MAPPING BROADBAND SPEEDS AROUND THE NATION

Elizabeth Belding holds a smart phone while performing a speed test to check the quality of her broadband connection. Accessing and interpreting crowd-sourced speed-test data is a key element of the three related research projects discussed here.

IN A TRIO OF PROJECTS, ELIZABETH BELDING SEEKS GRANULAR UNDERSTANDING OF COVERAGE

Reliable high-speed internet access and dependable cellular service have become essentials of modern life. Increasingly, we need them to access education, health care, financial services, travel opportunities, entertainment, emergency services, government information, directions to our destinations, and so much more.

But not everyone has the level of service they need to tap into those resources, and inequalities abound. For instance, The Markup — a nonprofit news organization that investigates "how powerful institutions are using technology to change our society" — looked at data speeds offered around the United States. They found that, often, around the country, slower speeds are provided to non-White people living in low-income neighborhoods. In one low-income area of New Orleans, AT&T customers had download speeds of only 1 Mbps or less, a rate that does not meet Zoom's recommended minimum for group video calls, doesn't approach the Federal Communications Commission's (FCCs) definition of broadband, and is more than 150 times slower than the median home internet speed in the U.S., of 167 Mbps. Meanwhile, AT&T offers residents in the nearby mostly White upper-income New Orleans neighborhood of Lakeview speeds that are almost 400 times faster — for the same price.

UC Santa Barbara computer science (CS) professor **Elizabeth Belding** is currently collaborating with researchers at the Georgia Institute of Technology on two projects intended to identify the quality of cellular service around the nation as precisely as possible. One of those projects was funded by the National Science Foundation (NSF) in August 2022 and the other by the Rockefeller Foundation. Both are intended to empower local communities to take actions to get better cellular coverage where they live. In a third, related project, Belding and UCSB professors **Arpit Gupta** (CS) and **Mengyang Gu** (statistics and applied probability) aim to analyze the quality and cost of fixed-broadband access around the country and study the relationship of that data to socioeconomic demographic variables in those same regions.

A GRANULAR LOOK AT INTERNET SERVICE INEQUITIES

The goal of the NSF-funded ADDRESS project is to use crowdsourced internet measurement data to estimate the quality of fixed internet in different areas and at different levels of geospatial granularity, ideally, down to the census-block or even street-address level. "The big picture is that we're trying to understand where broadband is and isn't and what the quality of current deployments is," she says of the three-year, $600,000 mapping collaboration, adding that, historically, "FCC maps have tended to grossly overstate coverage, because they are based only on provider-reported information."

Problems with current coverage estimates abound. "For instance, FCC rules said that if one home in a census block was covered, a provider could count that whole block as being covered," Belding explains. "But often the coverage is not that extensive, especially in rural areas, and that representation of access doesn't describe the *quality* of the coverage provided, whether you can, for instance, conduct a Zoom meeting, or stream a movie on Netflix or some other service."

One way to get a better understanding of one's service quality is to run a speed test on a device. Several organizations make that possible, and they aggregate the crowdsourced data from all the tests that people run, which is useful to Belding and her collaborators. "If, for instance, ten thousand people near you have completed speed tests, you can start to understand what coverage looks like in your community," she says.

Data used to ascertain internet availability is full of biases, however. "What if people with higher incomes tend to do speed tests more often than people with lower incomes, or people in urban areas do them more than people in rural areas?" Belding posits. "Or, there might be other types of skews. A slower result in an apartment building might occur if wifi networks there are interfering with each other. If the demographics of who does speed tests vary, then you can't just use the data at face value."

Belding knows this terrain. She, Gupta, Gu, fifth-year PhD student **Udit Paul,** and third-year PhD student **Jiamo Liu** presented research on biases and skews in crowdsourced speed-test data sets in a paper titled "The Importance of Contextualization of Crowdsourced Active Speed Test Measurements," which won the Best (Long) Paper Award at the 2022 Association for Computing Machinery Internet Measurement Conference.

"This work is timely given the recent focus on crowdsourced speed-test measurements for policy-related decision-making," Gupta says. "One significant roadblock in contextualizing these measurements is to accurately infer a user's subscription tier. A speed-test result of 10 Mbps is not bad if the user subscribed to a 10 Mbps plan, but it is problematic if they subscribed to a 100 Mbps plan. The team developed a novel methodology to infer the subscription tier for crowdsourced measurements."

The researchers will also use statistics and modeling to get a better idea of internet service in areas for which less data is available. For instance, there might be areas of a city where the data shows certain trends and other areas that have no data because no one has run speed tests there. Rural areas may also lack data. "So, we have to extrapolate from the data we've collected to make predictions about areas for which data is missing," Belding says. "That's where the modeling comes in."

THE CELLULAR-COVERAGE LANDSCAPE

The NSF-funded project called MapQ has goals similar to those of ADDRESS, the key difference being that it's for mobile, so it incorporates signal-quality information. This project, too, is based on global data aggregated from people taking speed tests — but from their phones — and brings together statistical modeling, machine learning, and mobile networking. Cellular propagation has its own characteristics, because it is not tied to a wire coming into a wall. "Again, though, we are asking how we can use the points where we know the speed-test data was collected and we also know the signal quality when the test was conducted to predict what coverage looks like in areas where we don't have data," Belding says.

About 85 percent of people in the U.S. own a smartphone and use it to access diverse resources. Sometimes the applications work well, and other times they do not. When they don't, the problem is often with the mobile broadband cellular network in the place and at the time of use. Currently, no one knows completely and accurately where high-quality access exists, or where regions of limited or no access are present.

The FCC has coverage maps, but, Belding explains, "They don't mean much. They're far less granular than what we'd like, and they show what we call binary coverage: the coverage is there or not there. A big thrust of our work is that it's more than just yes or no, service or none. There's a whole gradient. For instance, there might be some service, and maybe you can text and do voice calls but you can't websurf or sustain a Zoom call. Or maybe there's good coverage on a clear day but not when it rains. We're trying to show the quality of coverage over both space *and* time."

The researchers want to aggregate all the speed-test data and create a map that can be pulled up, perhaps when one is driving or in a new location, to see how the cellular quality is there. "That's particularly important for rural areas and other areas that might be underserved, to be able to say, as a community, 'Look at our mobile cellular service. It's substandard.'"

ROCKEFELLER: RESPONSIVE TO THE FCC CHALLENGE PROCESS

In this 1.5-year, $500,000 project funded by the Rockefeller Foundation, Belding will work with Georgia Tech researchers to develop open-source software, called CellWatch. The package will comprise a mobile application for taking network measurements, a community planning dashboard and map, and a cellular-quality prediction tool. The intention is to enable everyday people to take connectivity measurements and merge their data with others in their community. That will allow communities to build maps of coverage and challenge cellular providers' claims of coverage, which are often misrepresented across FCC maps and, preventing communities from applying for available federal funding.

The goal is to eventually have informed machine-learning algorithms and statistical analysis that can predict the quality of service in areas that have not yet been measured. It is intended that participating in the FCC-defined challenge process to show that coverage is substandard — and qualify to receive funding to improve it — will empower communities to ensure improvements to the cellular coverage they receive.

"The project is motivated directly by our previous work with Native American tribes," Belding says. "The cell companies come to them and say, 'We want to be your main provider; won't you promote us?" The tribe says, 'Well, your coverage is spotty.' And the provider says, 'No, it's not.' So who's right? We need data to determine that, and we're trying to get it."

Speed-test team (from left): Udit Paul, Jiamo Liu, Arpit Gupta, Elizabeth Belding, and Mengyang Gu