Internet Access Disparity

Aditya Kumar Bej
University of California, Davis
Davis, California
akbej@ucdavis.edu

Shivani Kalamadi
Department of Computer Science
University of California, Davis
Davis, California
skalamadi@ucdavis.edu

***Abstract*—Even in today's world with numerous technological advancements, the digital divide, in terms of Internet Access Disparity is a pressing issue that affects millions of people worldwide. The importance of having access to high-speed broadband had become more obvious during the COVID-19 pandemic. This paper presents a comprehensive analysis of the current state of the quality of Internet served and identifying access disparities in certain regions, focusing on the United States and India which rank in the top three highest number of active internet users. We highlight the multifaceted nature of Internet Access Disparity such as socioeconomic status, geographic location, infrastructure limitations, and political biases in the regions. We also try to understand and assess the impact of Internet Exchange Providers (IXPs) in bettering the quality of the network in the given region. We aim to provide an updated and comprehensive dataset of Internet Metrics along with relevant census information and discuss the implications of these findings to policymakers, infrastructure develop or any other relevant stakeholders.**

***Index Terms*—Digital Divide, Internet Access, Information Technology**

I. Introduction

Consistent with the approaches to what is considered to be human rights which includes equality, non-discrimination and covenants on civil, political, economic, social, and cultural rights, access to high-speed Internet is increasingly not seen just as a convenience, but as a necessity and more recently as a human right by the United Nations General Assembly in the Universal Declaration of Human Rights.

A total of roughly 5.35 billion people around the world use the Internet at the start of 2024, which is equivalent to 66% of the world's total population. While this figure suggests a significant level of global connectivity, there are a wide variety of factors which affect the quality of the internet served to the population. An often overlooked aspect of this disparity is the poor and unequal quality of internet service among those who are considered connected. This dimension of Internet Access Disparity (IAD) highlights that simply having access to the internet does not guarantee equitable participation in the digital world.

The importance of understanding and addressing IAD cannot be exaggerated. In the era where Internet is considered to be the backbone of global economies, education systems, and healthcare services, being disconnected or even having poor access means being disadvantaged. The implications of IAD extends beyond the inability to just browse the web or access social media; they include critical barriers to education, employment opportunities, and even life-saving health information. This disparity can have impacts on individuals and communities which can perpetuate exclusion in both urban and rural landscapes across developed and developing nations.

Addressing the issue of IAD requires a nuanced understanding that goes beyond merely counting who is an active internet user or not. This issue encompasses several key factors such as Bandwidth and Speed, Reliability and Latency, Data caps and affordability, the Internet Infrastructure, the geographic type (rural/urban), Population Diversity, and any relevant political biases in the region which affects implementing policies to make internet access more affordable and promoting the overall internet infrastructure in the given region.

To this end, we conduct two levels of analysis - a large scale analysis of all the US states and its counties and all the states and cities in India, and a low level analysis of certain regions from both the countries. We start with a broad analysis, collecting data on network performance from OOKLA [11] and M-LAB [10] which are crowd sourced measurements and relevant census information to spot patterns of consistent slow internet speeds and high latency. After identifying regions with lower performance, we zoom in for a detailed examination to understand why these areas are struggling. During this deeper dive, we consider several factors, including how many internet providers are available in that region, the common service plans people subscribe to, the presence of Internet Exchange Points nearby, the dynamic traceroute information from RIPE ATLAS [12] and any local government actions related to internet infrastructure and rules.

By integrating these two levels of analysis, this research work offers a holistic view of Internet Access Disparities, bridging the gap between macro-level trends and micro-level dynamics. Through our work, we anticipate uncovering actionable insights that can inform policy, practice, and infrastructure development, ultimately contributing to the reduction of the digital divide. The aim is to not only highlight the disparities that hinder equitable internet access but also the pave the way for any targeted interventions that ensure a more connected and inclusive digital future for all.

Summarizing our findings:

- We find evidence of Internet Access Disparities between regions having similar census metrics. We do an apples to apples comparison rather than an apples to oranges comparison when comparing different regions by taking into account the census metrics such as population size, the median household income etc.

- We find that the presence of IXPs does indeed improve the quality of network speeds in a given region versus a region with no IXPs.
- We find evidence pointing to regions experiencing inadequate internet services primarily due to insufficient funding for their network infrastructure. Additionally, historical instances of government-level corruption, exemplified by the notorious 2G scam in India as a case study [7], lend further support to our findings, highlighting how political mismanagement can significantly impact the quality and distribution of internet access. Various states in the USA have faced challenges in broadband expansion due to political, regulatory, and lobbying pressures [8]. For example, some states have laws that make it difficult for municipalities to create their own broadband services, often due to lobbying by private ISPs. This can limit options for expanding internet access in underserved areas.

II. Research Question

In this study, we focus on understanding the disparities in Internet Accesses across different regions within two countries. Our investigation is structured around three key research questions:

- **RQ1** : ***To what extent do differences in socioeconomic factors such as income level, geographic type (rural/urban), population diversity impact internet access?***
  We seek to explore how various socioeconomic factors, including income level, whether an area is rural or urban, and the diversity of its population, affect access to the internet.
- **RQ2** : ***Does the presence of IXPs in certain regions differentially impact internet access in terms of speed, reliability, and internet affordability***
  Our second question examines the influence of Internet Exchange Points (IXPs) on internet access, particularly looking at their impact on the speed, reliability, and affordability of internet services.
- **RQ3** : ***How does the presence of any inherent political biases influence internet access?***
  Lastly, we investigate how political biases and government policies affect internet access. This includes looking at whether certain regions or communities face disparities in internet access due to political decisions, regulations, or lack of initiatives aimed at improving digital infrastructure.

III. Hypothesis

A. ***Regions with lower socioeconomic status are served by lower-quality network providers, leading to significant disparities in internet access, speed, and reliability.***

Null Hypothesis (H0): There is no significant difference in the quality of network providers serving regions with different socioeconomic statuses. Variations in socioeconomic factors such as income level, geographic type (rural/urban), and population diversity do not significantly influence internet access, speed, and reliability.

Alternate Hypothesis (H1): Regions with lower socioeconomic status are served by lower-quality network providers, leading to significant disparities in internet access, speed, and reliability compared to regions with higher socioeconomic status. Variations in socioeconomic factors such as income level, geographic type (rural/urban), and population diversity significantly influence the quality of internet access.

B. ***The growth of Internet Exchange Points (IXPs) in specific regions over the last five years is likely to impact internet access disparities.***

Null Hypothesis (H0): The growth of Internet Exchange Points (IXPs) in specific regions over the last five years has no significant impact on internet access disparities among those regions.

Alternate Hypothesis (H1): The growth of Internet Exchange Points (IXPs) in specific regions over the last five years significantly impacts internet access disparities, potentially improving access quality and reducing disparities in those regions.

C. ***The presence of political biases significantly influences internet access***

Null Hypothesis (H0): The presence of political biases does not significantly influence internet access, including aspects such as availability, speed, and reliability.

Alternate Hypothesis (H1): Political biases stemming from government ideologies, regulatory bodies, and policy-making processes can lead to unequal distribution of internet access across different regions, communities, and socio-economic groups.

IV. Related work

To the best of our knowledge, prior work lacks an in-depth analysis of the way Internet trends and access disparities are analyzed especially when involving an entire country and all the states within it. Prior work was mostly focused on access disparities among ethnic groups and among specific communities and some studies are outdated [1]

We focus on studies which closely relate to our scope and note down the limitations for the studies which closely relate to our scope helps in evaluating gaps in current and prior research.

A. ***The type-of-internet-access digital divide and the well-being of ethnic minority and majority consumers: A multi-country investigation [2]***

The research paper "The type-of-internet-access digital divide and the well-being of ethnic minority and majority consumers: A multi-country investigation" by Bartikowski et al. (2018) explores the impact of mobile versus regular internet use on consumers' perceptions of their economic situation and life satisfaction. It highlights how these effects vary based on ethnic status (majority vs. minority) and the wealth of countries (richer vs. poorer). The study uses multi-level modeling and data from over 26,000 consumers across 21 countries to support its hypotheses.

While the study accounts for national wealth and ethnic status, there are myriad other socio-economic factors that could influence the relationship between internet access type and well-being, such as education levels, employment status, and urban vs. rural residency. These factors might require more detailed analysis using comprehensive census data.

B. *There is more to IXPs than meets the eye [3]*

The research paper "There is more to IXPs than meets the eye" by Chatzis et al. (2013) discusses the role of Internet Exchange Points (IXPs) in the Internet ecosystem, their operational and technical aspects, and their impact on innovation in Europe and globally. The paper highlights the critical role of IXPs in facilitating traffic exchange and reducing latencies.

The paper captures a snapshot of the IXP landscape as of May 2013. Given the dynamic nature of Internet infrastructure, with new IXPs emerging and existing ones expanding, the findings might not fully represent the current state or trends in IXPs' impact on network metrics and Internet performance. While the study offers substantial evidence highlighting the significance of Internet Exchange Points (IXPs), it does not delve into how IXPs in regions with suboptimal internet conditions might experience benefits from IXPs.

C. *Digital discrimination: Political bias in Internet service provision across ethnic groups [4]*

The research paper "Digital discrimination: Political bias in Internet service provision across ethnic groups" by Nils B. Weidmann et al. (2016) presents a comprehensive examination of the political bias in Internet service allocation among ethnic groups worldwide, demonstrating that politically excluded groups have significantly lower Internet penetration compared to those in power. The study shows a persistent political bias in internet access globally only among ethnic groups. The global expansion of the Internet is thought to enhance government transparency and democracy, but the study shows marginalized groups have significantly lower Internet access rates compared to those in power, highlighting a barrier to the liberating potential of technology. Nevertheless, the study is outdated, and due to its global scope, it lacks in-depth exploration of specific countries or regions. And moreover, the studies focuses on only ethnic groups which makes up a very small percentage of the entire population given a particular state or a county.

D. *The importance of contextualization of crowdsourced active speed test measurements [5]*

The paper "The importance of contextualization of crowdsourced active speed test measurements" by Udit Paul and colleagues (2022) addresses the need to contextualize crowdsourced speed test measurements to accurately understand network access and performance. This study contributes significantly to our understanding of the nuances behind speed test data which we have considered as part of our methodology when dealing with OOKLA and M-LAB network data.

This study contributes only on improving our methodology to analyze data and not contribute to other factors such as the actual analysis of the network metrics given a region with the census data as it is beyond the scope of the paper. Also, the study's context and findings are influenced by the specific policy and regulatory environment of the U.S. telecommunications sector. When applying these insights to other countries, differences in regulation, competition, and infrastructure development stages could lead to different implications.

E. *A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7) [6]*

The research paper "A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7)" by MacMillan et al. (2023) offers an in-depth evaluation of two widely used internet speed testing tools, focusing on their performance under various conditions and their impact on network metric analysis across different demographics.

While the paper provides a controlled, in-lab comparison of Ookla and NDT7 and extends the analysis to wide-area deployments, the variability of real-world internet usage scenarios and network configurations may not be fully captured. These scenarios can significantly influence the accuracy and representativeness of speed test results across different regions and demographics.

V. Theoretical Contribution

Summarizing our theoretical contributions:

- **Developing a general framework for a high level country wide analysis** We develop a general framework and parameters to consider when conducting a comprehensive analysis to identify internet trends and identify disparities in a given region
- **An exhaustive updated dataset** We contribute a novel dataset for evaluating the combined impact of socioeconomic factors and network provider quality on internet access disparities. By building upon existing tools like M-Lab Speed Test, OOKLA's Open Data, and RIPE Atlas data, we will create a geographically weighted, multi-layered dataset encompassing speed, reliability, network provider quality, and relevant socioeconomic indicators for select regions. This allows us to map internet access disparities with granular detail.
- **Historical trends of upto five years** We analyze and visualize the historical trend of the past five years based on the collected network data for each region.
- **Open source** We have open sourced our research work [9]

VI. Methodology

A. *Overview*

Figure 1 shows the high level architecture of how the research is conducted. The research methodology is crafted

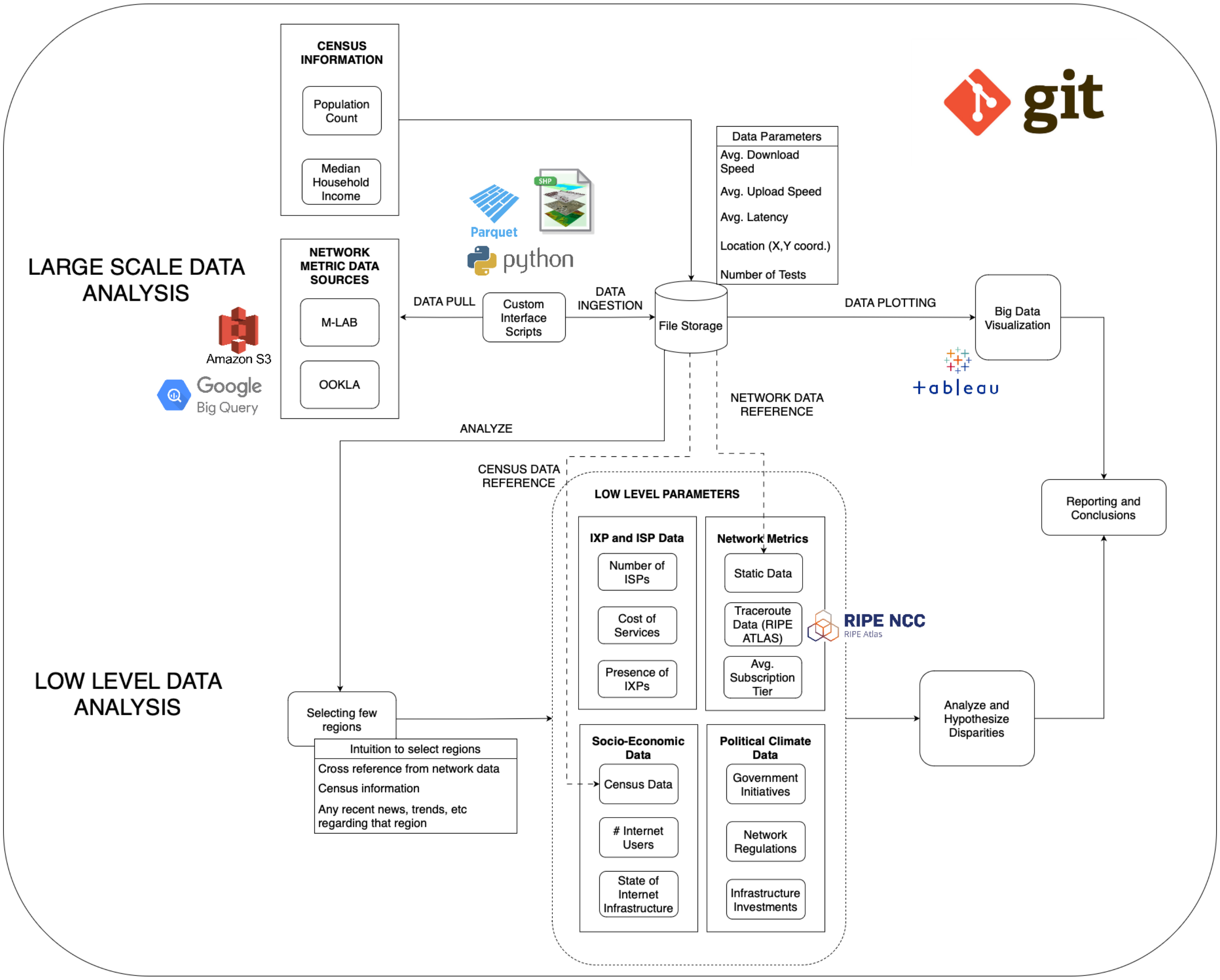

Fig. 1. Comprehensive Data Analysis Workflow

to tackle the challenge of analyzing large-scale internet access disparities, incorporating a variety of data sources and analytical tools like RIPE ATLAS to effectively explore research questions and hypotheses. The process begins with collecting data from leading network metrics providers, including MLAB, OOKLA, and RIPE Atlas, ensuring a solid foundation for analysis through a meticulous data preparation process involving extraction, correlation, cleaning, and storage, managed with Git for version control.

The methodology emphasizes data visualization using Tableau and Python to make complex data sets accessible and understandable, alongside detailed regional analyses that cross-reference network data with socio-economic and political information, utilizing resources like the International Telecommunications Union (ITU) [13], Census Data [14] [15], and Dynamic Traceroute information from RIPE Atlas for real-time network insights. This approach is enriched with Internet Exchange Point (IXP) data for geographical insights and ISP affordability assessments.

The culmination of the research is the production of comprehensive reports that synthesize the findings, supported by graphs that illustrate trends and patterns in internet access across different demographics, aiming to conclusively address the research hypotheses. Overall, this methodological framework is designed for its robustness, clarity, and ability to provide meaningful insights into internet access equality, potentially informing policy and practice.

B. Large Scale Data Analysis

Figure 2 outlines the general structured approach to collect and analyze network and census data as part of our High Level Bulk Analysis Methodology. The breakdown of the process is detailed in the below steps

- **Country-Level Breakdown** - We identify the country of interest and break it down into states or major administrative region (denoted as State 1 to State N)

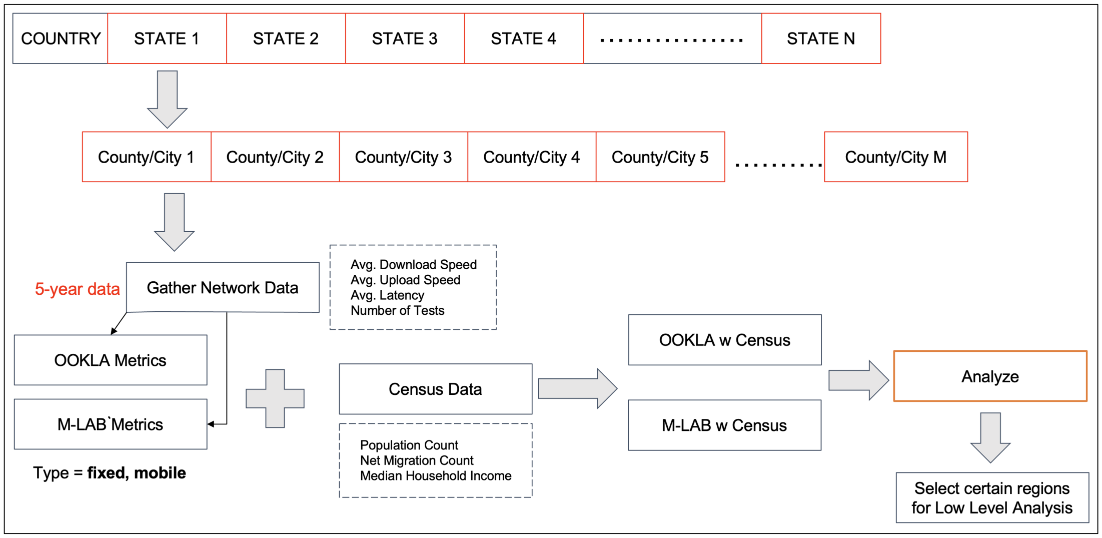

Fig. 2. High Level Bulk Analysis

- **Regional Segmentation** - Each state is further subdivided into smaller regions such as counties or cities. This allows for data collection at a more granular level.
- **Data Collection - Network Data** - Network data is collected for each county or city. The specific metrics gathered include - Average Download Speed, Average Upload Speed, Average Latency, Number of Tests conducted. These are similar metrics which are available in both OOKLA and M-LAB and the types of networks considered are fixed (wired connections) and mobile (wireless connections). We collect data of the past five years
- **Data Collection - Census Data** - Alongside network data, census data is also collected. This demographic information includes the Population Count, Net Migration Count, and the Median Household Income.
- **Data Integration** - The network data and census data are then combined and this integration allows for a comprehensive analysis that accounts for both technological and demographic factors
- **Analysis** - With the integrated dataset, an analysis is conducted to understand and interpret the combined data. The objective here is to derive insights from the fused datasets, aiming to identify correlations or patterns.
- **Low Level Analysis** - Following the high-level analysis, certain regions are selected for more detailed, low-level analysis. This step is a deeper dive into specific areas of interest that may emerged from the initial analysis such as lower download and upload speeds or higher latency.

C. Low Level Data Analysis

Figure 3 depicts the framework for conducting low-level analysis on identifying internet access disparities within select regions. This framework consists of a comprehensive set of parameters, each encapsulated within its domain, which when combined, offers a multi-faceted view of the internet ecosystem.

For each region, to conduct low-level analysis, gather the following information. The intuition of each segment is explained below.

- **IXP and ISP Data** - This segment focuses on analysis of Internet Service Providers (ISPs) and Internet Exchange Points (IXPs) within the region. Key metrics include the number of ISPs, which affects competition and potentially service quality; the cost of services; which influences accessibility and affordability for consumers; and the presence of IXPs, which can affect the speed and reliability of internet connectivity.
- **Network Metrics** - This segment analyzes the technical performance of internet services, considering both static measurements such as crowd-sourced data from M-LAB and OOKLA and dynamic metrics such as trace route information from RIPE ATLAS which gives the state of the internet in real time and tracks the path the data takes from a source to a destination over the internet so we can understand if any hops are through IXPs
- **Socio-Economic Data** - This component integrates socio-economic factors, with the data points such as the census data parameters which we have chosen for our research, the number of internet users, which reflects the penetration of internet usage; and the state of internet infrastructure, which is fundamental to service availability.
- **Political Climate Data** - This component delves into the political environment's impact on internet infrastructure, scrutinizing government initiatives that could promote or hinder the development of internet services, network regulations that may affect how services are provided and maintained, and infrastructure investment, which is critical for the expansion and upgrading of internet infrastructure.

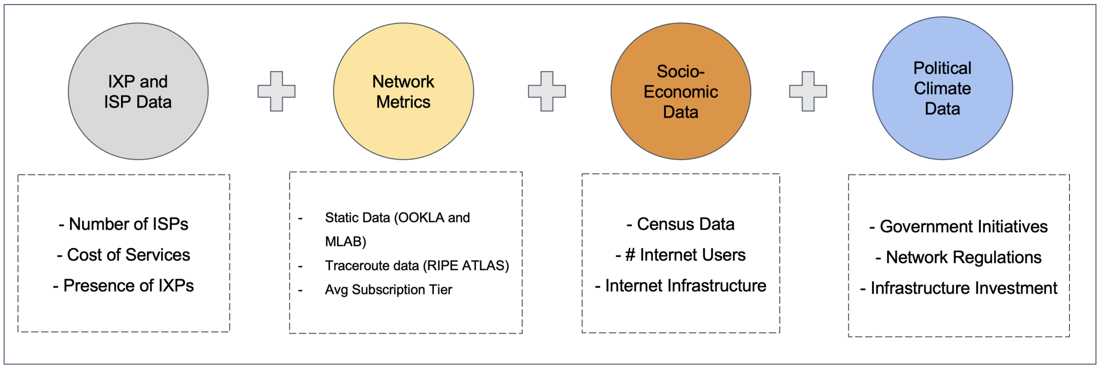

Fig. 3. Low Level Data Analysis Parameters

VII. Data Filtering

A. Criteria on Country Selection

Table I shows the number of Internet Users and Penetration rate of the top 3 countries with the highest number of internet users in the world [18]. We decided to choose United States and India since they rank among the top 3 countries with the most active internet users. We decided to skip China due to reasons such as less transparency in their census data and stricter network regulations which can bias our research work.

TABLE I
Internet Data (2021)

Country	# of Internet Users	Internet Penetration
China	1.8 trillion	73.1 %
India	692 million	48.7 %
USA	311 million	93.73 %

B. Contextualizing Datasets Used

Utilizing Crowdsourced data at the surface value is problematic and it is essential to contextualize these measurements to understand better what the attained network metrics truly measure. This was addressed in the paper "The importance of contextualization of crowdsourced active speed test measurements" [5]. We understand this need so we try to mitigate potential biases by the our own following measures done within the scope of a quarter:

- We **filter out** any biased data by considering the number of tests conducted for a particular network metric in a given region. We ensure the number of tests is greater than atleast 50. This is done based on the intuition that a larger number of tests contribute to the reliability and validity of the data and increases our confidence in the conclusions drawn.
- We introduce **dynamic traceroute information** during our low-level analysis as this provides us with near real-time information such as latency and packet loss. This can serve as a means to validate and corroborate the static crowdsourced data, ensuring the metrics used in the analysis are accurate and trustworthy.
- We **utilize both M-LAB and OOKLA datasets**, as opposed to relying on just one, as this enhances the overall richness of the dataset. This allows for cross-validation of results and since M-LAB and OOKLA use different methodologies for measuring network performance, combining them provides a more nuanced view that accounts for these methodological differences. Also, anomalies in one can be checked against the other if needed.

C. Data Consistency Across Sources

To maintain data consistency, we standardized the metrics across M-LAB and OOKLA, aligning data formats. For example, speeds are consistently measuremend in Mbps, and latencies in milliseconds. Data was transformed for OOKLA as it had measured the average download and upload speeds in kbps.

D. Data Synchronization with Census Information

We synchronized network metrics with the most recent census data based on GeoID for US and State Code for India. Each network metric and census data had a GeoID or State Code associated with it to identify which region it is measuring.

E. Handling any Outlier Data

To ascertain whether outliers are erroneous or valid, we cross-verify with alternative data sources for both network metrics and census related data. These sources can include government websites, ITU, or even NGO reports.

VIII. Measures

To assess the hypothesis proposed in our study in Section III on Internet Access Disparities, we employ a well-defined set of metrics as shown in Table II that captures the various dimensions of internet access. These measurements are selected to minimize ambiguity and ensure the comprehensive testing of our hypothesis, thereby reducing the scope of unanswered questions and accurately deciding the truth value of our hypotheses. While **M-LAB** offers randomly sampled tests for both mobile and fixed line users, **Ookla** splits its unsampled datasets into fixed line and mobile internet users.

TABLE II
METRICS CAPTURED

M-LAB and OOKLA Network Metrics (over 5 years)
Average Download Speed
Average Upload Speed
Average Latency
Location Data (X, Y coord.)
Number of Tests

US Census Metrics
Population Count
Median Household Income

India Census Metrics
Population Count
Net State Domestic Product (NSDP)
Gross State Domestic Product (GSDP)

Misc. Data for Low Level Analysis
Number of ISPs in a region
Number of IXPs in a region
Government Network Infrastructure Investment Data
Internet Penetration Rate of a Region

Analyzing this data **over a five-year period** is useful when examining historical trends in Internet Growth and correlating with the the related census data, we can identify how Internet Access Disparity has been in-place or introduced: The following metrics for network data are captured and analyzed for each County (US) and State (India)

1) **Average Download Speed:** Measures the speed at which data is transferred from the internet to the user's computer, indicative of network quality. Higher speeds suggest better internet access and are crucial for regions to support high-bandwidth applications.
2) **Average Upload Speed**: Reflects the speed at which data is sent from the user's computer to the internet. Essential for user-generated content, cloud services, and online gaming. This metric is significant for assessing the capacity for symmetrical internet usage.
3) **Average Latency**: Latency indicates the delay before a transfer of data begins following an instruction for its transfer. Lower latency is critical for real-time applications, such as VoIP calls and online gaming.
4) **Location Data (X, Y coord.)**: Provides geographical information, enabling the analysis of network performance in relation to specific regions. This aids in identifying geographic disparities in internet access.
5) **Number of Test**s: The volume of network tests conducted provides an indication of data reliability and can suggest user engagement and internet penetration in the area.

For Census data pertaining to United States of America: Each of the census data metrics provides a different perspective on the factors that may influence internet access:

1) **Population Count** - This metric gives insight into the demand side of the internet access equation. A higher population count typically indicates a greater need for robust internet infrastructure to support the connectivity needs of residents. In contrast, areas with sparse populations may experience less incentive for investment in digital infrastructure due to lower potential for return on investment for service providers.
2) **Median Household Income** - This economic indicator can be closely correlated with internet access disparities. Higher median household income levels are often associated with better access to a variety of services, including the internet, due to greater purchasing power and the likelihood of existing infrastructure in wealthier areas. Lower-income areas might have less access due to affordability issues and possibly lower levels of existing infrastructure, as service providers may not prioritize investment in these regions.

For Census data pertaining to India:

1) **Population Count**
2) **Net State Domestic Product (NSDP)** - It represents the average income produced per person in that region and is a useful indicator of the general standard of living and economic prosperity and is also indicative of the economic capability to afford internet services.
3) **Gross State Domestic Product (GSDP)** - It represents the total economic output or income generated within a state and is a broad measure of the state's economy. We can infer potential investment capacity for infrastructure, including internet services, in a state.

There are other several census data points which can be employed as part of future work for more precise analysis but for the sake of limiting the scope of the study, we have decided to not include these parameters as part of our high level study. However, these parameters can be considered when doing a low-level deep dive for a specific region to further understand the trends of Internet Access Disparities. A few examples of such parameters include Educational Attainment, Age Distribution, Race and Ethnicity, Language Proficiency, Housing Characteristics, Disability Status and Poverty Status.

In regards to the Miscellaneous Data captured for further Low-Level Analysis:

1) **Number of ISPs in a region**: More ISPs can indicate a competitive market, potentially leading to better services and lower prices for consumers.
2) **Number of IXPs in a region**: IXPs reduce the distance data must travel, improving speed and reliability. Their growth can be a proxy for improving network infrastructure.
3) **Government Network Infrastructure Investment Data**: Investments in infrastructure are crucial for enhancing internet access and quality. This metric helps assess the role of policy and government support.
4) **Internet Penetration Rate of a Region**: The percentage of individuals with internet access, directly related to the extent of internet disparities.

IX. Risks and Limitations

We identify certain risks/limitations and shown ways to reduce them for the scope of our research.

A. Subscription Tier Consideration

Lower-speed data points can be attributed to lower-tier subscriptions and not necessarily poor access. There can be multiple reasons for this which can include Consumer choice, economic constraints, Availability of ISP Plans and so on. There is no publicly available dataset which gives information on the subscription tier of each household in a region. For both OOKLA and M-LAB, to mitigate this risk, we do the following:

1) **When analyzing the OOKLA dataset** in our low level analysis, we look the complete picture of all the subscription plans from the ISPs in that area when making conclusions.
2) **For M-LAB**, we use the NDT (Network Diagnostic Tool) data and extract network metrics using BigQueries. By design, the value of NDT data is the aggregation of many connection test results. Such aggregation dilutes the variance introduced by users subscribed to different service tiers, thereby allowing us to identify trends and patterns that are more indicative of the overall network accessibility rather than the individual user choices.

B. Discrepancies between OOKLA and M-Lab Measurements

The median throughput reported by Ookla speed tests can be up to two times greater than M-Lab measurements for the same subscription tier, city, and ISP due to M-Lab's employment of different measurement methodologies [5] [6]. Hence to address this, we refrain from combining or taking an average of both the datasets into a single dataset. Instead we treat them as separate entities as seen in our high level methodology in Figure 2. We only use the two datasets as means to cross-validate the results if necessary as mentioned under sub-section A.

C. Inherent Incompleteness in crowd-sourced data

While not all regions or demographic groups may be equally represented in crowd sourced datasets, due to varying levels of participation and access to testing tools, these gaps can inform researchers about the digital divide and areas lacking sufficient data coverage. We recognize this limitation and propose future research work to deploy targeted data collection efforts in such underrepresented areas.

D. Lack of up-to-date census information for India

The last collected official census data for India was in 2011 [15]. So this was a hurdle during our large scale analysis. However, we mitigated this by calculating the projected census parameters we needed which was the population count of each state and acquired the Net State Domestic Product (NSDP) per-capita from the Ministry of Statistics and Programme Implementation [16] [17]

X. Execution

We have followed the detailed architecture of our Methodology referenced in Section VI, Figure 1 and frequently push our changes to our GitHub link [9]. Please refer to our GitHub for our execution and results. The proof of work is also present as part of our commit history. The results section in the Git project has 3 sections - mlab, ookla, and tableau.

XI. Results

For a comprehensive understanding of the methodologies and analyses that informed these conclusions, please refer to Section XII on page 8 and Section XIII on page 10.

- **RQ1** : ***To what extent do differences in socioeconomic factors such as income level, geographic type (rural/urban), population diversity impact internet access?***
  Drawing from our analysis presented in XIII-B in page 10 and XIII-C in page 10, we conclude that disparities in socioeconomic factors such as income level, geographic type, and population diversity significantly impact internet access.
- **RQ2** : ***Does the presence of IXPs in certain regions differentially impact internet access in terms of speed, reliability, and internet affordability***
  Drawing from our analysis presented in XIII-D in page 13, the presence of IXPs in certain regions typically enhances internet access by improving speed, reliability, and affordability.
- **RQ3** : ***How does the presence of any inherent political biases influence internet access?***
  Drawing from our analysis presented in XIII-C in page 10, the presence of inherent political biases can influence internet access by shaping policy decisions, funding allocations, and regulatory frameworks that determine how, where, and at what cost internet services are provided. Political biases may lead to unequal distribution of digital infrastructure, favoring certain regions or demographics over others, resulting in disparities in internet access, speed, and reliability.

XII. Data Presentation

To visualize the extensive datasets we collected, we utilize scatter plots to represent data points for US counties. This graphical representation shows average download and upload speeds for each region with respect to the median household income. Through this, we can discern if and how internet access disparity exists with respect to income levels of the region. Figure 4, 5 and 6 represent this.

Secondly, in Figure 7, 8 and 9 we represent the download and upload speeds by US states from 2019 to 2023 to determine historical trends. Since the trend of the whole country might not give us the clear picture and county wise data would be too granular, we resolve to states. Among these, we pick one or two states that show unique or abnormal trends for our analyses.

Figure 10 and 11 mimic the same methods but for India.

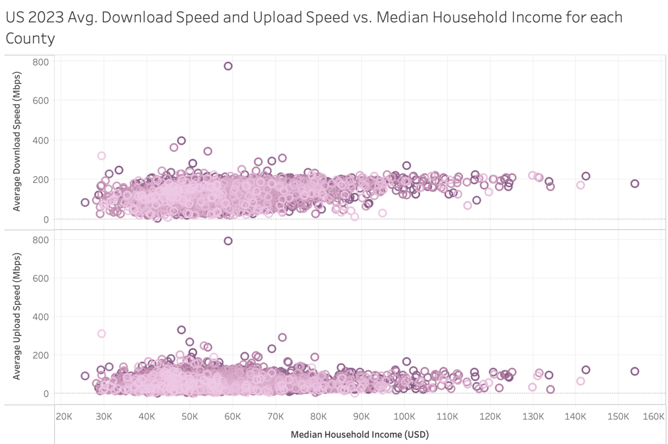

Fig. 4. US mLab 2023 Avg. Download & Upload Speeds vs. Avg. Median Household Income for each County

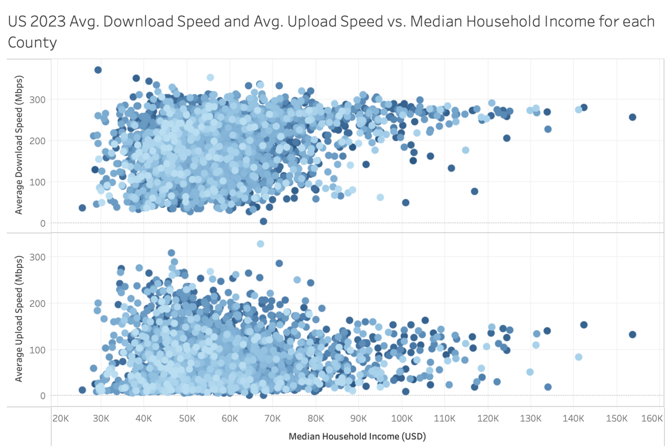

Fig. 5. US 2023 Ookla Fixed Line Avg. Download & Upload Speeds vs. Avg. Median Household Income for each County

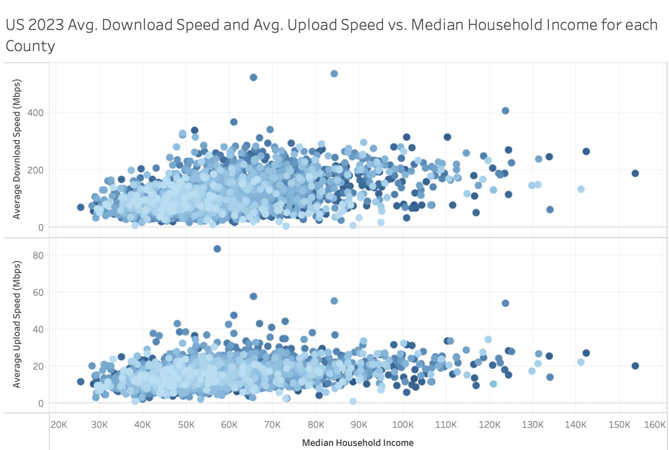

Fig. 6. US 2023 Ookla Mobile Line Avg. Download & Upload Speeds vs. Avg. Median Household Income for each County

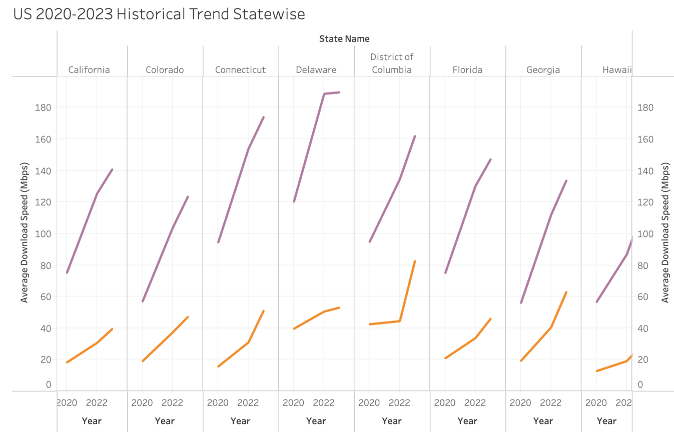

Fig. 7. US mLab Avg. Download (Purple) & Upload (Orange) Speeds by Year and State

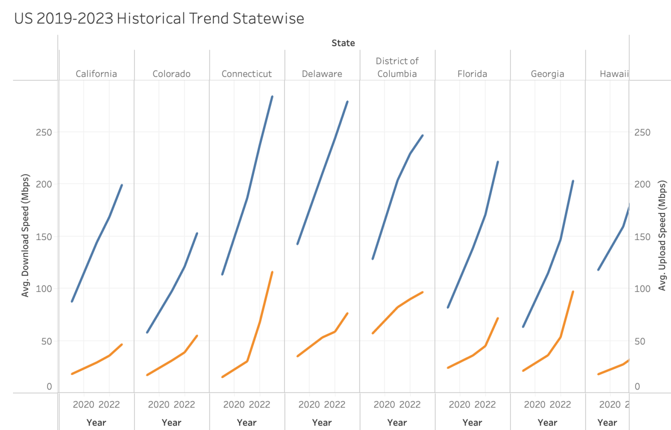

Fig. 8. US Ookla Fixed Line Avg. Download (Blue) & Upload (Orange) Speeds by Year and State

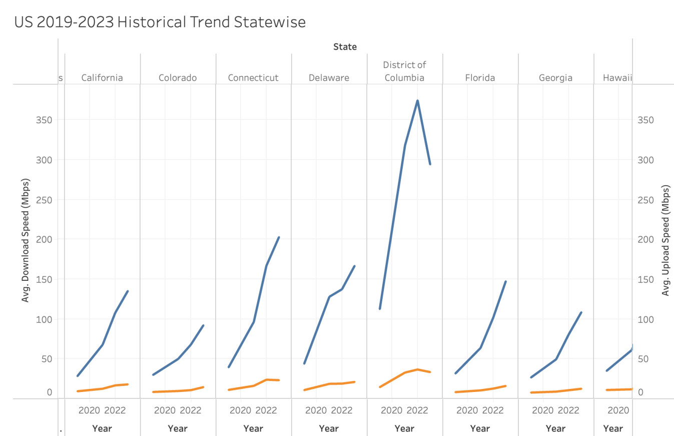

Fig. 9. US Ookla Mobile Line Avg. Download (Blue) & Upload (Orange) Speeds by Year and State

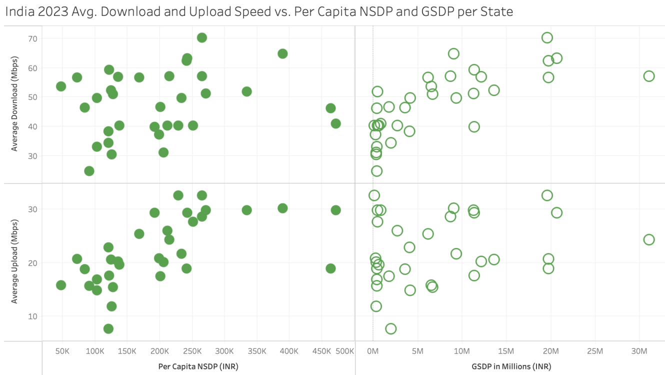

Fig. 10. India mLab Avg. Download & Upload Speeds vs. Per Capita NSDP and GSDP

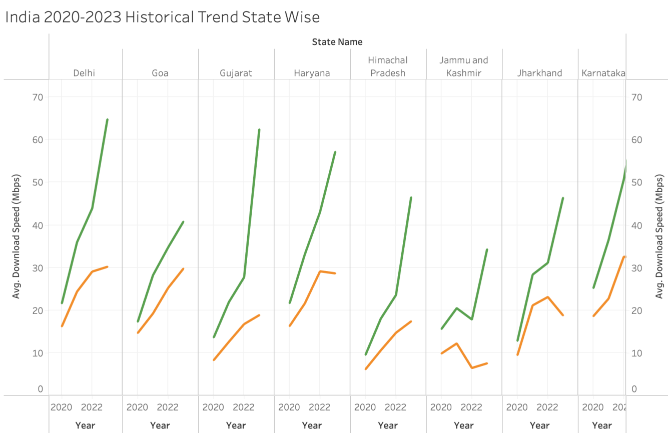

Fig. 11. India mLab Avg. Download (Green) & Upload (Orange) Speeds by Year and State

XIII. Analysis & Discussion

This section focuses on Analysis and Discussion of the data provided in Section XII along with details about the impact of IXPs and Political Climate in a region

A. Historical Trends Analysis and Discussion

As one would expect, the general trend for internet speeds according to both mLab and Ookla are constantly increasing yearly (Figure 7 & Figure 8). This is mostly due to technological or infrastructural advancements. District of Columbia stands out in Figure 6, not just because of a steady decline in the download speed in 2023, but also because its average download speeds have been and are still much higher than any other state's average download speeds even with the decline.

For India, Figure 11 shows abnormalities in this trend for Jammu & Kashmir and Jharkhand wherein upload speeds are lower than they were 2 years ago, which deems further investigation.

B. High Level Analysis and Discussion

Figure 4 represents data collected from mLab for US in the year 2023 against the median household income for each county. While we see that there's no apparent difference in the download speeds for low-mid income range households, there's a trend that forms for high income households averaged at approximately 200 Mbps if income is greater than $80,000 per year. Figure 5 strengthens this inference as it follows the same trend for download speeds for fixed line users @ approximately 300 Mbps. Please note that this trend might be due to better subscription tiers of households with higher incomes since Ookla doesn't average set of results from an area. Ookla's mobile lines however don't follow this trend strictly - Figure 6. This difference can potentially be influenced by a range of reasons like: better fixed line network infrastructure for regions with higher household incomes but since mobile towers serve a bigger area, they might be less biased. Please note that these are our inferences only, since the primary aim of this section is to outline potential trends for further research.

Figure 10 shows us somewhat inconclusive results for general internet speeds with respect to per capita NSDP and GSDP. We presume this result may be due to the lack of uniformity among the general income level across Indian states. Since per capita NSDP, GSDP are largely different for each state due to varying state size and population.

C. Low Level Analysis and Discussion

From analyzing the trends from the High Level Analysis, we select the following regions to conduct our low-level analysis on. For the scope of our research, we will be focusing on fixed networks (Broadband) for our low level analysis and will be focusing on a very small dataset.

We will follow the methodology as mentioned in Section XIII-C which is generalized to deal with any region to be analyzed

Justification is given for each region chosen.

TABLE III
Low Level Analysis Regions

US Regions to Analyze	Indian Regions to Analyze
Arizona Counties	Manipur
Kansas Counties	Jammu and Kashmir

1) Arizona Counties: **Understanding the geography of Arizona**

The state is the sixth largest in the U.S. and features several geographic zones: large desert areas, including the Sonoran Desert; forested mountain ranges, such as the White Mountains and the San Francisco Peaks; and, of course, the Grand Canyon. This diverse topography can present unique challenges and opportunities for internet infrastructure development.

Urban vs. Rural: The state has major metropolitan areas, such as Phoenix in Maricopa County and Tucson in Pima County, where most of the population resides. These urban centers often have better internet access due to higher population densities, greater demand for services, and more substantial infrastructure investment. ISPs are more likely to invest in areas where they can get a higher return on investment, which typically correlates with population density and economic potential.

Rural Areas: Rural and remote areas, such as those in Apache or Navajo counties, may face challenges due to their sparse populations and greater distances between communities. The cost of extending services to these areas is high, and the potential customer base is lower, which can result in less investment and lower service quality.

Tribal Lands: Arizona is home to several Native American reservations, such as the Navajo Nation, which is the largest in the U.S. by land area. These regions have historically been underserved in terms of internet connectivity due to a combination of geographic isolation, economic factors, and political complexities.

Fig-12 visualizes our captured data for the state and shows the parallelism between the average download speed, the median household income and the population count.

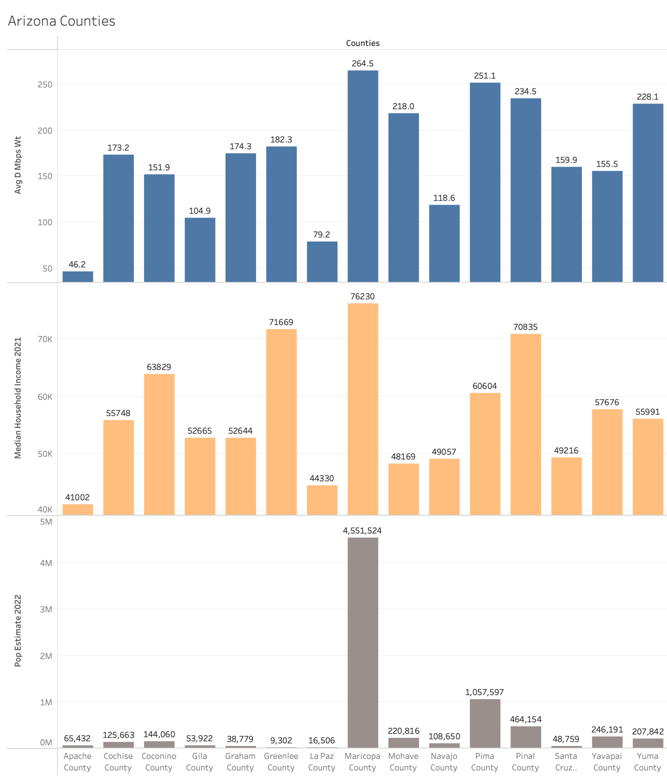

Fig. 12. Arizona County-wise metrics

The percentage of Arizonans who do not have broadband access is more in rural communities than urban. Even in some urban areas, the number of providers is limited. This highlights the **disparity** in the state. Refer to Fig 13

*2) **Kansas Counties**:* Kansas is situated in the Midwestern region of the United States and is characterized by its varied geography, which includes the flat expanse of the Great Plains in the west, rolling hills in the north-central region, and the more wooded and hilly eastern part of the state that stretches into the edge of the Ozarks.

Rural vs. Urban Internet Access: Kansas's vast rural areas, particularly in the western and central regions, pose challenges for broadband deployment. The cost of infrastructure per capita in sparsely populated areas is higher compared to densely populated urban areas like Wichita, Topeka, and the Kansas City metro area. This often results in slower internet speeds, less reliable connections, and fewer service providers for rural residents.

Economic Considerations: The state's economy is largely driven by agriculture, which may not always demand the same level of broadband infrastructure as industries that are more technology-dependent. However, as agriculture becomes increasingly digitized, the need for high-speed internet in rural Kansas grows.

Fig 14 visualizes our captured data for the state and shows the parallelism between the average download speed, the median household income and the population count.

Highlighting lack of adequate internet access in Kansas

Fig 15 provides a detailed breakdown of internet access across different geographic areas in Kansas, categorizing access into levels of adequacy and tying it to the presence of high-speed internet connections and devices. From urban to rural communities, the disparities are evident. Areas like Johnson County show a high percentage of residents with both a high-speed internet connection and a device, indicating a relatively affluent community with robust digital infrastructure. In contrast, regions such as Southeast and Southwest Kansas display a larger share of residents with inadequate internet access, suggesting that these areas may lack the necessary infrastructure or investment to support high-speed connectivity.

Problem with Inadequate funding allocation for Network Infrastructure growth

As per Fig 16 The funding allocation disparities between Kansas's Southwest and East Central counties, with \$10.8 million earmarked for the former encompassing over ten regions and \$8.2 million for the latter's two regions, may signal a potential oversight in addressing the true scope of network infrastructure needs. While larger geographic areas with lower population densities inherently face higher infrastructure costs, this unequal distribution raises questions about the criteria and metrics used to assess funding requirements. This can suggest **inherent political biases in play** These decisions must be critically examined to ensure they align with the objective of equitable network growth and that they do not inadvertently widen the digital divide.

*3) **Manipur**:* From Figure 10, we choose a state which has consistently lower download and upload speeds - Manipur. Next, we choose another state - Sikkim, which has nearly the same per capita NSDP and GSDP as Manipur but has consistently higher internet speeds (Figure 17). In fact, Sikkim's internet speed is approximately 3x higher than Manipur's. Since both are smaller states in North East India, they are often just summarised in comprehensive reports. For example, Telecom Regulatory Authority of India [26] reports their network data only in aggregation as North East. The articles [27] and [28] delve into the issue of internet

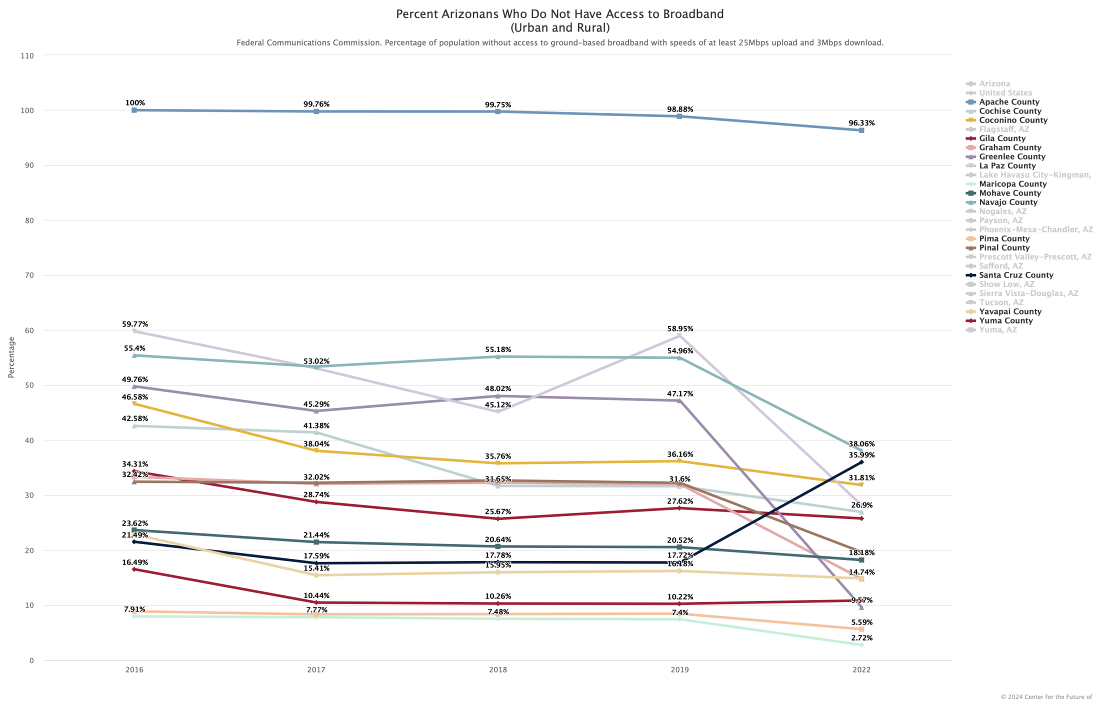

Fig. 13. Arizona Broadband Disparities [25]

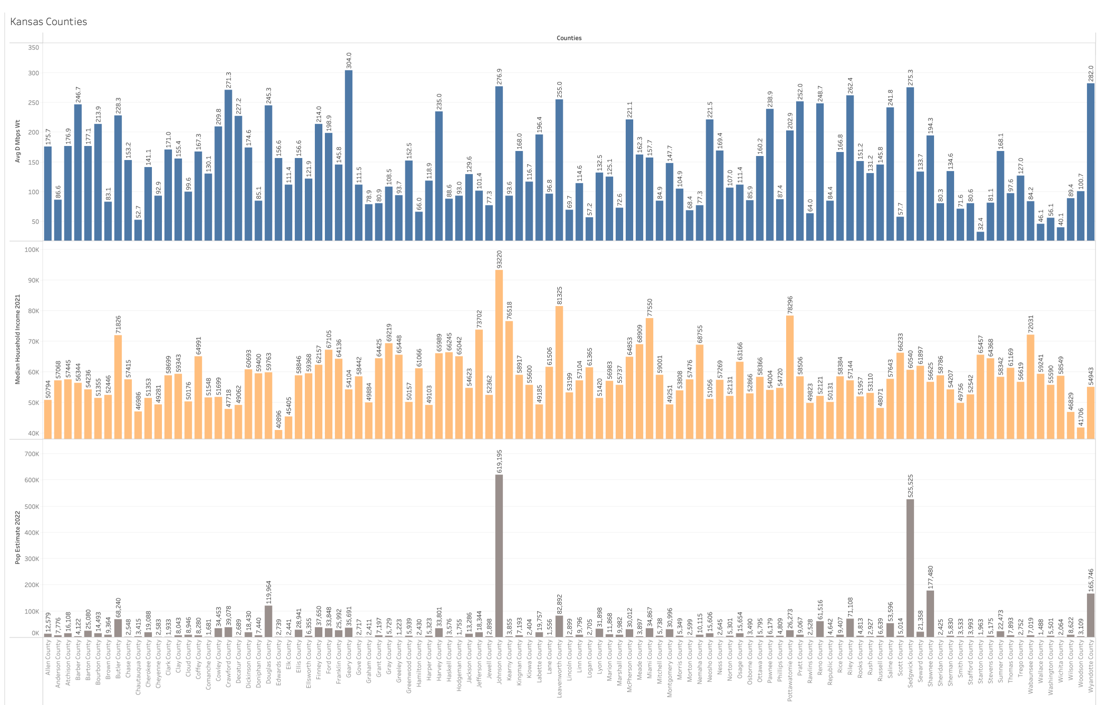

Fig. 14. Kansas County-wise metrics [25]

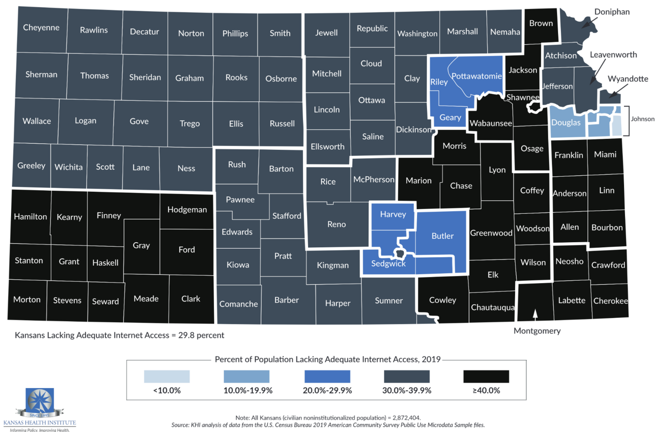

Fig. 15. Kansas Broadband Disparities [29]

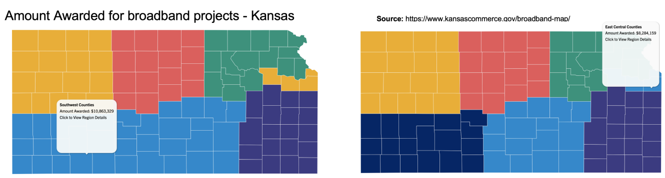

Fig. 16. Kansas Broadband Project funding [30]

shutdowns in Manipur amid internal disturbances and highlight its detrimental effects on fair internet access. Despite the state-wide censorship in the region, there's no accountability from the government. Even legal rulings from the public against such shutdowns have been largely ignored, leaving people resorting to strikes and protests. Moreover, the challenges in restoring internet access, particularly in remote areas of Manipur, underscore the need for infrastructure development.

*4) **Jammu and Kashmir**:* Jammu and Kashmir faces significant challenges in terms of internet access, primarily due to the sensitive nature of the region and government-imposed restrictions. Despite having a significant internet subscriber base, the government-imposed restrictions have led to slower speeds and limited connectivity for residents. Moreover, the

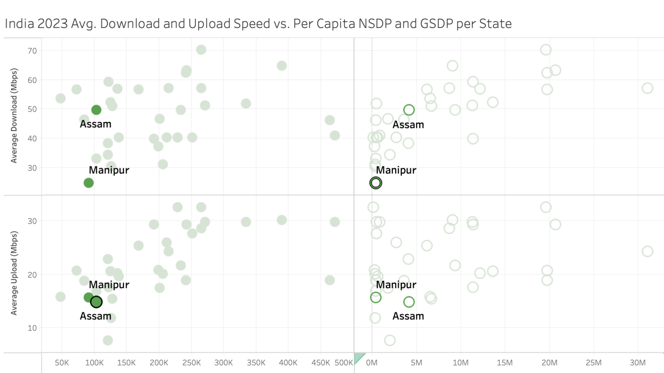

Fig. 17. Manipur and Assam in India's mLab data of Avg. Download & Upload Speeds vs. Per Capita NSDP and GSDP

region's mountainous terrain and inadequate network infrastructure further exacerbate these issues, resulting in poor internet access overall. It is important to note that only 10 of India's 1150 ISPs operate in Jammu and Kashmir. While Madhya Pradesh and Himachal Pradesh, which have similar per capita NSDP and GSDP as Kashmir, respectively, have much higher internet speeds. In addition, Madhya Pradesh has 29 ISPs and Himachal Pradesh has 11 ISPs, still having faster internet speeds.

Shockingly, Jammu and Kashmir experienced more internet outages than any other country in 2022 [31]. The region even endured the world's longest internet shutdown in a democracy, lasting from August 2019 to January 2020, further worsening the severity of the situation. These long and recurring internet blackouts have left residents isolated, forcing them to rely on outdated communication methods.

D. Impact of IXPs - Analysis and Discussion

For the purpose of strengthening our Hypothesis for Research Question 2, we will focus on a few case studies

*1) **Case Study - Kenya's IXP (KIXP)**:* Prior to Kenya's, there was no IXP on the African Continent between Morocco and South Africa. In early 2000, the association of Kenya's ISPs, called TESPOK, undertook to organize a neutral, non-profit IXP for its members. Its creation aimed to keep local internet traffic within Kenya, reducing reliance on international bandwidth, which was both costly and inefficient. [20]

Impact:

1) **Reduced Latency and Costs**: Before KIXP, Kenyan ISPs had to exchange traffic internationally, even when the origin and destination were within Kenya. KIXP dramatically reduced latency and the costs associated with international data transit. Reports indicate that ISPs saw up to a 90% reduction in operational costs related to international bandwidth.
2) **Growth of Local Content**: With improved access speeds and reduced costs, KIXP has encouraged the hosting and consumption of local content. It supported the development of local hosting services and content delivery networks, making the internet more relevant to Kenyans.
3) **Enhanced Network Resilience**: By facilitating easier data exchange within the country, KIXP has contributed to the resilience of Kenya's internet infrastructure, making it less susceptible to international outages and improving overall internet reliability.
4) **Economic Development**: KIXP has been a catalyst for the digital economy in Kenya, aiding the growth of e-commerce, online education, and digital services.

*2) **Case Study - Nigeria's IXP (IXPN)**:* The Internet Exchange Point of Nigeria (IXPN) was established in Lagos in 2006. Its goal was to keep Nigerian internet traffic within the country, improving the quality of internet access and reducing costs. [21]

Impact:

TABLE IV
HIGHLIGHTS FROM KENYA [23]

	Before IXP	After IXP
Peering networks	25	56
Internet users (%)	8.8	17.8
Mobile broadband subscribers (%)	0.4	41.9
Cost of 500MB prepaid data (USD)	5.92	2.42
Content Delivery Network	Google Global Cache	Multiple

1) **Reduced Dependency on International Bandwidth**: IXPN allowed for local routing of internet traffic, significantly reducing the need for costly international bandwidth. This has led to substantial cost savings for ISPs and improved speeds for end-users.
2) **Local Content and Services Growth**: With the establishment of IXPN, there was a noticeable increase in local content hosting. This was due to the improved economics of hosting content locally, leading to a richer and more diverse Nigerian internet content landscape.
3) **Improved Internet Quality**: IXPN has played a crucial role in enhancing the quality of internet services in Nigeria by reducing latency. This improvement has been beneficial for services requiring real-time data transmission, such as VoIP, online gaming, and streaming services.
4) **Digital Economy Boost**: By improving internet infrastructure, IXPN has contributed to the growth of Nigeria's digital economy, encouraging innovation and the creation of digital services that cater to the local market.

TABLE V
HIGHLIGHTS FROM NIGERIA [23]

	Before IXP	After IXP
Peering networks	30	71
Internet users (%)	16.1	42
Mobile broadband subscribers (%)	6.8	30.7
Cost for 500MB prepaid data (USD)	\$12.75	\$3.27
Content Delivery Network	Google	Multiple

*3) **Case Study - DE-CIX)**:* DE-CIX (Deutscher Commercial Internet Exchange) is one of the world's leading Internet Exchange Points (IXPs). Established in Germany, DE-CIX has expanded its services globally, including in key regions of India and the United States. The presence of DE-CIX has been associated with improved network quality in the regions it serves.

In **India**, DE-CIX operates in several key regions, including Mumbai, Delhi, Kolkata, Chennai, and Hyderabad. These locations have seen an enhancement in internet quality due to the increased efficiency of traffic exchange within the country. For instance, Mumbai, being one of the largest data center hubs in India, benefits significantly from the presence of DE-CIX in terms of reduced costs and improved service quality for local ISPs and their customers.

In the **United States**, DE-CIX is present in New York, Dallas, Chicago, and Richmond, among other locations. These IXPs serve a critical role in improving internet quality by providing a neutral, robust, and high-capacity platform for traffic exchange. For example, in New York, one of the largest internet hubs globally, DE-CIX has enhanced connectivity options for ISPs, content providers, and digital businesses, contributing to the area's status as a top-tier location for digital services.

Based on our research and study of the Annual reports from DE-CIX [24], we have found the following regional highlights for both the US and India

Regional highlights - US

The DE-CIX North American ecosystem has shown remarkable growth across various cities. The following are some key highlights:

- The DE-CIX North American ecosystem saw increasing demand for high bandwidth access, resulting in a 46% increase in 100GE ports.
  - **New York**
    - Overall connected customer capacity increased by 30%.
    - DE-CIX New York saw an increase in 100GE Ports of 33%.
    - The connected customer capacity grew by 26%.
  - **Dallas**
    - The number of 100GE ports increased by 23%.
    - DE-CIX Dallas almost doubled its data traffic throughput at peak times, reaching 625 Gbit/s by the end of the year.
  - **Chicago**
    - DE-CIX Chicago saw new demand in high-bandwidth 100GE ports, of which the first 11 were activated in 2022.
    - The number of connected networks increased by 33%.
  - **Richmond**
    - With an increase of 120%, DE-CIX Richmond more than doubled its connected customer bandwidth.
  - **Phoenix**
    - In its first 9 months of operation, connected customer capacity at DE-CIX Phoenix exceeded 600 Gbits.

Regional highlights - India

- The number of connected networks exceeded 500.
- The total connected customer capacity increased by 18% and grew to more than 8 Tbits.
  - **Mumbai**
    - At the region's flagship IX, data throughput at peak times increased by 18%, reaching 1.1 Tbit/s.
  - **Delhi**

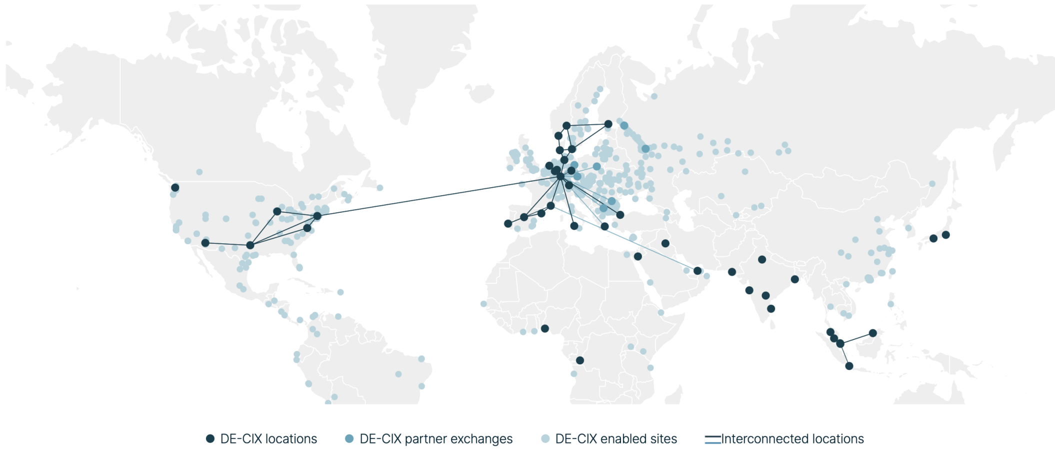

Fig. 18. DE-CIX Global Presence Map

  * DE-CIX Delhi saw a massive increase of data throughput at peak times, growing by 97% and amounting to 198 Gbit/s by the end of the year.
  * The number of 100GE ports grew by 43%.
  * The IX saw a 35% increase in connected networks.
- **Kolkata**
  * DE-CIX Kolkata enjoyed increased demand in 2022, with a 25% increase in connected networks.
  * Data throughput at peak times skyrocketed in Kolkata, increasing by a factor of 20 compared to the previous year.
- **Chennai**
  * DE-CIX Chennai saw 80% growth in connected networks in 2022.
  * The connected customer capacity increased by 27%.

Overall, the data indicates a positive correlation between the presence of IXPs and improved network quality. By reducing reliance on distant, often congested network routes, IXPs like DE-CIX provide a more optimized and efficient path for internet traffic.

XIV. Future Work

- Combining our approach with the Broadband Subscription Tier (BST) methodology [5] that associates a speed test data point with a residential broadband subscription plan can enhance our dataset even further. This will be a huge undertaking and can be done quicker with proper support from stakeholders such as M-LAB and OOKLA who can capture this tier information from the clients.
- Incorporating trace route information from RIPE ATLAS more deeply as part of our high level analysis can be beneficial.
- We can add in more countries to our generalized methodology for conducting a high level and low level analysis as part of future work to expand this study to more countries in the world. This will expand our case studies and help out researchers and policy makers in other regions of the world.

Acknowledgment

We would like to thank Dr. Alexander Gamero-Garrido since this study was a product of the final paper for both ECS 252: Computer Networks and ECS 289I: Internet Measurement & Policy. His support and direction throughout the process of this study was vital for getting it in the direction it has come.

References

[1] Mehra, B., Merkel, C., Bishop, A. P. (2004). The internet for empowerment of minority and marginalized users. New Media Society, 6(6), 781-802. https://doi.org/10.1177/146144804047513
[2] Boris Bartikowski, Michel Laroche, Ahmad Jamal, Zhiyong Yang, The type-of-internet-access digital divide and the well-being of ethnic minority and majority consumers: A multi-country investigation, Journal of Business Research, Volume 82, 2018, Pages 373-380, ISSN 0148-2963, https://doi.org/10.1016/j.jbusres.2017.05.033.
[3] Nikolaos Chatzis, Georgios Smaragdakis, Anja Feldmann, and Walter Willinger. 2013. There is more to IXPs than meets the eye. SIGCOMM Comput. Commun. Rev. 43, 5 (October 2013), 19–28. https://doi.org/10.1145/2541468.2541473
[4] Nils B. Weidmann et al.x, Digital discrimination: Political bias in Internet service provision across ethnic groups. Science 353,1151-1155(2016). DOI:10.1126/science.aaf5062
[5] Udit Paul, Jiamo Liu, Mengyang Gu, Arpit Gupta, and Elizabeth Belding. 2022. The importance of contextualization of crowdsourced active speed test measurements. In Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22). Association for Computing Machinery, New York, NY, USA, 274–289. https://doi.org/10.1145/3517745.3561441
[6] Kyle MacMillan, Tarun Mangla, James Saxon, Nicole P. Marwell, and Nick Feamster. 2023. A Comparative Analysis of Ookla Speedtest and Measurement Labs Network Diagnostic Test (NDT7). Proc. ACM Meas. Anal. Comput. Syst. 7, 1, Article 19 (March 2023), 26 pages. https://doi.org/10.1145/3579448
[7] Tripathi, Shivam. "2G SPECTRUM SCAM." GLOBAL FINANCIAL FRAUDS AND CRISES 1.1 (2018): 207-244.
[8] https://penntoday.upenn.edu/news/multi-layered-challenges-broadband-expansion
[9] https://github.com/AdityaKumarBej/Internet-Access-Disparities

[10] Phillipa Gill, Christophe Diot, Lai Yi Ohlsen, Matt Mathis, and Stephen Soltesz. 2022. M-Lab: user initiated internet data for the research community. SIGCOMM Comput. Commun. Rev. 52, 1 (January 2022), 34–37. https://doi.org/10.1145/3523230.3523236
[11] https://www.ookla.com/ookla-for-good/open-data
[12] Staff, RIPE Ncc. "Ripe atlas: A global internet measurement network." Internet Protocol Journal 18.3 (2015): 2-26.
[13] https://www.itu.int
[14] https://www.census.gov
[15] https://censusindia.gov.in
[16] wikipedia.org/wiki/List_of_Indian_states_and_union_territories_by_GDP
[17] https://pib.gov.in/Pressreleaseshare.aspx?PRID=1942055+
[18] https://www.statista.com/topics/1145/internet-usage-worldwide/topicOverview
[19] Shaping the Internet: 10 Years of IXP Growth arXiv:1810.10963 [cs.NI]
[20] https://www.tespok.co.ke/?page_id=11648
[21] https://ixp.net.ng
[22] https://www.internetsociety.org/resources/doc/ixpimpact/
[23] https://www.internetsociety.org/resources/doc/2020/ixp-report-2020/
[24] https://www.de-cix.net/en/about-de-cix/annual-report
[25] https://www.arizonafuture.org/progress-meters/infrastructure/broadband/
[26] https://www.trai.gov.in
[27] https://government.economictimes.indiatimes.com/news/technology/internet-partially-restored-in-manipur-broadband-allowed-but-no-mobile-internet/102141005
[28] https://thewire.in/rights/we-cant-look-away-from-internet-shutdowns-in-manipur
[29] https://www.khi.org
[30] https://www.kansascommerce.gov/broadband-map/
[31] https://qz.com/kashmir-saw-most-internet-outages-in-the-world-in-2022-1850004964