社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2007－ＱＡＩ－25（5）
2007／10／25

統計情報を利用したトラフィックバリエーションの見積もりに関する研究

原田　義明†　　岡村　耕二††

† 九州大学 システム情報科学府, 〒812-8581 福岡県福岡市西区元岡 744
†† 九州大学 情報基盤研究開発センター, 〒812-8581 福岡県福岡市東区箱崎 6-10-1
E-mail: †harada@ale.csce.kyushu-u.ac.jp, ††oka@ec.kyushu-u.ac.jp

あらまし　これまでインターネットの挙動を把握するための研究が多くなされてきたが，統計情報を用いた解析ではリアルタイムな異常検知ができないという問題点がある．しかし，相関関係などの解析により通信傾向や変化の前兆を把握することができれば，将来のデータを予測することができる．本論文では，フローデータを統計情報として用いて，変化が起こる際のトラフィックパターンを解析することでインターネットの挙動の把握を行っている．フローデータ解析は，ポート番号や利用プロトコル番号，通信対象 AS などに細分化して行い，より詳細な解析を行う．異常な通信の例として停電を挙げ，停電時のフローデータの挙動の解析を行っている．
キーワード　フローデータ，AS，トラフィック解析，不正検知

A Study on Prediction of Internet Traffic Variation using Statistics Information

Yoshiaki HARADA† and Koji OKAMURA††

† Graduate School of Information Science and Electrical Engineering(ISEE),
Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 092-802-3600 Japan
†† Research Institute for Information Technology,
Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
E-mail: †harada@ale.csce.kyushu-u.ac.jp, ††oka@ec.kyushu-u.ac.jp

Abstract　This paper analyzed the statistical information to assess the traffic behaivior in Internet. There is a problem with statistics analysis that it is impossible to anomaly detection in real time, but we can forecast a change in Internet traffic by the detailed analysis. We used the collected Flow Data stored in Hard Disk in Kyushu Univ. as statistics, and analyze the traffic behaivior by dividing Flow Data by port number, AS number, and country. In this paper, we show the tendency of Internet traffic in nomal traffic and some outage.
Key words　Flow Data, AS, Flow Analysis, Anomaly Detection

1.　まえがき

近年，急速にインターネットの普及が進み，莫大な数のホストがインターネットに接続されるようになった．安定したインターネット環境を提供するためには，インターネット上の通信分布やトラフィックの変化を把握し，ネットワークシステムを管理する必要がある．インターネットトラフィックは周期性や長期依存性などを有することが証明されており [1]，インターネット通信の遅延についての解析やトラフィック変化の特徴付けにおける解析など，より詳しくトラフィックを把握するための研究が数多く行われている [2]～[4]．しかし，ネットワークトラフィックは基本的に変動の大きいデータであるのに加えて，停電や不正なアクセスなど，様々な要因により変化する．その上，ネットワークの多様化とともに通信地域や通信サービスも複雑化してきているため，トラフィック変化を把握することは困難なものとなっている．そこで，従来のマクロなトラフィック解析手法では検知できなかった変化の検出を行うために、AS あるいはポートを指定したトラフィックに着目した，ミクロなトラフィックバリエーションの解析を行う．ここで，トラフィックバリエーションとは，ネットワークトラフィックの日常的な変化や変動の傾向，障害に伴うトラフィック変化のパターンなどを示す．従来のトラフィック解析手法は，一地点で通信フロー数やパケット数をマクロに解析するものが中心となっているが，国や地域ごとにインターネットトラフィックの傾向は異なって

おり，それぞれの変化が幾重にも重なり合いトラフィックの傾向が複雑化することで，マクロなトラフィックの把握が難しくなっている．例えば，不正な通信によってある地域から急激に大量な通信フローが流れたとしても，何らかの原因で通信フロー数が減少した地域があれば，不正な通信が他のフローに埋もれてしまい異常の検知は困難なものとなる．しかし，ポート番号や通信 AS に細分化してトラフィック変化の解析を行うと，ある程度の傾向を持っているため，将来的なトラフィック予測も可能となる．本論文では，統計情報を利用して解析を行うことで，ポート番号や通信 AS などに細分化したトラフィックバリエーションを把握し，将来のインターネットトラフィックを予測する手法を提案する．統計情報を用いた解析ではリアルタイムな異常検知ができないという問題点があるが，フローを細分化して解析を行い，通信傾向や変化の前兆などといったトラフィックバリエーションを把握することができれば，将来のデータを予測することができる．複雑に重なったトラフィックを地域情報などに分割することで簡単化し，個々のトラフィック変化を把握できれば，インターネットトラフィック全体を予測・把握することもできるようになる．また，不正アクセスには，同一の AS や IP アドレスから送信されるものや同一ポート番号を利用したものも多いため，通信地域情報や利用ポート番号に細分化して解析することで，不正アクセスや異常な通信の検出も容易なものとなる．本論文ではフローデータを統計情報として利用し，ポート番号や利用プロトコル番号，通信対象 AS や国情報などに細分化して解析し，従来の解析手法では発見できなかったトラフィック変化の検知と把握を行っていく．

本論文では，第 2 章においてインターネットプロトコルやインターネット経路制御などの，本論文を理解するにあたっての背景知識についての説明を述べる．第 3 章では，本研究の解析手法を述べ，どのように解析を進めていくのかを説明する．そして，第 4 章では，実際に得られた結果を示し，障害時のインターネット挙動の考察を行う．第 5 章で本研究のまとめと今後の課題について述べる．

2. 背　景

2.1 インターネットプロトコル

ネットワーク機器間で通信を行うために，様々な規約（プロトコル）が定められている．ネットワークを介してデータ交換を行うために，通信したいデータの前後に通信の内容や目的地の IP アドレス等を示すデータ部分（ヘッダ）を付加することで，目的地まで正しくデータが送信されている．現在，インターネット上で通信を行うために TCP/IP というプロトコルが用いられている．

TCP/IP においては，通信相手を特定するために利用される IP アドレスの他に，ポート番号が利用されている．IP アドレスから通信相手のホストを特定し，ポート番号からそのホストの利用しているプログラムを特定することができる．ポート番号は通信の種類毎にそれぞれ 0 から 65535（16bit）で割り当てられており，一般的に用いられるポート番号として 0 番から 1023 番（Well Known Port）が，登録済みポートとして 1024 番から 49151 番（Registered Port）が，自由に使用できるポートとして 49152 番から 65535 番（Dynamic and/or Private Ports）が割り振られている．本研究では，通信の用途を限定した解析を行う際にこのポート番号を用いる．よく利用されるプロトコルは，あらかじめ指定のポート番号が割り振られており，このポート番号を参照することで，現在利用されているサービスを識別することができる [5]．表 1 に，一般的に利用されるポート番号とプロトコル番号の対応表を示す．

ポート番号	プロトコル名	TCP/UDP	利用目的
20	ftp	TCP	ftp ファイル転送
22	ssh	TCP	リモート接続
25	smtp	TCP	メール送受信 (SMTP)
53	domain	UDP	DNS
80	http	TCP	WWW
443	http	TCP	TLS/SSL を利用

表 1　代表的なポート番号とプロトコル
Table 1　The List of Typical Port Number

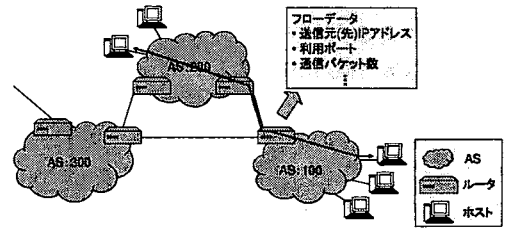

図 1　インターネットルーティング
Fig. 1　Internet Routing

2.2 インターネットルーティング

現在，インターネットには莫大な数のホストが接続されているが，これを管理するために AS（Autonomous System）という単位が利用されている．AS とは，共通のポリシーや同じ管理下におかれているルータやネットワークの集合のことである．インターネットでは，AS 単位に分割して管理することでネットワークの管理を容易にしている．表 1 が示すように，一つの AS には複数のホストが接続されており，AS 間と AS 内とでは別々に経路の制御がなされている．AS には，それぞれの AS を一意に識別するために AS 番号という 16bit の識別子が割り当てられている．AS や IP アドレスは ARIN [7] や RIPE-NCC [8]，APNIC [9] など，RIR（Region Internet Registry）と呼ばれる組織が管理しており，これらの組織がそれぞれ持っている WHOIS データベースに，ユニークな AS 番号と共に所有組織名や国情報等が保持，運用されている．WHOIS データベースを参照し，IP アドレスに対応するプレフィックスを検索することで，参照する IP アドレスがどの AS に所属しているかを識別することができる．

2.3 フローデータ

今回の研究では，解析データとしてフローデータを扱う．フローとは，送信元から宛先までの一連のパケットのことで，通信プロトコルや通信用途でまとめられている．フローデータ

srcIP	dstIP	router_sc	prot	srcPort	dstPort	octets	packets
133.5.[redacted] /24	220.58.144.38/18	0.0.0.6	6	110	36858	40	1
133.5.[redacted] /24	220.58.144.38/18	0.0.0.0	6	110	35858	114	2
133.5.[redacted] /24	220.190.223.148/18	0.0.0.0	6	110	60976	52	1
133.5.[redacted] /24	220.190.223.148/18	0.0.0.0	6	110	40376	76	1
211.168.234.201/18	133.5.[redacted]/24	0.0.0.0	6	2431	4882	227	2
210.166.237.201/18	133.5.[redacted]/24	0.0.0.0	6	80	3568	46	10
210.166.237.201/18	133.5.[redacted]/24	0.0.0.0	6	80	36187	3000	2
210.166.237.201/18	133.5.[redacted]/24	0.0.0.0	6	80	36187	675	2
133.5.[redacted]/24	211.14.34.8/18	0.0.0.0	6	22	60151	2808	2
133.5.[redacted]/24	211.14.34.8/18	0.0.0.0	6	22	60151	884	1

図 2　フローデータの一例

Fig. 2　Example of Flow Data

障害の種類	停電区域	期間
計画停電	伊都地区	03 月 17 日 08:00〜09:30
		03 月 24 日 05:00〜11:00
突発停電	伊都地区	03 月 24 日 14:48〜15:18
瞬間停電	箱崎地区	06 月 01 日 09:25

表 2　障害情報

Table 2　Outage Information

とはフローの集合を示し，インターネット上を流れる通信の経路選択を行う機器，ルータによって収集することができる．フローデータには，送信元と宛先の IP アドレスや通信プロトコル，ポート番号や，パケット数などが含まれる．このデータを解析することによって，インターネット上の通信傾向を把握することができる．フローデータの一例を図 2 に示す．

3.　解析手法

本研究では，フローデータと AS 情報を収集し，これらのデータからインターネットトラフィックの通信傾向の把握を行う．本章では，フローデータと AS，国情報の収集方法について述べ，具体的な解析手法や解析期間についての説明を行う．

3.1　フローデータと AS 情報の収集

今回の解析では AS2507（九州大学）のルータから抽出されたフローデータを利用している．九州大学から外部につながる通信はすべてこのルータを通るようになっている．フローデータに関しては，蓄積されているフローデータを使用する．また，全てのフローを収集するとルータに過負荷がかかってしまい，本来のルーティング操作に悪影響が出てしまうため，サンプリングされたデータを使用する．サンプリングレートは 10 分の 1 である．フローデータには，送信元アドレス，送信先アドレス，プロトコル番号，使用ポート番号やパケット数などが収集できる．

また，収集したフローデータの IP アドレスから AS 番号や国情報を割り当てるために，各種 RIR 機関から IP プレフィックスと AS 番号，国情報や地域情報の対応表をダウンロードし，世界中に割り当てられている全 IP アドレスと AS 番号との対応表を作成した．この IP アドレスと AS 番号，国や地域情報との対応表を用いることで，フローデータ中の IP アドレスを，地域ごとの情報に細分化し解析を行っている．

3.2　フローデータの解析手法

収集したフローデータを AS 番号，国や地域情報，ポート番号等に細分化して解析することで，インターネットトラフィックの傾向把握を行っている．AS 番号や地域情報に細分化して解析を行うことで，複雑なトラフィック変化を把握することができる．例えば，全体として通信フロー数が年々増加していたとしても，国や地域別に通信フロー数の変化を見ていくと，通信フロー数が年々増加しているや減少している国，変化の無い国も存在する．国や地域情報に細分化することで単純化されたフローの解析結果を把握し組み合わせることで，インターネットトラフィック全体の複雑な変化も把握することができる．また，不正アクセスの行われる国や地域はある程度偏っており，国ごとに流行しているウィルスや不正アクセスも違っている．ある地域から異常な通信が行われたとしても，全フローの解析では検出すべき変化が他の地域の通信の中に埋もれてしまい，異常トラフィックの検知が困難になる．それぞれの国や地域，AS 番号を抽出して解析を行うことで，他国や地域の影響を除去でき，精度の高い不正検知を行うことができる．また，通信傾向の把握にはポート番号も利用して解析を行う．通信用途を限定した解析を行うことができ，利用用途の不自然な通信が集中している場合などは不正なアクセスの可能性が高い．ポート番号の分布を見ることで不自然なポート利用，また一般的に利用されるポート番号の通信についても，通信傾向の変化を見ることで不正な通信の検知を行うことができる．

次にフローデータの解析を行うに当たって作成したプログラムについての説明を行う．本研究では AS や国単位での解析も行うため，フローデータ中のそれぞれの IP アドレスをプレフィックスと国情報の対応表と比較し，AS 番号を割り振る必要がある．解析を行うたびに AS を検索していては解析に必要以上に時間がかかってしまうため，まずフローデータの IP アドレスに AS を割り振り，その後必要なデータを抽出・保存するプログラムを作成した．フローデータの解析結果データベースは階層的に保存しており，トップダウン式に情報を参照していくことで，障害検知のコストを減らすことが出来る．特定のポート番号のフローデータを抽出する際には，送信元ポート番号は改ざんすることができるため，送信先のポート番号を参照している．また，異常な通信が発見された場合には，特定の AS 番号に対しての通信のみを抽出した解析も行う．

3.3　フローデータの解析期間と障害情報

九州大学のフローデータを統計情報として用いて，インターネットトラフィックの通信傾向の解析を行う．日常的な通信の傾向を解析すると共に，障害に伴うトラフィック変化の例として停電を選択する．工事や点検のため，事前に停電が行われることをアナウンスされている停電（計画停電），突発的に数十分の間配電が停まってしまった停電（突発停電），突発的に起こったが，一瞬で回復した停電（瞬間停電）の 3 種類の停電について解析を行う．また，障害に伴ったトラフィック変化の解析を行うため，停電以外に障害による影響だと思われるフローが発見された場合は，そのフローに対して詳細な解析を行う．日常的な通信の傾向を把握するために，通信フロー数に対して解析を行っている他，国情報ごとに細分化しての解析や，九州大学の利用ポート番号の分布等についての解析を行う．九州大学のルータから収集・蓄積されているフローデータの中から，2007 年 3 月 9 日から 2007 年 4 月 7 日まで，金曜，土曜のデータについて 5 週間分解析を行う．また，上記の期間以外に，瞬間停

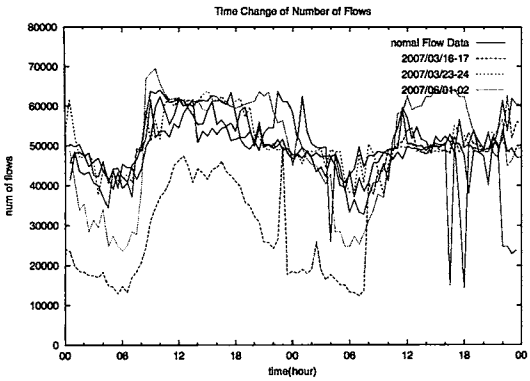

図 3　通信フロー数の時間変化
Fig. 3　Time Change of Flow Fata

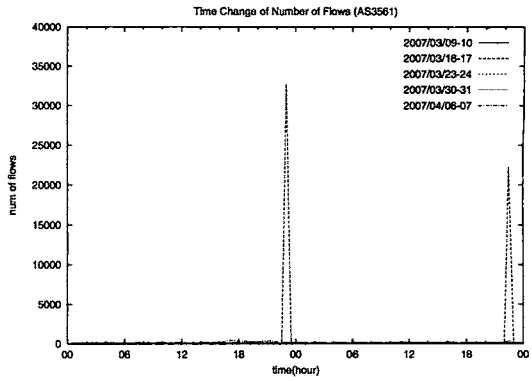

図 4　AS3561 における通信フロー数の時間変化
Fig. 4　Time Change of Flow Fata (for AS3561)

電の解析例として，2007 年 6 月 01 日から 6 月 2 日のフローデータの解析も行う．インターネットトラフィックには周期性が存在し，曜日によってそれぞれ通信傾向が異なっている．曜日ごとの通信傾向の差を吸収するため，金曜，土曜のみのデータを抽出して解析を行っている．停電情報については，九州大学の基盤センターで公開されているネットワーク障害情報を利用している．場合表 2 に今回解析を行った障害のイベントセットを示す

4.　解析結果

4.1　通信フロー数の解析

図 3 に九州大学の通信フロー数の時間変化を示す．全てのフロー数の変化を解析することで，大局的なデータ変化傾向を把握することができるが，様々なフローが複雑に重なり合っているため，微細な変化の検出は難しい．そのため，本節では大局的なトラフィックの通信傾向を調査する．図 3 を見てわかるように，インターネットトラフィックは，朝の時間帯には少なく，午前 6 時ごろから午後 8 時ごろまでに集中して流れる傾向にある．九州大学でインターネットを利用する時間は，開校時に集中するためである．計画停電時の挙動については，図 3 に示している通り，計画停電が発生した 3 月 17 日の午前 8 時までフロー数に減少傾向が見られる．計画停電が終了した午前 9 時 30 分以降は通常時と同様の通信傾向へと戻っている．計画停電に備えてサーバの電源を切る，インターネット自体の利用を控える等の対策が行われた結果，停電開始までのフロー現象につながったのではないかと考えられる．しかし，同様に計画停電の行われた 3 月 24 日の午前 5 時から 11 時までのデータに関しては，停電前に通信量が減少するような傾向は得られなかった．また，3 月 24 日の 14 時 48 分から 15 時 18 分にかけて伊都地区にて突発停電が起こっていたが，前後のフローデータには変化が見られなかった．インターネットフローに影響を与える主要なサーバやスイッチなどには，無停電電源装置 (UPS) が備えられていることが多く，停電が発生しても一定時間は電源を供給することができる．そのため，短時間の停電に関しての影響がでなかったのではないかと考えられる．6 月 2 日の瞬間停電についても突発停電のトラフィックと同様，変化が見られなかった．また，停電以外にも 3 月 16 日の 23 時 00 分に急激にフロー数が上昇している部分が見られた．フローデータを参照してみると，この通信はほとんどが AS3561 によるものだということがわかった．全フローに対する解析では通信フロー数の変化は変動が大きく，異常なアクセスとの自然な変動との判別が難しい．次節から，ミクロなトラフィック変化を検出するために，フローを細分化して解析を行っていく．

4.2　特定 AS に対するフロー解析

本節では，特定の AS 番号に限定したフローデータの解析を行っている．通信傾向は AS や地域ごとに異なっており，不正なアクセスには一つの AS や IP アドレスから集中して起こるものもある．特定の AS の通信を抽出して解析を行うことで，全フローに対する解析では様々な通信に埋もれて発見できなかった変化を検出することができる．図 4 では解析期間中のフローデータから異常な通信量の増加が見られた AS 番号 3561 の通信のみを抽出し，通信フローの時間変化を示している．AS3561 は savvis というアメリカのインターネットサービスプロバイダであった．図 4 を見てもわかるとおり，基本的に AS3561 から九州大学への通信はほとんど流れていないが，3 月 16 日の 23 時 00 分に急激に通信フロー数が増加している．これは九州大学のある一つの IP アドレスから，savvis の特定の IP アドレスへ行われた通信によるものであった．送信元ポート番号は 2830 番の固定であったが，送信先ポート番号が 40000 番以降をランダムで利用していたため，不正なアクセスであると考えられる．この通信は，フロー数のみでなく通信パケット数とバイト数に関しても急激な上昇が見られた．また，3 月 17 日の 22 時 30 分にも急激な通信フロー数が見られ，こちらは savvis 側のある特定の IP アドレスから，複数の九州大学側の IP アドレスにポート番号 22 番 (SSH) を利用したアクセスであった．それぞれの通信フロー増加に対して，前後のフローデータを細かく収集して解析を行ったが，不正アクセスの前兆は発見できなかった．しかし，特定の AS の通信を抽出することで，全フロー解析の際に発見できなかった 3 月 17 日の以上な通信量の変化を検知できた．特定の AS と 2 日間に複数の不正アクセス

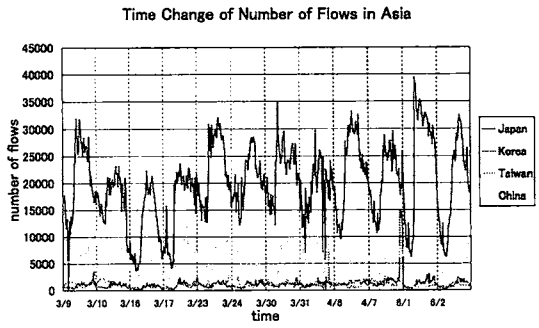

図 5 アジアにおける通信フロー数の時間変化
Fig. 5 Time Change of Number of Flows in Asia

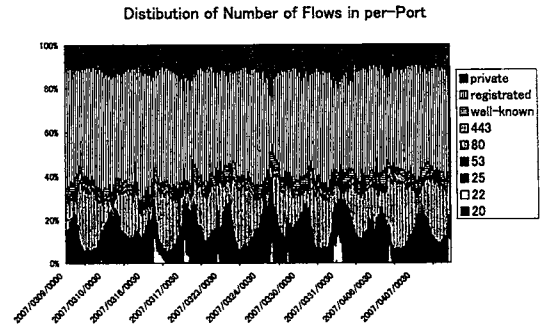

図 6 通信フローの利用ポート分布
Fig. 6 Distribution of Port Number

が発生しており，異常な通信や不正なアクセスが一つの AS に集中する傾向も発見できた．不正アクセスを解析する際に AS 番号に注目することで，より効率的に以上な通信の検知が行うことができることがわかった．

4.3 国情報，地域情報別のフロー解析

次に，国情報，地域情報別のフロー解析例を示す．通信量の増加傾向や減少傾向，利用ポートの傾向は国や地域ごとに異なっている．それぞれの国や地域について細分化して解析することで，インターネットトラフィックの複雑な変化を把握することができる．本研究ではアジアやヨーロッパ，北アメリカといった地域情報にも細分化して解析を行っているが，今回はより詳細に細分化した，国情報についての解析結果を示す．フローデータの国別解析の例として，図 5 に九州大学と通信を行っている，アジア諸国の中で通信フロー数の多かった 4ヶ国における通信フロー数の時間変化を示す．図 5 を見てもわかる通り，ほとんどの通信は日本と行われている．また，停電における 3 月 16 日，17 日の通信フロー数減少傾向は，日本以外の国ではあまり見られなかった．グラフを細かく見てみると，基本的には日本との通信傾向と同様に，周期的に通信量は増減しているが，ところどころ中国や台湾，韓国と通信フロー数が急激に増加している部分がある．台湾では，3 月 31 日 20 時 30 分に通常時の 10 倍を超える 8000 を超えるフローが，韓国では，4 月 7 日 20 時 00 分から 23 時 30 にかけて，通常時の 7 倍程度となる約 15000 のフローが流れている．AS 単位で解析を行うことでより詳細な解析を行うことができるが，全世界の AS 数は 30000 を超えており，単位が小さすぎるために不正アクセスのある AS を見つけるのが難しい．国や地域別の解析からトップダウン式に解析を行うことで，異常な通信のある AS を容易に見つけることができる．今回はそれぞれの通信フローの急増部分に対して詳細な解析は行っていないが，国別に情報を分割し急激なフロー増加部分を見ていくことで，全ての国に対する通信を見ていた時には発見できなかった変化も検出することができる．

4.4 通信フローの利用ポート番号解析

次に，九州大学のフローデータの利用ポート番号の分布を示す．図 6 は解析期間中の九州大学の通信フロー数をポート番号ごとに計算し，百分率で表したものである．解析するポートについては，一般的に使われているポート番号として表 1 に示したポート番号と，それ以外の Well Known Port，Registered Port，Dynamic and/or Private Ports に分けている．九州大学のネットワークは学術・研究機関として利用されているため，一般的に利用されているであろうメール（25 番）やウェブアクセス（80 番），それにと伴う DNS アクセス（53 番）を解析する．また，不正アクセスとして利用されやすい，端末へのリモートログインやリモートアクセス（22 番）についても解析を行う．図 6 より，ポート番号 25 番（SMTP），53 番（DNS），80 番（HTTP）で全通信の 40%の通信が利用されており，インターネットの通信用途には偏りがあることがわかる．また，表 2 で示した時間に各種停電が起こっているが，ポート番号の利用分布に目立った変化はない．3 月 16 日の計画停電前にはフロー数が減少する傾向が見られたが，全てのポート番号の通信において通信フロー数が減少しているため，利用ポート番号の分布には変化がなかったものと考えられる．停電の場合は通信フロー数が減少するが日本以外の国に対しては影響が少なく，利用ポート番号の分布に変化はない等，それぞれの解析結果を組み合わせることで，障害の特定を行うことができる．また，ポート番号 22 番（SSH）に注目すると，ときおり通信量が増加している部分がある．ポート番号 22 番は SSH を利用する際に用いられるポートで，不正のログインにおいては，DDos 攻撃の踏み台や不正なツールの保管場所として利用する等の目的でアタックが行われている．

図 7 は，図 6 と同様にポート番号ごとに通信フロー数を計算し，結果を通信フロー数で表しており，注目すべき変化のあったポートや代表的なポートのみを表示している．ポート番号ごとに通信傾向は異なっており，それぞれにおいて通信傾向を把握することで，総合的なフロー変化を予測し，またそのポート番号を利用した不正アクセス検知への比較対象として利用することもできる．ポート 22 番の通信に注目してみると，基本的には 1000 フローを超えることは稀だが，フロー数が増加している場合は数時間に渡り 1000 フロー以上の通信がされてお

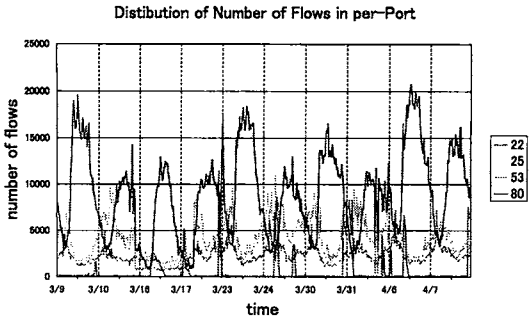

図 7　ポート番号ごとの通信フロー数の時間変化
Fig. 7　Distribution of Port Number

り，集中した通信が行われる時間帯もランダムである．一個人で 1000 フロー以上の SSH における通信を行うことは考え難く，SSH における不正アクセスが瞬間的にではなく，数時間に渡って起こっているものだと考えられる．最初に SSH アクセスの増加を検出した際に，以降の SSH アクセスに対して IP アドレスや AS 番号でアクセスブロックをする等の対策がとれれば，SSH における不正アクセスの被害を減少させることができると考えられる．また，通信フロー数の時間変化の傾向に対しては，図 3 で示した，全フローにおける通信傾向とは違ったものになっている．ポート 80 番は全フローにおける通信傾向と同様の結果になっているが，ポート 25 番とポート 53 番については，周期性はあるもののピークとなる時間が異なっている．ポート 25 番の通信は，朝 8 時前後と深夜に特に通信フロー数の増加傾向が見られる．メールのチェックなどは常時行っておらず，1 日の節目となる朝と夜の 2 箇所に集中してアクセスが行われているものと考えられる．また，ポート 53 番の通信は朝の 4 時前後に特に集中して行われている．DNS サーバはドメイン情報のキャッシュを 1 日周期で破棄・更新しているため，ユーザの動向に関わらず，周期的に集中的なフロー数の上昇傾向が現れたのだと考えられる．

5.　おわりに

本研究では，実際にインターネット上に流れているフローデータに対して日常的な通信傾向を把握するとともに，障害時のトラフィック変化として停電にも焦点を当てて解析を行った．計画停電に関しては，通信フロー数が減少する傾向が見られる場合もあったが，計画停電前に確実に通信フロー数が減少するわけはなかった．停電前の通信フロー減少は長期に渡る停電の影響を受けないようにサーバの電源を落とす，等の対策が練られているためだと考えられるが，停電の規模や地域によって影響も異なるため，障害情報とともにより詳細に解析を行う必要がある．また，数十分に渡る突発停電と瞬間停電についての解析では，フローデータに変化は見られなかった．主要なサーバやスイッチには無停電装置（UDP）等の対策が行われているため，短期間にわたる停電による影響が少なかったと考えられる．その結果，短期間の停電等の影響はインターネット構築の観点から見て影響が無いことがわかった．また，解析中に検出された異常なフローは AS3561 による通信であったが，前後のデータから不正なアクセスの前兆となるトラフィック変化を検出することはできなかった．しかし，AS3561 のように同一の AS と不正なアクセスが複数行われていることや，SSH によるフローの増加が数時間単位で渡っている等，不正なアクセスにある程度の傾向があった．異常なフローが検知された AS に注目することや，特定のポート番号を利用したフローに注目することなど，フローデータを細分化して解析を行うことが有効であることがわかった．また，ポート番号ごとに通信フロー数に違う傾向が表れており，ポート番号それぞれについて通信傾向を把握できるようにすることも有意義であると考えられる．

本研究では停電に注目して解析を行ったが，計画停電前にフローの減少傾向が見られる場合があるだけで，他の停電には特に目立ったトラフィック変化は検知できなかった．フローデータをトラフィック異常の検出に生かすためには，停電のみではなく，さまざまな障害や不正アクセス等の通信傾向を見出す必要がある．その際すべて手動で解析を行っていくのには限界があるため，トラフィック異常を自動的に検出する機構が必要となってくる．そのためには，通常時のトラフィックを定義し，解析を行いたいフローデータと比較することで異常の自動検出を行う必要がある．トラフィックの傾向はポート番号，国ごとに違うため，それぞれに対して通常時のフローを定義する必要がある．今後の課題として，通常時フローデータの平滑化手法の提案と，不正な通信を異常だと判断するための閾値の定義・定量化を行っていく．

文　献

[1] E. Laland, M.S. Taqque, W. Willinger, and D.V. Willson, "On the self-similar Nature of Ethernet Traffic (Extended Version)", Proc. IEEE/ACM Trans. Networking, vol2, no.1, pp1-15, Feb. 1994
[2] C.Estan and G.Varghese, "New Directions in traffic measurement and accounting", ACM SIGCOMM '02, 2002.
[3] Y.Zhang et al, "On the characteristics and origins of Internet flow rates", ACM SIGCOMM '02, 2002.
[4] Connie Logg, Les Cottrell, "Passive Performance Monitoring and Traffic Characteristics on the SLAC Internet Border", PAM2001 A workshop on Passive and Active Measurements, 2001
[5] Internet Assigned Numbers Authority (IANA), http://www.iana.org/assignments/port-numbers
[6] RFC 1772 - BGP-4 Application , http://www.ietf.org/.
[7] American Registry for Internet Numbers (ARIN), http://www.arin.net/.
[8] Resource IP Europeens Network Coordination Centre (RIPE-NCC), http://www.ripe.net/.
[9] Asia Pacific Network Information Centre (APNIC), http://www.apnic.net/.
[10] Japan Network Information Center (JPNIC), http://www.nic.ad.jp/.
[11] Korea Network Information Center (KRNIC), http://whois.nic.or.kr/.
[12] Arno Wagner, Martin May, Anukool Lakhina "Anomalies: Impact of packet sampling on anomaly detection metrics Daniela Braukhoff, Bernhard Tellenbach", 6th ACM SIGCOMM , pp 159-164, October 2006