

# something

Paper #0, 2 Pages

## Abstract

FCC reclassifying broadband and broadband speeds  
HD content, video traffic, CDNs  
How should ISPs progress?  
Want to know the current utilization during peak periods  
Better way to offer broadband

## 1 Introduction

- FCC hypothesis
- Answer from an ISP perspective
- Current usage

As part of this work, an experiment, 105 to 250 without telling users. Aim was to find out if their usage changes. – SG

FCC Reclassify – SG

## 2 Background and Motivation

Previous work states as capacity increases, usage increases, but we feel usage is a function of 3 entities: user, ISP, service provider, and with bb reclassification, the ISP is taking itself out of the loop? – SG

aim is to compare usage in high and low tier. So do an experiment with Comcast 105 to 250 without telling users – SG

### 2.1 FCC

report states [1] : (relevant questions come here) – SG

FCC hypothesis comes here – SG

- H1
- H2
- H3

### 2.2 User

broadband plan based on utilization? – SG

### 2.3 ISP

Questions asked - how much do users need now? – SG s

### 2.4 Content

Content providers

Rise in traffic video - refer FCC reports and sandvine  
Netflix CDN, Google CDN

pouring traffic into thin pipes - can't be free peering – SG

## 3 Data Source

Our dataset consists of network usage byte counters reported by Comcast gateways every 15 minutes from October 1, 2014 to December 29, 2014. There are two sets of broadband tiers that were used to collect this data: controlset, consisting of homes and businesses with a 105 Mbps access link, and the testset, consisting of homes and businesses that were paying for a 105 Mbps access link, yet were receiving 250 Mbps instead. Users in the test set were selected randomly and were not told that their access bandwidth has been increased. There were more than 15000 gateway devices in the control set, with varying usage over the three months, and about 2200 gateway devices in the test set. **TODO: confirm - these were reported by Comcast gateways right?**

Both the testand controlsets were collected from users in Salt Lake City, Utah, to avoid any biases in behavior based on location. Although this dataset corresponds to just one ISP, we believe that it is broadly representative of urban users in the US in the same, or higher broadband bandwidth tier ( 100 Mbps). Thus, we use this data to draw general conclusions about behavioral change with link capacity **TODO: (add more here...)**

Supplement the data with bismark? – SG

My Speed Test Usage Patterns data - Any chance? – SG

### 3.1 Data Description

Comcast splits the controlset into 8 separate pools on different date ranges and gateways **TODO: confirm if there is repeated device IDs in controll-8.** . Each dataset contains the following relevant fields: Device ID, sample period time, service class, service direction, IP address, and the bytes transferred in the 15 minute sample slot, as described in table 1. **TODO: find out more about service class name, and IP addresses being the same across all sets**

Field	Description
Device_number	Arbitrarily assigned CM device identifier
end_time	Fifteen minute sample period end time
date_service_created	Service start (not used in our analysis)
service_class_name	Used to differentiate data application
cmts_inet	Cmts identifier (derived from ip address)
service_direction	1-downstream, 2-upstream
port_name	Cmts port descriptor
octets_passed	Byte count
device_key	not used in our analysis
service.identifier	Service id (not used in our analysis)

**Table 1:** Field Descriptions for Comcast Dataset by Comcast

## 3.2 Data Processing

To process the large amount of data (more than 15000 unique devices, 96 time slots per day, 3 months, multiple service classes per time-slot), we first split the data by direction into uplink and downlink. The nature of the questions we ask in section 2 encourages us to concentrate on the downstream data, although we present similar results for upstream data in section 4. We also do not use the service class name identifier in our analysis, which is internal to Comcast. **TODO: Our analysis showed that ignoring the service class identifier does not change our conclusions on usage patterns and statistics.** Thus we only considered overall usage for each gateway device in each time slot, in uplink and downlink direction.

Note that 15 minute time slots were synchronized to the same time stamp. They were off only by a few seconds (under 30, as claimed by Comcast). As our analysis deals with aggregated patterns on a granularity of 15 minutes, we believe this time synchronization to be irrelevant.

## 3.3 Data Sanitization

correlated drops at times or devices that contributed only 2-3 days of the whole month – SG

remove machines that contribute to less than 0.8 of the time slots. – SG

**TODO: fig 1: Make common plot of availability – 8 control sets (half filtered) vs availability.**

final table of set, num gateways, date range – SG

after sanitization, we can either compare test to each control in the same time range and draw agg conclusions, or do it by month, or combo it into a huge control set and compare it with test set – SG

Present results for full sets, unless there is a significant difference in any trends or distributions – SG

## 4 Results

Q/A  
Plots

## 5 Discussion

results that are contradictory  
results that data suggests that is wrong  
- suggest that this may be due to representation  
a better way to measure and offer broadband based on utilization?

## 6 Related Work

fcc reports + sandvine on usage patterns  
papers studying broadband vs utilization  
this can go in the intro if lack of space

## References

- [1] Federal Communications Commission. Measuring Broadband America - 2014, April 2014. (Cited on page 1.)