Need, Want, Can Afford – Broadband Markets and the Behavior of Users

Zachary S. Bischof
Northwestern University

Fabián E. Bustamante
Northwestern University

Rade Stanojevic
Telefonica Research

ABSTRACT

We present the first study of broadband services in their broader context, evaluating the impact of service characteristics (such as capacity, latency and loss), their broadband pricing and user demand. We explore these relationships, beyond correlation, with the application of *natural experiments*. Most efforts on broadband service characterization have so far focused on performance and availability, yet we lack a clear understanding of how such services are being utilized and how their use is impacted by the particulars of the market. By analyzing over 23-months of data collected from 53,000 end hosts and residential gateways in 160 countries, along with a global survey of retail broadband plans, we empirically study the relationship between broadband service characteristics, pricing and demand. We show a strong correlation between capacity and demand, even though subscribers rarely fully utilize their links, but note a law of diminishing returns with relatively smaller increases in demand at higher capacities. Despite the fourfold increase in global IP traffic, we find that user demand on the network over a three year period remained constant for a given bandwidth capacity. We exploit natural experiments to examine the causality between these factors. The reported findings represent an important step towards understanding how user behavior, and the market features that shape it, affect broadband networks and the Internet at large.

Categories and Subject Descriptors

C.2.3 [**Computer Communication Networks**]: [Network Operations]; C.4 [**Performance of Systems**]: [Measurement techniques]

Keywords

Broadband access networks, User behavior, Causal inference, Natural experiments

IMC'14, November 5–7, 2014, Vancouver, BC, Canada.

ACM 978-1-4503-3213-2/14/11 ...$15.00.
http://dx.doi.org/10.1145/2663716.2663753 .

1. INTRODUCTION

As one of the most economically significant and fastest growing sectors of the Internet [19], broadband networks have attracted interest from researchers, network operators, and policy makers. Over the past decade, the number of broadband networks has increased rapidly. The latest "State of Broadband" reports that there are over 60 countries where fixed or mobile broadband penetration is above 25% and more than 70 countries where the majority of the population is online [4]. Providing broadband Internet access is known to be instrumental in social and economic development [35]. Several governments (including France, Finland and Spain) and the UN have even labeled broadband access a basic human right, similar to education and water.

While several recent and ongoing efforts have shed light on the performance and availability of broadband services [1, 2, 5, 12, 20, 28, 31, 33, 34], we lack a clear understanding of how these services are being used and how this use is impacted by the particulars of the market. The goal of our work is to examine broadband services in this broader context. How much bandwidth do people actually need? How does price affect usage? Do users in developing and developed countries impose different demands on their services? What is the impact of connection quality on usage?

We analyze over 23 months of information collected from 53,000 end hosts [30] and residential gateways [15] in 160 countries, along with a global survey of retail broadband plans [27]. We quantify the relationship between user demand on the network, retail price of available broadband services and the performance of the service to which the user subscribes. We observe a law of diminishing returns in the relationship between broadband capacity and the average/peak demand users put on their broadband link, implying that adding extra capacity on an already wide broadband line leads to a minor increment in user demand.

Looking at the longitudinal data, we find, somewhat surprisingly, that despite the fourfold increase in global IP traffic over the past five years [10], subscribers' demand in the same bandwidth capacity class remained constant, indicating that users "jump" to a higher service when their demand grows, rather than fully utilize their existing pipe. We study in depth the service upgrade dynamics and report our findings.

In a study such as this, controlled experiments are not feasible for studying the features of interest at scale. A key contribution of our work is thus a methodology for combining broadband measurement and retail price datasets along with the application of natural experiments to get to a problem otherwise impossible to tackle. We use natural experiments to examine the interaction between price, the quality of services available, and users' demands. We show that higher broadband prices increase demand when comparing users of similar capacities across markets. Additionally, we find that very high packet loss rates (over 1%) and latencies (above 500 ms) result in significantly lower usage.

Our study offers several insights on the interplay between user demand and broadband market features that are of value to the research community, network operators and policy makers. For network operators, an understanding of how user behavior changes with the network and broadband market can better inform network planning and operation. For policy makers, the work provides a firmer statistical footing for discussions on broadband incentives.

The rest of the paper is organized as follows. In Sec. 2, we explain our analysis methodology, describe our datasets, and summarize the performance of the broadband connections seen in our global dataset. We explore the impact of capacity on demand in Sec. 3, followed by a longitudinal study that investigates how demand changes over time in Sec. 4. In Sec. 5, 6, and 7 we study the impact of the price of broadband access, the cost of increasing capacity, and the connection quality on user demand, respectively. We then review related work in Sec. 8, summarize and discuss our findings in Sec. 9, and conclude in Sec. 10.

2. ANALYSIS METHODS AND DATASETS

In the following paragraphs we describe the three datasets we rely on for analysis, including a summary of key characteristics for the broadband connections they capture. We close the section with a brief discussion of the goals and methodology of our study.

2.1 Datasets

Our study builds on three datasets, two of broadband connections including: (*i*) measurements from residential gateways in the US, and (*ii*) detailed end-host collected data on broadband connections from around the world, and (*iii*) a compilation of retail broadband connectivity plans made available by Google [27]. We describe each of these in the following paragraphs.

Residential gateway data. Since 2010, the FCC, in collaboration with SamKnows, has publicly shared data collected from residential gateways distributed to broadband users around the US as part of the "Measuring Broadband America" effort [15]. Users that participate in this study were either selected to participate by their ISP or signed up through SamKnows' website. The data collected from these gateways includes measurements of link capacity, latency and packet loss as well as hourly recordings of the number of bytes sent and received over the WAN link.

End host data. Our end-host collected dataset comes from Dasu [30], a previously released network experimentation and broadband measurement client. Dasu is available as both an extension to BitTorrent and as a standalone client. As an incentive for adoption, Dasu informs users of their ISP's performance, providing detailed information on their home network configuration, the volume of network traffic sent and received by the localhost, the volume of detected cross traffic in the home network, and the results of performance measurements on their ISP (e.g. a comparison of their ISP's web browsing and DNS performance). Dasu records network usage data from the localhost and home network to account for cross traffic during characterization or the execution of network experiments.

Since its release, Dasu has been installed by over 100,000 users in over 160 countries, with the majority of clients using the BitTorrent extension. From this dataset, we select users that either have UPnP enabled on their home gateway device or those that were directly connected to their modem (thus their machine is the only device generating traffic). UPnP-enabled gateways provide byte counters that we use to measure activity on the link, taking into account issues with UPnP counters raised in other works [11, 29]. For users directly connected to their modem, we use byte counters available from `netstat` to monitor network usage (available by default on most popular operating systems). Traffic byte counters are collected at approximately 30 second intervals with some variations due to scheduling.

As it is the case with all observational studies, there is a concern about potential biases in our datasets, coming either from P2P or SamKnow's users (e.g., uniquely demanding users, early-adopters or "geek-effect" [5, 20, 25]. We account for some of these issues throughout our analysis by, for instance, focusing on measurements gathered when users are not actively downloading/uploading content on BitTorrent, restricting our users to those directly connected to a modem or wirelessly connected to a UPnP-enabled one, using neighbor matching with a caliper to ensure close matches. On the potential biases with our P2P users' data, we show in Sec. 3.1 that the average demand of Dasu users in the US – when not actively using BitTorrent – is comparable to that of participants in the FCC's study.

Connectivity plans. Our third dataset is a compilation of international retail broadband connectivity plans, made available by Google on their "Policy by the Numbers" blog [27]. This data was compiled by Communications Chambers, a consultant group, by visiting the websites of broadband service providers around the world. The dataset covers 1,523 service plans across 99 countries. It includes information on the upload and download speeds of each plan, the monthly traffic limits, and monthly cost in the local currency. We selected this dataset over those provided by the FCC, OECD, or ITU given the breadth of countries included and the depth of plans listed. The FCC and OECD datasets focus on the US and members of the OECD while the ITU dataset only includes a single service plan for each country. In a few cases, we expanded this dataset by manually visiting the websites of ISPs in countries where we had users but no broadband price data.

To directly compare the price of broadband plans across different economies, we convert the monthly cost to US dollars. We account for differences in relative purchasing power in each country by using the purchasing power parity (PPP) to market exchange ratio. In most cases, this is included in the broadband service survey provided by Google. When that is not the case, we use publicly available data from the International Monetary Fund's website[1]. All monetary figures throughout this work are normalized by purchasing power parity, including the GDP per capita data provided by the International Monetary Fund that we use later in our case study.

[1]International Monetary Fund. http://www.imf.org/

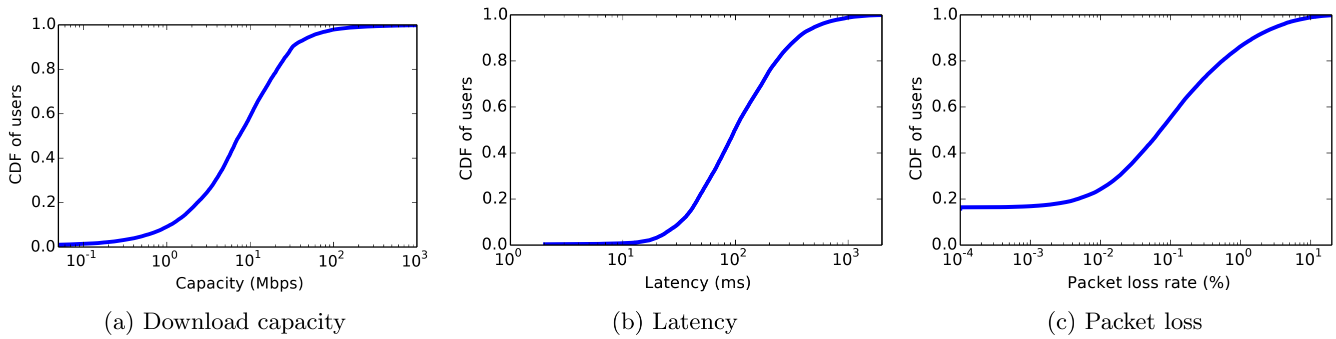


Figure 1: CDFs of the maximum download capacities, average latency to nearest available measurement server, and average packet loss rates measured for every network connections used throughout our analysis.

2.2 Broadband Networks Characteristics

We now describe the diversity of broadband connections in our global dataset, presenting distributions of their measured capacity, latency and packet loss. All the Dasu data were collected by running M-Lab's Network Diagnostic Tool (NDT) [23] within Dasu. NDT reports the upload and download capacity of a connection, as well as its end-to-end latency and packet loss rates.

Capacity. Figure 1a shows a CDF of the maximum download capacities, in Mbps, measured over each user's connection in our dataset. Our distribution has a median user download capacity of 7.4 Mbps and interquartile range of 14.3 Mbps (from 3.1 Mbps to 17.4 Mbps). Approximately 10% of users have download capacities below 1 Mbps, while the top 10% of users have capacities above 30 Mbps.

Latency. For latency, we measured the average latency to the closest NDT measurement server. Since measurement servers are hosted in a diverse set of networks of content providers (e.g. Google) and content distribution networks (e.g. Level 3), we believe such measurements provides a reasonable estimate of the latency to popular content. Figure 1b shows the distribution of measured latencies. We find that a "typical" user in our dataset has an average RTT of about 100 ms to the nearest NDT servers. The top 5% of users had an average latency above 500 ms. Based on the organization names that we found via `whois` lookups, the majority of connections with very high latencies appeared to be connecting over wireless modems or satellite providers.

Packet loss. Figure 1c shows the distribution of average packet loss rates reported by NDT tests. While the loss rate is relatively low for most users (less than 0.1%), approximately 14% of users saw an average loss rate above 1%. For the top 1% of users, average loss rates were above 10%. As was the case with high latency connections, the organization names of networks with very high packet loss rates indicated they were satellite or wireless (e.g. WiMAX, cellular) services.

2.3 Methodology

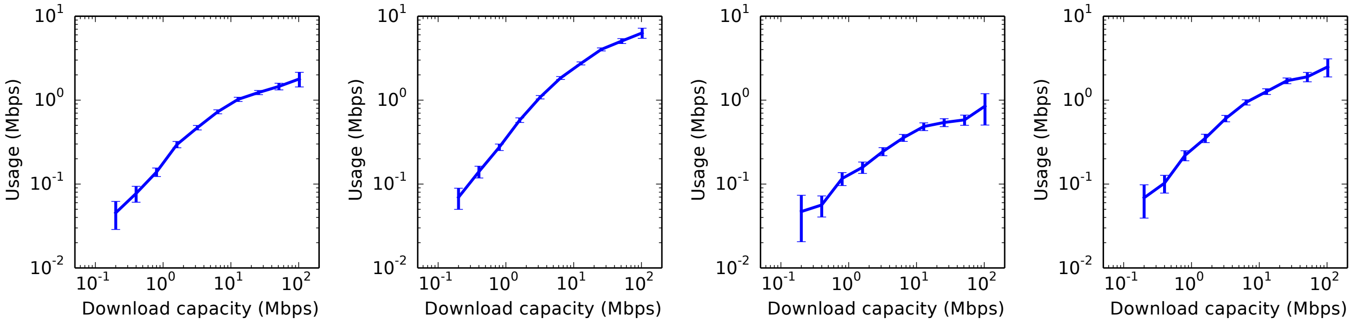
The main goal of our study is to provide insight into the impact of broadband service market characteristics on network usage. Specifically, we study the impact of the following market features: connection capacity, the price of broadband access, the cost of increasing capacity, and connection quality. While there are many other variables that can affect user behavior, this set covers the key characteristics of broadband service markets. Given the rapid pace of development in broadband and the reported growth in network traffic, we also conduct a longitudinal analysis of user demands on broadband services.

Beyond gathering a sufficiently large and diverse perspective of broadband connections, a key challenge for a macroscopic study such as ours is the nature of experiments one is able to conduct. Classical controlled experiments – where subjects in the study are randomly assigned to "treated" and "untreated" groups for comparison – are clearly not feasible at a global scale. It is also unlikely that the features we explore are independent, e.g., one would assume that price or service diversity can impact capacity and service quality. This has been a long, well understood problem in a range of fields, from epidemiology to sociology and economics. We address this challenge, as many studies do in those domains, by resorting to natural experiments in our analysis [14].

By using natural experiments [14] and related study designs, we remedy the fact that we cannot control the application of a treatment, Matching users in our treated group with similar users in the untreated group we simulate random or as-good-as-random assignment, manually ensuring that differences are evenly distributed between the two groups. This allows us to infer whether or not the relationship observed are likely to be causal. For example, to test if bandwidth capacity affects user demand, we pair users that are similar in terms of connection quality and broadband market. We then check if the user with higher capacity generates more traffic. If so, our hypothesis holds true for that pair. After testing this for each pair of similar users, we calculate the percentage of times that our hypothesis is correct.

If neither of the two variables under study – in this example, capacity and demand – have an impact on the other, then their interaction would be random. In our example this would mean, for instance, that lower capacity will result in lower (or higher) demand about 50% of the time. Significant deviations from this would suggest that a causal relationship is likely to exist between the two.

We use the one-tailed binomial test to measure the statistical significance of deviations from the expected distribution. As is common in many study designs, we consider a p-value that is less than 0.05 to be a strong presumption



(a) Mean
w/ BT
($r = 0.870$)

(b) 95th %ile
w/ BT
($r = 0.913$)

(c) Mean
no BT
($r = 0.885$)

(d) 95th %ile
no BT
($r = 0.890$)

Figure 2: Volume of download traffic generated by users versus their download link capacity. Users are grouped by their download capacity and each bin is averaged. The error bars represent the 95% confidence interval of the mean. In each case, usage is strongly correlated with link capacity.

against the null hypothesis (H_0). One potential issue with our application of the binomial test in this context is the known problem that given a large enough dataset, the test will consider even minor deviations to be significant. That is, with a large enough sample of throws, an unbiased coin could fail to pass a χ^2 test for fitting the predicted binomial distribution [26]. To account for this issue, we only consider deviations larger than 2% to be practically important. In other words the hypothesis holds at least 52% of the time.

3. IMPACT OF CAPACITY

The interplay between broadband service characteristics and user demand is complex [36]. For instance, while subscribers cannot directly affect the cost of their service, they have some freedom in choosing what package (capacity) they purchase and how much traffic they generate. On the other hand, although they come with needs and budgets when choosing a broadband plan, once acquired, their usage patterns are shaped by their selection. In addition, there is the potential impact of seemingly irrational and biased choices by subscribers [16, 32] that complicates any attempt at understanding and analytical modeling of the drivers of users' choices and demand. While we (or even most customers [17]) may not know the advertised service of a broadband connection, our study focuses on the impact of the actual maximum capacity provided to the user.

In this section, we begin to empirically explore the complex interactions between broadband service market features and user behavior by first studying the effects of capacity on user demand. When appropriate, we compare data collected from end hosts (via Dasu) and residential gateways (FCC/SamKnows).

3.1 Capacity vs. usage

We first explore the relationship between access link capacity and the demand users generate on the access network. To describe user demand, we rely on two metrics of usage: the average and peak volume of traffic generated. We define peak as the 95th-percentile value of the time series (sampled every 30-secs) of downlink demand for each user.

Figure 2 presents both the mean and peak demand, for different classes of users based on their measured downlink capacity. Given the range of services across the different markets we analyze, we split services into ten classes where every user in class (k) has a download capacity in the range of $(100Kbps * 2^{k-1}, 100Kbps * 2^k]$. We analyze usage both throughout the entire measurement period (Fig. 2a and 2b) and during periods when users are not actively uploading or downloading content on BitTorrent (Fig. 2c and 2d).

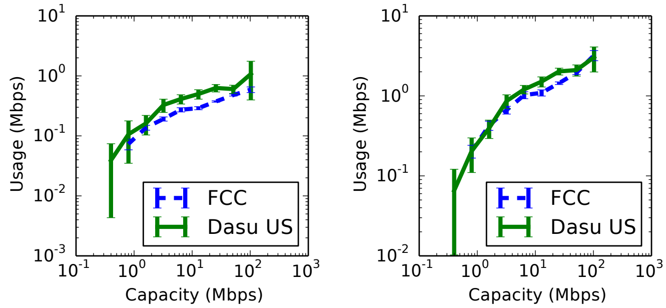
We also contrast Dasu' end-host collected data with that of users in the FCC study (gateway collected data), looking both at average and peak network usage. Figure 3 shows the mean and peak (95th percentile) demand of users, grouped by capacity, in the FCC and US-based Dasu dataset (when not using BitTorrent). Although the average usage is slightly higher for Dasu users, the peak usage is nearly identical for both groups. The difference in average demand is likely due to the fact that the FCC data is collected evenly throughout the 24-hour period, while Dasu usage (and thus our data) is partially biased towards peak usage hours.

We find that usage grows with capacity, as the plots of Fig. 2 and 3 clearly show. This is *despite* the fact that users rarely utilize their link (even at the 95th percentile, average utilization ranges between 10 and 48%). For both mean and 95th percentile usage, with and without BitTorrent traffic, we find that usage is strongly correlated with the group's link capacity ($r \geq 0.87$ for each).

Figure 2 also show that as capacity increases, usage begins to level off (related to the findings in Sundaresan et al. [34]). This would suggest that the relationship follows a law of diminishing returns: the relative increase in demand is greater for lower capacity connections than for higher capacity connections.

3.2 Inferring causality

While the access capacity and the demand users generate are strongly correlated, inferring the causality between these variables is significantly more challenging. For instance, demand might drive capacity (i.e. users sign up for faster services because they have access to applications with higher bandwidth needs, such as HD video streaming) or be driven by it, with users changing their behavior when given a higher



(a) Mean
($r = 0.915$)

(b) 95th %ile
($r = 0.905$)

Figure 3: Mean and peak (95th percentile) download traffic generated for FCC gateway users and Dasu users within the US when not using BitTorrent. The error bars represent the 95% confidence interval of the mean.

Metric	% H holds	p-value
Average usage	66.8%	1.94×10^{-25}
Peak usage	70.3%	1.13×10^{-36}

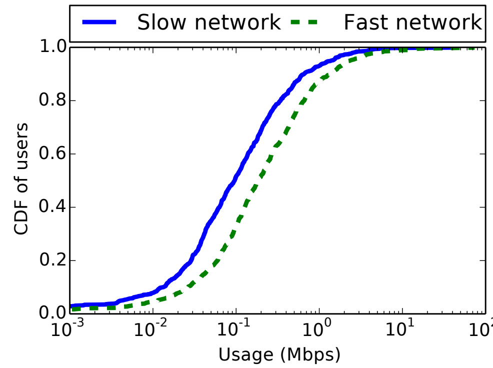
Table 1: Percentage of the time that an individual user's average and peak demand will increase when moving to a network with a higher capacity. In both cases, the control group is their behavior on the slower network and the treatment is their behavior on the faster network.

capacity and generating a higher demand. Additionally, there may be other factors that affect user demand such as the quality of the connection or the price of access.

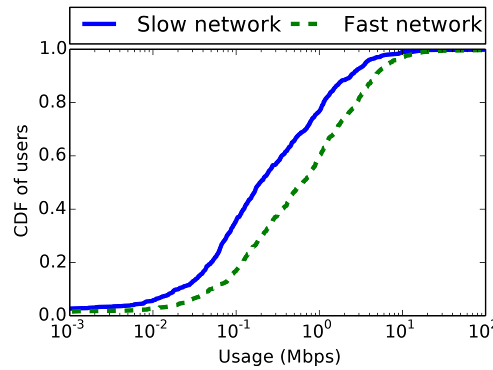
To explore a causal relationship between access capacity and demand, we first design a natural experiment to see if the behavior of individual users changes when switching between networks of different capacities. This let us test the idea that when given a higher capacity link, users will increase their demand on the network. We then compare the demand of users that are similar in terms of price of broadband access, cost to upgrade, and link quality but differ in terms of service capacity.

User upgrades. To determine if their relationship between capacity and demand is causal, we need to account for differences in usage patterns between different users. We do this by looking at how individual users change their network demand when switching to faster services, allowing us to determine if the relationship between capacity and demand is likely causal.

Figure 4 presents the CDFs of mean and peak download link usage for users switching between a "slow" and "fast" network. Both average and peak volume of traffic are when the client is not active on BitTorrent. Each network is identified by a tuple *(ISP name, network prefix, geolocated city)*. For both average and peak demand, we see that usage tends to be considerably higher on the faster network. For example, at the median, average usage doubles from 95 kbps to 189 kbps and peak usage (95th percentile) more than triples, from 192 kbps to 634 kbps.



(a) Mean



(b) 95th %ile

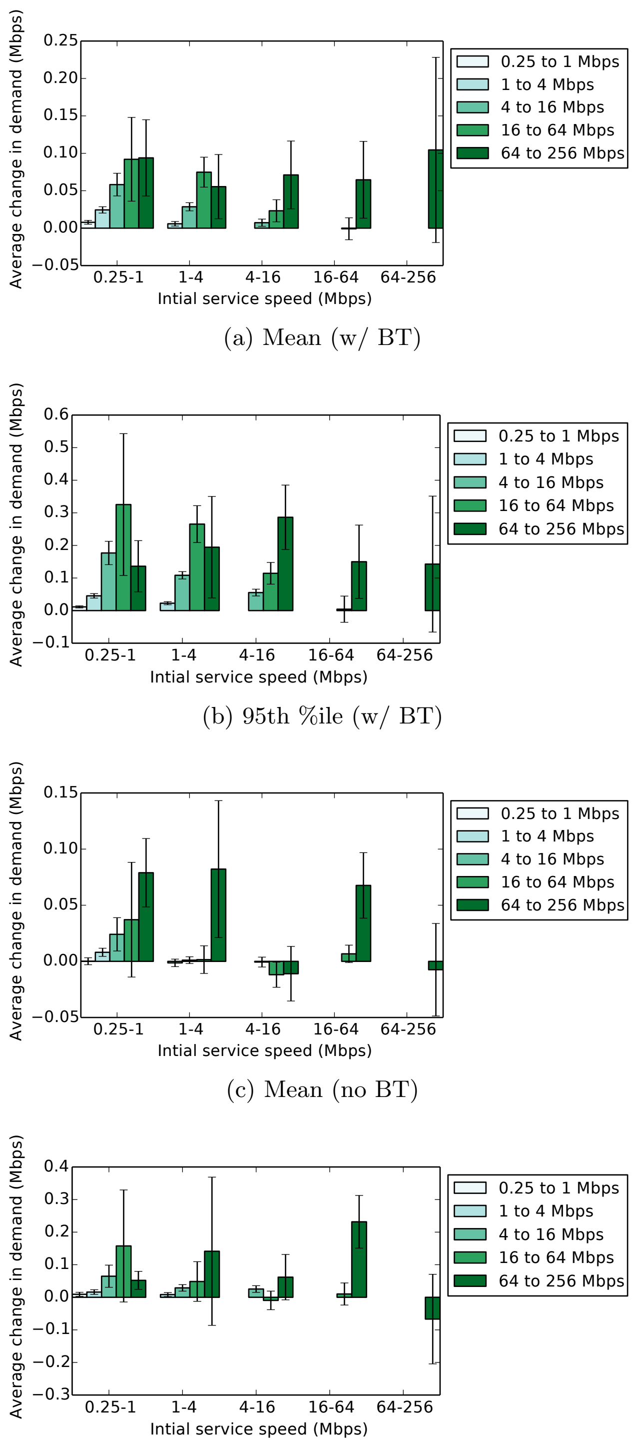
Figure 4: CDFs of the mean and peak download link usage for individual users on "slow" and "fast" networks when not using BitTorrent.

To validate this assertion we use a natural experiment. Our hypothesis (H) is that when a user moves from a slower to a faster service, demand will increase. As such, our null hypothesis (H_0) is that demand will not be affected by a change in capacity. We test this assertion for both the mean and peak demand and present our results in Table 1. As the table shows, our original hypothesis (H), is true 66.8% of the time when comparing average demand and 70.3% of the time for peak demand. For both metrics, we find very small p-values, leading us to reject the null hypothesis (H_0) that capacity does not affect the demand of individual users.

For these analysis we limit our data to that collected while the users were not generating BitTorrent traffic. Including BitTorrent traffic, we find an even higher increase in usage, and so is the percentage of the time that our hypothesis holds true. This is likely due to the fact that users are more likely to saturate their link for extended periods of time when using BitTorrent [9].

Impact of switching services. To further understand the interaction between capacity and demand, we explore the impact of service upgrades on demand, for different initial capacities of connections. Figure 5 shows the average change in demand, grouping upgrades by the "before" and "after" download capacities. The labels on the x-axis represent the capacity range of the initial service and each bar is the average change in demand when switching to a faster service in the respective capacity tier.

As the figure shows, for each metric, demand clearly increases when upgrading from slower services, particularly when looking at peak (95th percentile) usage. Increases in



(a) Mean (w/ BT)

(b) 95th %ile (w/ BT)

(c) Mean (no BT)

(d) 95th %ile (no BT)

Figure 5: Change in volume of traffic generated when switching to a faster connection. The x-axis corresponds to the initial service speed while each bar represents the average change for users switching to a faster service within that group. The error bars represent the 95% confidence interval.

Dasu data			
Control Group (in Mbps)	Treatment Group (in Mbps)	% H holds	p-value
$(0.1, 0.2]$	$(0.2, 0.4]$	75.2%	5.81×10^{-11}
$(0.2, 0.4]$	$(0.4, 0.8]$	63.4%	2.21×10^{-7}
$(0.4, 0.8]$	$(0.8, 1.6]$	59.9%	8.01×10^{-8}
$(0.8, 1.6]$	$(1.6, 3.2]$	59.3%	1.11×10^{-8}
$(1.6, 3.2]$	$(3.2, 6.4]$	53.3%	0.0166
$(3.2, 6.4]$	$(6.4, 12.8]$	57.5%	0.00707
$(6.4, 12.8]$	$(12.8, 25.6]$	56.8%*	0.0583
$(12.8, 25.6]$	$(25.6, 51.2]$	52.9%*	0.310
$(25.6, 51.2]$	$(51.2, 102.4]$	51.0%*	0.462
FCC data			
Control Group (in Mbps)	Treatment Group (in Mbps)	% H holds	p-value
$(0.4, 0.8]$	$(0.8, 1.6]$	66.4%	0.000223
$(0.8, 1.6]$	$(1.6, 3.2]$	58.1%	4.70×10^{-05}
$(1.6, 3.2]$	$(3.2, 6.4]$	56.2%	0.000487
$(3.2, 6.4$	$(6.4, 12.8]$	55.1%	0.00236
$(6.4, 12.8]$	$(12.8, 25.6]$	58.5%	2.54×10^{-7}
$(12.8, 25.6]$	$(25.6, 51.2]$	61.2%	6.76×10^{-17}
$(25.6, 51.2]$	$(51.2, 102.4]$	64.7%	0.00161

Table 2: Percentage of the time that increased capacity will increase demand when comparing similar users and each experiment's corresponding p-value. An asterisk denotes that a result was not statistically significant.

demand are less consistent when switching between already fast services, particularly above 16 Mbps, where there is a large variance on demand growth with capacity. In some cases, the large range in the 95% confidence interval shows that the upgrade likely had no significant impact on usage. These findings suggest that while capacities do drive demand, this is only true up to a certain point.

All users. We expand our comparison to all users in the datasets and use a matching study design to test the impact of increased capacity. As before, we place users into one of k bins, where $k = (100Kbps * 2^{k-1}, 100Kbps * 2^k]$. We then compare the usage of users in bins k and $k+1$. Our hypothesis (H) is that users in the "treated" group, $k+1$, will have a higher demand on the network due to their increased capacity. Our null hypothesis (H_0) is that the relationship is random and increased capacity will not result in higher demand.

To compare users from each group, however, we must ensure that each pair of users is similar in terms of connection quality (packet loss and latency), price of broadband access, and cost to upgrade capacity. For this and the remaining studies, we use nearest neighbor matching to pair similar users in "control" and "treatment" groups. We use a caliper to ensure that dissimilar users are not matched, requiring that users be within 25% of each other for each confounding factor. This means, for instance, that users with latencies of 50 and 62 ms and in regions where broadband Internet access costs \$25 and \$30 (USD) per month are considered to be sufficiently similar in terms of latency and cost of broadband access. Note the trade-off here, a tighter caliper will yield a potentially more accurate comparison, but will also reduced the number of comparisons we can perform.

Table 2 shows the experiment's results separated by the datasets used. For the Dasu data, increased capacity has the widest impact when comparing slower service groups. The increase in demand is statistically significant while comparing groups of users with capacities less or equal to 6.4 Mbps (though the achieved p-value when comparing groups $(6.4, 12.8]$ and $(12.8, 25.6]$ is very close to 0.05). When comparing users in bins above 12.8 Mbps, the difference tends to become random and our hypothesis holds about 50% the time. These results suggest that increasing capacity beyond ≈10 Mbps is less likely to have a significant impact on peak user demand.

For the FCC data, increased capacity tends to result in increased demand across all bins. We believe that this is largely due to the fact that the FCC vantage point set is comprised solely of users in the US, where higher capacity broadband services are available, but at a moderately higher price (this does not apply in many of the countries in our study). We also observed a similar trend of increased usage when studying Dasu users in the US, as we will show in Sec. 5.

4. LONGITUDINAL TRENDS IN USAGE

The last few years have witnessed a rapid growth on the capacity, coverage and affordability of broadband networks [4]. Concurrently, the volume of digital content and total IP traffic continue to grow at rapid pace. A recent Cisco report states that the total IP traffic has increased 18-fold since 2.4 exabytes in 2005 [10]. Meanwhile, the size of the "digital universe", the total amount of data created and replicated reported to be 2.8 zettabytes in 2012, doubles in size about every two years [18]. In this section, we look for changes in demand over time to see if these changes are reflected, and in what manner, in the network demand of broadband users.

To this end, we carry a longitudinal analysis of broadband connections in our dataset. We compare changing trends in usage relative to capacity, both average and peak, between 2011 and 2013. Figure 6 shows average and demand over this period, with and excluding BitTorrent traffic.

Trends in relative use are surprisingly different than what we expected. Despite the fourfold increase in global IP traffic, we find that subscribers' demand on the network remained constant at each speed tier. While we note a slight increase in demand for users with very fast connections (about 100 Mbps), overall the demand within the same broadband class has remained fairly consistent throughout the observation period.

Using a natural experiment, we are unable to find any significant change in demand at any given speed tier between 2011 and 2013. It would appear that growth in traffic is likely due to an increase in the number of broadband subscriptions and the increased capacity of available services.

5. PRICE OF BROADBAND ACCESS

This section examines the impact that the price of broadband access has on user demand. In recent years, we have seen significant growth in the number of people accessing the Internet [4]. While increased affordability has played a critical role in this growth, the price of broadband Internet access remains unaffordable in many parts of the developing world. In countries like Iran and Botswana, a 1 Mbps plan could cost as much as $150 USD per month, after accounting for purchasing power parity (PPP).

Control Group	Treatment Group	% H holds	p-value
($0, $25]	($25, $60]	63.4%	8.89×10^{-22}
($0, $25]	($60, ∞)	72.2%	5.40×10^{-10}

Table 3: Percentage of the time that increased price results in increased usage for pairs of similar users and corresponding p-values.

Contrast this with countries like Germany, Japan, and the US, where a 1 Mbps plan (or faster) are available for less than $25 per month.

We have seen how, up to a point, demand increases with capacity. If price is a factor that affects a customers' decision when selecting a broadband plan, then we would expect that higher prices will result in users signing up for lower capacity services despite their needs. Similarly, if two services with similar speeds are available at different prices in two markets, we would expect that the service in the more expensive market would experience higher network demand since subscribers are willing to pay more for it.

We design the following study to test this idea. We define our hypothesis (H) such that users in markets where broadband Internet access is more expensive will have higher demands on the network than users in less expensive markets. Our null hypothesis (H_0) then, is that increased price does not have an affect on network demand.

For this experiment, we first need to group users based on price of broadband access in their region. We define the price of broadband access in a country as the monthly cost (USD PPP) of the cheapest service with a capacity of at least 1 Mbps. We grouped users by the cost of broadband access using the following bins: less than $25 per month, between $25 and $60 per month, and over $60 per month. Users in countries such as Germany, Japan, and the US fit in the first bin (< $25 per month). Countries such as Mexico, New Zealand, and the Philippines had prices between $25 and $60 per month, while prices in counties such as Botswana, Saudi Arabia, and Iran were above $60 per month.

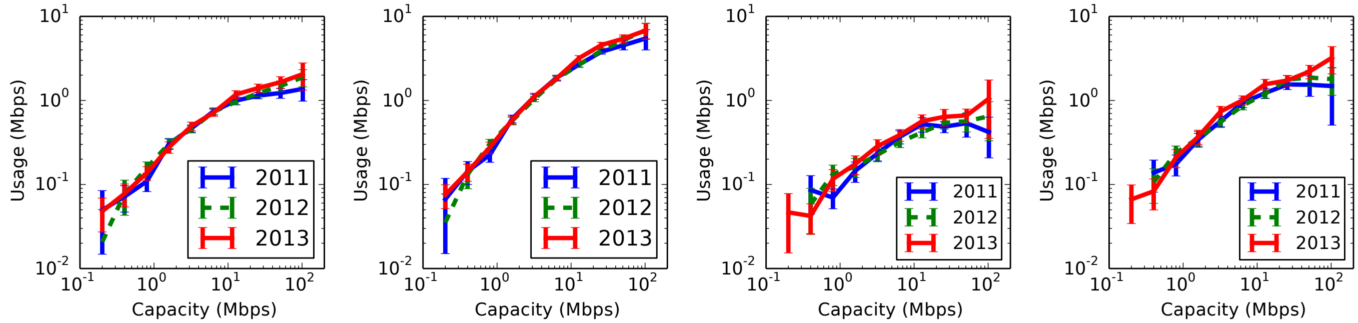
After placing users into groups based on the monthly price of broadband access, we compared the demand of otherwise similar pairs of users in each group. In these experiments, we use peak usage (when not active on BitTorrent) to measure demand.[2] For this experiment, users are "treated" with an increased cost, which our hypothesis says will increase demand. The results are shown in Table 3. We find that indeed, as price increases, more users have a higher demand than those with a similar connection where access is cheaper.

Case study. We now illustrate the impact that price can have on usage with a concrete example using four markets: Botswana, Saudi Arabia, the US and Japan. We selected these four as examples of the diversity of markets in our dataset.

We chose Botswana and Saudi Arabia given that both countries were among those with the highest broadband access costs, but differed in terms of typical service capacities. Since its independence Botswana has enjoyed one of the highest GDP growth rates in the world.[3] In recent years, the country has seen rapid growth in the percentage of citizens

[2]Results of experiments with and without BitTorrent for both average and peak demand were all comparable.

[3]CIA World Factbook. `http://www.cia.gov/library/publications/the-world-factbook/geos/bc.html`



(a) Mean (w/ BT) (b) 95th %ile (w/ BT) (c) Mean (no BT) (d) 95th %ile (no BT)

Figure 6: Peak and average usage versus capacity, grouped by year. The error bars represent the 95% confidence interval.

Country	Number of users in dataset	Median capacity (Mbps)	Nearest tier (Mbps)	Price in USD (PPP)	Annual GDP per capita (PPP)	Cost of Internet access as percentage of monthly GDP per capita
Botswana	67	0.517	0.512	\$100	\$14,993	8.0%
Saudi Arabia	120	4.21	4	\$79	\$29,114	3.3%
US	3759	17.6	18	\$53	\$49,797	1.3%
Japan	73	29.0	26	\$37	\$34,532	1.3%

Table 4: The "typical" price of broadband in each country. The "Median capacity" column lists the maximum download capacity for the median user. We then matched the median capacity with the nearest speed tier in our set of Internet services available in a country. The "Price" column shows the price of that service (converted into US dollars using the purchasing power parity conversion factor). This price is used to calculate the monthly cost in each country as a percentage of monthly income.

with access to the Internet (from 3% in 2005 to 12% in 2013[4]). The cost of Internet access in Botswana, however, remains comparatively high. A 1 Mbps service, including a phone line, from Botswana Telecom costs about \$150 per month after accounting for purchasing power parity. In contrast, a 1 Mbps service in the US would cost about \$20 per month.

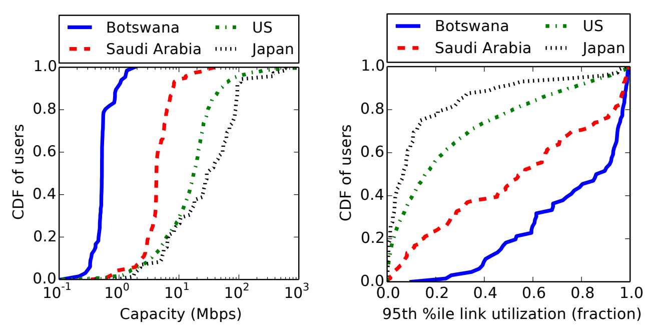
Over the past decade, Saudi Arabia has also experienced rapid growth in both GDP per capita (PPP) and the number of Internet subscribers. The percentage of the population using the Internet has tripled from just under 20% in 2007 to over 60% in 2013. However, according to the ITU only about 5% of the population with broadband subscriptions are on services faster than 10 Mbps (we see a similar percentage in our global dataset). A 1 Mbps connection is also relatively expensive in Saudi Arabia at about \$60 USD (PPP) per month, three times higher than a similar service in the US.

We include the US in our study as it presents another interesting case as one of the most diverse broadband service markets in terms of the available download capacities (from about 1 Mbps to over 100 Mbps). Japan, on the other hand, is one of the markets with widest availability of high-end broadband services. While the range of broadband service prices are similar to those in the US market, a larger fraction of users in Japan subscribes to high capacity services.

Table 4 summarizes the users and services seen in each market. We calculate the "typical" price of broadband in each of the country by matching the median capacity to the nearest service in our dataset. Compared to the US and Japan, customers in Botswana and Saudi Arabia are paying much more for slower services, especially as a fraction of monthly GDP per capita. Users in both Japan and the US appear to spend a similar fraction of monthly GDP per capita (1.3%). However, ISPs in Japan offered higher capacities at the same fraction of monthly income. As a result, users in Japan were more likely to subscribe to faster services.

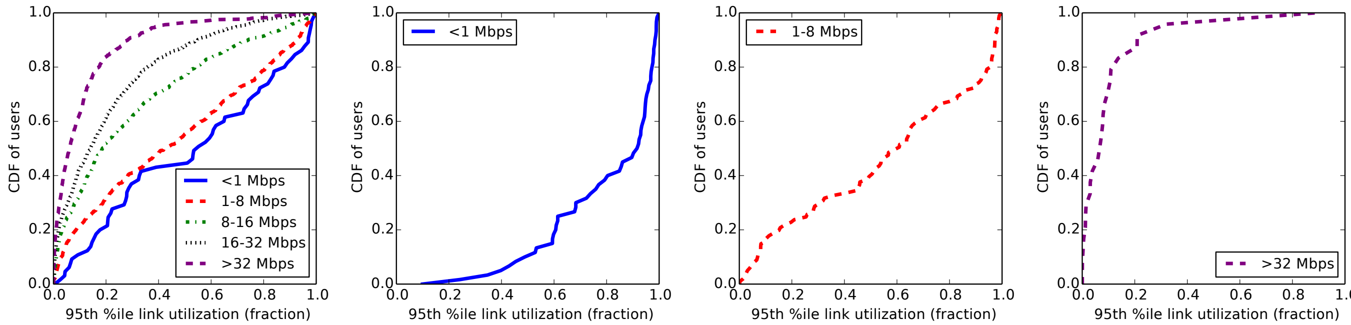
Figure 7a shows the maximum download throughput rates measured for connections in each country. The typical maximum download capacity increases across these markets

[4]All statistics on Internet access and growth are is from ITU. `http://www.itu.int`



(a) Download capacities (b) 95th %ile utilization

Figure 7: CDF of the download capacities and peak utilization for all users in each of the four markets.



(a) US (b) Botswana (c) Saudi Arabia (d) Japan

Figure 8: CDF of the 95th percentile link utilization for users in each country. Users are split into five different groups depending on their maximum download capacity.

(Botswana, Saudi Arabia, the US, Japan). We find a large number of Botswana users on a ≈ 512 kbps service while users in Saudi Arabia are heavily clustered around 4 Mbps. Both the US and Japan show a wider distribution across different service levels. In Japan, however, a higher fraction of users are on high-end services. The majority of users in Japan (60%) have download speeds of at least 25 Mbps, compared with over 71% of users in the US who are on services slower than 25 Mbps.

It is interesting to contrast maximum download throughout rates with the fraction of the link utilized during peak usage for each user in these four countries (Fig. 7b). The countries appears in exactly reverse order. Botswana shows the highest peak utilization while Japan shows the lowest. In Japan, and to some extent in the US, links tend to be very underutilized, even at the 95th percentile.

Based on our earlier findings, we expect that users in Botswana and Saudi Arabia will have higher network demands than users with similar services in the US, due to increased costs. On the other hand, users in Japan should have lower demand on the network than users with similar services in the US, due to lower service costs for the same capacity.

Unfortunately, at this point it is difficult to directly compare user demand in each market due to the large differences in service capacities. Therefore, we group users into different tiers of service based on their service capacity. We then compare usage within the same tier across markets. For this analysis, we selected the following tiers: below 1 Mbps, 1 to 8 Mbps, 8 to 16 Mbps, 16 to 32 Mbps, and above 32 Mbps. The selection of tiers was based on the speeds common among the broadband technologies in our dataset and the range of capacities in each country. In the following plots, we do not include data on a particular tier for a country with less than 30 users in our dataset.

Figure 8 shows the 95th percentile utilization of users, categorized by the aforementioned speed tiers. Figure 8a represents the utilization for users in the US. In this case, as customers sign up for faster services, they tend to be using less of the link during peak usage.

Note the higher link utilization in Botswana (Fig. 8b) compared to the utilization on the same tier in the US. In Botswana, the average 95th percentile link utilization was 80%; in the US, the average peak utilization was about 52%.

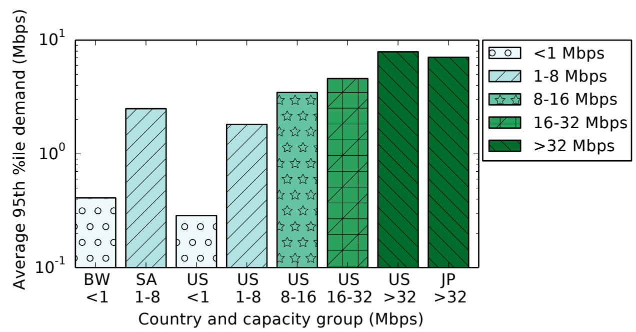


Figure 9: Average 95th percentile utilization for users in each country across each speed tier.

Such a significant difference could be explained by the much higher costs of faster service levels in Botswana where, for instance, a 2 Mbps plan costs about $200 (PPP) per month!

Figure 8c shows a similar, but less pronounced trend in Saudi Arabia. The large majority of users in Saudi Arabia have capacities around 4 Mbps, in the 1 to 8 Mbps download throughput-rate range. Compared to broadband users in the US on the same tier, we also find higher utilization of the link in Saudi Arabia,. Specifically, for users in the 1 to 8 Mbps group, the median link utilization increases from about 43% in the US to 60% in Saudi Arabia.

At the other end of the spectrum is Japan, shown in Fig. 8d. Here we find that, for the majority of users, links tend to be very under-utilized, with an average link utilization of 10%. Overall, the fraction of the link utilized is similar to the same tier in the US, though it is slightly higher, on average, in the US.

In Botswana, for example, users with less than 1 Mbps service used 410 kbps on average versus 286 kbps in the same tier in the US. Additionally, Fig. 9 shows that the demand on the network is 676 kbps (37%) higher in Saudi Arabia than on the same tier in the US. In fact, the average demand of the 8-16 Mbps tier in the US is only 39% higher than the 1-8 Mbps tier in Saudi Arabia, but is 90% higher than the 1-8 Mbps tier in the US. This difference supports our belief that the relatively high price of Internet access in the country, rather than user need, is preventing users in this market from signing up for faster services.

Similarly, users in the US with a service faster than 32 Mbps use 830 kbps more than users on the same tier in Japan. Despite the fact that the cost of broadband access is similar in both Japan and the US, the availability of faster services at a lower cost leads subscribers to sign up for services that will be less heavily used. We examine this trend in the next section. Figure 9 shows the average peak demand for different tiers in each country. We note that in the US, demand increases on each tier, despite the fact that the fraction of the link utilized decreased (shown in Fig. 8a). We also find that when comparing across markets at the same capacity tier, in addition to having higher link utilization, users in more expensive markets also tended to have a higher total demand.

6. COST OF INCREASING CAPACITY

Subscribers select broadband service based on their needs, the set of available plans and the plans' prices. Thus, given the diversity in service availability across markets, users with similar needs will end up choosing different broadband services, depending on what is available. In this section, we look at how the relative cost of alternative services impacts user demand.

Beyond price, broadband service markets differ in the relative cost of upgrading services. For example, according to our dataset of service plans, both Japan and the US have similar prices of broadband access with a connection of at least 1 Mbps costs less than \$25 per month. The two markets differ, however, in service availability and the cost of upgrading. In Japan, a 100 Mbps plan is considerably less expensive than in the US (\$40 per month instead of \$115 per month). Furthermore, in contrast to the US, the broadband service market in Japan has more options with capacities above 50 Mbps and fewer fixed-line services below 10 Mbps.

It is clear that the cost of upgrading capacity, similar to the cost of a particular service level, can have an impact on a demand users impose on their service. To explore this, we begin by generalizing the cost of increased capacity. To this end, we collect all service plans for each country, perform a linear regression analysis on each market, and measure the correlation between capacity and price. We find that, in the majority of these markets (66%) there is a strong correlation (> 0.8) between price and capacity and in 81% there is at least moderate correlation (> 0.4).

In markets where there is weak or no correlation, price is often affected by other factors. For example, in Afghanistan, it is possible to sign up for a dedicated (not shared) DSL connection that is slower and more expensive than alternatives, lowering the correlation coefficient between price and capacity. Whether or not a service is wireless or has a monthly traffic cap would also affect the relationship between price and capacity.

For markets where price and capacity are at least moderately correlated ($r > 0.4$), we use the slope of the linear regression line to estimate the cost of upgrades (the slope is measured in monthly price per Mbps increase in capacity). Figure 10 presents a CDF of the cost of increasing capacity by 1 Mbps for all markets in our dataset.

For illustration, we note in the figure where a few representative markets fall in the distribution. At the lower end of the curve (less than \$0.10 to upgrade), we find regions such as Hong Kong, Japan, and South Korea. Countries such as Canada and the US are at slightly above \$0.50 per Mbps increase. The higher end of the distribution is largely comprised of countries in Africa and the Middle East, like Ghana and Uganda.

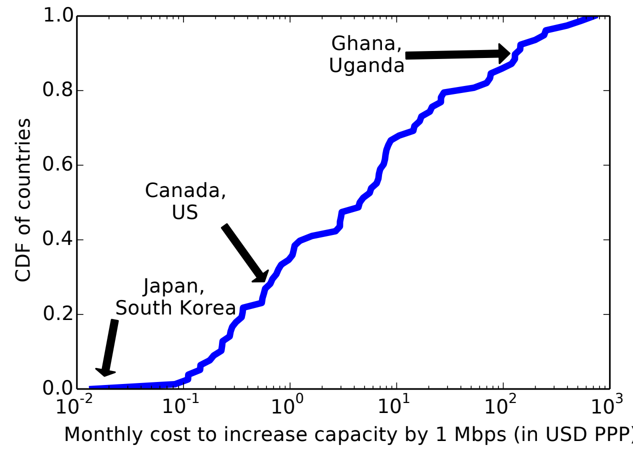


Figure 10: CDF of the monthly cost (after accounting for PPP) to increase broadband service capacity by 1 Mbps in a given country's broadband market. The arrows point out where the labeled countries were placed in the distribution.

Region	>\$1	>\$5	>\$10
Africa	100%	84%	74%
Asia (all)	67%	47%	33%
Asia (developed)	0%	0%	0%
Asia (developing)	83%	58%	42%
Central America/Caribbean	100%	86%	14%
Europe	10%	0%	0%
Middle East	86%	57%	43%
North America	0%	0%	0%
South America	78%	55%	33%

Table 5: The percentage of countries in each region where increasing capacity costs more than \$1, \$5, and \$10 per month for a 1 Mbps increase in capacity. We split Asia into two subgroups, developed and developing, given the diversity of economies within the area.[6]

As shown in Fig. 10, increasing capacity by 1 Mbps tends to cost less than \$1 per month in developed countries but can be well above \$100 (PPP) in some developing countries (e.g., Paraguay and Ivory Coast).[7] Table 5 summarizes this distribution by aggregated region, presenting the percentage of countries, per region, where the cost of increasing capacity by 1Mbps is above \$1, \$5 and %10 (PPP) per month. The trends are strikingly clear – for 74% of the countries in Africa and 43% of those in the Middle East, for instance, the costs of an additional 1Mbps is above \$10 per month.

To test for the impact of service upgrade on user demand we define a new study. For this experiment, our hypothesis (H) states that as the cost to upgrade increases, users are less likely to upgrade and will have higher network demand than users in markets where upgrading is cheaper. Our

[6] As defined by the International Monetary Fund.

[7] Two exceptions in Asia are India and China, where upgrading capacity cost less than \$1 per Mbps per month.

null hypothesis (H_0) is then that the price of upgrading will not affect demand. We use the cost of upgrade to split broadband markets into three classes: countries where the cost of increasing a service by 1Mbps is (i) below \$0.5, (ii) between \$0.5 and \$1 and (iii) above \$1.00 per Mbps.

Control Group	Treatment Group	% H holds	p-value
(\$0, \$0.50]	(\$0.50, \$1.00]	53.8%	0.00717
(\$0.50, \$1.00]	(\$1.00, ∞]	58.7%	0.0110

(a) Average demand w/ BitTorrent

Control Group	Treatment Group	% H holds	p-value
(\$0, \$0.50	(\$0.50, \$1.00]	52.2%*	0.0947
(\$0.50, \$1.00]	(\$1.00, ∞]	56.3%	0.0265

(b) Average demand w/o BitTorrent

Table 6: Percentage of the time that a higher cost to increase capacity (price per 1 Mbps increase) will result in higher network usage. An asterisk denotes that a result was not statistically significant.

We present the results of this experiment in Table 6, for average demand with and without including BitTorrent traffic. In general, increased upgrade prices do lead to higher demand. It is clear that users in developing countries tend to use more than similar users where faster service are more readily accessible. In cases where our results are statistically significant, we can reject the null hypothesis, and assert that the price of increasing capacity affects demand. Our results are inconclusive, i.e., p-value slightly higher than 0.05 when comparing demand (without BitTorrent) between markets where the cost of upgrade are (\$0, \$0.50] and (\$0.50, \$1.00].

We have already visited an example of the impact that the cost of increasing capacity can have on (Figs. 8 and 9). While both Japan and the US have similar monthly cost of broadband access, the costs of increasing capacity is over five times higher in the US explaining the observed higher demand in the US.

7. CONNECTION QUALITY

Previous works have shown that poor connection quality can have a negative impact on a user's quality of experience [13]. In this last section, we explore the potential impact that the quality of a connection, specifically latency and packet loss, has on user demand.

We hypothesize that a sufficiently poor quality of experience could lead to a decrease in demand on the broadband service. In the following paragraphs we test whether this is true by studying the impact of both long latencies and high packet loss rates. As we have done in our previous comparisons, we study the effects of these factors by comparing users that are similar in terms of link capacity and location. When testing the effects of increased latency, we require that average packet loss rates are similar between matched users and vice versa.

7.1 Latency

We first look at the impact of latency on user behavior. In this case, our hypothesis (H) is that decreasing latency will result in higher demand. Therefore, our null hypothesis

Control Group	Treatment Group	% H holds	p-value
(512, 2048]	(0, 64]	63.5%	0.00825
(512, 2048]	(64, 128]	63.4%	0.00620
(512, 2048]	(128, 256]	59.4%	0.00766
(512, 2048]	(256, 512]	56.3%	0.0330

Table 7: Percentage of the time that decreasing latency will result in higher 95th percentile usage (without BitTorrent). Very high latency (over 512 ms) to the nearest NDT server appears to result in lower demand than comparable users with lower latencies.

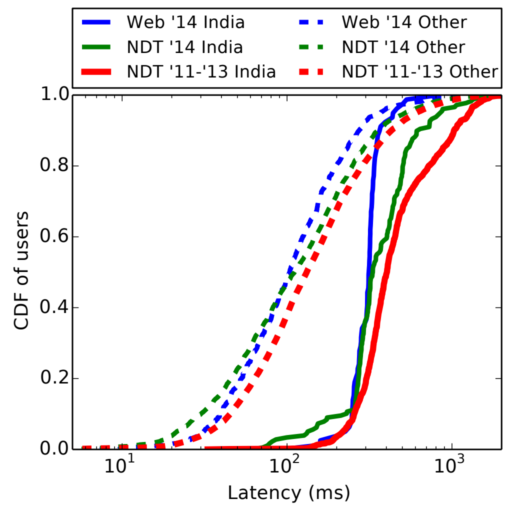


Figure 11: CDF of latency measurements for users in our dataset, grouped by location (India versus the rest of our sample population, labelled "Other"). "Web" represents each user's median latency to five of Alexa's Top Sites from our 2014 dataset. "NDT" represents the average latency to the nearest NDT server. We include NDT data from 2014 that was collected from the same set of users as the web 2014 data.

(H_0) is that decreasing latency does not affect demand and the interaction will be random.

We present the results of the study in Table 7. The table compares the peak demand (95th percentile usage when BitTorrent is not active) of users with problematically high latencies, above 512 ms in our dataset. Users are divided among exponentially increasing sized bins; our control and treatment groups in this case are the higher and lower latency groups, respectively. The results show that there is a significant increase in usage when switching from very high latency to any lower latency group, leading us to reject the null hypothesis.

While the case of latency impacting demand is visible in multiple countries, the impact of high-latency is clear when focusing on users in India. In our previous analysis broadband service plans, we find that the cost to increase capacity is similar in both the US and India (both are within 25% of each other). The cost of broadband access, however, is much higher in India (\$67 versus \$20). Thus, we would expect usage to be higher in India. When comparing

Control Group	Treatment Group	% H holds	p-value
$(0.1\%, 1\%]$	$(0, 0.01\%]$	55.4%	5.85×10^{-6}
$(0.1\%, 1\%]$	$(0.01\%, 0.1\%]$	53.4%	8.55×10^{-4}
$(1\%, 15\%]$	$(0, 0.01\%]]$	58.9%	2.16×10^{-5}
$(1\%, 15\%]$	$(0.01\%, 0.1\%]$	53.8%	0.0360

Table 8: Percentage of the time that decreasing packet loss will result in higher average usage (without BitTorrent). Very high packet loss (above 1%) appears to lead to lower demand than comparable users lower packet loss rates.

users in India to users with similar capacities in the US, we find, surprisingly, that users in India tend to impose lower demand 62% of the time (p-value < 0.001).

An analysis of NDT latencies shows that users in India report much higher latencies to NDT servers than users in other countries. The trends are not restricted to NDT servers but can also be seen when looking at latencies to the set of five globally popular websites: Facebook, Google, Windows Live, Yahoo, and YouTube. Four of these websites (Google, Facebook, YouTube, and Yahoo) accounted for the top five most popular websites in India[8] while Windows Live was ranked 26th.

Figure 11 describes these latency measurements, and compares them by user location (India versus the rest of our sample population). The lines labelled "Web" correspond to the median latency to the five popular websites while "NDT" is the average latency to the nearest NDT server (measured by NDT). We include data from two time periods – 2011 through 2013 (labelled "'11-'13") and from May 2014 to August 2014 (labelled "'14") to compare NDT and website latencies. [9]

Figure 11 shows that the distribution of latency measurements is similar for both NDT traces and the typical latency to the top Alexa sites. For the majority of users in India, we find much higher latencies to both NDT and popular websites compared to the rest of our sample population; nearly every user has a latency longer than 100 ms. Since we rejected the null hypothesis that latency does does not affect demand, we believe that the higher latency for users in India contributes to the fact that we see a decrease in network usage in India.

7.2 Packet loss

Next we examine the impact of packet loss on user demand. Our hypothesis (H) is that decreased packet loss rates result in higher demand. The results of this experiment are shown in Table 8. We find that when comparing users with very low packet loss rates to comparable users with very high packet loss rates, usage tended to be higher on connections with lower packet loss rates. This trend was most pronounced when comparing to connections with packet loss rates above 1%.

This impact of packet loss can be illustrated by looking at the behavior of users in India, as done in Sec. 7.1. We again found that users in India had much higher packet loss rates than the general population, as shown in Fig 12. As a result, we believe that the lower quality connections (both increased packet loss rates and latency) in India are the probable cause of lower demand on the network.

[8]Ranked by `http://www.alexa.com`

[9]We added the website latency experiment later in the study.

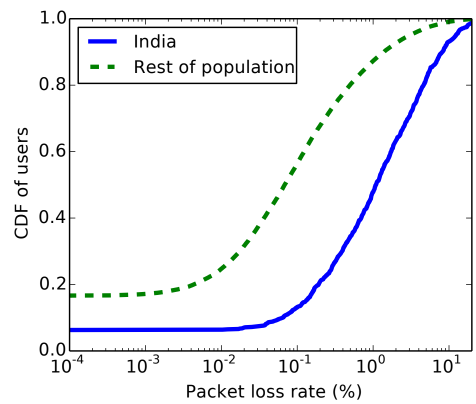


Figure 12: CDF of the average packet loss during measurements to the NDT servers for all users and users in India.

8. RELATED WORK

Broadband analysis has recently attracted much attention from the research community and the general public given its important business and policy implications. A number of efforts have focused on characterizing the availability and performance of broadband services around the world [1, 2, 5, 12, 20, 28, 31, 33]. The focus of our work is on exploring broadband services in their broader context, evaluating the complex interplay between broadband service characteristics, their market features and user demand.

Different aspects of the complex interplay between user behavior, network services and operation has been explore in previous work. Some recent studies have examined the relationship between user behavior, network services and the providers. In Dobrian et al. [13] the authors show that poor connection quality can have a negative impact on a user's quality of experience. Blackburn et al. [3] study how user behavior affects the economics of cellular operators. Chetty et al. [7] perform a user study to understand the effects of usage caps on broadband use. Other efforts have explored additional factors that may influence service demand, including the weather [6], service capacity [36] and the type of region [8].

The difficulty or outright impossibility of conducting controlled, randomized experiments of user behavior at Internet scale has been pointed out before. In his SIGCOMM 2011 Award presentation, Vern Paxson pointed to this issue and suggested the use of natural experiments to explore potential causal relationships with observational data. In a recent paper, Krishnan and Sitaraman [21] explore the use of related quasi-experimental design (QED) to evaluate the impact of video stream quality on viewer behavior and Oktay et al. [24] relies on it for causal analysis of user behavior in social media. We opted for natural experiments, rather than QED, as we consider the control and treatment groups to be sufficiently similar to random assignment.

9. SUMMARY AND DISCUSSION

The findings reported in this paper represent an import step towards understanding how user behavior, and the market features that shape it, affect broadband networks and the Internet at large. These findings should provide valuable insight to the research community, network operators and policy makers. For policy makers, there is a growing consensus that broadband access should be treated as a fundamental right and several efforts around the world are aimed at closing the digital gap. We believe that understanding digital inequality requires us to place broadband access in a broader context [22]. This work, to the best of our knowledge, is the first attempt to that end.

In our longitudinal study of usage between 2011 and 2013, we found that subscribers' demand remained relatively constant in a particular service class, despite the fourfold increase in global IP traffic over the past five years. Thus, we believe the growth in broadband traffic comes from a combination of increased service capacities and a rapidly increasing number of broadband subscribers, rather than higher demand at users' existing service levels.

We find a strong correlation between service capacity and user demand, despite the fact that users rarely fully utilize their links. We used two study designs to infer causality between capacity and usage by studying how individual users change behavior when switching to faster services and by comparing demand between users with different capacities that are otherwise similar. Their relationship also follows a law of diminishing returns; in both cases, we observed relatively lower increases in demand at higher capacities. This trend is particularly noticeable at approximately 10 Mbps, where usage begins to plateau for many users.

This suggest that as service capacities continue to increase, network operators can plan on higher over-provisioning rates. We did observe, however, larger increases in demand when including BitTorrent traffic in our analysis. Beyond capacity, we also showed the impact that the quality of a connection, in terms of latency and packet loss rates, has on user demand. For instance, we note that very long latencies (above 500 ms) and high packet loss rates (starting at 0.1%) clearly result in lower network usage. We speculate that the relationship between capacity, quality and demand will evolve with technology improvements and new applications with greater bandwidth requirements become widely available (e.g. 4k video streaming or telemedicine).

We examined how the price of broadband access affect user demand by comparing the behavior of users with similar broadband services located in different markets. We found that users in markets where broadband connections or additional capacity was more expensive, were more likely to impose higher bandwidth demands on their service than subscribers of comparable services in less expensive markets. For policy makers, this would imply that a focus on wider access to a medium, high-quality capacity service (around 10 Mbps) may have a more significant impact than a focus on increased service capacity. For operators, these trends may suggest a possible role for service pricing in network planning.

10. CONCLUSION

This paper is a first attempt at understanding broadband networks in their broader context, exploring the complex interplay between broadband service characteristics and user behavior. We combine data on broadband usage with relative broadband service pricing from around the world, and use alternative experimental designs to move beyond correlation analysis in our study of the relationship between user demand, broadband service retail prices, and connection characteristics.

There are a number of interesting research directions ahead, from further exploration of the interactions between market features and usage, to an expanded analysis capturing other contextual factors (from the weather and social events to economics). For instance, investigating the relative influence of levels of economic development or differences in broadband availability in rural and urban areas, will give us a more detailed understanding of the impact of different market features. Leveraging longitudinal data on broadband usage, it may be possible to explore the potential benefits of national broadband deployment plans [4], both on the market and on user behaviors. Finally, we have so far treated users as a homogeneous consumer group; it will be interesting to investigate how different categories of users (e.g., gamers, shoppers or movie-watchers) or more diverse households are impacted by different market and service features.

Acknowledgements

We thank our shepherd Vivek Pai and the anonymous reviewers for their invaluable feedback. This work was supported in part by the National Science Foundation through Award CNS 1218287 and by a generous Google Faculty Research Award.

11. REFERENCES

[1] Z. S. Bischof, J. S. Otto, and F. E. Bustamante. Up, down and around the stack: ISP characterization from network intensive applications. In *Proc. of W-MUST*, 2012.

[2] Z. S. Bischof, J. S. Otto, M. A. Sánchez, J. P. Rula, D. R. Choffnes, and F. E. Bustamante. Crowdsourcing ISP characterization to the network edge. In *Proc. of W-MUST*, 2011.

[3] J. Blackburn, R. Stanojevic, V. Erramilli, A. Iamnitchi, and K. Papagiannaki. Last call for the buffet: economics of cellular networks. In *Proc. of MobiCom*, 2013.

[4] Broadband Commission. The state of broadband 2013: Universalizing broadband. `http://www.broadbandcommission.org/Documents/bb-annualreport2013.pdf`.

[5] I. Canadi, P. Barford, and J. Sommers. Revisiting broadband performance. In *Proc. of IMC*, 2012.

[6] J. C. Cardona, R. Stanojevic, and R. Cuevas. On weather and Internet traffic demand. In *Proc. of PAM*, 2013.

[7] M. Chetty, R. Banks, A. Brush, J. Donner, and R. Grinter. You're capped: understanding the effects of bandwidth caps on broadband use in the home. In *Proc. of SIGCHI*, 2012.

[8] K. Cho, K. Fukuda, H. Esaki, and A. Kato. The impact and implications of the growth in residential user-to-user traffic. In *Proc. of ACM SIGCOMM*, 2006.

[9] D. R. Choffnes and F. E. Bustamante. Taming the torrent: A practical approach to reducing cross-ISP

traffic in peer-to-peer systems. In *Proc. of ACM SIGCOMM*, 2008.

[10] Cisco. Cisco visual networking index: Forecast and methodology 2012-2017.

[11] L. DiCioccio, R. Teixeira, M. May, and C. Kreibich. Probe and pray: Using UPnP for home network measurements. In *PAM*, 2012.

[12] M. Dischinger, K. P. Gummadi, A. Haeberlen, and S. Saroiu. Characterizing residential broadband networks. In *Proc. of IMC*, 2007.

[13] F. Dobrian, A. Awan, I. Stoica, V. Sekar, A. Ganjam, D. Joseph, J. Zhan, and H. Zhang. Understanding the impact of video quality on user engagement. In *Proc. of ACM SIGCOMM*, 2011.

[14] T. Dunning. *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press, 2012.

[15] FCC. Measuring broadband America. `http://www.fcc.gov/measuring-broadband-america`.

[16] T. Gilovich, D. Griffin, and D. Kahneman. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press, 2002.

[17] J. Horrigan and E. Satterwhite. Americans' perspectives on online connection speeds for home and mobile devices. `http://www.fcc.gov/encyclopedia/broadband-speed`, 2010.

[18] IDC. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east, 2013.

[19] ITU. The impact of broadband on the economy. `https://www.itu.int/ITU-D/treg/publications/bbreports.html`, 2012.

[20] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson. Netalyzr: Illuminating the edge network. In *Proc. of IMC*, 2010.

[21] S. S. Krishnan and R. K. Sitaraman. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. In *Proc. of IMC*, 2012.

[22] S. D. Meinrath, J. Losey, and B. Lennett. A growing digital divide: Internet freedom and the negative impact of command-and-control networking. *IEEE Internet Computing*, 2011.

[23] MLabs. Network diagnostic tool. `http://www.measurementlab.net/run-ndt/`.

[24] H. Oktay, B. J. Taylor, and D. D. Jensen. Causal discovery in social media using quasi-experimental designs. In *Proc. of the First Workshop on Social Media Analytics*, 2010.

[25] J. S. Otto, M. A. Sanchez, D. R. Choffnes, F. E. Bustamante, and G. Siganos. On blind mice and the elephant – understanding the network impact of a large distributed system. In *Proc. of ACM SIGCOMM*, 2011.

[26] V. Paxson. Strategies for sound Internet measurement. In *Proc. of IMC*, October 2004.

[27] Policy by the numbers. International broadband pricing study: Updated dataset. `http://policybythenumbers.blogspot.com/2013/05/international-broadband-pricing-study.html`.

[28] SamKnows. Accurate broadband information for consumers, governments and ISPs. `http://www.samknows.com/`.

[29] M. A. Sánchez, J. S. Otto, Z. S. Bischof, and F. E. Bustamante. Trying broadband characterization at home. In *Proc. of PAM*, 2013.

[30] M. A. Sánchez, J. S. Otto, Z. S. Bischof, D. R. Choffnes, F. E. Bustamante, B. Krishnamurthy, and W. Willinger. Dasu: Pushing experiments to the Internet's edge. In *Proc. of USENIX NSDI*, 2013.

[31] Speedtest.net. The global broadband speed test. `http://www.speedtest.net/`.

[32] R. Stanojevic, V. Erramilli, and K. Papagiannaki. Cognitive bias in network services. In *Proc. of HotNets*, 2012.

[33] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapè. Broadband internet performance: a view from the gateway. In *Proc. of ACM SIGCOMM*, 2011.

[34] S. Sundaresan, N. Feamster, R. Teixeira, and N. Magharei. Measuring and mitigating web performance bottlenecks in broadband access networks. In *Proc. of IMC*, 2013.

[35] World Bank. IC4D 2009: Extending reach and increasing impact. `http://go.worldbank.org/NATLOH7HV0`.

[36] M. Zheleva, P. SChmitt, M. Vigil, and E. Belding. The increased bandwidth falacy: Performance and usage in rural Zambia. In *Proc. of ACM DEV*, 2013.