Flow-Level Upstream Traffic Behavior in Broadband Access Networks: DSL versus Broadband Fixed Wireless

Amit Sinha, Kenneth Mitchell, Deep Medhi
School of Computing and Engineering
University of Missouri-Kansas City
Email: asinha@conrel.sice.umkc.edu, {mitchellke,dmedhi}@umkc.edu

Abstract— **In this paper, we present flow-level upstream traffic behavior based on data collected from Broadband Fixed Wireless (BFW) and Digital Subscriber Link (DSL) access services. The study involves data collected using Cisco's NetFlow tools on both access networks. The observations indicate that a number of applications based on peer-to-peer (P2P) concepts are creating most of the upstream traffic. The flows observed are mostly short-lived for both BFW and DSL access, with DSL being the lesser of the two. The inter-arrival time displays near-range correlation for the DSL flows. There is a significant periodicity observed in the inter-arrival time distribution for flows on BFW access, which is indicative of the influence of the underlying Medium Access Control (MAC) protocol on traffic behavior. These analysis of upstream traffic characteristics form the first step towards constructing a generalized parametric model for broadband access networks.**

Index Terms— **Traffic Measurement, Access Networks, DSL, Broadband Fixed Wireless.**

I. Introduction

The evolution of broadband solutions in the "last mile" access network has enabled users to use applications which predominantly require large bandwidth along with low delay and jitter. These applications include Streaming Audio/Video, online games, Internet Protocol (IP) Telephony, etc. The advent of technologies like Digital Subscriber Line (DSL), Broadband Fixed Wireless (BFW), and cable modems has made this possible.

Various studies have been conducted which delve into the nature of traffic observed on the network. These works have either been theoretical or empirical in nature (see [9], [10]). The work by Cano et al. described the traffic observed on Ciez@net, that provides Internet access to the users through N-ISDN technology [1]. Kilpi and Norros presented a test for Gaussian character of access network traffic for ISDN [8]. Oliver and Benameur in their work presented a flow-level traffic characterization for ADSL access link *downstream* (traffic from the ADSL service provider to the end-users) [11]. Vicari et al. in their paper [13] provided some information about the ADSL traffic through a field trial; though, as of now, we are not aware of any comparative study that provides the traffic profile among different access networking technologies.

This paper gives empirical results of the traffic behavior and content observed in a broadband fixed wireless (BFW) and in a DSL access environment, and provides a comparative assessment. We restrict ourselves to characterizing the flow in the *upstream* direction since both these accesses work in an asymmetric environment where the bandwidth available in the upstream direction (from end-user to network access point) is much less than the downstream (network access point to end-user) direction. The important distinction between BFW and DSL is that in BFW, the upstream access is shared, while in DSL, the upstream bandwidth is dedicated per end-user. The results shown in this paper are from services deployed in specific markets; thus, it is important to recognize that this may not be representative of traffic observed in a general broadband access environment. Furthermore, the traffic pattern can change over time due to availability of different applications over time. Nevertheless, our interest is to highlight the results from a comparative study of BFW and DSL, with the hope that this may be beneficial to other Internet providers and designers.

The following characteristics have been observed: (a) The flows are mainly short-lived for both BFW and DSL, (b) the popularity of peer-to-peer (P2P) applications such as *kazaa* and *gnutella*, with *www* traffic still being dominant, (c) the impact of the underlying Medium Access Control (MAC) protocol on the arrival process in BFW access is significant. There are additional interesting observations comparing the two broadband access schemes based on the distribution of the flow duration and flow volume for aggregate traffic, and the contribution of various applications to the total traffic on each access – these are highlighted in this paper.

The paper is organized as follows: Section II describes the broadband access technique and the data collection methodology. In Section III we present the upstream traffic data analysis for Fixed Broadband Wireless access. Section IV consists of the upstream data analysis for DSL access traffic. Section V gives a comparative analysis of the data observed in these two different broadband access techniques. Finally we conclude this paper in Section VI.

II. Access Technology and Data Collection

The Broadband Fixed Wireless service considered here uses Multi-Channel Multi-Point Distribution Service (MMDS) band as its wireless transport from customer premises back to the head end. MMDS operates in 2.1 GHz to 2.7 GHz band and can support distances up to 30 miles between sites. Fig. 1 presents the architecture of the BFW environment. Fig. 2 represents a typical DSL service structure. This paper considers measurements from two different markets where the two broadband data services are deployed, consisting primarily of best-effort Internet applications. Customer bases include both residential and business customers.

The data flow from end-user to the head-end (for BFW) or end-user to central office (for DSL) is referred to as *upstream* traffic. The market considered in this study allows an upstream speed of up to 256 Kbps in case of BFW and an upstream speed ranging from 96 Kbps to 640 Kbps in case of DSL. The upstream data consists mainly of new content requests by the end-user, which consists of a short message, although in certain circumstances this may not hold. Table I illustrates some typical traffic that can be observed in the upstream direction based on the application. The bulk-transfer applications that use Transmission Control Protocol (TCP) as the protocol at the transport layer typically has the two-by-one behavior after the connection establishment and the slow-start phase; this means for two data segments sent, an acknowledgement is sent in the other direction. For a web-page delivered to the user, this means that for two data segments that are sent in the downstream direction, an acknowledgement segment is sent in the upstream direction. On the other hand, for a file upload, two data segments are going upstream, and an acknowledgement is sent the downstream direction. As can be seen, uploads can cause a bottleneck in the access since the upstream bandwidth is limited; on the other hand, transfer of real data segments in the upstream direction can not be avoided due to user sending emails (especially with attachments). The P2P applications for which a user's computer can act as a server (without the user knowing it) for others causes files to be sent in the upstream direction. Due to the limitation of bandwidth in the upstream direction, it is important to understand the upstream traffic characteristics for two reasons: (1) information is helpful to IP-access network providers in operational domains in terms of configuration of the network, (2) it can lead to designing of better (prioritized) scheduling algorithms for medium access.

The flow-level data has been collected using Cisco's NetFlow at the head-end. It provides more fine grained data than SNMP, but not as detailed and high volume as packet sniffers. According to Cisco's documentation [2], a NetFlow identifies a flow based on the tuple (*source address*, *source port*, *source interface*, *destination address*, *destination port*, *IP protocol number*, *IP type- of-service*). Any undefined fields are set to 0 for each flow. Table II summarizes the various fields in a typical NetFlow record. A NetFlow enabled router exports aggregated flows to some predefined destination for collection using UDP. Cisco's collection of flow level data is based on the definition given by Claffy et al. in [3]. According to their definition, "a flow is active as long as observed packets that are meeting the flow specification are observed separated in time by less than a specified amount of time". Cisco's additional constraints on this definition for expiration of a flow are as follows:

- The flow exceeds a maximum duration that has been predefined.
- The flow contains a *FIN* or *RST* TCP flag.

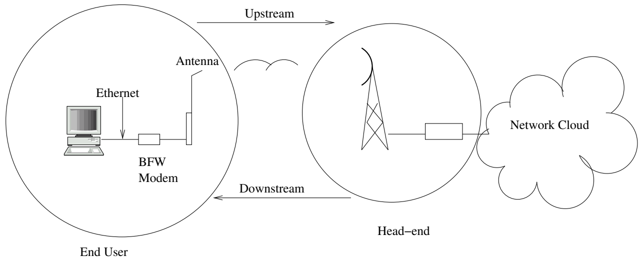

Fig. 1. A typical BFW service structure

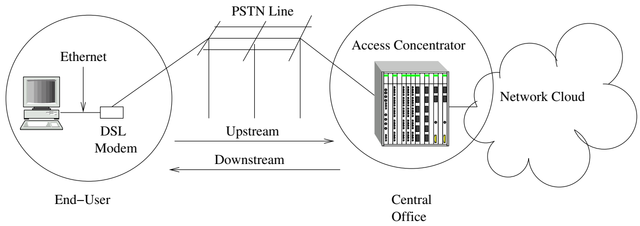

Fig. 2. A typical DSL service structure

TABLE I

TYPICAL UPSTREAM APPLICATION ACTIVITY

Application	Upstream Application Activity
HTTP	HTTP request message, HTTP file upload (include web-based mail sent out)
SMTP	Emails sent by the end-user especially with attachments
FTP	FTP connect request, GET request, file upload (if PUT command)
Streaming Video	Constant packet stream upstream

TABLE II

SUBSET OF FIELDS IN A TYPICAL NETFLOW RECORD

Field Name	Description
$Start$	Start time of the flow
End	End time of the flow
Sif	Source interface
$SrcIPaddress$	Source IP address
$SrcP$	Source Port
DIF	Destination interface
$DstIPaddress$	Destination IP address
$DstP$	Destination Port
P	Protocol Number
Fl	TCP control segment Flag
$Pkts$	Number of IP packets
$Octets$	Number of octets
tos	IP type-of-service

- The router's cache where the flow is temporarily stored before exporting gets exhausted.

In our paper, we refer to *flow* as defined by Cisco for the purpose of collecting data using NetFlow. The NetFlow data was collected for an hour duration on each type of broadband access network at different times. Since data are collected at the IP level, it is possible to identify and characterize different applications. Further, we have used Fullmer's *flow-tools* [7] to process the data collected from NetFlow.

III. Upstream Traffic Analysis for Broadband Fixed Wireless Access

In this section we present the upstream traffic analysis for BFW access. We observe the cumulative distributions of the flow duration and flow volume for aggregate traffic, and also for a few applications that had significant contribution to the traffic. The intent of this analysis is to quantify measurement, and to obtain some insights into application usage behavior.

A. Aggregate Flow on the Access

Fig. 3 shows the cumulative distribution of upstream flow packet volume, flow byte volume, and flow duration for the BFW access. Almost 30% of the upstream flows consist of single packets that are less than 100 bytes. The minimum size observed is 28 bytes (ICMP messages). The plot indicates that 80 percent of the flows consist of 10 packets or less, about 1 kilobyte of data or less, and about 9 seconds or less in duration. These measurements suggest that most flows are short.

We also observe a sudden peak for flows consisting of 3 packets and 144 bytes. On further analysis, we found that this was being caused by *www* traffic in the upstream, consisting mainly of TCP control segments such as SYN, FIN, and RST packets. The other sudden peak observed in the cumulative distribution of the bytes contained in a flow occurs at 40 bytes. These flows were mostly single packet *www* traffic, consisting of 76% RST, 12% FIN, and 12% of other TCP control messages.

B. Network Usage Information

The description of the flow characteristics shown thus far does not differentiate among various network applications. These applications significantly impact the duration and size of the flows. Fig. 4 shows the flow volume and duration distributions for eight common applications on the Internet. These applications are *UDP/dns*, *napster*, *smtp*, *kazaa*, *nntp*, *gnutella*, *www*, and *ftpdata*. The aggregate flow has been plotted as *All*. The impact these applications have on the aggregated flow behavior can be found in Table III in Section V, which details the contribution of each application as a percentage of total number of flows, packets, and bytes on the access network. For upstream traffic, 13.17% of flows and 1.60% of packets are from *UDP/dns*. 83% of the flows generated from this application are single packet flows. This is expected due to the short transactions that take place in the *dns* protocol. *www* traffic contributes the maximum number of flows (42.23%), packets (27.90%), and also a significant amount of bytes to the total access traffic. This again does not come as a surprise, due to the immense popularity of the web. What comes as a surprise though is that *www* contributes 12.94% of the total bytes observed in the aggregate flow, which is only second to *kazaa*. The average volume of the flow is around 1 Kilobyte with a maximum of approximately 7 Megabytes. Contrary to the expectation that upstream *www* traffic consists of small TCP control messages, we find larger byte volume in the flows. This is indicative of HTTP file uploads being carried out by the end-user, which includes web-based emails. Therefore, this application can influence the behavior of the aggregate flow observed in the upstream direction.

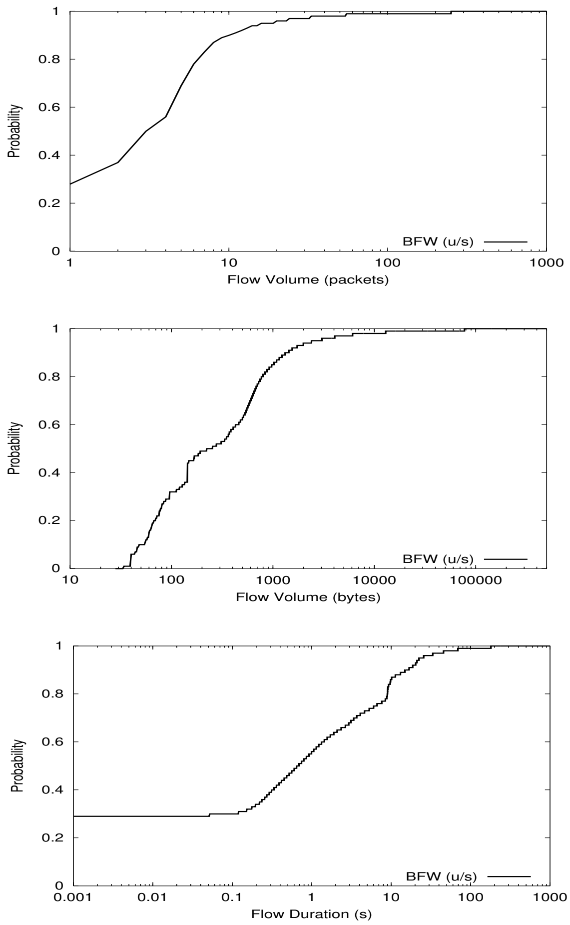

Fig. 3. Cumulative distribution of the upstream flow packet volumes, flow byte volumes, and flow duration for BFW access

The element of interest here is *kazaa* and *gnutella* (P2P applications) that are used for file sharing (like the mp3 files) over the Internet. The contribution of these applications toward total flow and packet is moderate, while the byte contribution is highest as compared to other applications being analyzed. Together they form 6% of the total flows, but contribute approximately 22.5% of the total bytes observed upstream. The average volume for *kazaa* is 7.6 Kilobytes per flow with a maximum of 22.34 Megabytes, while that of *gnutella* is 14.32 Kilobytes per flow with a maximum of 4.8 Megabytes. This has significant impact on the access in the upstream direction. Access links are designed assuming asymmetric flow in the upstream and downstream direction. Any imbalance in that assumption could lead to poor performance of the access protocol.

The other application that consists of a small number of flows but has a significant contribution to the number of bytes carried is *smtp*. This application falls behind the P2P applications and *www* only in the number of bytes. The reason for this is obvious, as the emails sent by the end-users (that may carry huge file attachments) travel in the upstream direction. Flows attributed to *ftpdata* seem to be insignificant, showing that not much data is being uploaded using *ftp* by the end-users.

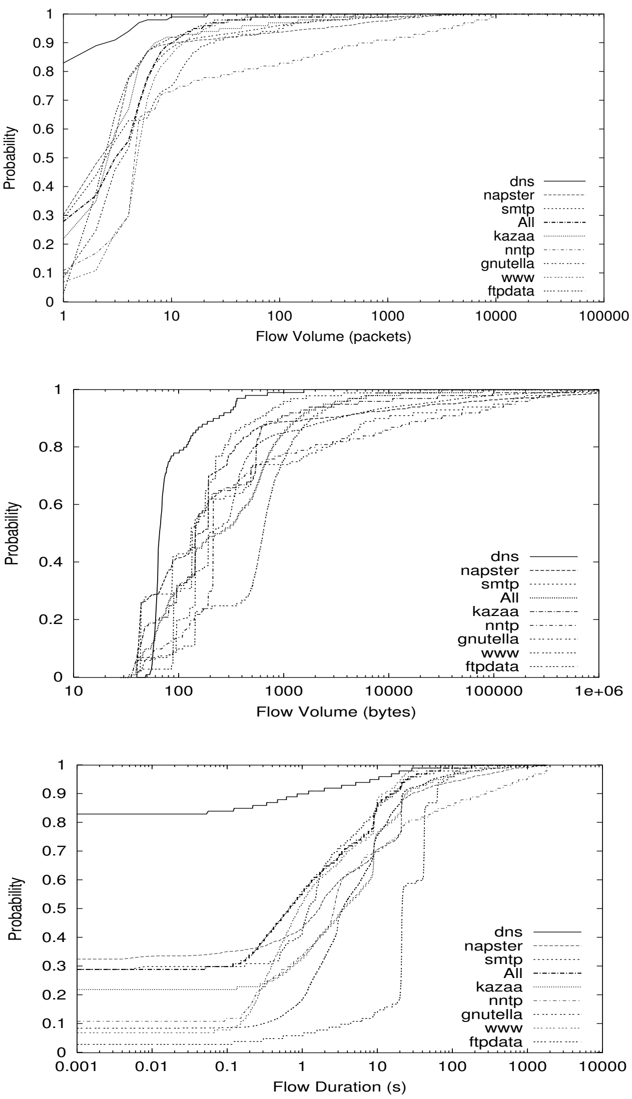

Fig. 4. Cumulative distribution of the upstream flow packet volumes, flow byte volumes, and flow durations of eight network applications for BFW access

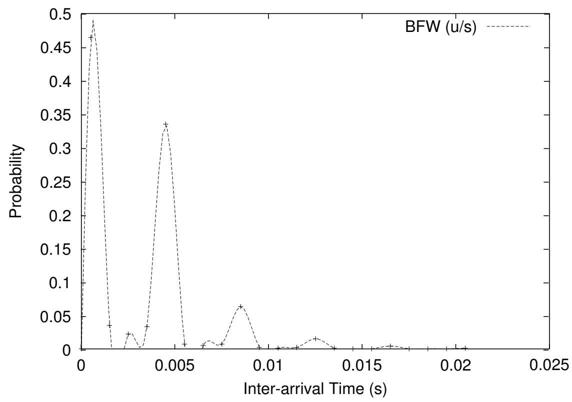

Fig. 5. Probability density for inter-arrival time on the BFW access

C. Arrival Process

In order to analyze the inter-arrival process for connections on the broadband access, we filtered all the upstream TCP flows that had the *SYN* flag set, since the arrival of a new TCP connection is done using a three-way handshake using the *SYN* (Synchronize) flag in the TCP packet header. As for the non-TCP flows, the start time of every such flow in the data set is assumed to be a new arrival. Moreover, since the granularity of the measured data is 1 msec, an arrival at time t msec actually arrives somewhere between t and $t + 1$ msec. We assume the arrivals to be evenly distributed in the interval $t, t+1$, and therefore shift all the inter-arrival times by 0.5 msec, i.e., to the mean of each interval. Therefore, the inter-arrival time of t msec from the trace is actually $t + 0.5$ msec.

Fig. 5 shows the probability density function (*pdf*) of the connection arrival process on the BFW access. The arrival process in BFW is influenced by the underlying proprietary MAC protocol that runs between the end-user and the head-end. This protocol, based on contention, polling, and dedicated channels, decides when a flow is accepted from any particular end-user. This causes periodicity that can be observed in the flow inter-arrival time distribution. Three significant periods observed are between 0-2 msec, 3-6 msec, and 7-9 msec. The mean inter-arrival time was found to be 0.003286 seconds with a variance of 0.000038. The squared coefficient of variation (c^2) for the process is 3.55. Hence, the traffic on BFW access is bursty.

An important aspect of this analysis is to help improve the underlying MAC protocol. Tilakawardana and Tafazolli in their work [12] looked into the effect of application models on MAC performance. Fig. 5 looks at the reverse, i.e., the impact of MAC protocol on the arrival process, which in turn would affect the performance of various applications. This analysis is outside the scope of the current paper.

IV. Upstream Traffic Analysis for DSL

So far, we looked into the characteristics of upstream traffic on the BFW access. Now, we delve into the traffic characteristics for DSL in order to create a platform for comparison. We look at the cumulative distributions for the flow-duration of the traffic, and also at various applications that contribute to this traffic. We then look into the dominant application in this access, along with the behavior of the connection arrival process.

A. Aggregate Flow on the Access

Fig. 6 shows the cumulative distributions of upstream flow packet volume, flow byte volume, and flow duration for the DSL access. Almost 38% of the upstream flows consist of single packets, with less than 100 bytes. The plot indicates that 80% of the total upstream flows consist of 8 packets or less, about 0.8 kilobytes of data or less, and are about 3 seconds or less in duration. These measurements suggest that most flows on the DSL access are short. The sudden peak observed at 5 packets is due to *www* flows, 88% of which contain both SYN and FIN TCP flags indicating the request for start and the request to end the flow respectively.

B. Network Usage Information

Fig. 7 shows the flow volume and duration distributions for the six most common applications being used on DSL access.

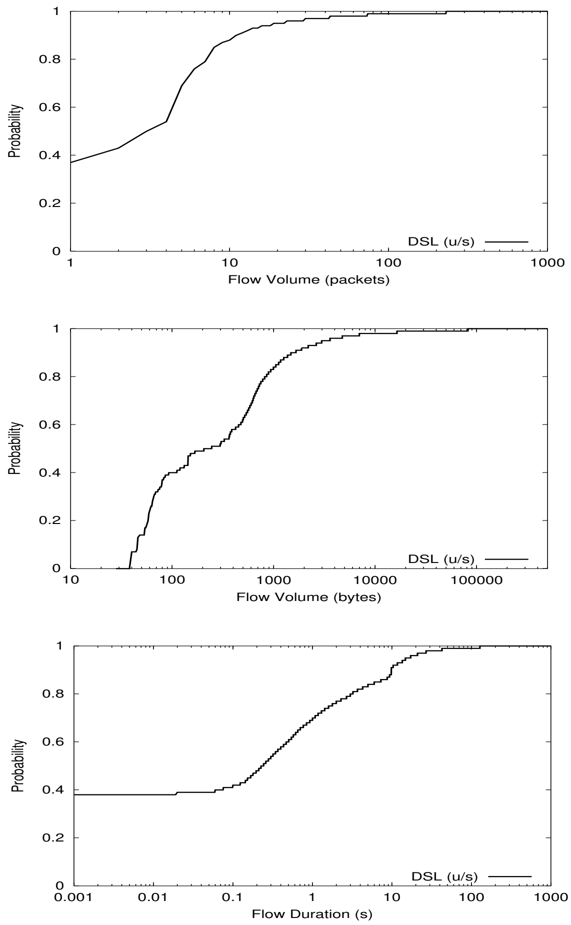

Fig. 6. Cumulative distribution of the upstream flow packet volumes, flow byte volumes, and flow durations for DSL access

These applications are *UDP/dns*, *ftpdata*, *smtp*, *www*, *kazaa*, and *gnutella*. The total aggregate flow has been plotted as *All*. Table III details the contribution of various applications to the aggregate flows observed.

Here again, we find that *UDP/dns* contributes a large majority of small flows, as 95% of the *dns* flows consist of a single packet. The *www* application contributes the maximum number of flows (37.08%) and packets (20.09%), and also a significant amount of bytes (9.62%) to the total access traffic. This reveals that *www* is popular amongst users, irrespective of the type of access they are using.

The contribution of *kazaa* toward the total flow is insignificant (1.62%), while packet contribution is moderate (4.60%). However, the byte contribution is highest (17.48%) as compared to other applications being analyzed. *smtp* has a moderate contribution to the number of flows, packets and bytes on the access. The data indicates that *ftp* is rarely being used for uploading files.

The sudden jump observed in the flow duration of Fig. 6 at 10 seconds can now be attributed to the significant peak observed in *smtp*, *kazaa* and *www* flows at this duration.

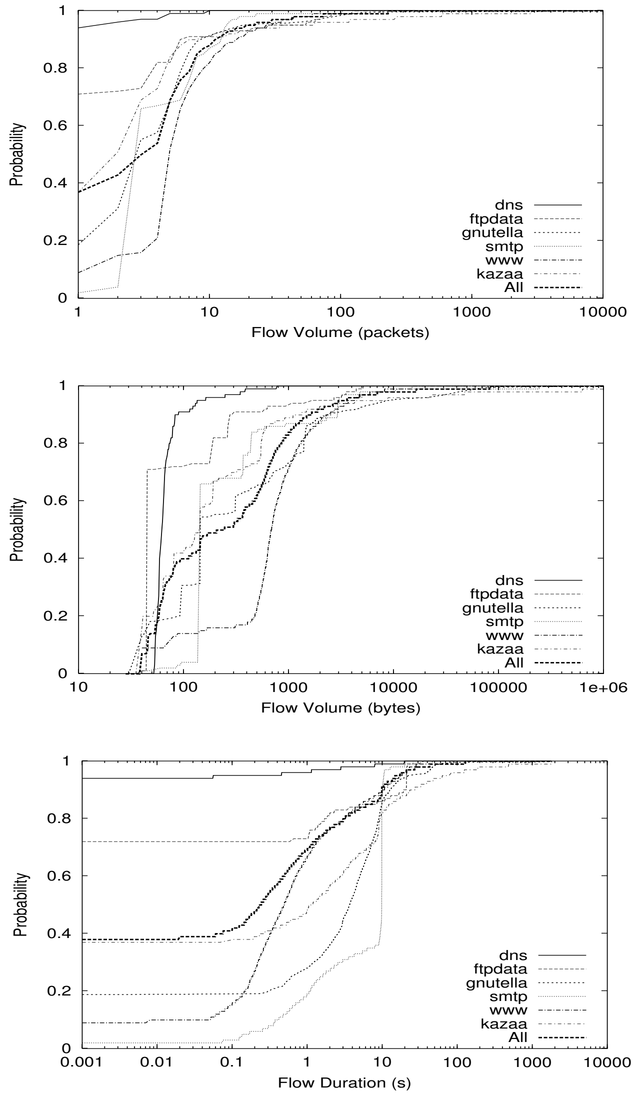

Fig. 7. Cumulative distribution of the upstream flow packet volumes, flow byte volumes, and flow durations of six network applications for DSL access

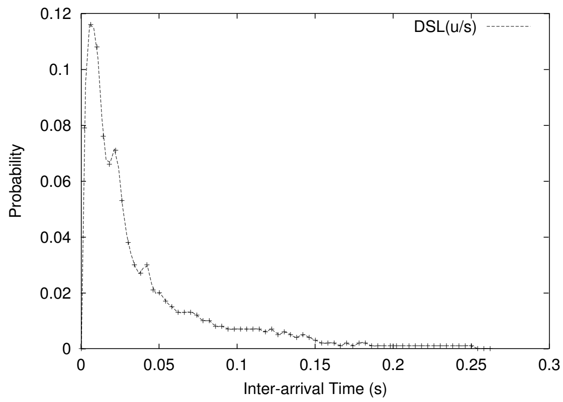

Fig. 8. Probability density for inter-arrival time in DSL access

C. Arrival Process

In order to analyze the inter-arrival process for connections on DSL, we go through the same process of filtering TCP flows containing SYN flags and shifting the inter-arrival time computed by 0.5 msec, as described in Section III-C.

The mean inter-arrival time was found to be 0.045361 seconds with a variance of 0.004389. The squared coefficient of variation (c^2) for the process is 2.133. Since, the value of c^2 is greater than 1, we expect the arrival process to the network access point in the DSL to be moderately bursty.

Fig. 5 shows the probability density function (*pdf*) of the connection arrival process on the DSL access. The peak is observed to occur between 4 and 20 msec, while the distribution spreads from 0 to 250 msec. Further analysis of the arrival process is beyond the scope of the current paper.

V. Comparing the Upstream Traffic Characteristics

The important distinction between BFW and DSL is that upstream access works in a shared mode in BFW while in DSL, upstream bandwidth is dedicated per user. Fig. 9 shows the comparison between the upstream flow characteristics with respect to flow volume and duration for the DSL and BFW access. Though the traffic belongs to different markets, with different customer bases, there is a striking similarity between the two access networks in the flow volume aspect of the aggregate traffic. Though we observe that there is a higher percentile of flows consisting of single packets in the DSL access, this is expected because of the greater *UDP/dns* application domination.

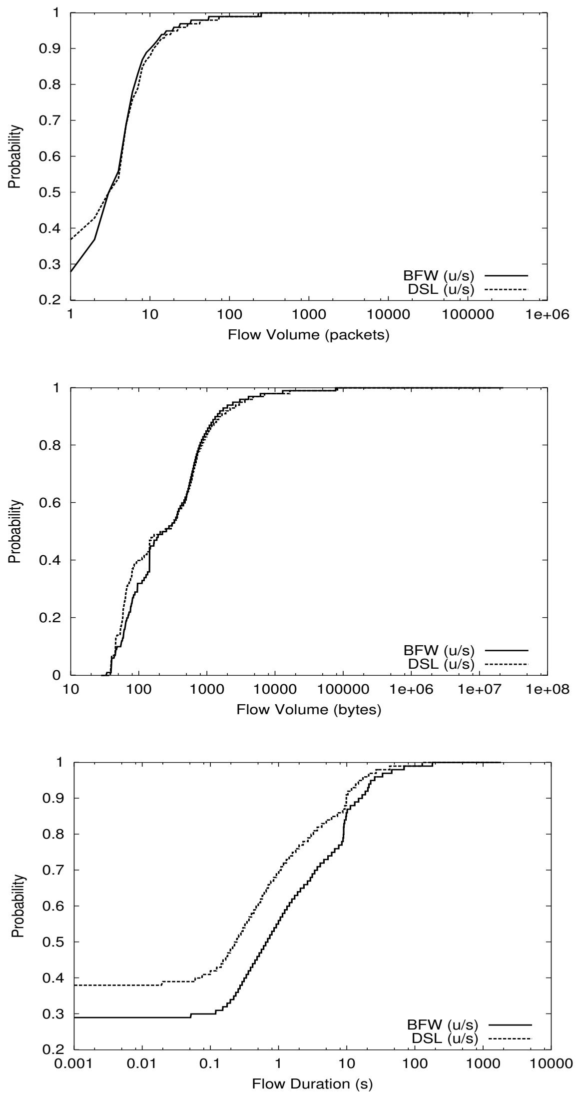

Fig. 9. Comparison of cumulative distribution of the upstream flow packet volumes, flow byte volumes, and flow durations between DSL and BFW access

DSL access consists of flows with durations shorter than that of BFW access. A possible reason again is the domination of *UDP/dns* which has 95% flows with single packets in DSL access as compared to 83% in BFW. Moreover, this application contributes to 22.49% of the total flow on DSL, as compared to 13.17% of the entire flow on BFW. Therefore 80% of the flow is equal to or less than 10 seconds in BFW, as compared to 3 seconds in DSL.

Table III presents the contribution from various applications to the total traffic upstream in terms of number of flows, bytes, and packets. The trend of the relative contribution of each of these applications is very similar in both access schemes. Note that the usage of a particular application is also dependent on the time of the day that is being observed. Therefore, the trend observed here cannot be an exact reflection of the user behavior.

Fig. 10 indicates the lag-k autocorrelation coefficient in the inter-arrival time process of both DSL and BFW. The interesting thing to note here is that the autocorrelation coefficient starts around 0.20 for lag-1 and dies quickly for DSL, therefore possibly indicating near-range correlation, while for BFW, there is no significant correlation even for lag-1. This is an interesting observation in BFW access since it was found earlier that the underlying MAC protocol causes periodicity in the inter-arrival time distribution. This issue forms an important aspect of our future work.

TABLE III

% Contribution of various Application in DSL and BFW

Application	% Bytes		% Packets		% Flow	
	DSL	BFW	DSL	BFW	DSL	BFW
UDP/DNS	0.44	0.44	1.50	1.60	22.49	13.17
FTP-DATA	0.05	0.06	0.27	0.27	0.83	0.40
SMTP	4.59	7.05	4.36	2.61	5.96	0.91
WWW	9.62	12.94	20.09	27.90	37.08	42.23
Kazaa	17.48	14.59	4.60	8.53	1.62	3.97
Gnutella	3.37	7.97	3.55	6.49	2.45	2.09
Napster	-	1.10	-	0.46	-	0.09
NNTP	-	0.80	-	4.66	-	0.12

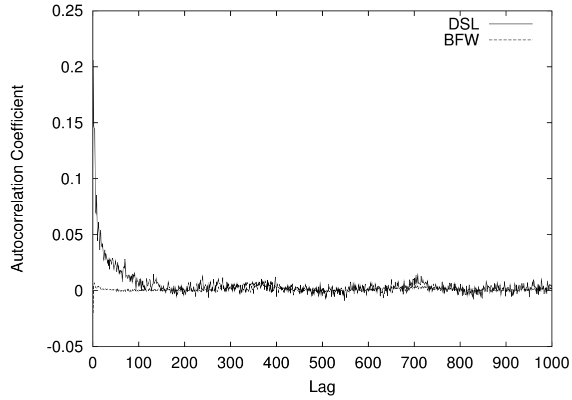

Fig. 10. Correlation for inter-arrival time

VI. Summary and Future Work

In this paper, we presented the preliminary analysis of flow-level upstream traffic behavior in Broadband Fixed Wireless and DSL access. The analysis of the aggregate flows indicate that most flows are short-lived, with the flow duration of DSL being smaller than that of BFW. We found that although *www* flows still dominate the traffic, the maximum load generated per flow is from *kazaa* and *gnutella* (P2P applications). These applications dominate both types of broadband access that was studied.

Moreover, there was a possible near-range autocorrelation observed in the arrival process on the DSL access. There was a significant periodicity observed in inter-arrival time distribution on the BFW access, which underlines the impact of its MAC protocol on the traffic behavior.

This study has brought up some interesting facts in profiling the behavior of flows on different broadband access. This analysis of upstream traffic characteristics forms the first step toward constructing a generalized parametric model for broadband access networks.

Acknowledgement

We thank Sprint Corporation for making the representative data set available for this study; in particular, we thank Nick Gerber, and Sheldon Fisher for their help.

References

[1] M. Cano, J. Malgosa-Sanahuja,, F. Cerdan and J. Garcia-Haro, "Internet Measurements and Data Study over the Regional Network Ciez@net," *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 2001.

[2] Cisco System Inc., "NetFlow Services and Applications - White Paper," `http://www.cisco.com/warp/public/cc/pd/iosw/ioft/neflct/tech/napps_wp.htm`.

[3] K. Claffy, H. Braun and G.C. Polyzos, "A Parameterizable methodology for Internet Traffic Profiling," *IEEE Journal on Selected Areas in Communications*, 1995.

[4] D. Comer, *Internetworking With TCP/IP: Principles, Protocols And Architectures*, Prentice-Hall, 1995.

[5] R. Epsilon, J. Ke and C. Williamson "Analysis of ISP IP/ATM network traffic measurements," *ACM SIGMETRICS Performance Evaluation Review*, Volume 27, 1999.

[6] J. Farber, S. Bodamer and J. Charzinsky, "Measurement and Modelling of Internet Traffic at Access Networks," *Proceedings of the EUNICE 98 Open European Summer School on Network Management and Operation*, August 1998, Munchen, Germany.

[7] M. Fullmer's flow-tools, `http://www.splintered.net/sw/flow-tools/`.

[8] J. Kilpi and I. Norros, "Testing the Gaussian character of access network traffic," *ACM SIGCOMM Internet Measurement Workshop*, 2002.

[9] B. Mah, "An Empirical Model of HTTP Network Traffic," *In Proceeding of IEEE INFOCOM*, 1997.

[10] A. Mena and J. Heidermann, "An Empirical Study of Real Audio Traffic," *In Proceeding of IEEE INFOCOM*, 2000.

[11] P. Oliver and N. Benameur, "Flow level IP traffic characterization," *Seventeenth International Teletraffic Congress ITC-17*, 2001.

[12] S. Tilakawardana, and R. Tafazolli, "Effect of Service Modeling on Medium Access Control Performance," *IEEE PIMRC'01*, October 2001.

[13] N. Vicari, S. Kohler and J. Charzinsky, "The Dependence of Internet User Traffic Characteristics on Access Speed," *In Proceedings of the LCN'2000*, November 2000, Tampa, USA.