

2 ブロードバンドトラフィック

2.1 はじめに

本レポートでは、IIJが運用しているブロードバンド接続サービスのトラフィックを分析し、その結果を報告します。

過去5年間の、インターネットのトラフィック量の伸びは、国内及び世界的にも、比較的安定していることが報告されています(参考文献[3][4][5]P23参照)。国内ブロードバンドトラフィックの総量については、年率30%程度で増加していて、これは国内バックボーントラフィック全体の約60%にあたります。個人のインターネット利用者の多くはブロードバンド接続を利用しており、ブロードバンドトラフィックの傾向を知ることが、全体のトラフィックを理解するためにも重要です。(参考文献[1][2]P23参照)。

本レポートでは、利用者の1日のトラフィック量やポート別使用量等をもとに、最近のブロードバンドトラフィックの傾向を見ていきます。P2Pファイル共有等を使うヘビーユーザのトラフィック量は、依然、量的には支配的ですが、あまり増えていません。その一方で、一般利用者のトラフィック量は、ビデオ系コンテンツの増加やウェブサイトのリッチコンテンツ化で着実に増えています。

2.2 データについて

今回利用した調査データは、個人及び法人向けのブロードバンド接続サービスについて、ファイバーとDSLによるブロードバンド顧客を収容するルータで、Sampled NetFlowにより収集したものを使っています。ブロードバンドトラフィックは平日と休日で傾向が異なるため、一週間分のトラフィックを解析することとし、2009年5月25日から31日の一週間を期間としています。また、比較のため2005年2月21日から27日のデータも用いています。2005年時点ではまだ、YouTubeやニコニコ動画などのビデオ共有サービスは登場していません。

各利用者の使用量は、利用者に割り当てられたIPアドレスと、観測されたIPアドレスを照合して求めています。また、NetFlowでは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のいずれかに設定されています。観測された使用量に、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。サンプリングによって、使用量の少ない利用者のデータには少し誤差がでますが、ある程度使用量のある利用者に対しては、統計的に意味のある数字が得られます。

ファイバーとDSLの観測された利用者数を見ると、2005年ではほぼ同数でした。しかし、2009年にはファイバーへの移行が進み、観測ユーザ数の84%はファイバー利用者で、トラフィック量全体の90%を占めるまでになっています。

なお、本レポート中のトラフィックのIN/OUTは、ISPから見た方向を表し、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となります。

2.3 利用者の1日の使用量

まずはブロードバンド利用者の、1日の利用量をいくつかの切口から見ていきます。ここでの1日の利用量は、各利用者の1週間分の1日平均です。

図-1は、利用者の1日の平均利用量の分布(確率密度関数)を、アップロード(IN)とダウンロード(OUT)に分け、利用者のトラフィック量をX軸に、その出現確率をY軸に示しています。X軸はログスケールで、 10^4 (10KB)

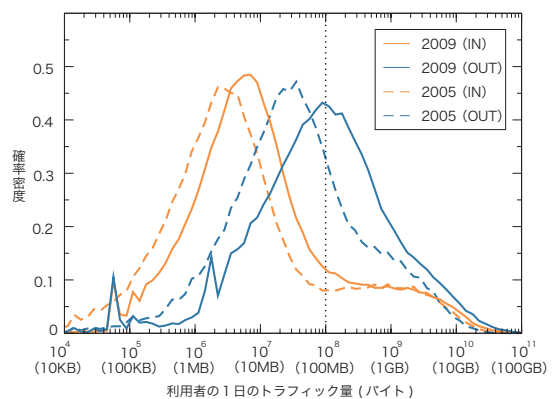


図-1 利用者の1日のトラフィック量分布

から 10^{11} (100GB) の範囲を示しています。一部の利用者はグラフの範囲外にあり、最も使用量の多い利用者は200GBにもものぼりますが、概ね 10^{11} (100GB) までの範囲に分布しています。2009年のグラフ左側に少しヒゲが出ていますが、これはトラフィックの増加に伴い、サンプリングレートが粗くなった影響のノイズです。

INとOUT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。これはリニアなグラフで見ると、左端近くにピークがあり右へなだらかに減少する、いわゆるロングテールな分布です。OUTの分布はINの分布より右にずれていて、ダウンロード量がアップロード量より、ひと桁程大きくなっています。平均値はグラフ右側のヘビーユーザの使用量に引っ張られるので、INの平均は2005年で430MB、2009年では556MB、OUTの平均は2005年で447MB、2009年では971MBにもなります。

INの分布の右端を見ると、もうひとつ小さな分布の山があるのが気になります。実はOUT側にもメインの分布に重なって、同様の分布の山があります。これらの分布は、INとOUTでほぼ同じ位置にあり、IN/OUT量が対称なヘビーユーザを示しています。そこで便宜上、大多数のIN/OUT非対称な分布を「クライアント型利用者」、右側の小数のIN/OUT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の、分布の最頻出値を2005年

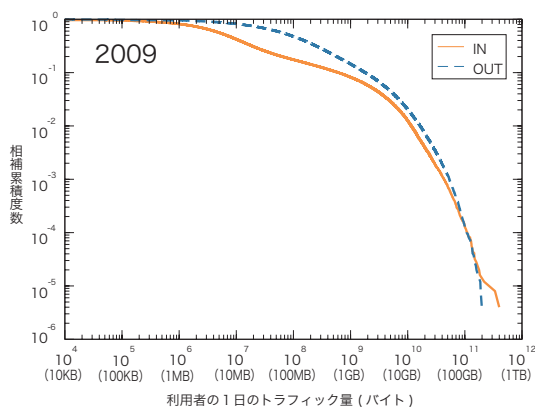


図-2 利用者の1日のトラフィック量の相補累積度分布

と2009年で比較すると、INでは3.5MBから6MBに、OUTでは32MBから114MBに増えていて、各利用者のトラフィック量が、特にダウンロード側で大きく増えていることがわかります。それに対して、ピア型利用者の分布では、最頻出値の位置は2005年、2009年ともに2GB付近で大きな変化はありません。つまり、一般利用者の使用量が大きく増えているのに対して、ヘビーユーザの使用量は、一定しています。

図には示していませんが、ファイバーとDSLでそれぞれ同様の分布を調べると、同じ年ではクライアント型とピア型の、分布の位置はほぼ同じですが、ファイバーではピア型利用者の割合が多くなっています。つまり、それぞれの分布の典型的な利用量に差はないのですが、ファイバーではヘビーユーザの割合が多くなっています。なお、ピア型分布の最頻出値の1日に2GBという数字は、ビット/秒換算で185kbpsとなります。

図-2は、利用者の1日のトラフィック量を、相補累積度分布にしたものです。これは、使用量がX軸の値より少ない利用者の、全体に対する割合をY軸に、ログ・ログスケールで示したもので、ヘビーユーザの分布を見るのに有効です。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わかります。ヘビーユーザは統計的に分布していて、決して一部の特殊な利用者ではないと言えます。

図-3は、利用者間のトラフィック使用量の偏りを示し

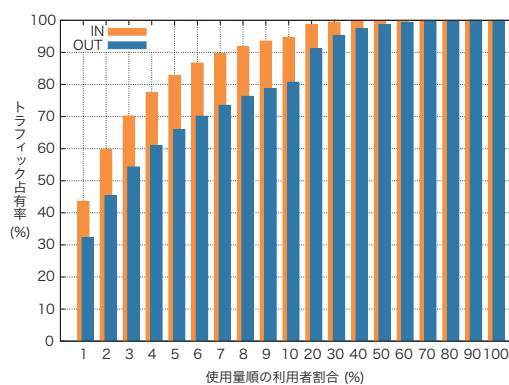


図-3 利用者間のトラフィック使用量の偏り

ます。使用量上位X%の利用者が、全体トラフィック量のY%を占めることを表します。使用量には大きな偏りがあり、結果として全体は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の80%、INの95%を占めています。さらに、上位1%の利用者がOUTの30%、INの40%を占めます。上位の利用者の使い方により若干の変動は生じますが、このような偏り方は2005年からほとんど変わっていません。これ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な傾向です。例えば、ピア型利用者を除いた偏りを見ても、やはり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決して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で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、使用量の分布はベキ乗則に従っていて、多様なユーザが幅広く存在していることが分かります。

図-4は、利用者ごとのIN/OUT使用量を2005年と2009年で5,000人をランダムに抽出してプロットしています。X軸はOUT(ダウンロード量)、Y軸はIN(アップロード量)で、ともにログスケールです。利用者のIN/OUTが同量であれば対角線上にプロットされます。

ここでは、2つのクラスタが観測できます。対角線の下

側に対角線に沿って広がるクラスタは、ダウンロード量がひと桁多いクライアント型の一般ユーザです。もう一方のクラスタは、右上の対角線近辺に広がるピア型のヘビーユーザです。しかし、ふたつのクラスタの境界はあいまいです。これは、実際には、クライアント型の一般ユーザでも、skypeなどのピア型のアプリケーションを利用し、また一方のピア型のヘビーユーザもウェブ等のダウンロード型のアプリケーションを利用しているからです。つまり、多くの利用者は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るのです。また、各利用者の使用量やIN/OUT比率にも大きなバラツキがあり、多様な利用形態が存在することが伺えます。

2005年と2009年を比較すると、クライアント型のクラスタは中心が右上に移動していること、ピア型のクラスタは幅が広がり密度が低くなっていることが確認できます。

2.4 ポート別使用量

つぎにトラフィックの内訳を、ポート別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には、双方が動的ポートを利用するものが多く、また、多くのクライアント・サーバ型アプリケーションが、ファイアーウォールを回避するため、HTTPが使う80番ポートを利用します。大雑把に分けると、双方が1024番以上

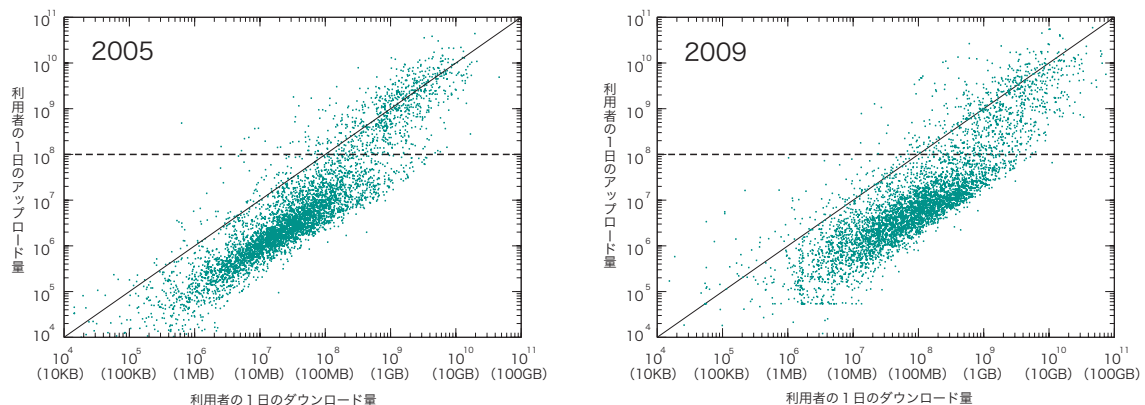


図-4 利用者ごとのIN/OUT使用量 2005年(左) 2009年(右)

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・サーバ型のアプリケーションの可能性が高いと言えます。そこで、TCPとUDPで、ソースとデスティネーションのポート番号の小さい方を取り、ポート番号別の使用量を見えます。

また、全体トラフィックは、ピア型のヘビーユーザのトラフィックに支配されているので、クライアント型の一般利用者の動向を見るために、少し荒っぽいですが、1日のアップロード量が100MB未満のユーザを抜き出して、これをクライアント型利用者として見ます。これは図-1では、IN側のふたつの分布の中間にあたり、図-4では、IN = 100MBにある、水平線の下側の利用者にあたります。

図-5はポート使用の概要を、全体とクライアント型利用者について、2005年と2009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。

トラフィックの95%以上はTCPです。さらに、全体で見るとほとんどはTCPの動的ポートで、2009年には総量の78%が、双方とも動的ポートを使ったトラフィックとなっています。動的ポートでの、個別のポート番号の割合は僅かで、最大のもので総量の1.1%に過ぎません。80番ポートの割合は、2005年の9%から14%に増加しています。

一方、クライアント型利用者に限ると80番ポートが多く、2005年には51%であった割合が2009年には67%にまで増加しています。逆に、動的ポートの割合は、36%から18%に減少しています。また、2番目に多いのは554番ポートです。これはReal-Time Streaming Protocol (RTSP) で使われるポートで、ビデオコンテンツの増加と関連しています。

これらのデータから、TCP80番ポートのトラフィック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等も含まれているた

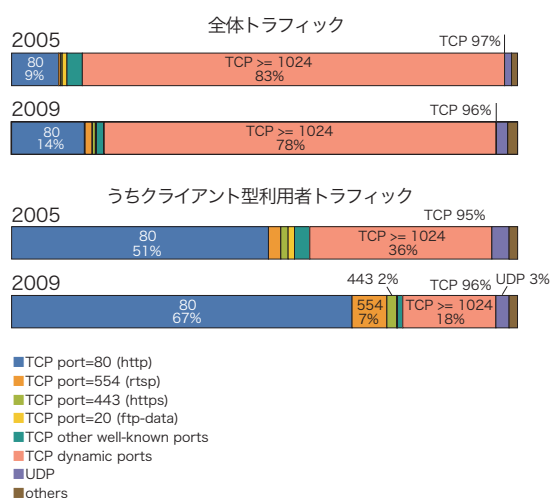


図-5 ポート別使用量概要

protocol port	2005		2009	
	total (%)	client type	total (%)	client type
TCP *	97.43	94.93	95.80	95.73
(<1024)	13.99	58.93	18.23	77.31
80 (http)	9.32	50.78	14.46	67.30
554 (rtsp)	0.38	2.44	1.48	6.89
443 (https)	0.30	1.45	0.64	1.91
20 (ftp-data)	0.93	1.25	0.19	0.17
(>=1024)	83.44	36.00	77.57	18.42
6346 (gnutella)	0.92	0.84	1.10	0.60
6699 (winmx)	1.40	1.14	0.70	0.24
1935 (rtmp)	0.20	0.81	0.36	1.51
7743 (winny)	0.48	0.15	0.25	0.03
UDP *	1.38	3.41	2.24	2.60
53 (dns)	0.03	0.14	0.03	0.07
ESP	1.09	1.35	1.87	1.55
GRE	0.07	0.12	0.07	0.08
ICMP	0.01	0.05	0.02	0.05

表-1 ポート別使用量詳細

め、コンテンツタイプの特定はできませんが、クライアント・サーバ型の通信量が増えていることが伺えます。

図-6は、全体トラフィックのTCPポート利用の週間推移を、2005年と2009年で比較したものです。ここでは、TCPに関して、ポート利用を80番、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、それぞれの推移を示しています。トラフィックの絶対量は開示できないので、ピーク時の総トラフィック量を1として正規化して表しています。全体では動的ポートが支配的で、そのピークは23:00~1:00、土・日には昼間のトラフィックが増加していて、家庭での利用時間を反映しています。

図-7は、クライアント型利用者について同様に、TCPポート利用の週間推移を示します。こちらは、2005年では80番は動的ポートより少し多い程度でしたが、2009年には80番が支配的になっています。また、ピーク時間は21:00~23:00と少し早くなっていて、土・日は朝からの利用が増えています。

全体とクライアント型との推移を比較すると、夜中を過ぎた後の、トラフィックの減り方に違いがあるのが分かります。80番は夜中を過ぎると急激に減少し、4時頃に最小となります。これに対して、動的ポートの

流量は朝にかけて徐々に減少し、朝8時頃に最小となります。この理由として、動的ポートを使うP2Pファイル共有で、夜にマニュアルでアップロードされたファイルが、一晩かけて拡散することや、目的のファイルのダウンロードが完了したら、アプリケーションを終了する利用者の行動を反映しているものと推測しています。

2.5 おわりに

これまで見てきたように、P2Pファイル共有に代表されるピア型トラフィックは、量的には依然支配的ですが、2005年から大きな増加は見られません。その原因として、利用者がP2Pファイル共有から、より使い易く魅力あるビデオ共有サイト等の、サービスへ移行していることが挙げられます。また、一時期急速にトラフィックを増やした、P2Pファイル共有が問題視されてきた結果、P2Pファイル共有の仕組みが、過度な帯域使用をしないように修正されてきたことや、ISPによる過度の利用制限の導入等で、利用者の意識が変化してきた影響もあるでしょう。

一方で、一般利用者の使用量が、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

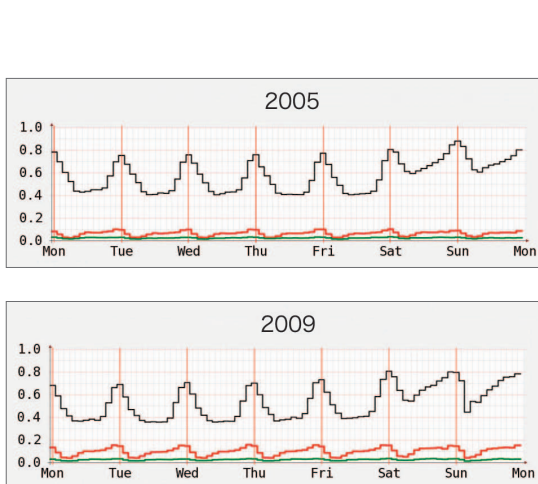


図-6 全体トラフィックのTCPポート利用の週間推移
2005年(上) 2009年(下)

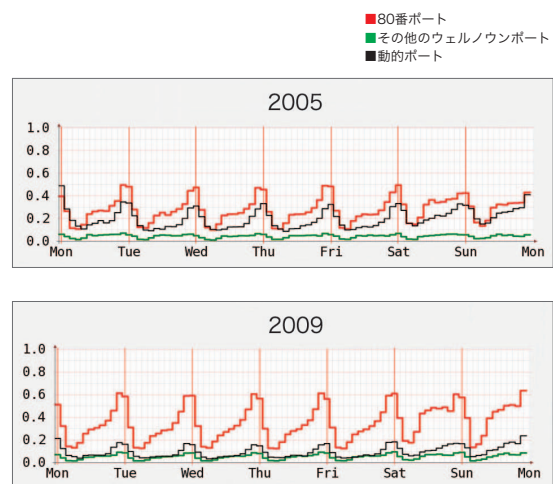


図-7 クライアント型利用者のTCPポート利用の週間推移
2005年(上) 2009年(下)

してきています。ビデオコンテンツだけではなく、ユーザが明示的にクリックしなくても、自動的にさまざまな情報を切替えて提供する、あるいはバックグラウンドで、次にユーザが利用する可能性のあるコンテンツを、先回りして取得するようなウェブサイトが増えていて、それに従ってトラフィック量も増えているのです。

2004年頃には、P2Pファイル共有の登場でトラフィックが急増し、回線の逼迫が予測されていました。その後5年間のトラフィック増加は、年率30%程度で安定して増加しています。その一方、バックボーン等のネットワークの容量も、年率50%程度で増強が進んでいると言われています(参考文献- [6])。そのため、現状マクロレベルでは、回線容量に余裕が生じていると考えられています。

しかし、インターネットのトラフィックに関して、過去のデータをもとに将来を予測するのは困難です。これ

は、一部のヘビーユーザの挙動の影響が大きいため、彼らの行動に変化があると、予測が大きくずれることになるからです。また、ユーザのインターネットの利用の仕方は、技術的な要因だけでなく、経済的要因、社会的要因、政治的要因等に大きく影響を受けます。そして、過去にウェブやP2Pファイル共有が登場して、トラフィックが激変したように、新しい技術の登場により、インターネットの使われ方が大きく変わる可能性は常に存在します。これまでは5年ないし10年周期で、トラフィックに大きな変化があったことを考えると、最近のトラフィック成長はある意味、安定し過ぎているとも言えます。近いうちにまた大きな変化が起こるのかも知れません。

IJでは、インターネットの利用形態の変化に素早く対応できるよう、継続的なトラフィックの観測を行っています。今後も、定期的にこのようなレポートをお届けしていく予定です。

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 副所長。インターネットがシンプルかつ柔軟な、安定した通信インフラとして発展するため、ネットワークの複雑な挙動の解析、QoS通信、オペレーティングシステムのネットワーク機能の研究等を行っている。WIDEプロジェクトボードメンバー、北陸先端科学技術大学院大学客員教授

参考文献

- [1] K.Cho, K.Fukuda, H.Esaki, and A.Kato.
The impact and implications of the growth in residential user-to-user traffic.
In ACM SIGCOMM2006, Pisa, Italy, Aug. 2006.
- [2] K.Cho, K.Fukuda, H.Esaki, and A.Kato.
Observing Slow Crustal Movement in Residential User Traffic.
In ACM CoNEXT2008, Madrid, Spain, Dec. 2008.
- [3] Cisco. Visual Networking Index - Forecast and Methodology, 2007-2012. June 2008.
- [4] Cisco. Approaching the zettabyte era. June 2008.
- [5] A.M. Odlyzko. Minnesota Internet traffic studies.
<http://www.dtc.umn.edu/mints/home.html>.
- [6] TeleGeography Research. Global Internet Geography. 2008.