Characterizing and Improving the Reliability of Broadband Internet Access

Zachary S. Bischof[†‡] Fabián E. Bustamante[‡] Nick Feamster[⋆]
[†]IIJ Research Lab [‡]Northwestern University [⋆]Princeton University

Abstract

In this paper, we empirically demonstrate the growing importance of reliability by measuring its effect on user behavior. We present an approach for broadband reliability characterization using data collected by many emerging national initiatives to study broadband and apply it to the data gathered by the Federal Communications Commission's Measuring Broadband America project. Motivated by our findings, we present the design, implementation, and evaluation of a practical approach for improving the reliability of broadband Internet access with multihoming.

1 Introduction

Broadband availability and performance continue to improve rapidly, spurred by both government and private investment [14] and motivated by the recognized social and economic benefits of connectivity [35, 63]. The latest ITU "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 [13]. According to Akamai's "State of the Internet" report, over the last four years, the top four countries in terms of average connection speed (South Korea, Hong Kong, Romania, and Japan) have nearly doubled their capacity [18].

Although providing access and sufficient capacity remains a challenge in many parts of the world [34, 66], in most developed countries, broadband providers are offering sufficiently high capacities (i.e., above 10 Mbps [18]) to encourage consumers to migrate services for entertainment, communication and home monitoring to over-the-top (OTT) alternatives. According to a recent survey, nearly 78% of U.S. broadband households subscribe to an OTT video service [53]. Enterprises are following the same path, with over one-third opting to use VoIP phones instead of landline ones [31].

The proliferation of high-capacity access and the migration to OTT services have raised users' expectations of *service reliability*. A recent survey on consumer experience by the UK Office of Communication (Ofcom) ranks reliability first—*higher than even the speed of connection*—as the main reason for customer complaints [44]. Figure 1 illustrates the impact that low reliability can have on video Quality of Experience in a lab experiment using HTTP Live Streaming [4]. The plot shows the draining of the buffer - in blue - during two service interruptions (gray bars) and the drop on video quality (right axis) as a result. While the buffer is quick to refill after the first outage, the draining of it causes a drop in quality. Our empirical study of access-ISP outages and user demand corroborates these observations, showing the effects of low reliability on user behavior, as captured by their demand on the network (§2). Researchers and regulators alike have also recognized the need for clear standards and a better understanding of the role that service reliability plays in shaping the behavior of broadband users [10, 27, 41]. Despite its growing importance, both the reliability of broadband services and potential ways to improve on it have received scant attention from the research community.

In this paper, we introduce an approach for characterizing broadband reliability using data collected by the many emerging national efforts to study broadband (in over 30 countries [57]) and apply this approach to the data gathered by the Measuring Broadband America (MBA) project, which is operated by the United States Federal Communications Commission (FCC) [28]. Motivated by our findings, we present the design and evaluation of a practical approach to improve broadband reliability through multihoming using a prototype implementation built as an extension to residential gateways and a cloud-supported proxy.

We make the following contributions:

- We demonstrate that poor reliability can affect user traffic demand well beyond periods of unavailability. For instance, we find that frequent periods of high packet loss (above 1%) can result in a decrease in traffic volume for 58% of users *even during periods of no packet loss* (§2).
- We present an approach to characterize broadband service reliability. We apply this approach to data collected from 7,998 residential gateways over four years (beginning in 2011) as part of the US FCC MBA deployment [28]. We show, among other findings, that current broadband services deliver an average availability of at most two nines (99%), with an average annual downtime of 17.8 hours (§3).
- Using the FCC MBA dataset and measurements collected by over 6,000 end-host vantage points in 75 countries [46], we show that multihoming the access link

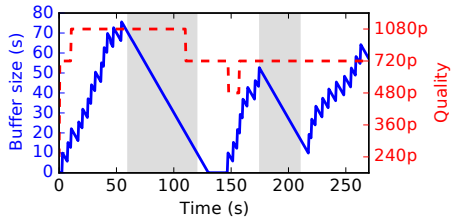


Figure 1: *The effect of service interruption on video Quality of Experience. Even small outages (highlighted in gray) can have clear detrimental effects on video quality, with buffers draining (on the left) to the point of service interruption and the service dropping to a lower quality stream (on the right) in response.*

at the home gateway with two different providers adds two nines of service availability, matching the minimum four nines (99.99%) required by the FCC for the public switched telephone network (PSTN) [40] (§4).

- We present the design, implementation, and evaluation of *AlwaysOn*, a readily deployable system for multihoming broadband connections (§5).

To encourage both reproducibility and use of the system, we have publicly released our dataset, analysis scripts, and AlwaysOn prototype [2].

2 Importance of Reliability

In this section, we motivate the importance of reliability by studying the relationship of service reliability and user behavior, as measured by fluctuations in user traffic demand over time. Studying this relationship at scale requires us to design several *natural experiments* [23] using a subset of the data collected by the FCC MBA effort [28]. We begin by describing this dataset and our experimental methods followed by a discussion of the results.

2.1 Dataset

Since 2011, the FCC has been conducting broadband service studies using data collected by custom gateways in selected users' homes. The collected data includes a rich set of metrics, such as the bytes transferred per unit time, as well as the loading times of popular websites.[1] This data has been primarily used to create periodic reports on the state of broadband services in the US as part of the MBA initiative [28]. For this analysis, we use two sets of measurements from this dataset: UDP pings and traffic byte counters, for the FCC reports from 2011 through August 2016 (all of the publicly available data).

UDP pings continually measure round-trip time to at least two (typically three) measurement servers hosted by either M-Lab or the ISPs. Every hour, the gateway sends probes to each server at regular intervals, fewer if the link is under heavy use for part of the hour. The typical number of probes sent per hour changed from 600 per hour in 2011 to approximately 2,000 in mid-2015.

[1] A full description of all the tests performed and data collected is available in the FCC's measuring broadband technical appendix [26].

For each target server, the gateway reports hourly statistical summaries of latency measurements, the total number of probes sent and the number of missing responses (i.e., lost packets). We use the latter two values to calculate the average packet loss rate over the course of that hour. Since measurements target multiple target servers throughout each entire hour, we select the server with the lowest loss rate each hour for our analysis. This prevents a single failed target server from increasing the calculated loss rate.

The traffic byte counters record the total number of bytes sent and received across the WAN interface over each previous hour. They also record counters for the amount of traffic due to the gateway's active measurements, which we subtract from the total volume of traffic.

2.2 Method

Understanding how service outages affect user behavior requires us to accurately assess user experience in a way that is quantifiable, meaningful, and applicable across the thousands of subscribers in the FCC dataset. For this study we leverage *natural experiments*, a class of experimental design common in epidemiology, the social sciences, and economics [23].

An alternative approach to explore the relationship between reliability and user behavior would involve controlled experiments with random treatment—randomly assigning participants to a diverse set of connections and measuring differences in user behavior and their experiences. Conducting controlled experiment at scale such as this is impractical for a number of reasons. First, subjecting users to a controlled setting may cause deviations from normal behavior. Second, we want to monitor users' network use at home under typical conditions, but doing so would require us to have control of the quality of their access link and the reliability of their broadband service. Intentionally degrading users' connections at the gateway would require access to their routers (which we do not have) and would also likely deter users from participating; even if we had access to users' home routers, changing network conditions without their knowledge introduces complicated ethical questions.

Using natural experiments, rather than control the application of a treatment to the user, we test our hypotheses by measuring how users react to network conditions that occur *spontaneously* while controlling for confounding factors (e.g., service capacity, location) in our analysis [11, 39]. For example, to test whether high loss rates result in decreased user demand, we compare the demand of users with low average packet loss (our *control* set) to the demand of users of otherwise similar services with high average packet loss (our *treatment* set). To accept the hypothesis, application of the treatment should result in significantly lower network utilization. On the other hand, if user demand and reliability are not related, we expect the number of cases where our hypothesis holds to be about 50% (i.e., random).

We use network demand as a measurable metric that may reflect user experience. Recent work [7, 11] suggests that this metric acts as a suitable proxy for user experience. A

Treatment group	% H holds	p-value
(0.5%, 1%)	48.1	0.792
(1%, 2%)	57.7	0.0356
$>2\%$	60.4	0.00862

Table 1: *Percentage of the time that a higher average packet loss rates will result in lower usage. Users in the control group have similar download capacities with an average packet loss rate between 0% and 0.0625%.*

significant change in network usage (e.g., bytes transferred or received) can be interpreted as a response to a change in the user's experience.

We use a one-tailed binomial test, which quantifies the significance of deviations from the expected random distribution, to check the validity of each hypothesis. We consider a p-value less than 0.05 to be a reasonable presumption against the null hypothesis (H_0). To control for the effects of large datasets on this class of studies, potentially making even minor deviations significant, we only consider deviations larger than 2% to be practically important [51].

2.3 Experiment results

Several possible experiments can shed light on how service reliability affects user behavior. Although we expect that usage will drop around a single outage, we aim to understand how poor reliability over longer periods of time affects user behavior. Our experiments test the effects on user demand of connections that are consistently lossy and connections that have frequent periods of high loss.

High average loss. To understand how consistently lossy links affect user demand, we calculate the average packet loss rate over the entire period during which the user is reporting data. We then group users based on their average packet loss rate. We select users from each treatment group and match[2] them with users in the same region with similar download and upload link capacities (within 10% of each other) in the control group. Users in the control group have an average loss rate of less than 0.0625%. Our hypothesis, H, is that higher average packet loss rates will result in lower usage, due to a consistently worse experience. Our null hypothesis is that average packet loss and user demand are not related. Table 1 shows the results of this experiment.

The results show that usage is significantly affected even for average packet losses above 1% — 57.7% of our cases show a lower volume of traffic with a p-value of 0.0356. This leads us to reject the null hypothesis.

This experiment shows that a consistently lossy connection – one with high average packet loss – can affect user demand. However, it is unclear if this is caused by a change in user demand or the result of protocols reacting to lost packets (e.g., transfer rates decreasing after a packet is dropped or switching to lower quality streams). We attempt to address this with our next experiment.

[2]In observational studies, matching tries to identify subsamples of the treated and control units that are "balanced" with respect to observed covariates.

Control group	Treatment group	% H holds	p-value
(0.5%, 1%)	(1%, 10%)	54.2	0.00143
(0.1%, 0.5%)	(1%, 10%)	53.2	0.0143
(0%, 0.1%)	(1%, 10%)	54.8	0.000421
(1%, 10%)	$>10\%$	68.3	3.65×10^{-05}
(0.5%, 1%)	$>10\%$	70.0	6.95×10^{-06}
(0.1%, 0.5%)	$>10\%$	70.8	2.87×10^{-06}
(0%, 0.1%)	$>10\%$	72.5	4.34×10^{-07}

Table 2: *Percentage of the time that users with more frequent high-loss hours ($\geq$ 5% packet loss) have lower network usage.*

Frequent periods of high loss. In this experiment, we test if more frequent periods of high packet loss affects the traffic demands of users during hours of no loss.

To understand the effects of frequent periods of high loss on user behavior we calculate, for each user, the fraction of hours where the gateway measured more than 5% packet loss. We group users based on how frequently periods of high loss occurred. For example, users that recorded loss rates above 5% during 0% to 0.1% of measurements were placed in a group that we used as one of the controls. We then compared the network demands during peak hours with no packet loss between each pair of user groups. In this case, our hypothesis, H, is that groups with a high frequency of high loss rates (treatment group) will have lower usage than groups with a low frequency of high loss rates (control group). Table 2 shows the results of this experiment.

We find that users with high packet loss rates during more than 1% of hours, tend to have lower demand on the network during periods of no packet loss. As the difference between the frequency of high loss rates periods increases, the magnitude of this effect increases, with larger deviations from the expected random distribution.

Previous studies have discussed the importance of broadband service reliability [41], and surveys of broadband users have shown that reliability, rather than performance, has become the main source of user complaints [44]; our findings are the first to empirically demonstrate the relationship between service reliability and user traffic demand.

3 Characterizing Reliability

We now present an approach for characterizing broadband service reliability that can apply to the datasets that many ongoing national broadband measurement studies are collecting. The decision to make our analysis applicable to the existing national datasets introduces several constraints on our analysis method, including the type and granularity of metrics and the placement of vantage points. At the same time, our approach is applicable to the various available datasets. Ultimately, our work can motivate future experiment designs to better capture all aspects of broadband service reliability. We describe ongoing broadband measurement efforts before presenting our methods and metrics for characterizing service reliability. We then discuss our findings concerning the reliability of broadband services in the US.

3.1 Approach

Available data. Over the last decade, the number of governments with national broadband plans has increased rapidly [36], and several of these governments are funding studies to characterize the broadband services available to their citizens. Two prominent examples are the efforts being carried out by the UK Ofcom and the US FCC in collaboration with the UK company SamKnows. In the few years since their initial work with Ofcom, SamKnows has begun working with at least six additional governments including the US, Canada, Brazil, the European Union and Singapore. Data for these efforts is typically collected from modified residential gateways distributed to participants in a range of service providers.

We use the FCC's dataset for our characterization of broadband reliability, as it is the only effort that currently publishes copies of its raw data. In addition to using the techniques in Section 2.1 to clean the data, we also attempt to validate a user's ISP by looking at the gateway's configured DNS IP addresses, making sure they are consistent with subscribing to that provider (e.g., a user listed as Comcast user should be direct to a DNS server in Comcast's network). We also remove any gateways that have been marked for removal by the FCC's supplemntary unit metadata.

Metrics. To analyze the data from these efforts, we use a number of conventional metrics to quantify the reliability of broadband service. These metrics are defined based on an understanding of what constitutes a failure. We define the *reliability* of a broadband service as the average length of time that the service is operational in between interruptions and *availability* as the fraction of time the service is in functioning condition.

We adopt several well-accepted metrics from reliability engineering, including *Mean Time Between Failure* (MTBF) and *Mean Down Time* (MDT). MTBF is the average time that a service works without failure; it is the multiplicative inverse of Failure Rate, formally defined as

$$MTBF = \frac{\text{Total uptime}}{\text{\# of failures}}$$

To characterize the length of time a service is unavailable during each failure, we use MDT, which is defined as

$$MDT = \frac{\text{Total downtime}}{\text{\# of failures}}$$

We can now define availability (A) as the probability that at any given point in time, the service is functioning/operational. Unavailability is the complement of availability. More formally

$$A = \frac{\text{MTBF}}{\text{MTBF} + \text{MDT}}$$
$$U = (1 - A).$$

Definition of a failure. What constitutes a failure or outage in the context of broadband services is a critical issue tightly coupled to the collected metrics. Although the definition of failure is obvious in many systems, it is less clear in the context of "best-effort" networks.

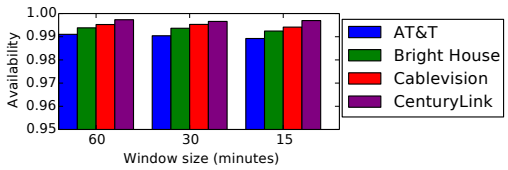


Figure 2: *Service availability for four ISPs across multiple observation window sizes.*

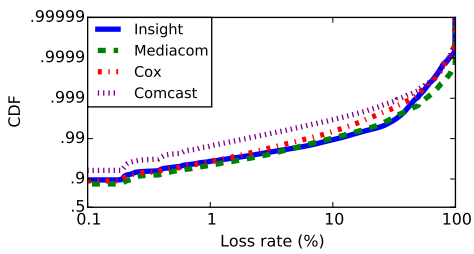


Figure 3: *Hourly loss rates measured from gateways of four cable providers. Lower curves indicate a less available service; curves crossing over each other implies that different loss-rate thresholds would yield different rankings.*

We choose to identify connections failures by detecting significant changes in lost packets. It is unclear what packet loss rate (or rates) should be used as thresholds for labeling failures. Achievable TCP throughput varies inversely with the square root of loss rate [42, 49] and even modest loss rates can significantly degrade performance. Xu et al. showed that video telephony applications can become unstable at a 2% bursty loss [65], with significant quality degradation occurring around 4% in some cases. In our analysis, we use three thresholds for classifying network failures – 1%, 5%, and 10%.

While the FCC MBA dataset is currently the largest publicly available dataset on broadband service performance, relying on it for our analysis means we are only able to measure loss rates at a one-hour granularity. To evaluate the impact of monitoring granularity, we rely on a platform installed in 6,000 end-hosts to measure loss by sending packets approximately every five seconds and use this data to calculate loss rate using different sizes and loss rate thresholds. Figure 2 shows the availability of four ISPs in our dataset using the 10% loss threshold. We found that changing the window size has little impact on our calculation of availability and the relative ranking of ISPs.

The distribution of loss rates are quite different for different broadband technologies, and can vary even across providers with the same technology at different loss rate thresholds. Figure 3 shows the CDF of loss rate of four cable providers, with the y-axis showing the cumulative fraction of all hourly time intervals. Although two providers may offer the same MTBF for a particular loss rate threshold, considering the difference in loss rate distributions, a different definition of "failure" could result in a different ranking. For instance,

Technology	% of participants
Cable	55%
Cable (business)	1%
DSL	35%
Fiber	7%
Satellite	1%
Wireless	1%

Table 3: *Percentage of the sample population in the FCC's dataset using each access link technology.*

defining a failure as "an hour with $>1\%$ packet loss" yields a similar MTBF for both Cox and Insight Cable ($\approx$ 27.5 hours), using a 10% loss rate threshold, but results in a MTBF over 50% higher for Cox ($\approx$ 150 hours) than for Insight ($\approx$ 94 hours).

The assessment of broadband reliability could focus on different aspects, ranging from the reliability of the network connection, the consistency of performance, and the availability of services offered by the ISP, such as DNS servers and email [41]. The primary focus of this work is on broadband service reliability, under which we include both the availability of connection itself as well as that of the ISP's DNS service. From the perspective of most users, failures in either are indistinguishable. We plan to study other aspects of service reliability, such as performance consistency, in future work.

3.2 Characterization of service reliability

We apply the approach presented in the previous section to characterize the reliability of broadband services in the US using the FCC MBA dataset. We first provide a short summary of the population of participants in the SamKnows/FCC study. In our study, we seek to understand the role that a set of key attributes of a subscriber's connection play in determining its reliability: (1) How does reliability vary across different providers? (2) What is the impact of using different access technologies or subscribing to different tiers of service? (3) Does geography affect reliability? (4) How reliable is the provider's DNS service?

Sample population. As part of the MBA dataset, the FCC also provides metadata about each participant including the user's service tier (i.e., subscription speed), service technology (e.g., cable or DSL), and geographic location. Combining this information with the loss rate data described in Section 3.1, we compare the reliability of broadband services across different axis.

The list of ISPs covered in the sample population includes both large, nationwide ISPs and smaller, regional ISPs. Since the number of devices per ISP is weighted by the number of subscribers, most devices (71%) are located in larger ISPs (AT&T, Comcast, and Verizon).

The FCC's dataset includes a diverse set of technologies, including satellite and fixed wireless providers. Table 3 shows a summary of the distribution of participants by access technology. "Wireless" access refers to fixed wireless (not mobile) from providers such as Clearwire, where users connected there FCC-provided device to a wireless modem. Additional information, such as the process used for selecting participants, can be found in the technical appendix of the FCC's report [26].

ISP	Average availability			Average annual downtime (hours)		
	1%	5%	10%	1%	5%	10%
Fiber						
Frontier (Fiber)	98.58	99.47	99.77	124	46.8	20.3
Verizon (Fiber)	**99.18**	99.67	99.80	72	29.2	17.8
Cable						
Bright House	98.21	99.28	99.58	156	62.8	36.7
Cablevision	98.33	99.53	99.70	146	41.4	25.9
Charter	97.84	99.29	99.59	189	62.5	36.1
Comcast	98.48	99.45	99.66	134	48.0	29.7
Cox	96.35	98.82	99.33	320	103.0	58.4
Insight	96.38	98.31	**98.94**	318	148.0	93.0
Mediacom	95.48	98.34	99.03	396	146.0	85.3
TimeWarner	98.47	99.48	99.69	134	45.9	26.9
DSL						
AT&T	96.87	99.05	99.42	274	83.3	51.1
CenturyLink	96.33	98.96	99.39	322	90.9	53.7
Frontier (DSL)	93.69	98.18	**98.87**	553	160.0	98.7
Qwest	98.24	99.24	99.51	154	66.7	42.8
Verizon (DSL)	95.56	98.43	99.00	389	137.0	88.0
Windstream	94.35	98.72	99.42	495	112.0	50.6
Wireless						
Clearwire	88.95	96.96	**98.13**	968	266.0	164.0
Satellite						
Hughes	73.16	90.15	94.84	2350	863.0	453
Windblue/Viasat	72.27	84.20	96.37	2430	1380.0	318.0

Table 4: *Average availability and annual downtime for subscribers, per service, for three different loss-rate thresholds. Verizon (fiber) is the only service providing two nines of availability at the 1% loss rate threshold. Clearwire is able to reach performance close to Frontier (DSL) and Insight at the 10% threshold.*

To understand the relative importance of the different attributes, we calculated the information gain—the degree to which a feature is able to reduce the entropy of a target variable—of each attribute of a subscriber's connection (ISP, download/upload capacity, region, and access technology). We found the subscriber's ISP to be the most informative feature, with access link technology as a close second, for predicting service availability. In the rest of this section we analyze the impact of these attributes on service reliability. We close with an analysis of DNS and ISP reliability.

3.2.1 Effect of ISP

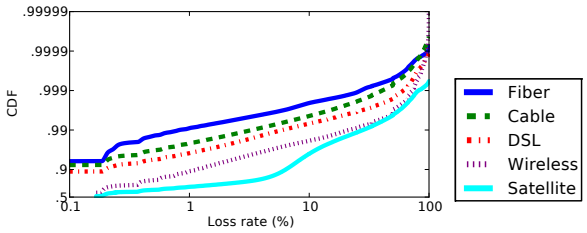
We first characterize service *availability*— the probability that a service is operational at any given point in time—for each provider in our dataset. Table 4 lists the average availability per ISP, as well as the provider's unavailability, described as the average annual downtime (in hours). We evaluate both metrics in the context of the three loss rate thresholds for network failures measured over an hour. For comparison, five nines is often the target availability in telephone services [43].

We find that, at best, some providers are able to offer two nines of availability. Verizon's fiber service is the only one with two nines of availability at the 1% threshold. At 5%, about half of the providers offer just over two nines. The satellite and wireless services from Clearwire, Hughes, and Viasat provide only one nine of availability, even at the 10% loss rate threshold.

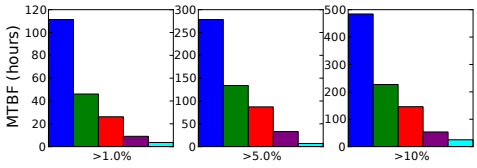
Because broadband users are more likely to be affected by outages in the evening, we also measured availability during peak hours (from 7PM to 11PM, local time), as shown in

ISP	A	% change in U	A	% change in U
	1%		10%	
Satellite				
Hughes	60.97	+45.4	91.38	+66.9
Wildblue/ViaSat	69.44	+10.2	94.14	+61.2
Wireless				
Clearwire	86.35	+23.6	97.57	+29.9
DSL				
Windstream	89.17	+91.8	99.13	**+50.4**
Frontier (DSL)	87.98	+90.4	98.42	+39.9
Verizon (DSL)	93.95	+36.2	98.90	+9.9
CenturyLink	94.19	+58.2	99.35	+6.9
AT&T	95.85	+32.4	99.38	+5.4
Qwest	97.92	+18.5	99.51	+1.2
Cable				
Cablevision	97.76	+34.2	99.64	**+22.6**
TimeWarner	98.03	+28.5	99.69	+1.3
Insight	95.31	+29.4	98.98	-3.9
Charter	97.75	+4.2	99.61	-6.4
Mediacom	94.52	+21.1	99.09	-7.0
Comcast	98.39	+5.3	99.70	-11.7
Brighthouse	98.15	+3.5	99.63	-11.8
Cox	96.30	+1.3	99.42	**-13.3**
Fiber				
Frontier (Fiber)	98.56	+1.4	99.78	-4.6
Verizon (Fiber)	99.11	+8.7	99.83	**-14.7**

Table 5: *Average availability (A) and percent change in unavailability (U) for subscribers of each ISP during peak hours. Some providers had significantly higher unavailability at the 10% threshold during peak hours, including Windstream and Cablevision, as well as satellite and wireless services. Cox and Verizon (fiber) had the largest improvement in availability during peak hours, as outages were concentrated during early morning or mid-day.*



(a) *Hourly loss rates*



(b) *MTBF*

Figure 4: *Hourly loss rates and MTBF for each type of access technology. There is a clear separation between technology for both metrics.*

Table 5. Although all providers show a lower availability at the 1% loss rate threshold compared to their full-day average, most cable providers actually performed better at a 10% loss rate threshold. We expect that some of these providers may perform planned maintenance, which would introduce extremely high periods of loss ($> 10\%$), during the early morning or midday. Overall, Cox and Verizon (fiber) had the largest decrease in unavailability during peak hours.

On the other hand, DSL, wireless, and satellite providers continued to have lower availability at the 10% threshold during peak hours, as compared to their average availability over all time periods. Of all the cable providers, only two had an increase in unavailability during peak hours, with Cablevision having the biggest change. Windstream and Frontier (DSL) also had a much larger increase in unavailability during peak hours compared to other DSL providers.

We also analyzed the MTBF for each provider, which represents the average time between periods with high packet loss. Most ISPs appear to maintain a MTBF of over 200 hours ($\approx$ 8 days), but a few experience failures every 100 hours, on average. ClearWire, Hughes, and Viasat again have notably low MTBF: 73.8, 26.0, and 4.78 hours, respectively. CenturyLink and Mediacom offer the two lowest MTBFs for DSL and cable providers, respectively. These network outages are resolved, on average, within one to two hours for most ISPs. The main exception is satellite providers—more specifically Viasat— with a MDT (mean downtime), close to 5.5 hours.

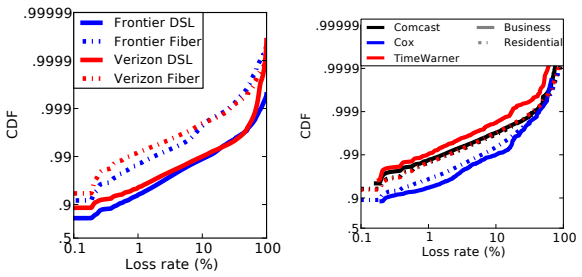
In general, most ISPs ranked similarly across both MTBF and MDT, with a few exceptions. For instance, Verizon's fiber service had the highest MTBF, but its periods of downtime were often over 2.5 hours. Frontier's DSL service, on the other hand, had frequent failures, but these periods of failure were relatively short.

3.2.2 Effect of access technology

Next, we study the impact of a subscriber's access technology. Figure 4a shows a CDF of packet loss rates for each access technology. As expected, we find that fiber services provide the lowest loss rates of all technologies in our dataset with only 0.21% of hours having packet loss rates above 10%. Stated differently, fiber users could expect an hour with 10% packet loss to occur approximately once ever 20 days. Cable and DSL services are next in terms of reliability, with periods of 10% packet loss only appearing 0.44% and 0.68% of the time, respectively. Periods with packet loss rates above 10% were almost a full order of magnitude more frequent for wireless (1.9%) and satellite (4.0%) services.

We compare the average interval between hours with loss above the different loss-rate thresholds, shown in Figure 4b. For each threshold, fiber performs significantly better, with cable and DSL again showing relatively similar performance.

Other factors that affect the reliability may in fact be related to access technology; for example, network management policies of a particular ISP might be correlated with the ISP's access technology and could hence play a role in determining network reliability. To isolate such effects, we compare the difference in service reliability within the same provider, in the same regions, but for different technologies. Only two providers offered broadband services over more than one access technology: Frontier and Verizon, both of which have DSL and fiber broadband services. Figure 5a shows a CDF of the loss rates measured by users of both services. Although there are differences across the two providers, in general, subscribers using same access technology tend to experience similar packet loss rates. Verizon and Frontier DSL customers



(a) *Fiber vs. DSL.* **(b)** *Business vs. residential.*

Figure 5: *The hourly loss rates for subscribers of each service. Technology, rather than provider, is the main determinant of availability and service tiers has little effect.*

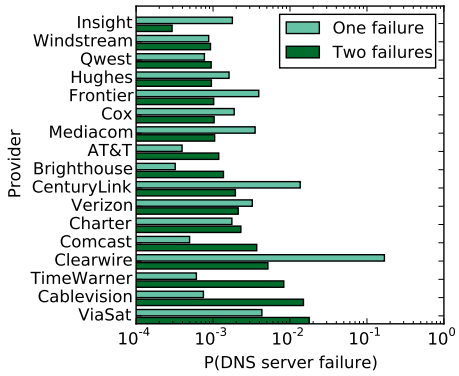


Figure 6: *Probability that one (or two) DNS servers will be unavailable for each ISP's configured DNS servers. We consider the two cases independently (i.e., "one failure" reflects the event that exactly one server fails to respond to queries).*

measured high loss rates (above 10%) during 1.56% and 1.82% of hours, while Verizon and Frontier fiber customers saw high loss rates during 0.33% and 0.53% of hours.

3.2.3 Effect of service tier

In addition to offering broadband services over multiple access technologies, a number of ISPs offer different service tiers on the same access technology. For example, Comcast, Cox, and Time Warner all have a "business class" service in addition to their standard residential service. We explored how reliability varies across different service tier offerings within the same provider.

Figure 5b shows a CDF of the loss rates reported by users of each provider's residential and business class Internet service. In general, the service class appeared to have little effect on the reliability of a service. The differences in packet loss rates are small compared to the difference between access technologies in the same provider. Comcast business subscribers see about the same loss rates as the residential subscribers, while Time Warner's business subscribers report slightly lower packet loss rates. On the other hand, Cox business subscribers actually report a slightly higher frequency of packet loss when compared to residential subscribers. In particular, there are occasionally anecdotes that providers might be encouraging subscribers to upgrade their service tier by offering degraded service for lower service tiers in a region where they were offering higher service tiers; we did not find evidence of this behavior.

3.2.4 Effect of demographics

We also explored the relationship between population demographics and the reliability of Internet service. For this we combined publicly available data from the 2010 census with the FCC dataset to see how factors such as the fraction of the population living in an urban setting, population density and gross state product per capita relate to network reliability.

We looked at service reliability and urban/rural population distributions per state using the classification of the US Census Bureau with "urbanized areas" ($>$ 50,000 people), "urban clusters" (between 2,500 and 50,000 people), and "rural" areas ($<$ 2,500 people) [62]. We also explore the correlation between failure rate and a state's gross state product (GSP) per capita.

Overall, we found a weak to moderate correlation between failure rates and both percent of urban population ($r = -0.397$) and GSP per capita ($r = -0.358$), highlighting the importance of considering context when comparing the reliability of service providers (the direction of the causal relationship is an area for further study).

3.2.5 ISP and DNS reliability

We also include a study of ISPs' DNS service availability in our analysis of broadband reliability. Previous work has shown that DNS plays a significant role in determining application-level performance [46, 64] and thus users' experience. Additionally, for most broadband users, the effect of a DNS outage is identical to that of a faulty connection.

For DNS measurements, the gateway issues an hourly A record query to both the primary and secondary ISP-configured DNS servers for ten popular websites. For each hostname queried, the router reports whether the DNS query succeeded or failed, the response time and the actual response. Every hour, the FCC/SamKnows gateway performs at least ten queries to the ISP-configured DNS servers. For this analysis, we calculate the fraction of DNS queries that fail during each hour. To ensure that we are isolating DNS availability from access link availability, we discard hours during which the gateway recorded a loss rate above 1%. This corresponds to less than 3% of hours in our dataset. We classified hours where the majority of DNS queries failed (over 50%) as periods of DNS unavailability.

Figure 6 shows the probability of each provider experiencing one and two DNS server failures during a given hour. We sort providers in ascending order based on the probability that two servers will fail during the same hour.

Surprisingly, we find that many ISPs have a higher probability of two concurrent failures than a single server failing. For example, Comcast's primary and secondary servers are al-

most an order of magnitude more likely to fail simultaneously than individually.[3]

As one might expect, a reliable access link does not necessarily imply a highly available DNS service. For example, in our analysis of the reliability of access link itself, Insight was in the middle of the pack in terms of availability, offering only one nine of availability (Table 4), yet the results in Figure 6 show Insight having the lowest probabilities that queries to both DNS servers would fail simultaneously.

3.2.6 Longitudinal analysis

We close our analysis of broadband reliability with a longitudinal analysis of ISPs' reliability. With the exception of satellite, which started with very frequent periods of high loss, service reliability has remained more or less stable for most providers over the years.

Figure 7 shows the longitudinal trends for four ISPs in our dataset.[4] Though there were some year-to-year fluctuations, we did not find any consistent trends in terms of reliability over the course of the study.

AT&T showed some of the widest variations for DSL providers across multiple years. Other DSL providers, such as CenturyLink, Qwest, and Verizon (DSL) tended to be more consistent below the 10% loss rate threshold.

Cable providers tended to be the most consistent. Comcast, shown in Figure 7b, was particularly consistent over the years. Other cable providers such as Cablevision, Charter, and Cox were similar, though some did have a one year during which it recorded a slightly higher frequency of high loss rates.

The fiber services, including Verizon's fiber service (shown in Figure 7c), tended to be consistently more reliable than most other providers using other technologies, but did show some year-to-year variations. That said, there did not appear to be a trend (i.e., services were not getting consistently more or less reliable over time).

Both satellite providers in our dataset did tend to get better over time. Figure 7d shows the annual trend for Viasat. After having issues in 2011 and 2012, service reliability becomes much more consistent.

With our increasing reliance on broadband connectivity, the reliability of broadband services has remained largely stable. This highlights the importance for studies such as this and the need for techniques to improve service reliability.

4 Improving Reliability

Our characterization of broadband reliability has shown that even with a conservative definition of failure based on sustained periods of 10% packet loss, current broadband services achieve an availability no higher than two nines on average, with an average downtime of 17.8 hours per year. Defining availability to be less than 1% packet loss (beyond which many applications become unusable) leaves only a single provider of the 19 ISPs in the FCC dataset with two nines of availability.

Motivated by our findings of both poor reliability and the effect that this unreliabilty has on user engagement, we aim to improve service reliability by two orders of magnitude. This will bring broadband reliability to the minimum four nines required by the FCC for the public switched telephone service. Our solution should *improve resilience at the network level*, be *easy to deploy* and *transparent to the end user*.

- **Easy to deploy:** The solution must be low-cost, requiring no significant new infrastructure and the ability to work despite the diversity of devices and home network configurations. It should, ideally, be plug-and-play, requiring little to no manual configuration.
- **Transparent to the end user:** The solution should transparently improve reliability, "stepping in" only during service interruption. This transition should be seamless and not require any action from the user.
- **Improve resilience at the network level:** There have been proposals for improving the access reliability within particular applications, such as Web and DNS (e.g., [3, 50]). A single, network-level solution could improve reliability for all applications.

Towards this goal, we present a multihoming-based approach for improving broadband reliability that meets these requirements. Multihoming has become a viable option for many subscribers. The ubiquity of broadband and wireless access points and the increased performance of cell networks means that many subscribers have multiple alternatives with comparable performance for multihoming. In addition, several off-the-shelf residential gateways offer the ability to use a USB-connected modem as a backup link.[5] While the idea of multihoming is not new [1, 59], *we focus on measuring its potential for improving the reliability of residential broadband.*

We use active measurements from end hosts and the FCC's Measuring Broadband America dataset to evaluate our design. We find that (1) the majority of availability problems occur between the home gateway and the broadband access proider (§4.1); (2) multihoming can provide the additional two nines of availability we seek (§4.2); and (3) multihoming to wireless access points from neighboring residences can often dramatically improve reliability, even when the neighboring access point is multihomed to the same broadband access ISP (§4.3).

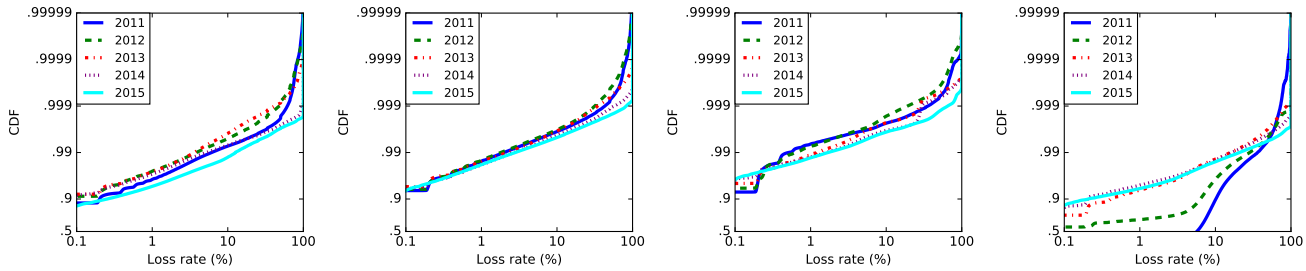
4.1 Where failures occur

We first study where the majority of broadband connectivity issues appear. We deployed a network experiment to approximately 6,000 endhosts running Namehelp [46] in November and December 2014. For each end host, our experiment ran two network measurements, a ping and a DNS query, at 30-second intervals. We chose to target our measurements to

[3] One possible explanation is the reliance on anycast DNS. We are exploring this in ongoing work.

[4] At the time of publication, only the first 8 months of 2016 were available on the FCC's website.

[5] For example, a wireless 3G/4G connection or a second fixed-line modem as in the case of the Asus' RT-AC68U [5].



(a) *AT&T* **(b)** *Comcast* **(c)** *Verizon (fiber)* **(d)** *Windblue/ViaSat*

Figure 7: *Longitudinal analysis of loss rates for four different ISPs.*

Farthest reachable point in network	Percent of failures
(1) Reached LAN gateway	68%
(2) Reached provider's network	8%
(3) Left provider's network	24%

Table 6: *Farthest reachable point in network during a connectivity issue, according to traceroute measurements.*

Google's public DNS service (i.e., 8.8.4.4 and 8.8.8.8). For this experiment, we considered this to be a sufficient test of Internet connectivity.

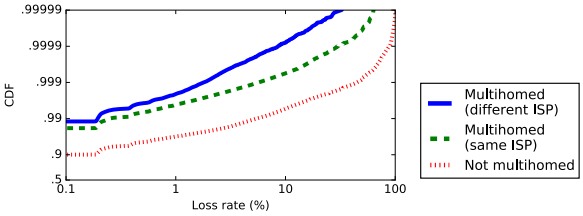
If neither ping nor a DNS query received a response, we immediately launched a traceroute to the target. If the traceroute did not receive a response from the destination, our experiment recorded the loss of connectivity and reported the traceroute results once Internet access had been restored. As in previous work [29], we used this traceroute data to categorize the issue according to how far into the network the traceroute's probes reached. Table 6 lists the farthest reachable point in the network during a connectivity interruption.

We find that most reliability problems occur between the home gateway and the service provider. During 68% of issues, our probes were able to reach the gateway, but not the provider's network. We cannot determine whether there was a problem with the access link, the subscriber's modem, or the gateway configuration, but in each case, we ensure that nothing had changed with the client's local network configuration (e.g., connected to the same access point and has the same local IP address) and that the probes from the client reached the target server during the previous test. Another 8% of traces were able to reach the provider's network, but were unable to reach a network beyond the provider's. The remaining 24% left the provider's network, but could not reach the destination server.

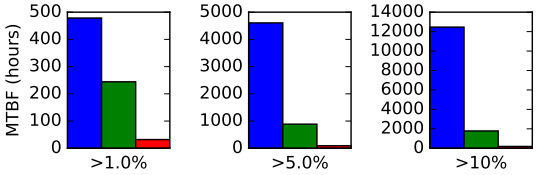
4.2 Broadband multihoming

Because the majority of service interruptions occur between the home gateway and the service provider, we posit that a second, backup connection—multihoming—could improve service availability.

To estimate the potential benefits of broadband multihoming for improving service reliability, we use the FCC dataset and group study participants by geographic region based on



(a) *Hourly loss rates*



(b) *MTBF*

Figure 8: *Hourly loss rates, and MTBF measured from each gateway and simulated multihomed connection.*

their Census Bureau block group. A Census block is the smallest geographical unit for which the Census Bureau publishes data, such as socioeconomic and housing details. Blocks are relatively small, typically containing between 600 and 3,000 people.[6] Unfortunately, we are not able to study trends at a finer granularity.

We identify blocks with at least two users online during the same time period. For each pair of users concurrently online in a region, we simulate a multihomed connection by identifying the minimum loss rate between the two connections during all overlapping time windows. We distinguished between simulated multihomed connections depending on whether both users subscribed to the same ISP.

Figure 8a shows the results of this experiment as a CDF of the loss rates reported for each simulated multihomed connection. As a baseline for comparison, we include the original reported loss rates for the same population of users, labeled "Not multihomed". For both types of simulated multihomed connections (same and different ISP), high packet loss rates are at least an order of magnitude less frequent. Furthermore,

[6] https://www.census.gov/geo/reference/gtc/gtc_bg.html

the benefits of multihoming with different ISPs as opposed to using the same ISP increase as the loss rate threshold increases. For example, using a 1% threshold as a failure, both scenarios provide two nines of reliability (99.59% when using the same ISP, 99.79% when using different ISPs). However, at 10% loss, multihoming on the same ISP provides only three nines (99.94%), while multihoming on different ISPs provides four nines (99.992%).

Figure 8b shows the average interval between periods of high packet loss rates, with thresholds of 1%, 5%, and 10%. Although both types of multihomed connections improve availability, as the loss rate threshold increases, the difference between connections multihomed on the same ISP and connections multihomed on different ISPs increases: with a 10% packet loss rate threshold, a multihomed connection using different ISPs provides four nines of availability, versus three nines for a connection multihomed on the same provider, and about two nines on a single connection.

4.3 Neighboring networks to multihome

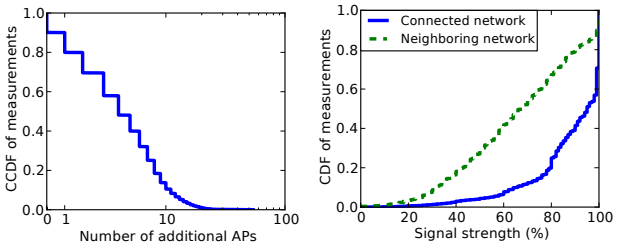
There are multiple ways that broadband subscribers could multihome their Internet connection. One possibility would be for users to subscribe to a cellular network service, adding to their existing wireless plan. This approach would be straightforward to implement, as users would only need to add a 4G dongle to their device. However, the relatively high cost per GB of traffic would likely be too expensive for most users, preventing them from using network-intensive services, such as video streaming.

An alternative, and cheaper, realization of our approach could adopt a cooperative model for multihoming between neighbors either through a volunteer model [24, 30] or a provider's supported community WiFi [37].[7]

To show the feasibility of this model, we used Namehelp clients to measure wireless networks between December 7, 2015 and January 7, 2016. For each user, every hour we recorded their current wireless configuration and scanned for additional wireless networks in the area using OS X's `airport` and Windows' `netsh.exe` commands.

One challenge to estimating the number of available APs is that, in many cases, an individual AP device will host multiple networks (e.g., 2.4 Ghz, 5 Ghz, and/or guest networks) using similar MAC addresses. To avoid overestimating the number of available APs, we used multiple techniques to group common SSIDs that appeared in our wireless scans. We first grouped MAC addresses that were similar to each other (i.e., string comparisons showed they differed in four or fewer hexadecimal digits or only differed in the 24 least significant bits). We then manually inspected these groups and removed any with an SSID that clearly did not correspond to a gateway, such as network devices and WiFi range extenders (e.g., SSIDs that contained "HP-Print", "Chromecast",

[7]Providers offering such services include AT&T, Comcast, Time Warner, British Telecom (UK) and Orange (France).



(a) **(b)**

Figure 9: *Number of additional APs available (a) and signal strengths for the current and strongest alternative AP (b).*

or "EXT"). We consider the AP groups remaining as gateway devices.

Figure 9a shows the CCDF of the number of additional unique groups seen across all measurements. Since we combine our findings at the end of this section with those of the previous section (§4.2), we only include measurements collected from clients within the US in our analysis. In 90.2% of cases, one or more *additional* wireless APs are available to the client. In approximately 80% of cases, two or more additional APs are available. These results highlight the potential for using nearby APs to improve service availability via multihoming.

The availability of neighboring AP is a necessary but not sufficient condition; a remaining concern is whether clients would actually be able to connect to these APs. Figure 9b shows a CDF of the signal strength percentage of both the AP to which the client is currently connected as well as the signal of the strongest available alternative network ("Neighboring network"). While the signal strengths of the neighboring networks are typically lower than that of the home network, it is still sufficiently strong in most cases, with a signal strength of 40% or higher for 82.7% of measurements.

Last, to estimate the potential improvement in service availability of using a neighboring AP as a backup connection, we infer the ISP of an AP by analyzing its SSIDs. For example, we found a large number of APs advertising SSIDs that clearly identify the provider, such as those starting with "ATT" and "CenturyLink". Similarly, we classified APs that hosted an "xfinitywifi" network in addition to other SSIDs as neighboring networks that belonged to Comcast subscribers. We were able to infer the ISP of at least one neighboring AP in 45% of all scans. Of these, 71% of APs appeared to belong to subscribers of an ISP different from that of the client.

In conjunction with the results from Section 4.2, these findings suggest that if clients used these additional APs for backup connections, service availability would improve by two nines in at least 32% of cases and by one nine in at least an additional 13% of cases. Since many APs advertised user-defined or manufacturer default SSIDs (e.g., "NETGEAR" or "linksys"), this is a lower bound estimate of the potential of improving service availability through multihoming. The next section presents a prototype system, *AlwaysOn*, that is based on this insight.

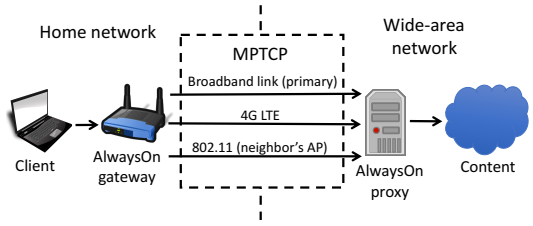
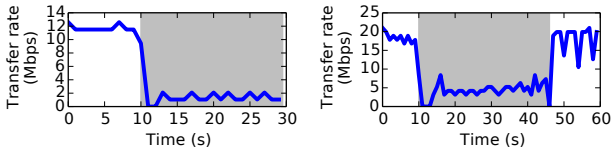


Figure 10: *Two* AlwaysOn *configurations using a neighboring AP or a 4G hotspot. The black solid line represents a client's normal path while the gray lines represent possible backup routes.*



(a) *Comcast / ATT (75 Mbps / 3 Mbps)*

(b) *RCN / VZW (150 Mbps / 4G LTE)*

Figure 11: *Throughput using* `iperf` *using* AlwaysOn *in two different network deployments. Each figure lists the service providers and speeds for the primary and secondary connections. The gray shaded section of the graph timeline represents the time during which the we simulated an outage on the primary link.*

5 AlwaysOn

In this section, we discuss various challenges associated with broadband multihoming; we describe how we address these concerns in our prototype service, *AlwaysOn*; and evaluate its performance.

5.1 Design Challenges

Multihoming a residential broadband connection presents different challenges than conventional multihoming. Whether failing over to a neighbor's wireless access point or to a 4G connection, a naïve implementation may interrupt the clients' current open connections and require re-opening others, because switching connections will cause the client to have a different source IP address for outgoing traffic. A broadband multihoming solution should be able to seamlessly switch between the primary and secondary connections without interrupting the user's open connections.

Broadband multihoming also introduces concerns related to usage policies and user privacy. First, some backup connections (e.g., 4G) may have data caps. A common broadband use like streaming ultra HD videos may have to be restricted over those connections considering their cost. Neighbors sharing connections with each other may also prefer to impose limits on how they share their connection, for instance, in terms of available bandwidth, time or total traffic volume per month. Second, in locations where there is more than one alternative backup connection, users may want to state their preference in the order of which networks to use based on factors such as the structure of their sharing agreement or the amount of wireless interference on a network's frequency. Finally, there are privacy concerns for both parties when multihoming using a neighbor's network. Users that "borrow" their neighbor's network would, by default, be allowing their neighbor to capture their unencrypted traffic. Conversely, neighbors who are "loaning" access to their network should not have to compromise their own privacy to do so.

5.2 AlwaysOn Design & Implementation

AlwaysOn has two components: a client running in the gateway and a proxy server deployed as a cloud service. A diagram of this deployment is shown in Figure 10. The additional lines in this figure represent a backup paths via a neighboring AP and another through a 4G hotspot.

AlwaysOn uses Multipath TCP (MPTCP) [47] to transition connections from primary to secondary links without interrupting clients' open connections. The AlwaysOn Gateway creates an encrypted tunnel to the MPTCP-enabled AlwaysOn Proxy. All traffic from the private LAN is routed via this tunnel. Our current implementation uses an encrypted SOCKS proxy via SSH; a virtual private network (VPN) tunnel would be an alternate implementation. Using an encrypted tunnel ensures that user traffic remains confidential when it traverses the backup path. The MPTCP-enabled proxy can be shared between multiple users, with each user being assigned (after authentication) to a unique tunnel.

Additionally, the AlwaysOn gateway sends the traffic via a guest network that is isolated from the private LAN when sharing its connection; many commodity residential access points already offer guest network configuration to enable connection sharing while limiting a guest's access to the local network. Deploying AlwaysOn requires an MPTCP-enabled kernel; although home network testbed deployments (e.g., BISmark) do not run such a kernel, OpenWrt can be built with mptcp support.

AlwaysOn allows users to configure traffic shaping settings to facilitate resource sharing across connections. Options that concern traffic shaping must be synchronized between the gateway and proxy. The gateway can shape outgoing traffic, but incoming traffic must be shaped on the AlwaysOn Proxy. Our current prototype uses `tc` and `iptables` to enforce traffic management policies. For outgoing traffic, the AlwaysOn gateway can throttle traffic traversing the neighboring access point, as well as traffic on its own guest network. Each user has unique port number at the gateway to use for their tunnel, and the IP address serves to identify traffic to or from secondary links. Using `iptables`, the gateway and proxy mark traffic according to whether it corresponds to a primary or secondary connection. They then use `tc` to apply the appropriate traffic shaping policy.

A user must currently manually configure policies such as link preference and traffic shaping at the AlwaysOn Gateway and proxy; this manual configuration poses a problem when a user needs to configure settings such as link preference on devices that they may not necessarily control. We are exploring alternatives to realize this through a third-party service that accepts, encodes, and enforces such policies on outgoing and incoming traffic.

5.3 AlwaysOn Evaluation

We ran multiple experiments to evaluate the AlwaysOn prototype in various network settings. We instantiated an AlwaysOn proxy server on a university network. We aim to evaluate AlwaysOn in two operating modes: (1) during the failure of the primary link; and (2) during normal operation. Routing traffic via the AlwaysOn proxy should not affect performance during normal operation, considering that even the least reliable service still has about 36 hours of downtime each month. In addition, to limit the effect of outages on user quality of experience, AlwaysOn should respond quickly and route traffic via the backup link as soon as a failure is detected.

Reaction to network failures. In our first experiment, we test AlwaysOn's ability to react to network failures. We ran `iperf` for 30 seconds from a client behind the AlwaysOn Gateway, recording `iperf`'s measured throughput rate each second. We emulated different outages, represented in the plots by time periods highlighted in gray.

We ran this experiment in two different scenarios for our evaluation, shown in Figure 11. In the first scenario (Figure 11a), we used a Comcast 75 Mbps service as the primary connection and a 3 Mbps AT&T service as the secondary connection. In the second scenario, shown in Figure 11b, we used an RCN 150 Mbps service as the primary, and a Verizon Wireless 4G LTE hotspot as the secondary connection. For this test, the primary connection was re-enabled after approximately 35 seconds. Once the primary connection was reestablished, AlwaysOn switches traffic back to the RCN connection.

In each case, AlwaysOn can recover relatively quickly once it realizes the primary link is no longer working, and does not require the connection to be reestablished. We also ran `iperf` over each connection between the same client and server without the AlwaysOn gateway and proxy and consistently measured similar throughput rates. The relatively slow performance compared to the access link speed in Figures 11a and 11b is likely due to other limiting factors such as end-to-end latency, congestion on the path, and only using a single TCP connection.

Application performance during access network failures. We also evaluated how AlwaysOn was able to handle service outages while clients used an OTT service. We ran tests using both Netflix and HTTP Live Streaming (HLS) services. For each service, we tested how outages affected playback of the stream for a non-multihomed and multihomed configurations (using AlwaysOn and two 150 Mbps RCN connections).

Figure 12 shows the playback of a Netflix stream for both configurations. In this test, we streamed a video for one minute, allowing the stream's buffer to stabilize at ≈220 seconds. We then simulated a five minute outage on the primary link. Without multihoming, any outage lasting longer than about 3.5 minutes resulted in the buffer being drained completely, interrupting playback of the stream. Once the connection is restored, the non-multihomed host is able to quickly

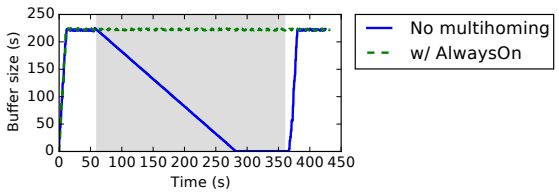
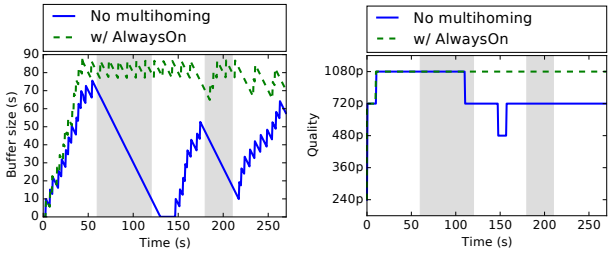


Figure 12: *Two Netflix video streaming sessions, one without any multihoming and one while using AlwaysOn. The gray highlighted section represent simulated network outages on the primary link.*



(a) *Buffer size* **(b)** *Video quality*

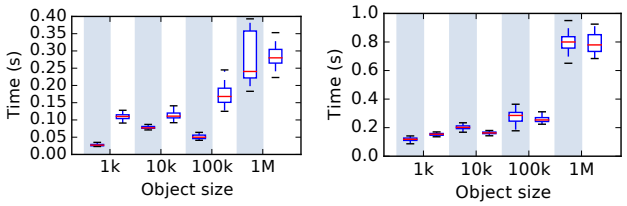
Figure 13: *Comparison of two HLS video sessions, one without multihoming and one while using AlwaysOn. The gray highlighted sections represent simulated network outages on the primary link.*

resume the stream. In the experiment using AlwaysOn, the stream is uninterrupted as the traffic was able to seamlessly be routed via the secondary connection.

Figure 13 shows the result of this experiment using HLS. In this test, we evaluated the impact of shorter outages. Figure 13a shows the size of the buffer; Figure 13b shows the current video quality being displayed to the user. In both configurations, we simulated two shorter outages that were one minute apart, the first lasting for 60 seconds and the second lasting for 30 seconds.

In both scenarios, the buffer is able to fill quickly during the initial 60 seconds of the stream, with playback quality quickly increasing to 1080p. Without AlwaysOn, the buffer is completely drained during the first outage and the stream is interrupted. It then resumes the stream at 480p for a short period. The second outage then forces the stream to stay at 720p for the remainder of the stream. In contrast, the AlwaysOn configuration maintains quality at 1080p shortly after initializing the stream, even during the outage period.

Performance overhead of AlwaysOn. To see how our AlwaysOn proxy affected network performance, we also measured the time to fetch objects hosted on Akamai's CDN, both when using and not using the proxy. For this test, we downloaded files of varying sizes (1 kB, 10 kB, 100 kB, and 1 MB) 100 times. The box plots shown in Figure 14 summarize the distribution of download times for objects of each size while using the RCN (Figure 14a) and Verizon Wireless (Figure 14b) connections. The highlighted box plots show the fetch times while using the proxy for each respective file size. Clients who forwarded traffic via AlwaysOn experienced similar performance; in some cases, we found that download times actually improved while using the AlwaysOn proxy, since clients were directed to a much faster replica when using the proxy.



(a) *RCN (150 Mbps)*

(b) *VZW (4G LTE)*

Figure 14: *Box and whisker plots showing the time to fetch objects hosted by Akamai while using (highlighted) and not using the AlwaysOn proxy. for RCN and Verizon Wireless (VZW). Each box and whisker represents the median, interquartile range (IQR), and 1.5 IQR for each experiment configuration.*

6 Related Work

Previous work on broadband networks has often focused on characterizing services in terms of performance (e.g., link capacity and latency) from a range of platforms and vantage points, including customized home gateways [56, 60], applications in end-user devices [12, 21, 46, 58], Web-based tests [15, 38, 54], and well-provisioned measurement nodes outside the access networks [22]. We use longitudinal data collected by two of these efforts, the FCC's MBA initiative and Namehelp [46], to study broadband reliability.

Reliability of phone networks. A number of efforts have tackled reliability characterization in other contexts, such as telephone and cellular networks. Thanks in part to the pioneer work at Bell Labs [20] by the end of the 20th century, public switched telephone networks (PSTN) had become so reliable that AT&T expected no more than two hours of failure over a 40-year period [40]. Today the FCC requires that PTSN providers document and report outages affect more than 30,000 users or last longer than 30 minutes [25] which corresponds with at least four nines of availability.

Effect of network factors on user behavior. Recent work has explored the effect of network factors on user experience with applications, including VoIP [16], Web [6] and Internet video [7]. Rather than deriving a model for user experience based on multiple factors, our work focuses on the effect of reliability on user demand. Others have started to explore the use of alternative experimental designs to evaluate user experience. Krishnan et al. [39] apply quasi-experimental design to evaluate the effect of video stream quality on viewer behavior, Oktay et al. [45] relies on it for causal analysis of user behavior in social media. Bischof et al. [11] explores the effect of contextual factors such as price and competition on user demand. We apply similar methods to understand the effect of service reliability on user behavior.

Reliability of broadband providers. Baltrūnas et al. [8] presented a study of the reliability of four mobile broadband providers in Norway using the Nornet Edge dedicated infrastructure. This work illustrates the value of end-to-end measurements to identify failures and performance events not always captured by the operators' monitoring systems. Broadband reliability has received little attention until recently. Lehr et al. [41] discuss some of the challenges of characterizing reliability and their economic and policy implications and identify three different ways in which the "reliability" of broadband services can be measured: (1) the reliability of the service itself; (2) the reliability of network services offered by the ISP (e.g., DNS); and (3) the consistency of the service's performance. We focus on characterizing reliability in terms of the former two categories and study the third in previous works [9].

Multihomed access networks. Beyond improvements in access link technology, one way to enhance the reliability of access networks is through redundancy. Gummadi et al. [32] propose a detouring approach to recover from Internet path failures. Andersen et al. improved web availability with their system, MONET, an overlay network of multihomed proxy servers [3]. We have seen the recent introduction of consumer-grade residential gateways that support a second WAN connection (such as a 3G or 4G modem) [5] and some work exploring the performance benefits of the on-loading of broadband traffic using a 3G connection [55]. Detal et al. presented MiMBox, a system for translating between TCP and MPTCP connections at the middlebox [19]. Other works have explored the benefits of using MPTCP on mobile devices [17, 48] and the possibility of bonding multiple access links (such as DSL and cable) to increase performance [33, 52, 61].

7 Conclusion

As broadband performance and availability continue to improve and users migrate services to over-the-top alternatives, reliability will become the dominant feature in the evaluation of broadband services.

We empirically demonstrated the importance of broadband reliability on users' quality of experience. We presented an approach for broadband reliability characterization using data collected by common national efforts to study broadband and discussed key findings from applying it to four-year dataset collected through the FCC's MBA program. Motivated by our findings on both the importance of reliability and the current reliability of broadband services, we presented the design, implementation, and evaluation of AlwaysOn, practical approach to improving broadband reliability through multihoming. There are a number of promising research directions for future work from alternative metrics for broadband service reliability to alternative approaches to study the effect of reliability on user behavior. To facilitate reproducibility and use of the system, we have publicly released our dataset, analysis scripts, and *AlwaysOn* prototype [2].

8 Acknowledgements

This research was supported in part by the National Science Foundation through Award CNS 1619317. This work was also partially supported by the JSPS fellowship program.

References

[1] A. Akella, B. Maggs, S. Seshan, A. Shaikh, and R. Sitaraman. A measurement-based analysis of multihoming. In *Proc. of ACM SIGCOMM*, 2003.

[2] AlwaysOn. Broadband reliability project. https://github.com/bb-alwayson/alwayson.

[3] D. G. Andersen, H. Balakrishnan, M. F. Kaashoe, and R. Rao. Improving web availability for clients with MONET. In *Proc. of USENIX NSDI*, 2005.

[4] Apple. HTTP Live Streaming. https://developer.apple.com/streaming/.

[5] Asus. RT-AC68U. https://bit.ly/1Bx3Pj5.

[6] A. Balachandran, V. Aggarwal, E. Halepovic, J. Pang, S. Seshan, S. Venkataraman, and H. Yan. Modeling web quality-of-experience on cellular networks. In *Proc. of MobiCom*, 2014.

[7] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang. Developing a predictive model of quality of experience for Internet video. In *Proc. of ACM SIGCOMM*, 2013.

[8] D. Baltrūnas, A. Emokashfi, and A. Kvalbein. Measuring the reliability of mobile broadband networks. In *Proc. of IMC*, 2014.

[9] Z. S. Bischof, F. E. Bustamant, and R. Stanojevic. The utility argument – making a case for broadband SLAs. In *Proc. of PAM*, March 2017.

[10] Z. S. Bischof and F. E. Bustamante. A time for reliability - the growing importance of being always on. In *Proc. of ACM SIGCOMM*, 2014. Poster.

[11] Z. S. Bischof, F. E. Bustamante, and R. Stanojevic. Need, want, can afford – broadband markets and the behavior of users. In *Proc. of IMC*, November 2014.

[12] 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.

[13] Broadband Commission. The state of broadband 2012: Achieving digital inclusion for all. https://bit.ly/2nE8Nde.

[14] Broadband Commission. The state of broadband 2014: broadband for all. https://bit.ly/1rqThxU.

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

[16] K.-T. Chen, C.-Y. Huang, P. Huang, and C.-L. Lei. Quantifying Skype user satisfaction. In *Proc. of ACM SIGCOMM*, 2006.

[17] Y.-C. Chen, Y.-s. Lim, R. J. Gibbens, E. M. Nahum, R. Khalili, and D. Towsley. A measurement-based study of multipath TCP performance over wireless networks. In *Proc. of IMC*, 2013.

[18] E. David Belson. The state of the Internet. Technical report, Akamai, 2015.

[19] G. Detal, C. Paasch, and O. Bonaventure. Multipath in the middle(box). In *Proc. of HotMiddlebox*, 2013.

[20] B. Dhillon. *Reliability, Quality, and Safety for Engineers*. CRC Press, 2004.

[21] L. DiCioccio, R. Teixeira, and C. Rosenberg. Characterizing home networks with HomeNet profiler. Technical Report CP-PRL-2011-09-0001, Technicolor, September 2011.

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

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

[24] Electronic Frontier Foundation. Open Wireless Movement. https://openwireless.org/.

[25] P. Enriquez, A. Brown, and D. A. Patterson. Lessons from the PSTN for dependable computing. In *Proc. of Workshop on Self-Healing, Adaptive and self-MANaged Systems*, 2002.

[26] FCC. 2013 measuring broadband America February report. https://bit.ly/2KWkcON.

[27] FCC. In the matter of reliability and continuity of communication networks. PS Docket 11-60.

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

[29] N. Feamster, D. G. Andersen, H. Balakrishnan, and M. F. Kaashoek. Measuring the effects of internet path faults on reactive routing. In *Proc. ACM SIGMETRICS*, 2003.

[30] Fon. How it Works — Fon. https://bit.ly/1qDhUr5.

[31] Frost & Sullivan. The advanced capabilities of VoIP and SIP trunking exceeds traditional voice services, February 2014. http://tinyurl.com/z349mxt.

[32] K. P. Gummadi, H. V. Madhyastha, S. D. Gribble, H. M. Levy, and D. Wetherall. Improving the reliability of internet paths with one-hop source routing. In *Proc. USENIX OSDI*, 2004.

[33] A. Habib, N. Christin, and J. Chuang. On the feasibility of switching isps in residential multihoming. In *Proc. of the IEEE International Workshop on Quality of Service*, 2007.

[34] S. Isaacman and M. Martonosi. The C-LINK system for collaborative web usage: A real-world deployment in rural Nicaragua. In *Workshop on Networked Systems for Developing Regions*, 2009.

[35] ITU. The impact of broadband on the economy. https://bit.ly/2nE7WJy, 2012.

[36] ITU. New global broadband study: national plans and competitive markets are crucial. https://bit.ly/2KWl2ep, July 2013.

[37] D. Jha, J. P. Rula, and F. E. Bustamante. exploring xfinity: A first look at provider-enabled community networks. In *PAM*, 2016.

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

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

[40] D. R. Kuhn. Sources of failures in the public switched telephone network. *IEEE Computer*, pages 31–36, April 1997.

[41] W. Lehr, S. Bauer, M. Heikkinen, and D. Clark. Assessing broadband reliability: Measurement and policy challenges. In *Proc. of the Research Conference on Communication, Information and Policy Challenges*, 2011.

[42] M. Mathis, J. Semke, J. Mahdavi, and T. Ott. The macroscopic behavior of the TCP congestion avoidance algorithm. 27(3), 1997.

[43] S. Networking. The Holy Grail of five-nines reliability. https://goo.gl/3JZRPG.

[44] Office of Communication (Ofcom). The consumer experience. Technical report, Ofcom, London, UK, January 2015.

[45] 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.

[46] J. S. Otto, M. A. Sánchez, J. P. Rula, and F. E. Bustamante. Content delivery and the natural evolution of dns: Remote DNS trends, performance issues and alternative solutions. In *Proc. of IMC*, 2012.

[47] C. Paasch and O. Bonaventure. Multipath TCP. *Queue*, Feb. 2014.

[48] C. Paasch, G. Detal, F. Duchene, C. Raiciu, and O. Bonaventure. Exploring mobile/wifi handover with multipath TCP. In *Proc. of CellNet*, 2012.

[49] J. Padhye, V. Firoiuy, D. Towsley, and J. Kurose. Modeling TCP throughput: A simple model and its empirical validation. In *Proc. of ACM SIGCOMM*, 1998.

[50] K. P. Park, V. S. Pai, L. Peterson, and Z. Wang. CoDNS: Improving DNS performance and reliability via cooperative lookups. In *Proc. USENIX OSDI*, 2004.

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

[52] B. Peirens, Proximus, G. Detal, S. Barre, O. Bonaventure, and Tessares. Link bonding with transparent multipath TCP. https://bit.ly/2nBAYcK.

[53] PricewaterhouseCoopers. Global entertainment and media outlook 2016-2020. http://www.pwc.com/us/em/outlook?_ga=1.63986735.1233613438.1485541217.

[54] A. Ritacco, C. Wills, and M. Claypool. How's my network? A Java approach to home network measurement. In *Proc. of IEEE ICCCN*, 2009.

[55] C. Rossi, N. Vallina-Rodriguez, V. Erramilli, Y. Grunenberger, L. Gyarmati, N. Laoutaris, R. Stanojevic, K. Papagiannaki, and P. Rodriguez. 3GOL: Power-boosting ADSL using 3G onloading. In *Proc. ACM CoNEXT*, 2013.

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

[57] SamKnows. Our regulatory clients. https://www.samknows.com/regulators.

[58] 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.

[59] P. Smith. Bgp multihoming techniques. goo.gl/gVKYwm, October 2004.

[60] 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.

[61] N. Thompson, G. He, and H. Luo. Flow scheduling for end-host multihoming. In *Proc. IEEE INFOCOM*, 2006.

[62] United States Census Bureau. 2010 census urban and rural classification and urban area criteria, 2010. https://bit.ly/1qtZ9Is.

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

[64] A. K. Xiao Sophia Wang, Aruna Balasubramanian and D. Wetherall. Demystifying page load performance with wprof. In *Proc. of USENIX NSDI*, 2013.

[65] Y. Xu, C. Yu, J. Li, and Y. Liu. Video telephony for end-consumers: measurement study of Google+, iChat, and Skype. In *Proc. of IMC*, 2012.

[66] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. The increased bandwidth fallacy: performance and usage in rural Zambia. In *Proc. of the 4th Symposium on Computing for Development*, 2013.