Analysis of Impartial Quality Measurements on Indian Broadband Connections

TSRK Prasad*, *Dhruv Shekhawat†, *Sukanto Guha‡, Neena Goveas§ and Bharat Deshpande¶
Department of Computer Science & Information Systems
BITS Pilani - K K Birla Goa Campus, Goa, India - 403726
{*tsrkp, †f2012110, ‡f2012106, §neena, ¶bmd} @goa.bits-pilani.ac.in

***Abstract*—Broadband Internet user base is growing at an exponential pace in India. The market pressures are forcing the Internet service providers (ISPs) to sell their service offerings aggressively, resulting in a growing discrepancy between consumer expectations and service offerings from the ISPs. This discrepancy is often evident in experienced vs offered Internet connection service.**

We study the Quality of Service (QoS) variations for Internet connections of the Indian broadband users. We use the Network Diagnostic Test (NDT) data set provided by the Measurement Lab (M-Lab) for the study. Our study shows a significant increase in the maximum and the average Internet connection throughput values. We present a statistical summary of the four QoS parameters (throughput, jitter, latency and packet loss) for the Indian broadband connections. Our study shows a lack of diversity in the QoS characteristics of the Indian broadband connections.

I. Introduction

Broadband consumers demand high speed, low latency, low packet loss etc., i.e., a good Quality of Service (QoS) on their Internet connections; consumers are also willing to pay a premium for good QoS. The Internet service providers (ISPs) claim good QoS experience on their networks; yet dissatisfied consumers have contested the QoS claims of the ISPs. The main problem is the lack of objectivity in assessing the QoS parameters of an Internet connection.

Network measurement is an objective and scientific approach that we can apply to the problem at hand. An objective assessment of the existing situation would enable the stakeholders (customers, ISPs and regulators) to evolve some fact-based, realistic technical solutions to improve the Internet connection experience of the customers.

A simple measure of the QoS of a connection is to perform a speed test with respect to websites such as *Speedtest.net*. Researchers sometimes deploy custom software such as *iperf* to measure the performance of a connection. Results of such tests are neither comprehensive nor representative. The measurement data generated by these platforms is also not available on the world-wide web for third-party verification.

To objectively assess the difference between the claimed vs the experienced conditions on broadband connections, we utilize the data set generated by the network diagnostic test (NDT) tool hosted on the measurement lab (M-Lab) platform [1]. The M-Lab was set up in 2008 as a platform for performing active Internet measurements [2]. Participation in the project is voluntary and the measurement data is made available to researchers, regulators and general public at no cost. The M-Lab is a distributed measurement infrastructure primarily hosted on dedicated servers spread across the world. M-Lab servers collectively host measurement tools written by researchers and perform network measurements on Internet. One such tool is NDT [3], [4]; NDT is an active measurement tool fashioned in client-server paradigm. NDT helps measure the performance of a network connection. Volunteer users perform NDT test between their computer and an M-Lab server [5].

978-1-5090-2361-5/16/$31.00 © 2016 IEEE

In this work, we use the NDT data set for the years 2009–2014 to perform QoS analysis [1]. Apart from the QoS analysis for a representative sample of the Indian broadband consumers, we also provide an ISP-specific throughput analysis. Furthermore, we make inferences on QoS policies deployed on the backbone networks of ISPs. Our analysis code base is publicly available [6].

The remainder of this paper is organized as follows. Section II provides literature review. Section III-A contains description of the NDT data set. In section III-B, we analyze the NDT data set of the Indian volunteers for the years 2009-2014. In section III-C, we correlate various QoS parameters of the NDT data set and try to draw conclusions based on the correlation values between QoS parameters. Section IV lists the limitations of the NDT data set; Section V contains summary of our work.

II. Literature Review

Distributed Internet measurements are performed by prominent organizations such as *M-Lab*, *CAIDA*, *WAND*, *Route Views* and *RIPE*. Researchers take either an active or a passive approach to Internet measurements [7]. Route Views predominantly performs passive measurements where as WAND, RIPE and CAIDA perform both. Lehr et al. lists popular platforms for broadband speed measurements and the discrepancies among the competing platforms [8]. M-Lab specializes in active Internet measurements. M-Lab provides the largest collection of the Internet measurement and performance data. The M-Lab data sets are available through either cloud storage in raw format or Google cloud platform.

We selected the NDT tool for our network measurement and analysis task. NDT relies on the Kernel Instrument Set (KIS) which was developed as a part of the Web100 project [9]. The

Web100 project enables a passive per-connection monitoring of the TCP state [10]. Much of the NDT test data set comes from the KIS probes inserted into the Linux kernel.

Apart from NDT, other tools like *iperf*, *Speedtest.net* and *grenouille.com* can also perform throughput measurement. Attempts have also been made to perform customized client-side measurements with devices such as *SamKnows* and *BISMark* [11]. Among all the other tools, *Netalyzr* comes closest to NDT in functionality [12].

Quality of Service (QoS) as a network management technique has not seen wide-spread deployment across the Internet [13], [14]. Of all the QoS techniques, Differentiated Services (DiffServ) is the only standards-compliant technology that has seen widespread deployment [15]. To incentivize the deployment of guaranteed QoS across Internet, there have been proposals by Huston et al. and Norton to change the Internet's revenue sharing model from Sender Keep All (SKA) to a graded payment model; In graded payment model all the intermediaries get a cut in the revenue [13], [16]. This graded payment model has been suggested as a monetary motivation to implement guaranteed QoS for customers. We use the NDT data set to look at the ground reality of broadband customer's QoS experience.

III. Data Set and Analysis

A. Description of Data Set

The NDT data set contains test results for all the countries [2]. Country specific parts of the NDT data set have been used by researchers to create influential technical reports and policy guidelines [17].

We perform India-centric analysis on NDT data set for the years 2009-2014. Even though India-centric data set is not as large as US-centric data set, the data set is representative. The number of times a user repeats a test is verified by checking the number of repetitions in the client IP addresses. Figure 1 shows the repetitiveness of users (IP addresses) in the NDT data set for the year 2012. For the sake of compact plot, each user IP address from NDT data set has been mapped to a unique number in an ascending order number series starting at number 1; these numbers are placed along the x-axis. The number of times a client IP address appears in the data set is indicated on the y-axis. As Figure 1 clearly shows, most of the clients appear just once in the NDT data set over the sample period of one year, thus attesting to the diversity of the NDT data set.

In the sample space for the calendar year 2012, only 2.2% of the IP addresses appeared more than twice in the entire year, and 88.6% of the values appeared exactly once over the duration of the entire year.

All the major ISPs of India are present in the data set. The data points received per month are in the range of 200-500 for smaller ISPs (Ex.: In2Cable, Sify Broadband). The data points for larger ISPs like Airtel and BSNL are in the range of 2000-10000 per month. We do not consider ISPs with fewer than 200 data points over any four month period.

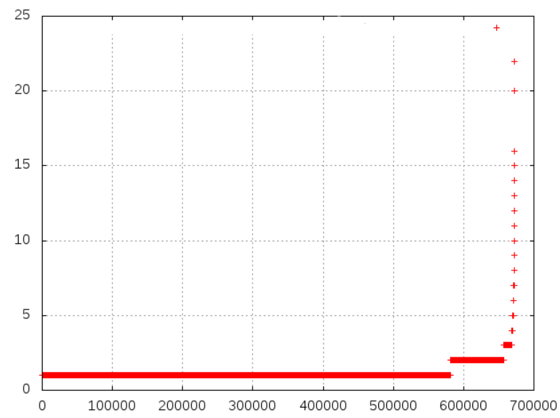


Fig. 1: Representativeness of NDT data set for the calendar year 2012. X-axis enumerates the IP addresses of the clients. Y-axis counts the number of appearances of a client IP address in the data set for the whole year (2012).

All the NDT tests performed by the M-lab platform are archived in the M-Lab data set table of the Google BigQuery database [18]. Google BigQuery database itself is hosted on the Google cloud platform. All fields of the M-Lab data set table can be classified into two categories: connection-specific information and test-specific information. The fields related to connection-specific information record the IP addresses, host names, ISP identity and geolocation information (latitude and longitude) of the client. The test-specific information fields record each of the web100 kernel instrumentation set (KIS) variables as one field in the database table. The names of the BigQuery table and the fields used in our research are listed in the Appendix.

B. Throughput Analysis

The average throughput of a broadband connection is calculated using (1) [19].

$$Throughput = \frac{DataOctetsOut}{8 * (T_{recv} + T_{cwnd} + T_{snd})} \quad (1)$$

Where,
T_{recv} = Receiver Limited Transitions
T_{cwnd} = Congestion Limited Transitions
T_{snd} = Sender Limited Transitions

The unit for throughput numbers mentioned in this paper is Mbps, unless stated otherwise.

We selected the India-specific data points from the NDT data set. Each of the data points in the NDT data set contains latitude, longitude, throughput information along with the date of measurement. We use September as the reference month for the annual comparison of the throughput metric. The India-specific NDT data points with valid latitude and longitude information are represented on the map of India. For the selected 2009-2014 time period, we get six geo-specific throughput graphs.

The generated throughput graph sequence shows stark contrast between the initial and final phases of the selected five

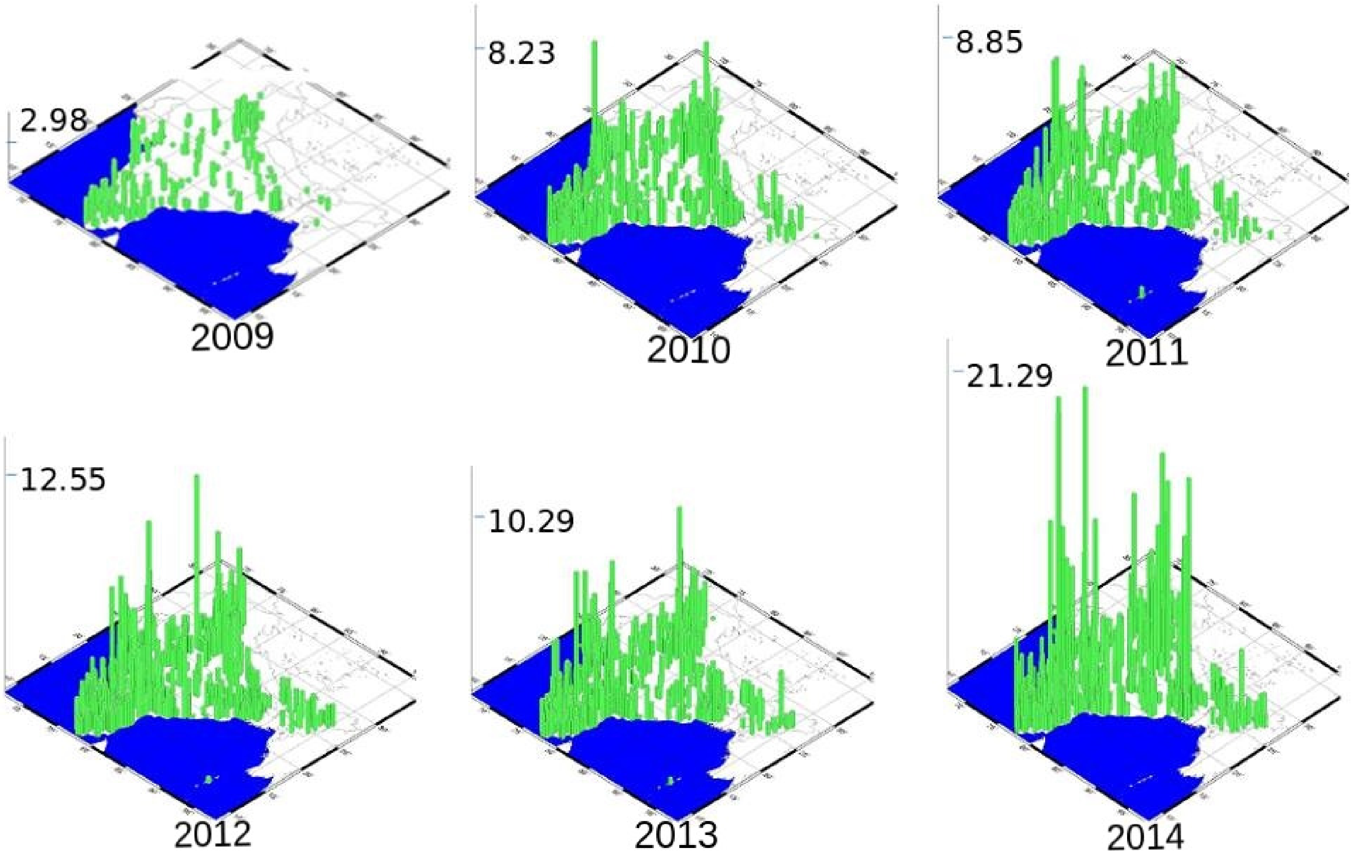


Fig. 2: Average throughput of the Indian broadband users for the month of September during the years 2009 - 2014.

TABLE I: Annual average throughput for the years 2009 - 2014.

Year	Average Throughput (Mbps)
2009	0.437
2010	0.532
2011	0.548
2012	0.644
2013	0.543
2014	0.711

year period (2009 - 2014), an example of which is displayed in Figure 2. An interesting observation over the same period is the considerable increase in the number of subscribers in non-metropolitan areas. In addition to that, we observe that the maximum throughput value rises from 2.98Mbps in September 2009 to 21.29Mbps in September 2014.

Furthermore, Table I shows the average throughput for the years 2009-2014. A steady growth has been observed during the years 2009-2012 and 2013-2014. According to the Telecom Regulatory Authority of India (TRAI), the speed of broadband is largely dependent upon three factors: bandwidth utilization, latency and contention ratio [20]. Except for four months (May-August) in 2013, we observe a month-on-month increase in average throughput during 2009 - 2014 period.

Shown in Figures 3 and 4 are the graphs for Bharti Airtel Pvt. Ltd. and the comparison charts between ten major ISPs respectively. We generate a timelapse for the period 2009 -

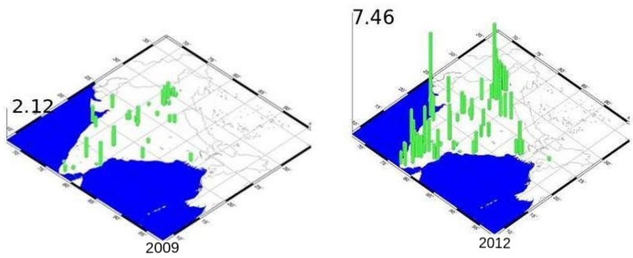
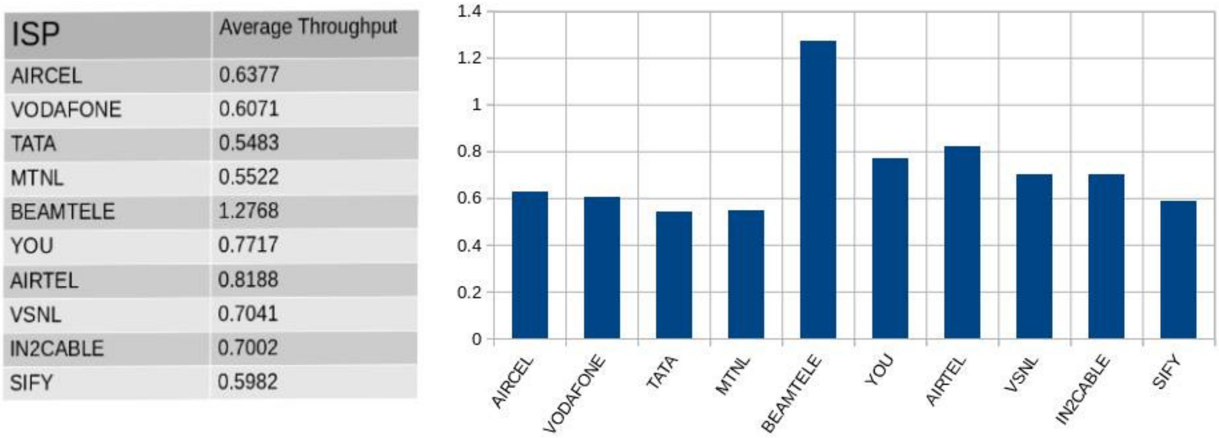


Fig. 3: Bharti Airtel : September 2009 vs August 2012.



ISP	Average Throughput
AIRCEL	0.6377
VODAFONE	0.6071
TATA	0.5483
MTNL	0.5522
BEAMTELE	1.2768
YOU	0.7717
AIRTEL	0.8188
VSNL	0.7041
IN2CABLE	0.7002
SIFY	0.5982

Fig. 4: Average throughput for different ISPs for the month of October, 2012.

2012 using monthly throughput plots of Bharti Airtel users. We also generate an all India timelapse for the 2009 - 2014 period. We observe that both the maximum value of throughput as well as the geographical spread of the test increased drastically for the country as a whole. The generated timelapse is available at [6].

The maximum throughput value saw a 250 percent increase in the case of Bharti Airtel, and other ISPs showed a similar rise. This further proves that all of India's major service providers increased their broadband service speeds during 2009-2012.

The geographical spread of test data points shows a pronounced increase in the case of Bharti Airtel and Tata Communications. Other ISPs such as Beam Telecom, MTNL and Aircel do not show a significant rise in the geographical spread.

We observe a marked increase in the number of Internet consumers in rural areas during 2009-2014. This positive outcome is due to a financial grant received by the BSNL from the Universal Service Obligation (USO) fund. The two year grant amounted to Rs. 2750 crore (approximately $500 million) and was given to support rural wireline connections [21]. We also observe a substantial increase in the maximum throughput experienced by the metropolitan users. We define the average maximum throughput as average of maximum throughput numbers obtained by clients of all the ISPs. In some metropolitan cities, the average maximum throughput shows an increase of more than 500 percent.

The increase in broadband throughput can also be attributed to policy recommendations by Department of Telecommunications (DoT) and TRAI. TRAI has revised the minimum broadband speed from 256 Kbps (as per Broadband Policy 2004) to 512 Kbps (National Telecom Policy 2012). Our analysis shows that all ISPs conformed to the guidelines set down by TRAI. Even with the above mentioned achievements, India still ranks very poorly in the global broadband speed rankings [22].

C. *Quality of Service (QoS)*

The jitter, latency and reliability values have been computed from the kernel instrumentation set (KIS) variables of the web100 project using the following equations.

$$\begin{aligned} Jitter &= RTTVar \\ Latency &= \frac{SumRTT}{CountRTT} \\ Reliability &= \frac{OctetsRetrans}{HCDataOctetsOut - OctetsRetrans} \end{aligned} \tag{2}$$

TABLE II: Statistical summary of the four QoS parameters for the month of November, 2010.

	Mean	Std. Dev.
Throughput	0.5Mbps	0.3734Mbps
Jitter	57.282ms	49.167ms
Latency	427.18ms	218.22ms
Packet Loss (Reliability)	4.62%	7.37%

We consider four Quality of Service (QoS) parameters, namely throughput, latency, jitter and reliability of a broadband connection. All the four parameters are measured separately by NDT. NDT records reliability as the percentage of packets dropped during a test, thus a smaller number indicates a more reliable connection. We tabulated the statistical summary of the four aforementioned QoS parameters in Table II. We calculate correlations between the four QoS parameters. The results are shown in Table III.

We observe from Table III that throughput has a weak negative correlation with all the other parameters. This could be because higher throughput is generally associated with better and more expensive broadband connections, and thus latency and jitter would be low and reliability high. Furthermore, we observe that even the correlation factor is constant(~ -0.3) between throughput and two other parameters, namely jitter and latency.

TABLE III: Correlation coefficients between different QoS parameters for the month of November, 2010.

	Throughput	Jitter	Latency	Reliability
Throughput	1	-0.296	-0.306	0.012
Jitter		1	0.817	-0.033
Latency			1	-0.024
Reliability				1

Another revelation is the strong correlation between latency and jitter. In India, low latency seems to guarantee timely packet arrival (low jitter) and vice versa. Though a correlation between latency and jitter is expected, it is perhaps surprising that only latency and jitter parameters are strongly linked and not reliability.

Putting all of these facts together, we can draw two conclusions. First conclusion is, there is probably a significant deployment of DiffServ across the ISP networks in India. DiffServ traffic classes make the separation of traffic into four classes possible and can also explain the strong correlation between latency and jitter [23]. The high priority packets are serviced first using expedited forwarding of Diffserv [24]; such a packet scheduling algorithm reduces the jitter experienced by high priority (i.e., low latency) packets.

Second conclusion is, throughput determines the DiffServ traffic classes. Throughput (in layman's terms, connection speed) is the most visible and marketed term in broadband connectivity. Broadband consumers who purchase high throughput connection are placed in higher DiffServ classes. Better treatment for premium customers could explain the negative correlation between throughput and the other three QoS parameters.

Reliability does not have any correlation with either latency or jitter, in fact, it is almost totally independent. In India, reliability seems to be only a weak function of the throughput, otherwise, there is really no way of guaranteeing high reliability.

IV. Limitations of the Data Set

In a significant number of cases, we found that the geographical location of the IP addresses was not available in the NDT data set, so data points from those IP addresses were not mapped. However, we noticed that the percentage of unmapped addresses decreased from 38 percent of the total number of rows in 2009 to 33 percent in 2014, and hope that this decreasing trend continues in the future. The name of ISP field in the NDT data set table is not available for some of the clients. Hence performing an ISP-specific analysis on the NDT data set doesn't always provide a comprehensive picture.

M-Lab tests have a biased user base - the users are mostly technical experts. Due to this, there is also inadequate geographical coverage of the tests. However we expect that as user awareness increases, this user profile bias will be reduced. The test results depend on the test time - connecting at non-peak times leads to significantly better results for most ISPs.

NDT requires Reno-type TCP congestion algorithms, and packet coalescing to be disabled, otherwise the test's heuristics may not be accurate. In addition, the tests are server and client dependent - the spread of users of the NDT is not uniform across all geographical and economical barriers.

V. Conclusions

We utilize the NDT data set for the years 2009-2014 to observe the patterns in broadband connectivity among Indian consumers. There has been sustained increase in the broadband penetration and connection throughput among the Indian broadband consumers.

We also utilize NDT data set to deduce QoS policies of ISP networks. Of the four QoS parameters (throughput, jitter, latency and reliability), only latency and jitter are strongly correlated. There is a general trend suggesting that a faster connection also has better jitter, latency and reliability metrics. As expected, the service providers are having a difficult time migrating to a well-nuanced QoS-enabled broadband offerings. More work is needed from all stakeholders to cultivate varied QoS approaches through diverse service level agreements (SLAs) between ISPs and customers.

VI. Acknowledgements

We thank the researchers at *Internet2* for creating the NDT tool. We thank the M-Lab for hosting the NDT tool and making the data set publicly available. We thank Nathan Kinkade and Collin Anderson of M-lab for facilitating our access to the M-lab data set.

Code Base

Our analysis code base is available for download at: *https://github.com/prasadtalasila/NDT*.

References

[1] "The MLab NDT Dataset," January 2009 - December 2014. [Online]. Available: http://measurementlab.net/tools/ndt

[2] "Google cloud platform M-Lab data set," 2014. [Online]. Available: https://cloud.google.com/bigquery/docs/dataset-mlab#schema

[3] R. Carlson, "Network Diagnostic Tool," 2014. [Online]. Available: http://software.internet2.edu/ndt/

[4] R. Carlson, "Developing the Web100 based network diagnostic tool (ndt)," in *Proceedings of the 4th International Conference on Passive and Active Measurement (PAM)*, 2003.

[5] "Network Diagnostic Test," 2015. [Online]. Available: http://www.measurementlab.net/tools/ndt

[6] D. Shekhawat, S. Guha, and P. Talasila, "NDT," 2014. [Online]. Available: https://github.com/prasadtalasila/NDT

[7] M. Crovella and B. Krishnamurthy, *Internet Measurement: Infrastructure, Traffic and Applications*. New York, NY, USA: John Wiley & Sons, Inc., 2006.

[8] W. Lehr, S. Bauer, and D. D. Clark, "Measuring Performance when Broadband is the New PSTN," *Journal of Information Policy*, vol. 3, pp. 411–441, 2013.

[9] M. Mathis, J. Heffner, and R. Reddy, "Web100: Extended TCP instrumentation for research, education and diagnosis," *SIGCOMM Comput. Commun. Rev.*, vol. 33, no. 3, pp. 69–79, Jul. 2003.

[10] M. Mathis, J. Heffner, and R. Raghunarayan, "TCP Extended Statistics MIB," RFC 4898 (Proposed Standard), Internet Engineering Task Force, May 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4898.txt

[11] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè, "Broadband internet performance: A view from the gateway," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 134–145, Aug. 2011.

[12] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, “Netalyzr: Illuminating the edge network,” in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC ’10. New York, NY, USA: ACM, 2010, pp. 246–259.
[13] G. Huston, “Interconnection, peering, and settlements,” in *proc. INET*, vol. 9, 1999.
[14] P. Ferguson and G. Huston, “Quality of service in the internet: Fact, fiction, or compromise?” *AUUGN*, p. 231, 1998.
[15] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services,” RFC 2475 (Informational), Internet Engineering Task Force, Dec. 1998, updated by RFC 3260. [Online]. Available: http://www.ietf.org/rfc/rfc2475.txt
[16] W. B. Norton, “The Emerging Video Internet Ecosystem (v0.9) EARLY DRAFT,” 2015. [Online]. Available: http://drpeering.net/white-papers/Emerging-Video-Internet-Ecosystem.html[Lastaccessed:27.09.2015]
[17] FCC, “A report on Consumer Wireline Broadband Performance in the U.S.” 2014. [Online]. Available: http://www.fcc.gov/reports/measuring-broadband-america-2014/
[18] M-Lab, “M-Lab Dataset - Bigquery Schema,” 2014. [Online]. Available: https://github.com/m-lab/mlab-wikis/blob/master/BigQueryMLabDataset.md
[19] “Documentation of variables for the Web100 TCP Kernel Instrumentation Set (KIS) project,” 2014. [Online]. Available: http://www.web100.org/download/kernel/tcp-kis.txt
[20] Telecom Regulatory Authority of India, “Bandwidth required for ISPs for better connectivity and improved quality of service,” 2009. [Online]. Available: http://www.trai.gov.in/WriteReaddata/ConsultationPaper/Document/cpaper15jan09.pdf
[21] Telecom Regulatory Authority of India, “Recommendations on Support for Rural Wire-line Connections Installed before 01.04.2002,” 2012. [Online]. Available: http://www.trai.gov.in/WriteReadData/Recommendation/Documents/Binder1.pdf[Lastaccessed:06.11.2014]
[22] Akamai Technologies, *Akamai’s State of the Internet Report*, 2015, vol. 8, no. 3.
[23] K. Nichols, S. Blake, F. Baker, and D. Black, “Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers,” RFC 2474 (Proposed Standard), Internet Engineering Task Force, Dec. 1998, updated by RFCs 3168, 3260. [Online]. Available: http://www.ietf.org/rfc/rfc2474.txt
[24] B. Davie, A. Charny, J. Bennet, K. Benson, J. L. Boudec, W. Courtney, S. Davari, V. Firoiu, and D. Stiliadis, “An Expedited Forwarding PHB (Per-Hop Behavior),” RFC 3246 (Proposed Standard), Internet Engineering Task Force, Mar. 2002. [Online]. Available: http://www.ietf.org/rfc/rfc3246.txt

APPENDIX
RELEVANT DETAILS OF BIGQUERY DATABASE

Table name: `plx.google:m_lab.yyyy_mm`
where *yyyy* is the placeholder for the year in four digit format and *mm* is the placeholder for the month in two digit format.

NDT connection-specific information
For each NDT test, the following test-specific information gets recorded.

Longitude	`connection_spec.client_geolocation.longitude`
latitude	`connection_spec.client_geolocation.latitude`
ISP Name	`connection_spec.client_hostname`
Country	`connection_spec.client_geolocation.country_code = 'IN'`
IP Address	`web100_log_entry.snap.RemAddress`
NDT Data set	`project =0`

NDT test-specific information
From here on, all the relevant test-specific information fields of BigQuery table are shown in a truncated format for brevity. The common prefix of all the fields, `'web100_log_entry.snap.'` is not shown. A proper way to reconstruct the name of a field such as `SndLimTransRwin` is to prefix the field with the string `'web100_log_entry.snap.'` to obtain `web100_log_entry.snap.SndLimTransRwin` as the field name.

Throughput
The following BigQuery fields get used in the throughput computation.

$DataOctetsOut$	`HCDataOctetsOut`
T_{recv}	`SndLimTransRwin`
T_{cwnd}	`SndLimTransCwnd`
T_{snd}	`SndLimTransSnd`

Latency
The following BigQuery fields get used in the latency computation.

$SumRTT$	`SumRTT`
$CountRTT$	`CountRTT`

Reliability
The following BigQuery fields get used in the Reliability computation.

$OctetsRetrans$	`OctetsRetrans`
$HCDataOctetsOut$	`HCDataOctetsOut`

Jitter
The jitter value is available in the `RTTVar` field.