Broadband Deployment Data - Moving Beyond Form 477

Henning Schulzrinne

(with Columbia DSI capstone teams: Aaron Sadholz, Aman Varma Mantena, Anna Zhou, E.K. Ito, Robert (Davis) DeRodes, Zixuan (Armstrong) Li, Yao Li, Bing Xu)

December 13, 2018

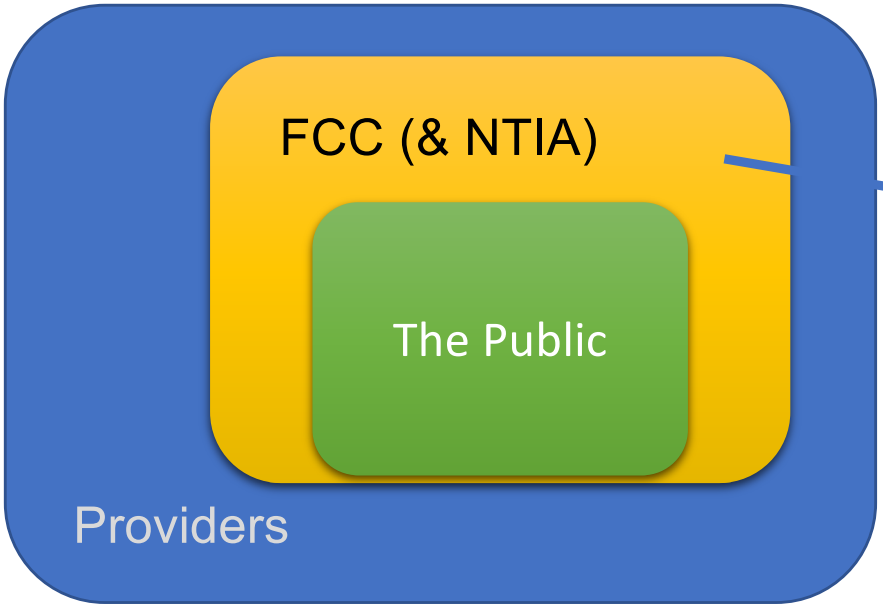
It’s the law (Section 706 aka 47 USC 1302)

(a) In General: The Commission and each State commission with regulatory jurisdiction over telecommunications services shall encourage the deployment on a reasonable and timely basis of advanced telecommunications capability to all Americans (including, in particular, elementary and secondary schools and classrooms) by utilizing, in a manner consistent with the public interest, convenience, and necessity, price cap regulation, regulatory forbearance, measures that promote competition in the local telecommunications market, or other regulating methods that remove barriers to infrastructure investment.

We have questions

- Where is broadband available?
  - Not just residential, also business-grade (e.g., >= 1 Gb/s)
  - What predicts deployment – we found that road miles/population and elevation differences are good predictors
- How well does it work?
  - Reliability, actual performance
  - including for home Wi-Fi → often effectively limits performance to 30 Mb/s
- Where would it get deployed on its own, "naturally"? By whom?
  - Can we predict this?
- How effective are USF subsidies?
- How much competition is there?
- What is the average data usage, for different types of users?
  - mobile, satellite, wireless, wireline, …
- How much does it cost?
  - Including in various bundles
- Who is adopting fixed (wireline and wireless) broadband? Who is not and why not?

The Russian- doll information model



Report the total number of in-service connections for each … unique combination of technology of transmission, downstream bandwidth, and upstream bandwidth.

- Form 477: Provider, technology, max. speed at census block level
- MBA: roughly 100 nodes per service tier (goal) for performance
- ACS: broadband usage (5 years, tract)
- Pew Internet surveys

Form 477

1654124,30510,0004325205,Monmouth Telephone & Telegraph,Monmouth Telephone & Telegraph,Monmouth Internet Corporation,170067,Monmouth Internet Corporation,NJ,340030280022002,30,0,0,0,1,1.5,1.5
1654125,30510,0004325205,Monmouth Telephone & Telegraph,Monmouth Telephone & Telegraph,Monmouth Internet Corporation,170067,Monmouth Internet Corporation,NJ,340030280022002,50,0,0,0,1,100,100
7479256,31677,0003316692,Verizon New Jersey Inc.,Verizon New Jersey Inc.,Verizon Communications Inc.,131425,Verizon Communications Inc.,NJ,340030280022002,50,1,940,880,1,0,0
11543559,32487,0025646373,"Charter Communications, Inc.",Charter Communications Inc,Charter Communications,130235,Charter Communications,NJ,340030280022002,42,1,300,20,1,0,0
18892016,33149,0004963088,"ViaSat, Inc.",ViaSat Inc,"ViaSat, Inc.",290111,"ViaSat, Inc.",NJ,340030280022002,60,1,25,3,1,0,0
55287474,39920,0001568880,GCI Communication Corp.,GCI Communication Corp.,"General Communication, Inc.",130534,"General Communication, Inc.",NJ,340030280022002,60,0,0,0,1,0,0
55447503,33379,0012369286,"HNS License Sub, LLC",HughesNet,"dishNET Holding, LLC",130627,"dishNET Holding, LLC",NJ,340030280022002,60,1,25,3,1,0,0
55607532,30279,0018756155,"VSAT Systems, LLC",Skycasters,"VSAT Systems, LLC",300167,"VSAT Systems, LLC",NJ,340030280022002,60,1,2,1.3,1,2,1.3

Example 1: we can predict deployment

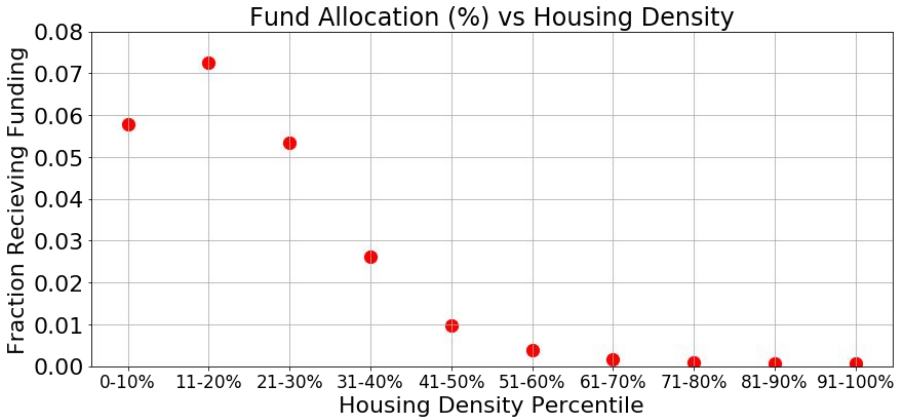


Figure 2. Fraction of census blocks receiving broadband funding across US housing density. Rural blocks are most likely to receive funding.

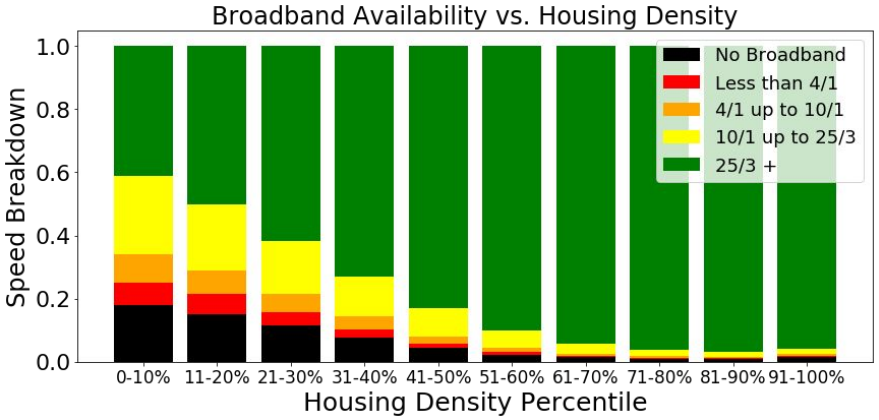
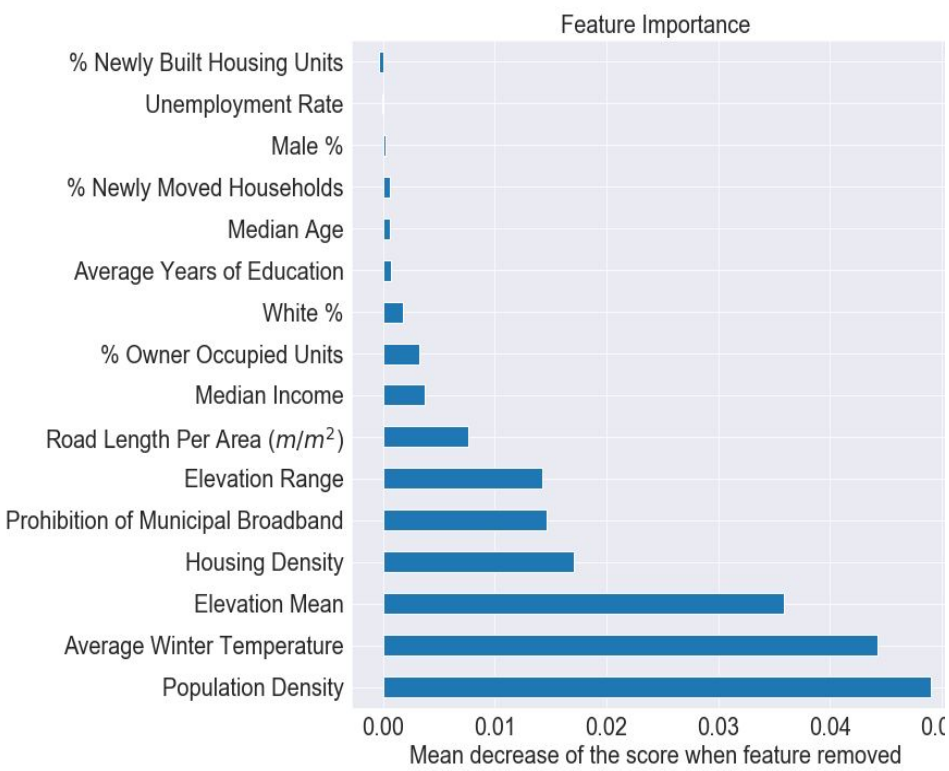
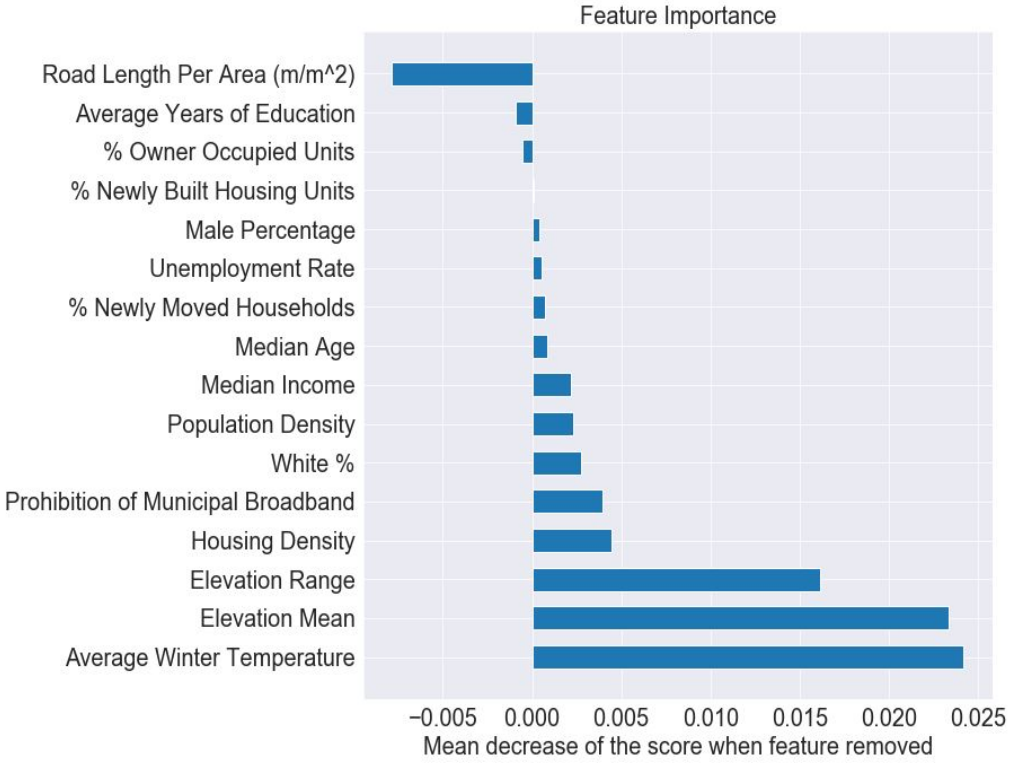


Figure 1. Broadband Speed & Access Across US housing density (download/upload speed in Mbps). 4/1 is considered the minimum viable speed.

Example 1: Organice vs. Funded Expansion

Organic vs. Funded Expansion Models





Figures 3 & 4. Feature importance of the organic (left) & funded (right) expansion models using permutation importance method.

- 165 GB = 810 million rows of data → Google BigQuery
- ROC AUC scores of 0.85 for organic (gradient boosting) & 0.83 for funded (random forest)

Example 2: Broadband correlates with income gains & median home values

Economic Indicator	Model	Percent Increase	p-value
Median Home Value	Linear	3.04%	<.0005
Median Home Value	Fixed Effects	0.30%	0.009
Median Home Value	Mixed Effects	0.11%	<.0005
Median Household Income	Linear	2.53%	<.0005
Median Household Income	Fixed Effects	0.20%	0.187
Median Household Income	Mixed Effects	0.98%	<.0005

Figure 4. Impact of Access to Broadband with at least 10 Mbps Download Speed on Rural Block Groups

Economic Indicator	Model	Percent Increase	p-value
Median Home Value	Linear	17.46%	<.0005
Median Home Value	Fixed Effects	1.30%	<.0005
Median Home Value	Mixed Effects	5.42%	<.0005
Median Household Income	Linear	9.93%	<.0005
Median Household Income	Fixed Effects	0.62%	<.0005
Median Household Income	Mixed Effects	5.89%	<.0005

Figure 5. Impact of Access to Fiber Broadband on Urban Block Groups

The current data is problematic

- Form 477 (broadband availability):
  - Guaranteed not to underestimate availability
    - not just rural – apartment buildings may not allow entrants or FiOS
  - One connection at highest speed → whole block (11,166,336 total)
    - some are quite large (8,500 sq miles); median: 6.4 acres
  - Form 477: Weird effects – broadband disappears, then reappears → data consistency analysis
  - May not actually have availability (DSLAM full)
  - Only starts in 2014, with earlier data not comparable
  - Mapping providers and locations from USAC to Form 477 not easy
  - Provider names change year-over-year
- Census ACS:
  - 2013-2017 5-year estimates → data quality problems
  - Broadband subscribership fraction down to census tract
  - but no speed tiers
- MBA data:
  - limited sampling for smaller geographic regions (4,545 samples for 2016)
  - only large providers (14), but covers 80%+ of consumers
  - data reported with significant delay (Sept. 2016 published now)
  - mobile data never published

Example: ACS (Bergen County, NJ)

	Percent
Cellular data plan	50.7
Broadband such as cable, fiber optic, or DSL	81.7
Satellite	2.9
Dial-up alone	0.3

	Percent
Cellular data plan	53.7
Broadband such as cable, fiber optic, or DSL	82.5
Satellite	2.0
Dial-up alone	0.4
Other service alone	0.1

No pricing data

- There **is** pricing data (sampled) for cable TV (mandated)
- Unclear how factored into CPI (BLS hedonic model)
- Some bill sampling available commercially
- Cannot readily model influence of cost on adoption
  - or pricing by different types of providers (cable vs. REC)
  - or impact of competition

Table 4
Historical Price Series
2006–2017

Year	Basic Service Price	Expanded Basic Service		
		Price	Channels	Price per Channel
2006	$14.59	$45.26	71.0	$0.650
2007	$15.33	$47.27	72.6	$0.670
2008	$16.11	$49.65	72.8	$0.680
2009	$17.65	$52.37	78.2	$0.710
2010	$17.93	$54.44	117	$0.560
2011	$19.33	$57.46	124.2	$0.569
2012	$20.55	$61.63	149.9	$0.505
2013	$22.63	$64.41	159.6	$0.484
2014	$22.78	$66.61	167.3	$0.496
2015	$23.79	$69.03	181.3	$0.456
2016	$25.40	$71.37	181.0	$0.469
2017	$25.06	$75.21	195.1	$0.487

Data vision

- Integrated data: availability, pricing, usage, performance
  - common timing
  - overlapping tools
- Performance data is not that hard → build into home routers
  - remote-control sampling → reachability during large-scale events
- Gather data on billing from representative sample
  - at least annually
- Actual availability
  - need street address data
  - easy (or easier): provider provisioning data
    - some engineering uncertainty
  - harder: self-selected survey (see PA) or door-to-door sampling
  - or competitive challenge (e.g., if below 50% served in tract)
    - prove no service → prove service availability