

## P2Pファイル共有からWebサービスへシフト傾向にあるトラフィック

これまで年率30%程度で比較的安定して伸びてきたインターネットのトラフィック量が2010年1月初めに20%近く急減しました。その原因は、改正著作権法、いわゆるダウンロード違法化にあると言われています。ここでは、2010年5月24日から一週間のトラフィック量やポート使用量を用い、2009年のデータと比較することで、トラフィック量が減少した原因を探っていきます。

### 3.1 はじめに

本レポートは、IIR Vol.4でのレポート[1]の続報であり、IIJが運用するブロードバンド接続サービスでのトラフィックを分析して、その結果をまとめたものです。前回のレポートでは、過去5年間で国内外のインターネットのトラフィック量が年率30%程度の伸びを示し、比較的安定していると報告しました。ただし、トラフィック量には、一部のヘビーユーザの挙動が大きく影響したり、技術以外の社会的要因等でユーザの利用の仕方が変わったりするため、過去のデータを元に将来を予測することが難しいことも述べました。

実際2010年1月には、トラフィック量の伸びがマイナスになるという状況が生まれました。図-1に示すように、ブロードバンドトラフィック量が1月初めに2割近く減りました。これまでもトラフィック量の変動はありましたが、このような大きな減少が半年も継続したことはありませんでした。今回の減少は、2010年1月に施行された改正著作権法、いわゆるダウンロード違法化の影響だろうと言われています。しかし、罰則規定のないダウンロード違法化の施行が、ここまで影響したことは予想外でした。

本レポートでは、前回と同様に、利用者の1日のトラフィック量やポート別使用量などを元に、1年前と比べて実際に何が変わったのかを検証していきます。

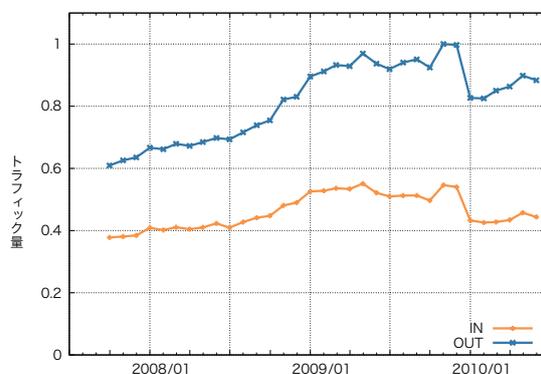


図-1 過去3年間のブロードバンドトラフィック量の推移 (2009年11月のOUT値を1として正規化)

### 3.2 データについて

今回利用した調査データは、前回と同様に、個人および法人向けのブロードバンド接続サービスでファイバーとDSLのブロードバンド顧客を収容しているルータから、Sampled NetFlowによって収集しました。ブロードバンドトラフィックは平日と休日でその傾向が異なるため、一週間分のトラフィック量を解析の対象にしています。今回は、2010年5月24日から30日の一週間のデータを使用しました。また、比較のために、前回解析した2009年5月25日から31日の一週間のデータも用いています。

各利用者の使用量は、利用者に割り当てられているIPアドレスと、観測したIPアドレスを照合して求めました。また、Sampled NetFlowでパケットをサンプリングして統計情報を取得しています。サンプリングレートは、ルータの性能や負荷に応じて、1/1024、1/2048、1/4096、1/8192のいずれかを設定しています。観測した使用量にサンプリングレートの逆数を掛けることで、全体の使用量を推定しています。このようなサンプリング方法を採用したため、使用量が少ない利用者のデータには誤差が生じています。しかし、使用量がある程度以上である利用者のデータでは、統計的に意味のある数字を得ることができます。

なお、ファイバーとDSLで観測する利用者数は、2005年頃にはほぼ同数でしたが、その後ファイバーへの移行が進み、2010年に観測したユーザ数の85%がファイバー利用者で、全トラフィック量の92%を占めるまでになっていました。また、本レポート内でのトラフィックのIN/OUTは、ISP側から見た方向で表しています。INは利用者からのアップロード、OUTは利用者へのダウンロードになります。

### 3.3 利用者の1日の使用量

まず、ブロードバンド利用者の1日の使用量をいくつかの視点から見ていきます。ここでの1日の使用量は、各利用者の1週間分のデータを7で割った1日平均です。

図-2に、利用者の1日のトラフィック量を相補累積度分布で示します。これは、1日の使用量がX軸の値より大きい利用者の割合を両対数グラフで示したもので、全体に対するヘビーユーザの分布を見るときに有効です。グラフの右側が直線的に下がっていて、ベキ分布に近いロングテールな分布であることが分かります。また、ヘビーユーザは全体に分布していて、決して一部の特殊な利用者ではないとも言えます。

IIR Vol.4で示した2009年での相補累積分布と比較してみると、ヘビーユーザの割合がIN（アップロード）で若干減少しています。例えば、1日に $10^8$  (100MB)以上をアップロードしているユーザの割合は、2009年に全体の8.2%でしたが、2010年に6.5%に減っています。これは、全体に対して1.7ポイントの減少に過ぎませんが、ヘビーユーザの数で見ると約20%も減少したことになります。一方、グラフ右端の部分では、2009年での分布よりも2010年のほうが伸びていて、極端なヘビーユーザは逆に増えていることも分かります。

図-3に、利用者間のトラフィック使用量の偏りを示します。このグラフは、使用量が上位X%である利用者によって全トラフィック量のY%が占められている事を表しています。使用量には大きな偏りがあり、結果として全体が一部利用者のトラフィックで占められています。例えば、上位10%の利用者がOUT（ダウンロード）

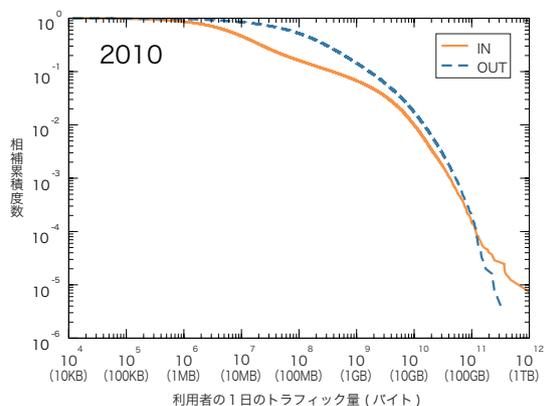


図-2 利用者の1日のトラフィック量の相補累積度分布

ド)の78%、IN(アップロード)の96%を占めています。また、上位1%の利用者がOUT(ダウンロード)の33%、IN(アップロード)の51%を占めています。

IIR Vol.4で示した2009年での使用量の偏りと比較すると、IN(アップロード)での偏りが大きくなっています。これは、ヘビーユーザの総数は減ったが、極端なヘビーユーザが増えた結果です。OUT(ダウンロード)では、ほとんど変化がありませんが、上位3%以上で若干偏りが小さくなっています。

しかし、このような偏りの傾向は、ロングテールな分布の特徴で、インターネットデータに共通したものです。例えば、使用量が上位である利用者を除いて再度偏りの分布を見ても、ほぼ同様な偏りが観測されます。このような偏りは、インターネット以外でも珍しいものではなく、単語の出現頻度や富の分布など大規模で複雑な統計によく現れることが知られています。

利用者間のトラフィック量の偏りは、一見すると一部のヘビーユーザとそれ以外のユーザという二極化が起きている印象を受けるかもしれませんが、使用量の分布がベキ乗則に従っているため、実際には多様なユーザが幅広く存在しています。

図-4に、利用者の一日の平均利用量の分布(確率密度関数)を示します。ここでは、IN(アップロード)とOUT(ダウンロード)に分け、X軸に利用者のトラフィック量、Y軸にその出現確率をそれぞれ示しています。また、X軸は対数表示で、その範囲は $10^4$ (10KB)から $10^{11}$ (100GB)です。今回の調査では、最も使用量が多い利用者のトラフィック量は2TBにも及び、一部の

利用者がグラフの範囲外になりますが、おおむね $10^{11}$ (100GB)までの範囲に分布しています。なお、グラフ左側に突起が現れていますが、これはサンプリングレートの影響によるノイズです。

IN(アップロード)とOUT(ダウンロード)の各分布は、片対数グラフ上で正規分布となる、対数正規分布に近い形をしています。これは、線形グラフで見ると、左端近くにピークがあり右方向になだらかに減少する、いわゆるロングテールな分布になります。OUTの分布がINの分布より右にあり、ダウンロード量がアップロード量よりも一桁ほど大きくなっています。平均値は、グラフ右側にあるヘビーユーザの利用量の影響を受け、IN(アップロード)で469MB、OUT(ダウンロード)で910MBです。2009年では、それぞれ556MBと971MBでしたので、利用量は減少しています。

IN(アップロード)の分布の右端に注目してみると、もう1つの小さな分布の山があることに気づきます。実際には、OUT(ダウンロード)にも、メインの分布に重なっていますが、同様の分布の山があります。これらの分布は、IN(アップロード)とOUT(ダウンロード)でほぼ同じ位置にあり、IN/OUTのトラフィック量が対称であるヘビーユーザの存在を示しています。そこで、ここでは便宜上、大多数を占めIN/OUTのトラフィック量が非対称な分布を「クライアント型利用者」、右側にある少数でIN/OUTのトラフィック量が対称なヘビーユーザの分布を「ピア型利用者」と呼ぶことにします。

クライアント型利用者での最頻出値を2009年と2010年で比較すると、IN(アップロード)で6MBから7MBに、OUT(ダウンロード)で114MBから145MBにそ

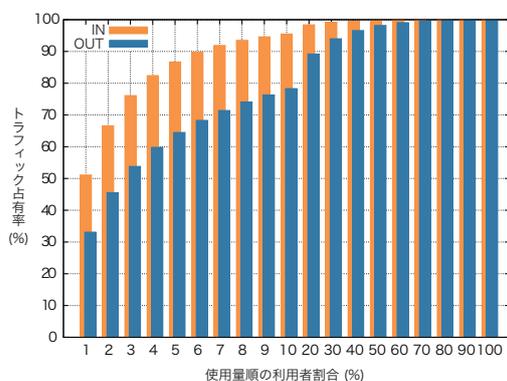


図-3 利用者間のトラフィック使用量の偏り

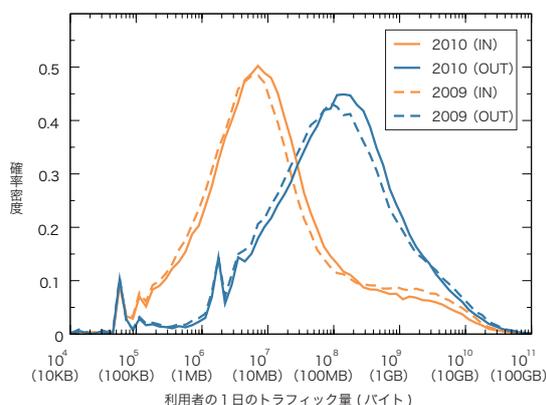


図-4 利用者の1日のトラフィック量分布

それぞれ増え、各利用者のトラフィック量が特にダウンロード側で増えていることが分かります。これに対して、ピア型利用者の出現確率は、わずかですが低下しています。つまり、一般利用者の利用量が着実に増えているのに対して、量的に大勢を占めるヘビーユーザの利用量が減ったため、それがトラフィック総量の減少に繋がったと言えます。

図-5では、ランダムに抽出した利用者5,000人のIN/OUT利用量をプロットしています。X軸にOUT(ダウンロード)、Y軸にIN(アップロード)をそれぞれ採り、両対数グラフで表しています。IN/OUTのトラフィック量が同量である利用者は、グラフの対角線上にプロットされます。

ここでは、2つのクラスターが観測できます。対角線の下側にあり、対角線に沿って広がるクラスターは、ダウンロード量がアップロード量より桁多いクライアント型利用者です。もう1つのクラスターは、右上の対角線あたりを中心に広がるピア型利用者です。しかし、この2つのクラスターの境界はあいまいです。これは、実際には、クライアント型利用者である一般ユーザもSkypeなどのピア型アプリケーションを利用しますし、ピア型利用者であるヘビーユーザもウェブ等のダウンロード型アプリケーションを利用しているためです。つまり、多くの利用者が両タイプのアプリケーションを異なる割合で使用しています。また、各利用者の利用量やIN/OUT比率のバラツキも大きく、多様な利用形態が存在することがうかがえます。この傾向は、2009年と比較しても、ほとんど違いがありませんでした。

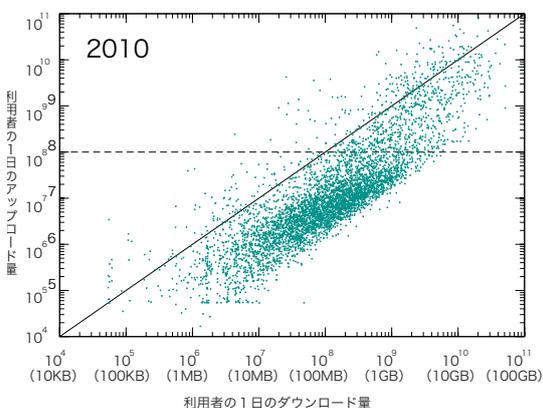


図-5 利用者ごとのIN/OUT使用量

### 3.4 ポート別使用量

次に、トラフィックの内訳をポートごとの使用量から見ていきます。最近では、ポート番号からアプリケーションを特定することは困難です。P2P系アプリケーションは、双方で動的ポートを使うものが多く、また、クライアント・サーバー型アプリケーションの多くは、ファイアウォールを回避するために、HTTPで使われる80番ポートを利用しています。大雑把に分けると、双方が1024番以上の動的ポートを使っていればP2P系アプリケーションの可能性が高く、片方が1024番未満のいわゆるウェルノウンポートを使っていればクライアント・サーバー型アプリケーションの可能性が高いと言えます。そこで、ここでは、TCPとUDPでソースとデスティネーションのうちポート番号の小さいほうを採り、ポート番号ごとの使用量を見ることにします。

また、トラフィックの全量がピア型利用者であるヘビーユーザのトラフィックに支配されていることから、クライアント型利用者である一般利用者の動向を知るために、少々乱暴なやり方ですが、1日のアップロード量が100MB未満のユーザを抜き出し、それらをクライアント型利用者としてしました。これは、図-4でIN(アップロード)の2つの分布の中間あたり、図-5でIN = 100MBにある水平線の下側の利用者になります。

図-6は、ポートの使用状況を、全体とクライアント型利用者に分けて2009年と2010年で比較したものです。また、表-1に、その詳細を数値で示します。

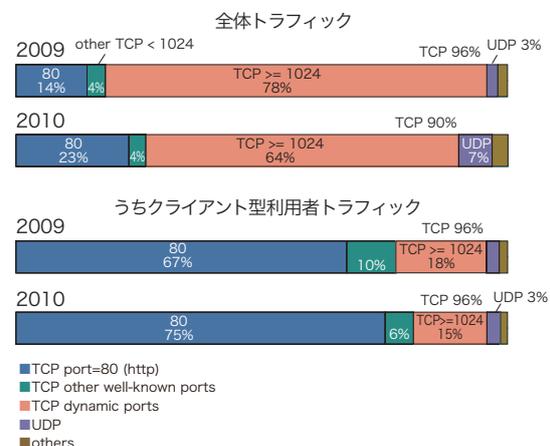


図-6 ポート別使用量概要

2010年のトラフィック量の90%がTCPです。全体のトラフィック量では、2009年に総量の78%だった1024番以上のTCPの動的ポートの割合が2010年には64%にまで減少しています。動的ポートでの個別のポート番号の割合はわずかで、最大でも総量の1%にしか過ぎません。一方、80番ポートの割合は、2009年の14%から23%に増加しています。つまり、2010年のトラフィック減少分を考慮すると、動的ポート同士の通信が約25%減少し、そのうちの1/3程が80番ポートに移行したと考えることができます。

また、クライアント型利用者に注目してみると、2009年に67%を占めていた80番ポートの利用割合は、2010年に75%まで増加しています。一方、動的ポートの割合は、18%から15%に減少しています。

これらのことから、依然として、TCP80番ポートのトラフィック量が増加傾向にあると言えます。80番ポートにはビデオコンテンツやソフトウェアアップデートなども含まれるためコンテンツタイプは特定できませんが、クライアント・サーバ型の通信量が増えていると言えます。

protocol port	2009		2010	
	total (%)	client type	total (%)	client type
TCP *	95.80	95.73	90.09	95.82
(<1024)	18.23	77.31	26.46	80.87
80 (http)	14.46	67.30	23.00	75.12
554 (rtsp)	1.48	6.89	1.15	2.45
443 (https)	0.64	1.91	0.98	2.28
20 (ftp-data)	0.19	0.17	0.18	0.07
(>=1024)	77.57	18.42	63.63	14.95
1935 (rtmp)	0.36	1.51	1.04	2.91
6346 (gnutella)	1.10	0.60	0.86	0.33
6699 (winmx)	0.70	0.24	0.65	0.17
8084	0.00	0.00	0.61	0.00
UDP	2.24	2.60	6.79	2.76
ESP	1.87	1.55	2.91	1.30
GRE	0.07	0.08	0.14	0.06
IP-IP	0.01	0.00	0.04	0.01
ICMP	0.02	0.05	0.02	0.04

表-1 ポート別使用量詳細

図-7は、全トラフィックによるTCPポート利用状況の週間推移を2009年と2010年のそれぞれで示したものです。ここでは、TCPでのポート利用を、80番ポート、その他のウェルノウンポート、動的ポートの3つに分けています。また、トラフィックの絶対量が公開できないため、ピーク時の総トラフィック量を1として正規化しています。全トラフィックでは、動的ポートが支配的であり、そのピークが23:00～1:00にあります。また、土日には昼間のトラフィック量も増加していて、家庭での利用時間が反映されています。2009年のデータと比較してみると、80番ポートの割合が増えていることがわかります。

図-8は、図-7と同様にクライアント型利用者によるTCPポート利用状況の週間推移を示したものです。ここでも、2010年には80番ポートの割合が増加しています。また、ピーク時間は21:00～23:00で、全トラフィックの利用状況を示す図-7に比べて2時間ほど早くなっています。さらに、土日は朝から利用が増えています。

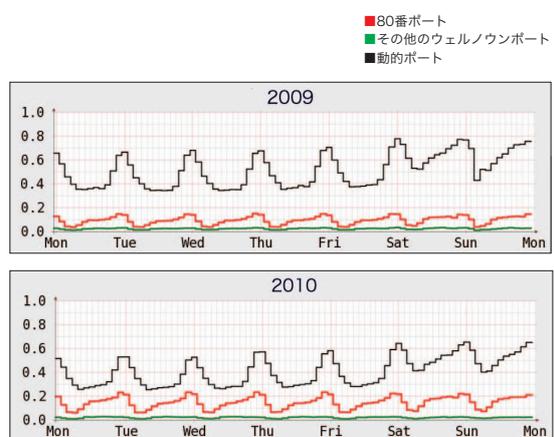


図-7 TCPポート利用の週間推移

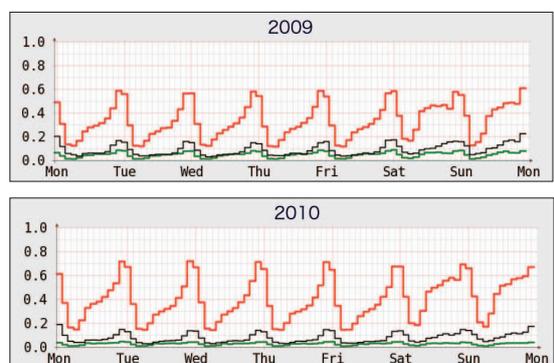


図-8 クライアント型利用者のTCPポート利用の週間推移

### 3.5 おわりに

前回のレポートでは、P2Pファイル共有に代表されるピア型トラフィックが、依然として量的に支配的であるものの、2005年からあまり増加していないことを示しました。また、利用者がP2Pファイル共有から、より使いやすく魅力的なビデオ共有サイト等のWebサービスへ移行している傾向を伝えました。その一方で、一般利用者の利用量が、ビデオコンテンツや他のWeb 2.0系のリッチコンテンツによって着実に増加していることも示しました。

今回の調査でも、これらの傾向に変化はありませんが、ヘビーユーザ部分の変動がこれまでよりも大きかったと言えます。ただし、ヘビーユーザや動的ポート同士の通信が単純に減ったわけではありません。ヘビーユーザ数は20%程減少しましたが、一方で極端なヘビーユーザのトラフィック量は増えています。また、ポートごとの通信量では、動的ポート同士の通信量が25%程減少しましたが、そのうちの約1/3は80番ポートに移行しています。これらのことから、前回のレポートで一般利用者に顕著だったP2Pファイル共有からWebサービスへの移行がヘビーユーザにも広がっていると言えます。

したがって、トラフィック量の減少と改正著作権法の影響に関しては、「改正著作権法でP2Pファイル共有が減った」というよりは、むしろ「以前からあったP2Pファイル共有からWebサービスへの移行という流れ

が、改正著作権法を契機に進んだ」と考えるべきだと思います。例えるならば、地震で地滑りが起こったのは、地盤が緩んでいたことが本当の原因であり、地震はあくまでトリガーに過ぎない、ということです。

世界的にも、トラフィックは、P2Pファイル共有からWebサービスへシフトしてきています[2、3、4]。また、今回取り上げた改正著作権法と同様に、他の国でも著作権に関する規制や法整備、あるいは違反者の検挙などで、トラフィックが減少したという事例が報告されています。しかし、ここまで示してきたように、個別の事例の背景には、著作権侵害のリスクに対する社会的認識の変化や、P2Pファイル共有の代替技術の成熟に伴う全体的な利用形態の変化が関係していることがあります。

今回取り上げた2010年初頭からのトラフィックの減少は日本だけの現象です。日本では、ファイバーアクセスが普及して帯域に余裕があるため、他の国に比べてヘビーユーザのトラフィック利用量の比率が高く、P2Pファイル共有でのトラフィックの割合も大きくなっています。その分、今回示したように、ヘビーユーザの挙動変化が全体に及ぼす影響が大きくなります。

IJでは、インターネットの利用形態の変化に迅速に対応できるよう、継続的にトラフィックの観測を行っています。今後も、定期的にレポートを提供していく予定です。

#### 参考文献

- [1] 長 健二郎. ブロードバンドトラフィック: 増大する一般ユーザのトラフィック. Internet Infrastructure Review. vol.4. pp18-23. August 2009.
- [2] G.-Maier, A.Feldmann, V.Paxson, and M.Allman. On Dominant Characteristics of Residential Broadband Internet Traffic. *IMC2009*. Chicago, IL, Nov. 2009.
- [3] C.Labovitz, D.McPherson, and S.Iekel-Johnson. 2009 Internet Observatory Report. *NANOG47*. Dearborn, MI. October, 2009.
- [4] Cisco. Visual Networking Index — Forecast and Methodology, 2009-2014. June 2010.

執筆者:

長 健二郎 (ちょう けんじろう)

株式会社IJイノベーションインスティテュート 技術研究所