

この1年間のトラフィック傾向について

この1年間のブロードバンドトラフィックには大きな変化はなく、安定傾向にあります。

今年、10月1日から施行される改正著作権法により、ブロードバンドトラフィックにどのような変化が出るのか注目されます。

3.1 はじめに

本レポートは、IIR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て、その結果を報告しています*1*2*3。今回も、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基に、この1年間のトラフィック傾向の変化を報告します。

3.2 データについて

今回も前回までと同様に、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により収集した調査データを利用しています。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、1週間分のトラフィックを解析しています。2012年5月28日から6月3日までの1週間分のデータを、前回解析した2011年5月30日から6月5日までの1週間分と比較します。

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を考慮して、1/8192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、ある程度以上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。

ここ数年、DSLからファイバーへの移行が進み、2012年には観測されたユーザ数の91%はファイバー利用者で、トラフィック量全体の95%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。

図-1は、ブロードバンド全体の過去5年間の月平均トラフィックです。一昨年の記事で報告したように、2010年1月にトラフィックが減少しました。これは、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と考えられています。それ以降、ダウンロード量(OUT)が増えている一方で、アップロード量(IN)は横ばいとなっていて、P2Pファイル共有のトラフィック割合が減っていることが伺えます。この1年のトラフィック量は、INは0.2%微増、OUTは8.8%増加しています。

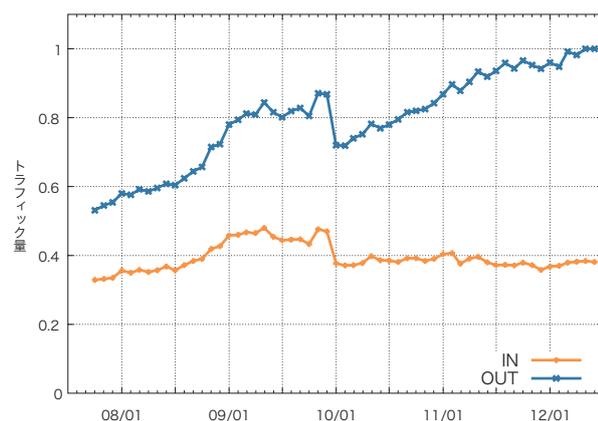


図-1 過去5年間のブロードバンドトラフィック量の推移

*1 IIR Vol.12「ブロードバンドトラフィックレポート:マクロレベルな視点で見た、震災によるトラフィックへの影響」を参照のこと (http://www.iij.ad.jp/company/development/report/iir/pdf/iir_vol12_report.pdf)。

*2 IIR Vol.8「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」を参照のこと (http://www.iij.ad.jp/company/development/report/iir/pdf/iir_vol08_report.pdf)。

*3 IIR Vol.4「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」を参照のこと (http://www.iij.ad.jp/company/development/report/iir/pdf/iir_vol04_traffic.pdf)。

3.3 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週間分のデータの1日平均です。

図-2は、利用者の1日の平均利用量の分布(確率密度関数)を示します。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軸に示していて、2011年と2012年を比較しています。X軸はログスケールで、10KB(10^4)から100GB(10^{11})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800GBにもなりますが、概ね100GB(10^{11})までの範囲に分布しています。

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、ひと桁大きくなっています。2011年と2012年で比較すると、INとOUT共に分布の山が右に少し移動して、利用者全体のトラフィック量が増えていることがわかります。

INの分布の右側は、分布の裾が広がっています。以前は、ここによりはっきりした山がINとOUT両方にあり、IN/

OUT量が対称なヘビーユーザを示していました。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」と呼んできました。今回もその慣習に従います。2011年と2012年の比較でも、ピア型利用者の山は小さくなっているのが分かりますが、これは、ヘビーユーザの割合が減少していることを示しています。グラフ左側に少しヒゲが出ていますが、これはサンプリングレートの影響によるノイズです。

表-1は、平均値と、分布の山の頂点にある最頻出値の推移を示します。分布の最頻出値を2011年と2012年で比較すると、INでは8.5MBから14MBに、OUTでは223MBから282MB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増えていることがわかります。一方、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、2012年には、INの平均は410MB、OUTの平均は1,026MBと、最頻出値よりかなり大きな値になります。2011年では、それぞれ432MBと1,001MBだったので、平均値で見ると、INが減少して、OUTが増えています。

図-3は、利用者ごとのIN/OUT使用量から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、共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。

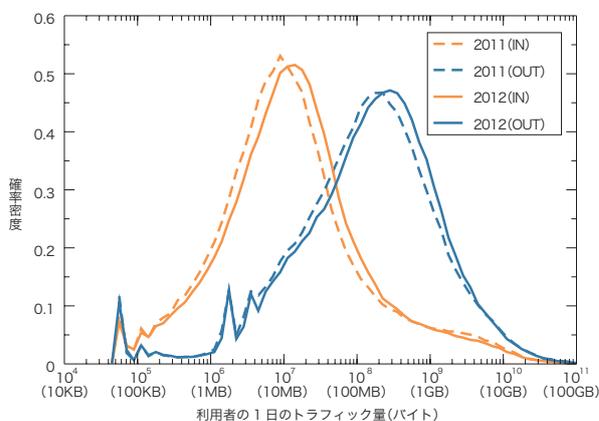


図-2 利用者の1日のトラフィック量分布2011年と2012年の比較

年	IN (MB/day)		OUT (MB/day)	
	平均値	最頻出値	平均値	最頻出値
2005	430	3.5	447	32
2007	433	4	712	66
2008	483	5	797	94
2009	556	6	971	114
2010	469	7	910	145
2011	432	8.5	1,001	223
2012	410	14	1,026	282

表-1 利用者の1日のトラフィック量の平均値と最頻出値の推移

対角線の下側に対角線に沿って広がるクラスは、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。以前は、右上の対角線上あたりを中心に薄く広がるピア型のヘビーユーザのクラスがはっきり分かっていましたが、今では識別が難しくなっています。便宜上、クライアント型とピア型に分けましたが、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Webなどのダウンロード型のアプリケーションを利用しているので、その境界はあいまいです。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ます。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。ここでは、2011年と比較しても、ほとんど違いは確認できません。

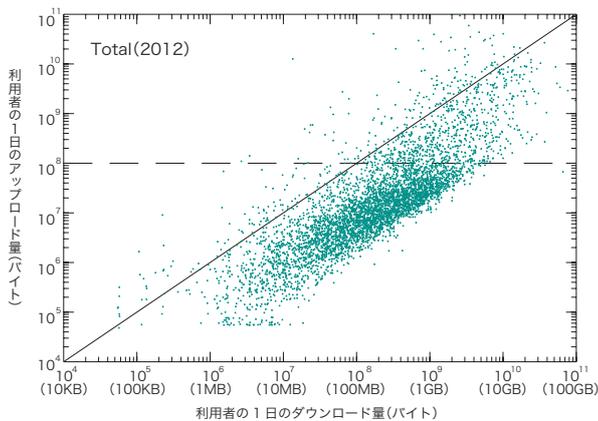


図-3 利用者ごとのIN/OUT使用量

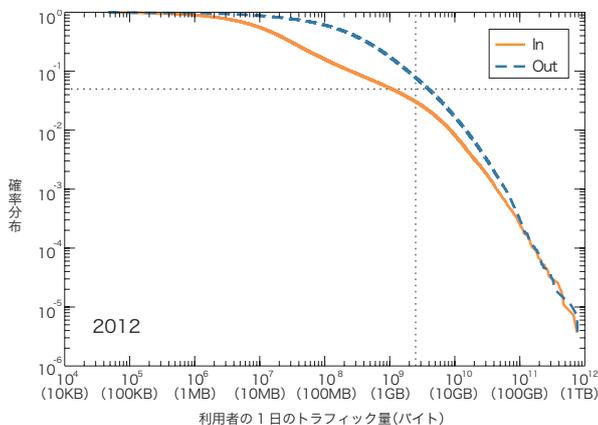


図-4 利用者の1日のトラフィック量の相補累積度分布

図-4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軸の値より多い利用者の、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分かります。2011年と比較すると、グラフ右端のテイル部分が若干延びていて、ヘビーユーザのトラフィック量も増えていますが、直線から右側に外れるような極端なヘビーユーザは逆に減っています。いずれにせよ、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-5は、利用者間のトラフィック使用量の偏りを示します。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73%、INの95%を占めています。更に、上位1%の利用者がOUTの33%、INの59%を占めています。

3.4 ポート別使用量

次に、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使うものが多く、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

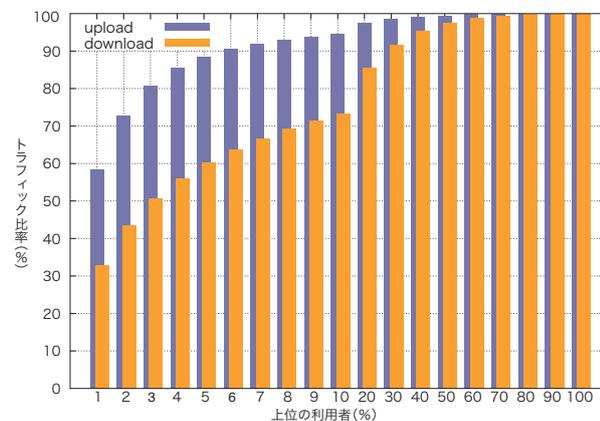


図-5 利用者間のトラフィック使用量の偏り

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番以上の動的ポートを使っていればP2P系の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てみます。

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します。これは、図-3では、IN=100MBにある水平線の下側の利用者にあたります。

図-6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、2011年と2012年で比較したものです。また、表-2にその詳細を数値で示します。

2012年のトラフィックの82%はTCPです。更に、全体で見ると、2011年には総量の50%だったTCPの動的ポート

が、2012年には41%にまで減少しました。動的ポートでの個別のポート番号の割合は僅かで、Flash Playerが利用する1935番が最大で総量の約2%ありますが、後は0.5%未満となっています。逆に、80番ポートの割合は、2011年の32%から36%に増加しています。

一方、クライアント型利用者に限ると、2011年には67%を占めていた80番ポートが、2012年には79%に増加しています。2番目に多いのは、HTTPSで使われる443番ポートで、2011年の2%から3%に増えています。昨年2番目に多かった554番ポートは、Real-Time Streaming Protocol (RTSP) で使われるポートで、2011年の7%から1%に減りました。また、動的ポートの割合は、11%から10%に減少しています。

これらのデータから、TCP80番ポートのトラフィック増加傾向が一般利用者だけでなく、ヘビーユーザにも広がってきたという傾向が、引続き進行していることが確認できます。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ているため、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のは確かです。

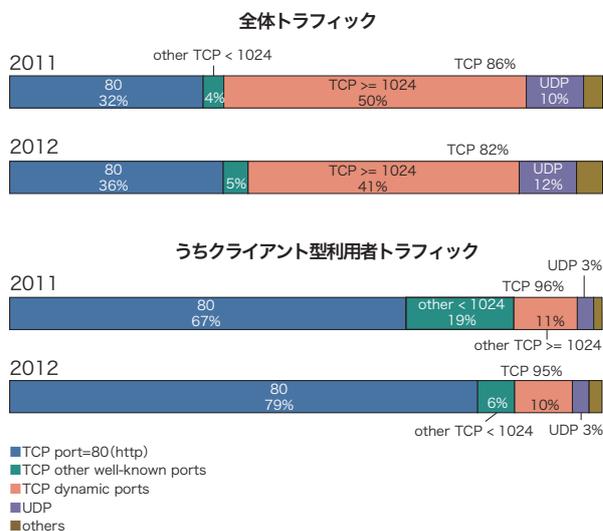


図-6 ポート別使用量概要

protocol port	2011		2012	
	total (%)	client type	total (%)	client type
TCP *	85.95	96.28	81.86	95.09
<1024	36.24	85.69	41.23	85.25
80(http)	32.10	67.30	36.22	79.39
443(https)	1.33	1.91	2.45	3.43
554(rtsp)	1.33	6.89	0.77	1.01
22(ssh)	0.27	0.17	0.22	0.06
(>=1024)	49.71	10.59	40.63	9.84
1935(rtmp)	1.58	1.51	2.12	3.91
7144(peercast)	0.38	0.00	0.44	0.04
6346(gnutella)	0.68	0.60	0.37	0.09
8080	0.26	0.14	0.30	0.17
UDP	10.01	2.61	12.38	2.94
ESP	3.56	1.02	5.29	1.79
GRE	0.15	0.05	0.16	0.14
L2TP	0.13	0.00	0.14	0.00
IP-ENCAP	0.10	0.01	0.09	0.01

表-2 ポート別使用量詳細

図-7は、全体トラフィックにおけるTCPポート利用の週間推移を、2011年と2012年で比較したものです。ここではTCPのポート利用を、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おり、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。2011年と比較すると、全体でも80番ポートの割合が更に増え、動的ポートの割合に匹敵する程になっているのが分かります。全体のピークは21:00～1:00、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-8は、クライアント型利用者とピア型利用者について、TCPポート利用の週間推移を示します。クライアント型利用者では、ほとんどが80番ポートで、ピーク時間は21:00～23:00です。ピア型利用者では、依然動的ポートが多いことが確認できます。

3.5 まとめ

これまで見てきたように、この1年間のブロードバンドトラフィックには大きな変化はなく、安定傾向にあります。

全体として、ダウンロード量は9%程度増えましたが、アップロード量はほぼ横ばいで、TCP80番ポートの割合が更に増えています。以前から報告しているWebサービスへの移行が一層進んだことが確認できます。

しかし、この安定傾向が今後も続くかは不透明です。今年の6月20日には、違法ダウンロードの刑事罰とDVDリッピングなどの違法化を盛り込んだ改正著作権法が国会で成立し、10月1日から施行されることになりました。2010年1月のダウンロード違法化の施行時には大きくトラフィックが減りましたが、今回も同様の影響が出る可能性があります。その一方で、前回のレポートで議論したように、ダウンロード違法化はすでにあった流れを加速するトリガーとなったに過ぎないと考えると、今回はあまり影響がでない可能性もあります。法的強制措置の効果を検証する意味で、今年の10月にブロードバンドトラフィックにどのような変化が出るのか注目されます。

IJでは、ユーザ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、定期的にレポートをお届けしていく予定です。

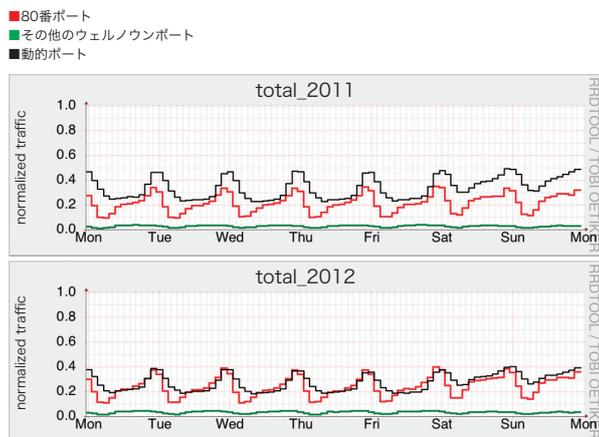


図-7 TCPポート利用の週間推移2011年(上)と2012年(下)

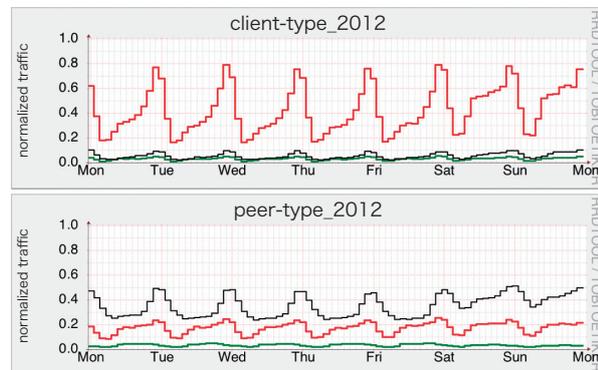


図-8 クライアント型利用者(上)とピア型利用者(下)のTCPポート利用の週間推移

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 所長。トラフィック計測やデータ解析などのインターネット研究に従事。慶應義塾大学 環境情報学部 特別招聘教授。北陸先端科学技術大学院大学 情報科学研究科 客員教授。