管轄する AS と他の AS との間で,どの程度の量の通信がやりとりサれているかという情報である．本論文では外部ネットワークとの間の通信記録におけるデータ転送量の総計がトラフィックの量的情報にあたる．通信時間帯ごとにトラフィックの量的情報を関しすることによって,どのような時間帯にトラフィックが増加するかといった,ネットワークトラフィックの全体像を文政することが可能になる．さらに中・長期的にトラフィックデータを蓄積,分析することにより,トラフィックの増減からネットワークの異常を発見することが可能となる．

2.2.5 通信先 AS

本論文における通信先 AS とは,内部ネットワークにある端末の通信相手がいる AS をさす．

2.2.6 フロー

本論文では，送信元 IP アドレス・宛先 IP アドレス・送信元ポート番号・宛先ポート番号が全て一致する一連のパケット群をフローと定義する．

この定義を使うことにより，通信記録をパケット単位ではなく，コネクション単位で扱うことが可能になる．

2.2.7 通信履歴

通信履歴とは，通信記録から，管理ネットワークにある端末の IP アドレス・通信先 AS・送信元ポート番号・宛先ポート番号・データ転送量・パケット数・通信時間帯をそれぞれフロー単位で記録したものである．

個人のプライバシー保護の観点を踏まえ，本研究において保存する通信履歴は，TCP・UDP ならびに IP 層におけるヘッダに存在する情報のみである．

2.3 既存研究・技術

2.3.1 MRTG(The Multi Router Traffic Grapher)[4]

エージェントから取得したデータを加工してグラフとして可視化するプログラムである．MRTG では，監視対象の機器に対して SNMP リクエストを送信し，グラフを出力する．監視対象のシステムから収集したすべてのデータは過去 2 年間分保持され，過去 1 日間・1 週間・1 ヶ月間・1 年間のトラフィックのグラフ生成に利用する．出力されるグラフの例として，過去 1 日間のトラフィック量のグラフを図 1.1，過去 1 週間のトラフィック量のグラフを図 1.2 に示す．グラフ化できるデータとしては，SNMP で取得可能なトラフィック量・CPU Load Average・Disk 使用率・メモリ空き容量などがあげられる．MRTG では HTML 形式のページを生成するため，HTTP デーモンが動作しているサーバで利用することによって，Web ブラウザ経由で閲覧可能である．

MRTG を用いることにより，総通信量の増減を視覚的に把握することが可能である．加えて週，月，年単位のグラフにより，過去の増減の把握，現在の増減との比較をすることも可能である．

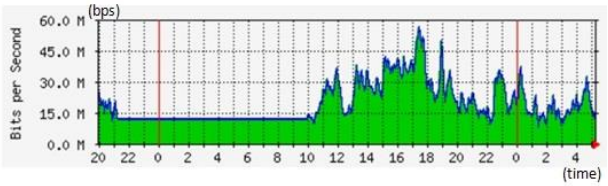

図 1.1 (過去 1 日間のトラフィック量のグラフ)

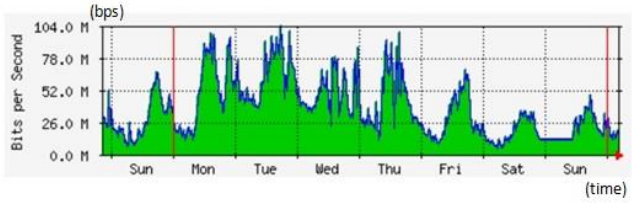

図 1.2(過去一週間のトラフィック量のグラフ)

2.3.2 ネットワークトラフィックの可視化による通信データの監視および制御に関する研究

長坂らの研究[5]では，ネットワークに接続された複数端末の通信を 1 つの管理用端末が収集し，各監視対象端末の情報を可視化して監視するシステムを構築した．この研究では，パケットキャプチャを行い通信データの監視をする．キャプチャしたパケット内のヘッダの種類により，パケットの種類を判断する．ヘッダから取得される情報は，送信元 IP アドレス・宛先 IP アドレス・送信元ポート番号・宛先ポート番号・プロトコルの種類・データ転送量・パケット数・ウィンドウサイズである．

これらの情報を用いて，監視対象端末ごとの 1 秒間あたりに届いたパケット数および，データ転送量(bps)をパケットの種類別にグラフとして可視化する．図 2.1 における横軸は監視対象端末の IP アドレスである．このグラフによって，監視対象端末がどのようなデータをどれくらい送受信しているかを確認できる．監視対象端末ごとのデータ転送量の比較も可能である．さらに各監視用端末におけるアプリケーションごとのデータ転送量を表すグラフ図 2.2 も作成する．このグラフでは通信相手 IP アドレス・プロトコルの種類・ポート番号などの項目別に，データ転送量およびパケット数を表示する．このグラフでは，端末ごとのさらに詳細な監視が可能となる．

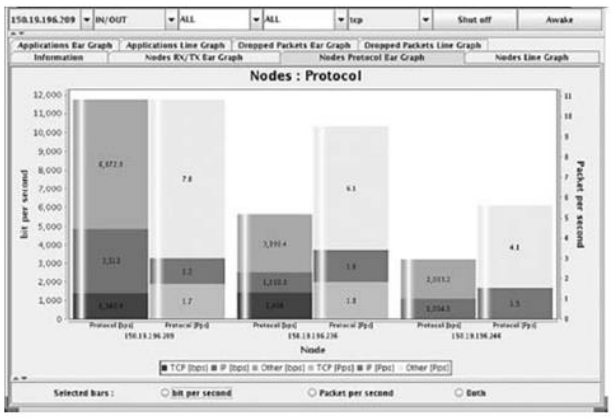

図 2.1 監視対象端末毎のデータ転送量のグラフ

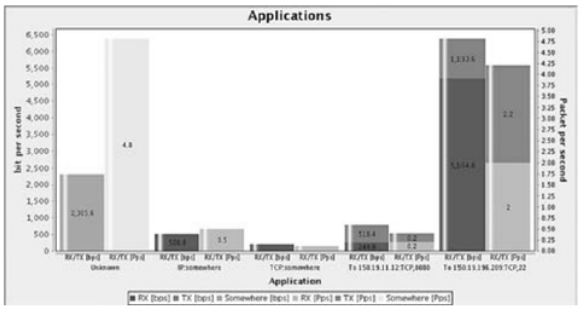

図 2.2 監視対象端末におけるアプリケーション毎のデータ転送量のグラフ

2.3.2 既存研究の問題点

MRTG はトラフィックの量的情報のみを監視している．個別端末，接続先，使用ポート情報等のトラフィックの監視に必要な詳細情報を見ることは出来ない．\\

長坂らの研究は，端末単位でのリアルタイム監視である．障害が発生した時点で，管理者が監視していなければ，それ以降の障害の検知および原因究明は困難となる．
加えて，監視対象端末の数に比例してグラフが大きくなる為，ネットワークの規模に比例して使いやすさが低下する．

3. 研究目的

本研究では，トラフィックの通信先 AS を用いた分類手法による可視化を提案する．
TCP・UDP ならびに IP 層におけるヘッダの情報のみを使うことにより，ユーザのプライバシなどの問題を回避する．
加えて，トラフィックの正常な状態を把握しやすい形で管理者に提示することにより，障害発生時の感知補助，
通信履歴を利用した障害の原因究明の補助が可能になると考える．

次章でその手法について述べ，実装したシステム構成，家庭規模の環境での運用実験を用いて，既存技術・研究との比較評価をおこなう．

4. 提案手法

本研究における AS の分類方法であるが， BGP ルータの経路情報を取得することが出来た為，それを利用することにより分類の補助とする．

4.1 AS の分類

4.1.1 HyperGiants と日本における通信量上位 AS 群

全世界的にインターネット・トラフィックのかなりの部分を占める，HyperGiants と呼ばれる AS 群．
Alexa[6]を参考に抽出した日本での通信量上位を占める AS 群．

以上の 2 つよりなるこのカテゴリーは，通信量，変動量ともに他の 3 つを大きく上回る．以下分類 1 とする．

4.1.2 SINET から経路広告される AS 群

学術情報ネットワーク（SINET4）と呼ばれる，全国の大学や研究機関，並びに多くの海外ネットワークと接続しているネットワークがある．
本学は SINET4 と接続している為，BGP の経路情報に SINET 経由の経路が広告される．
この経路の先の AS は，国内の学術・研究機関である．
以下分類 2 とする．

4.1.3 ブラックリスト AS 群

一般に公開されているセキュリティ関係の情報[7]より手動で拾い上げたもの．
海外オンラインストレージからダウンロードするソフト[8]が対応している対象等をさす．

ウィルスが通信をする相手を内在する AS 群や，違法アップロードの問題が大きいオンラインストレージが存在する AS 群である．以下分類 3 とする．

4.1.4 それ以外の AS 群

それ以外のインターネットに存在する AS のうち、前述 3 つの AS を除いたもの．
大多数のインターネット/コンテンツ ・サービス・プロバイダーなどはここに属する.以下分類 4 とする．

4.2 管理者への提示

上記の分類を行ったトラフィックの量的情報を，折れ線グラフとして可視化する．

提案システムを利用して，管理者が障害を発見し原因を究明する流れは以下の通りである．

1. グラフによるネットワークトラフィックの監視
2. 異常トラフィックの発見
3. 障害の原因と思われる端末の絞り込み
4. IP アドレス検索で障害の原因と思われる各端末の詳細情報を取得
5. 原因端末の特定および，原因究明

5. 提案システム

提案するシステムは，データベース部・データ取得部の 2 つから構成される．提案システムのシステム構成および機能を図 4 に示す．

提案システムのデータベース部では，通信履歴の保存をおこなう．ここで保存される通信履歴は 1 日分と 1 週間分の情報である．データ取得部では，データベース部において保存された通信履歴を，1 日分と 1 週間分のトラフィックの量的情報に加工し取得する．さらに，データ取得部で取得したトラフィックの量的情報から，監視ネットワークをグラフとして可視化する．作成されたシステムは，Web ブラウザで閲覧可能とした．

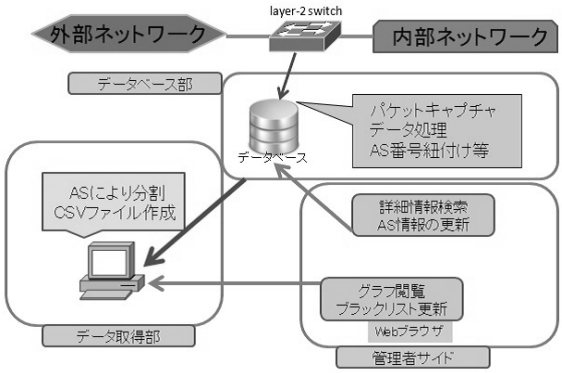

図.4 提案システム

5.1 データベース部

データベース部では，通信のパケット情報を通信履歴として保存する．内部ネットワークと外部ネットワークの境界部分にあるスイッチからポートミラーリングによって取得した情報を 5 分ごとにデータベースに格納する．提案システムでは，incoming および outgoing について，それぞれキャプチャをおこなっている．

データベース部において，作成される通信履歴は以下の 8 種類である．1 日分の通信履歴は，過去 1 日の通信履歴を 5 分間隔で格納する．1 週間分の通信履歴では，過去 1 週間の通信履歴を 30 分間隔で格納する．

以下の 8 種がデータベースに保存される情報である．
・1 日分の TCP/UDP 通信（incoming）
・1 日分の TCP/UDP 通信（outgoing）
・1 週間分の TCP/UDP 通信（incoming）
・1 週間分の TCP/UDP 通信（outgoing）

5.2 データ取得部

データ取得部では，データベース部でデータベースに格納された通信履歴から，過去 1 日分と過去 1 週間分のトラフィックの量的情報を取得し，CSV ファイルとして保存する．

保存される情報は，通信時間帯および，その通信時間帯における incoming のデータ転送量の総計，outgoing のデータ転送量の総計である．ここで取得される通信時間帯は，パケットキャプチャをおこなった時間と同義である．

データ取得部で作成される CSV ファイルの種類は以下の 4 種である．
・1 日分の TCP/UDP 通信におけるデータ転送量
・1 週間分の TCP/UDP 通信におけるデータ転送

5.3 データ表示部

5.2 で作られた CSV ファイルを表示する．

Web ブラウザ上でグラフを表示するにあたり，インタープリタ型のプログラミング言語である JavaScript を使用した．主要な Web ブラウザのほとんどにエンジンが搭載されているという点から選択した．グラフの作成には，オープンソースのウェブウィジェットである，SIMILE Widgets の Timeplot[8]を使用した．Timeplot は CSV 形式のテキストファイルに対応しており，指定したテキストファイルを自動で読み込みグラフを作成することができる．1 つのグラフ内に複数のグラフを表示することも可能である．

本システムでは，データ取得部において作成される CSV ファイルを用いて，以下のグラフを作成し，それぞれに incoming と outgoing を表示させる．

1 日分の TCP/UDP 通信における 5 分平均の
データ転送量[kbps]
1 週間分の TCP/UDP 通信における 30 分平均の
データ転送量[kbps].

図 3.2 はある一日の TCP 通信のグラフである

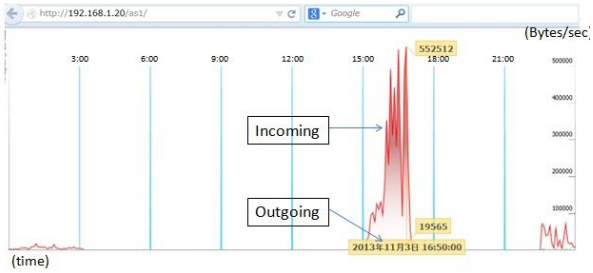

図 5 1 日分の TCP 通信における 5 分平均(分類 1)

6. 運用実験

6.1 実験環境

提案システムを，学外特定環境に導入する運用実験を実施した．環境は図 6.1 のようになっている．本実験では，リピータハブを用いる事によりパケットのキャプチャをおこなった．運用実験は，2013 年 11 月 3 日の 0 時 0 分から 2013 年 11 月 9 日の 23 時 59 分までの 1 週間実施した．意図的に Youtube 等を利用することにより，ピーク時で 5Mbps 以上の通信を発生させている．

6.2 検出結果および検証

本実験では，2013 年 11 月 3 日に分類 3 に当たる通信が検出された．

結果および検証を，次項で述べる．

6.3 ブラックリスト反応の検証結果

発見されたブラックリストの反応は図 6.1 に見ることが出来る．

グラフによって示された時間に該当する通信履歴を精査した．結果はブラックリストに乗せた AS に存在する広告を，ブラウザで表示した，ブラウザの履歴を用いて確認を取った．

この通信はこちらの想定した障害の原因ではなく，正規の通信であることが判明した．該当する AS との間の通信ログを MySQL より SQL 文で取得したものが図 6.2 である．

11 月 3 日の反応以外ブラックリストに乗せた AS に対する通信は存在しなかった．

図 6.1 TCP 通信における 5 分平均の分類 3 の転送量

端末IP(Int表記)	通信先AS番号	送信元ポート番号	通信量(byte)
3232235781	[redacted]	80	20689
3232235781	[redacted]	80	18905
3232235781	[redacted]	80	14271

図 6.2 SQL 文による出力結果

7. 評価・考察

7.1 MRTG との比較評価

提案システムは，情報を保持する期間は劣るが，特定の通信先を強調する事により視覚的な把握が容易になっている．加えて，通信履歴を活用することによって MRTG の問題点を克服した．

前章の運用実験においても，異常の可能性があるトラフィックを把握し，端末を特定，原因を把握することが可能であった．

以上のことから，提案システムは MRTG に比べ監視できる期間は短いが，障害の感知，原因究明において優れているといえる．

7.2 長坂らの研究との比較評価

提案システムは，長坂らの研究に比べ端末単位に対するリアルタイム監視という点では劣るが，過去の事象に対する管理者の感知というと端末数に依存するという問題点を克服した．

以上の点により，提案システムは長坂らの研究に比べ，端末単位の詳細情報という点では及ばないが，障害の感知，原因究明において優れているといえる．

7.3 今後の課題

本節では，本研究の今後の課題について述べる．

7.3.1 ブラックリストの運用基準

現在は，ブラックリストに対する通信全てに対してグラフ表示を行っているが，実験の結果正当な運用であってもブラックリストに対する通信が発生する事が分かった．

この問題に対しては，時間と通信量の 2 つの要素に閾値を設ける事により解決できると考える．

7.3.2 中長期に渡る通信履歴のよりよい保存法

現在の構成では，通信履歴は多大な容量を必要とするため，長期に渡って情報を蓄積するのが難しいという欠点がある．この部分に関しては，必要なデータの絞り込みと長期的な可用性を両立し得るデータの選別が不可欠である．

今後さらに運用実験を通じて最適化していきたいと考える．

7.3.3 実環境での長期運用

現時点では学外特定環境における実験であるが，これを本学の対外線において運用実験をおこない中長期的なデータの蓄積をおこないたい．

4.1.1 HyperGiants と日本における通信量上位 AS 群

参考文献

[1] 次期情報セキュリティ基本計画に向けた第 1 次提言
2008 年 6 月 19 日，情報セキュリティ政策会議 基本計画検討委員会
http://www.nisc.go.jp/active/kihon/pdf/jiki1teigen.pdf

[2] 秋山浩一，“@IT - アットマーク・アイティ ネットワーク運用管理入門”，2003 年 9 月
http://www.atmarkit.co.jp/ait/articles/0311/14/news001.html

[3] あきみち，“インターネットの形を変えて行く Google,Facebook,Akamai...”，あきみち，2009 年 10 月
http://www.geekpage.jp/blog/?id=2009/10/20/1

[4] “MRTG:The Multi Router Traffic Grapher”
http://www.mrtg.jp/doc/

[5] 長坂康史，福田宏見 “ネットワークトラフィックの可視化による通信データの監視および制御に関する研究”
広島工業大学紀要研究編 第 44 巻（2010），pp.223-227

[6] “Alexa The Web Information Company”，参照 2014 年 1 月
http://www.alexa.com/topsites/countries;0/JP

[7] “DNS-BH Malware Domain Blocklist”参照 2014 年 1 月
http://www.malwaredomains.com/wordpress/?page_id=66
小島肇，“セキュリティホール memo - 各種 OS のセキュリティホールの備忘録”，龍谷大学理工学部
http://www.st.ryukoku.ac.jp/~kjm/security/memo/
Jeff Miller,
“Hacker News 100”，参照 2014 年 1 月
http://feeds.feedburner.com/newsyc100

[8] “RapGet”，参照 2014 年 1 月
http://www.rapget.com/en/index.html