Whale Watching in Inland Indonesia: Analyzing a Small, Remote, Internet-Based Community Cellular Network

Matthew Johnson
matt9j@cs.washington.edu
Paul G. Allen School
Seattle, Washington, USA

Jenny Liang
jliang9@cs.washington.edu
Paul G. Allen School
Seattle, Washington, USA

Michelle X. Lin
linm23@cs.washington.edu
Paul G. Allen School
Seattle, Washington, USA

Sudheesh Singanamalla
sudheesh@cs.washington.edu
Paul G. Allen School
Seattle, Washington, USA

Kurtis Heimerl
kheimerl@cs.washington.edu
Paul G. Allen School
Seattle, Washington, USA

ABSTRACT

While only generating a minuscule percentage of global traffic, largely lost in the noise of large-scale analyses, remote rural networks are the physical frontier of the Internet today. Through tight integration with a local operator's infrastructure, we gather a unique dataset to characterize and report a year of interaction between finances, utilization, and performance of a novel, remote, data-only Community LTE Network in Bokondini, Indonesia. With visibility to drill down to individual users, we find use highly unbalanced and the network supported by only a handful of relatively heavy consumers. 45% of users are offline more days than online, and the median user consumes only 77 MB per day online and 36 MB per day on average, limiting consumption by frequently "topping up" in small amounts. Outside video and social media, messaging and IP calling provided by over-the-top services like Facebook Messenger, QQ, and WhatsApp comprise a relatively large percentage of traffic consistently across both heavy and light users. Our analysis shows that Internet-only Community Cellular Networks can be profitable despite most users spending less than $1 USD/day, and offers insights into the unique properties of these networks.

CCS CONCEPTS

• **Networks** → **Network measurement**; *Network manageability*; • **Information systems** → *Traffic analysis*.

KEYWORDS

Community Network; Community Cellular; Rural Access; Network Traffic; Measurement

ACM Reference Format:
Matthew Johnson, Jenny Liang, Michelle X. Lin, Sudheesh Singanamalla, and Kurtis Heimerl. 2021. Whale Watching in Inland Indonesia: Analyzing a Small, Remote, Internet-Based Community Cellular Network. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3442381.3449996

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3449996

1 INTRODUCTION

At the close of 2020, the UN's International Telecommunications Union (ITU) estimates that 51% of the global population is "using the Internet," and 93% of the earth's population is covered by a 3G, 4G, or 5G mobile broadband network [27]. Yet as has been well-documented over the last 20 years, the exact nature and definition of use can vary widely [15]. Furthermore, while 93% coverage is a great achievement, this leaves over 500 million people still unconnected, mostly in rural, remote, and hard to serve areas.

The GSMA published in 2017 that traditional mobile networks have largely expanded their footprint to *all* areas where service is profitable, that new deployments are slowing, and that new paradigms are needed if Internet access is to reach those who remain unconnected [5]. Community Networks, networks owned and operated by the community members that they serve, are a promising paradigm that changes the economics of sustainable deployment [8, 16, 25, 39]. Community *Cellular* Networks, Community Networks based on mobile access technologies like GSM, UTMS, LTE, (and now 5G-NR), offer advantages relative to technologies like WiFi mesh networks in performance or TV-Whitespace in device availability, but at the cost of more complex licensing and deployment [23, 30, 38, 46]. Community Cellular Networks are particularly well suited to remote rural use cases, where spectrum is readily available (if secondary use is allowed) and wide-area coverage is desirable. Due to their remote nature, these networks often operate with satellite-based backhaul links, limiting performance in terms of both latency (500ms RTT, although this is changing–see 6.3.1) and throughput (1-10Mbps aggregate for all users). While these conditions are constrained, they are the reality of "access" to the demographics currently on the frontier of the Internet.

In this work, we explore a modern, extremely remote, data-only, Community LTE Network in Bokondini, Indonesia. The network is run by a nonprofit affiliated with a school situated in the town center, and covers most of the town. We were able to tightly integrate with the operator's systems to gather a dataset integrating both technical and business information, allowing us to analyze user traffic as well as how users purchase and spend prepaid balance in the network. Our data collection spans over a year of the network's operation, covering 53 weeks from midnight March 10, 2019 to midnight March 15, 2020 (local time UTC+9). We ask whether a prepaid, satellite-based, data-only, mobile network can be profitable without subsidies, what applications will be used in such a network, how frequently will users interact with the network, and how will users manage their prepaid credit?

We find the network is profitable, but that average-based metrics fail to capture on-the-ground reality with a small population size. We observe individual use is highly intermittent; users elect to purchase many small amounts of data rather than one lump sum, and many are offline for days at a time. Use is also unequal. Heavy "whale" users contribute disproportionately to the network's financial sustainability, and light and heavy users differ both in the way they use the network as well as by how much. Video dominates the backhaul link, but video consumption is driven primarily by a small subset of users. Large platforms (Google & Facebook) source the majority of traffic and reach all users, but we find local trends can also have strong effects on the network.

We analyze how the network can remain profitable despite a relatively small number of users and an Average Revenue Per User (ARPU) of less than $1USD per day, and hope to offer a detailed characterization of the unique and understudied properties of an extremely remote network at the outer reaches of the mobile Web.

2 RELATED WORK

2.1 Community Networking

This research builds on a long history of work on community networking. Community networks, networks owned and operated by users in some sort of collaborative way, have long been viewed as a promising mechanism for increasing access among rural and disadvantaged populations [20]. Community networks can operate using a variety of technologies including standard telephony [2], 802.11 WiFi [8], or cellular protocols [25] and examples exist in both rural [23] and urban environments [39], and developing [23, 25, 31] and developed countries [8, 16, 39, 42].

With the variety of community networks, there are similarly a variety of engagements with them in the networking literature. For example, the team behind Guifi.net, operating the preeminent community network with over 35000 nodes in Catalonia [8], has explored a wide swath of issues including topologies [53], cloud services [45], and sustainability [7]. Works related to other networks have explored appropriate network architectures [23, 38], licensing [42], and many other myriad issues in community networking. A notable consistent thread is the importance of human factors in the operation of the network. This is echoed in Jang et al. which explored leveraging local actors to conduct maintenance and repair [28] and Moreno et al. discussing the importance of user engagement for sustainability of community networks [41].

Most similar to this paper, there is a body of measurement studies in community networks. Heimerl et al. [25] presented measurements of inbound versus outbound traffic in an Indonesian 2G community network, finding outbound traffic was much more significant. They also explored the sustainability of the network, finding that it was economically viable. Follow-on research explored phone adoption in the same community [26]. Cerdà-Alabern et al. explored the financials of the Guifi network [12]. Lertsinsrubtavee et al. [36] recorded Web usage in a Thai community network, finding that usage behavior was similar to that of commercial networks in significant ways, such as a focus on social networks. Unfortunately, they also found that user apps performed similarly (e.g. downloading lots up updates) despite the limited bandwidth available. This is supported by Johnson et al. [29], where a rural Zambian WiFi network saw similar behavior. Our work expands this literature to explore 1) the specifics of service utilization in data-only LTE networks, 2) service use *combined with* service utilization, 3) addition of techniques for mitigating the operational difficulties of data collection behind limited backhaul for studies like these, and 4) a more modern (2020) look into the operation of these networks.

2.2 Rural and Developing Networks

Outside of community networks (comprised of both developing and developed regions) [12, 25, 29, 36], there is also a body of network measurement literature in rural and developing regions. Often utilizing traces from telecoms, ISPs, or regional IXPs, these works explore the unique circumstances of rural networks. These include studies on broadband performance and adoption in Nepal [32], Pakistan [6], or South Africa [13]; cellular performance in India [48] and Pakistan [4]; censorship in Pakistan [1]; mobile phone properties in a Pakistani cellular network [3], and web latency in Ghana [57]. These have scaled up to continent-wide analyses, such as IPv6 adoption [37], interdomain routing [17], and inter-country latency [19] in Africa. World-wide studies exist as well, such as Schinkler et al.'s work on the performance of Facebook's edge caching [43].

The massive diversity of these works, inclusive of multiple continents, scales, technologies, populations, and venues, demonstrates the value of breadth in network measurement research. While each individual research agenda is not (and does not claim to be) "representative" of the Internet in whole, together they provide a holistic, diverse perspective on Internet use throughout the world. Our work contributes to this whole with the perspective of an extremely remote, data-only cellular network. In addition, we provide new insights by analyzing network traces together with records of user spending and the network's finances.

2.3 Small-scale Network Measurements

Lastly, a body of research focuses on small-scale networks, servicing households or small groups of people but not in an explicitly community-oriented fashion. One example is Maitland et al.'s exploration of Internet use in a refugee camp [44], where the network is run by the UN Refugee Agency. They found a wide range of experiences with the Internet in the camp, contributing to a set of barriers to access. Another set of works focus on tribal Internet, with Vigil et al. [54] explicitly focusing on failures in the use of apps like YouTube in the context of a TVWS deployment in US tribal lands. Even some of the community networking work cited above (notably Hasan et al. [23] and Martinez-Fernandez et al. [38]) involved partnerships with outside organizations, such as telecoms for spectrum, limiting the extent of the community participation. These works similarly broaden the range of measurements possible and inform elements of the novelty of our work.

We also note a large body of home and consumer Internet measurement [21, 22, 52]. While relevant in that community networks often leverage consumer Internet connectivity, their needs and expectations differ greatly from networks serving customers directly.

3 CONTEXT

3.1 Deployment Location

Bokondini, Indonesia, is a community of ~2,000 located two hours drive (on a rough muddy road) from Wamena in the highlands of Papua. Bokondini, a central location for missionary activities for decades, still has limited infrastructure. There is a small airstrip, but no community-wide water or reliable power. The local regency government moved to Bokondini two years ago, and the town has been growing quickly. A national carrier has provided 2G coverage via a satellite-based small cell through a universal access program for about 4 years, but there was no Internet access until recently. During the final weeks of this study the carrier began offering LTE service nearby, still served by satellite. Mobile phones are common, with a critical number of LTE-capable devices already present [47].

3.2 Network

The network profiled in this paper is an instance of a *Community Cellular Network*, owned and operated since 2013 by a missionary group whose primary function is running an elementary school in the area. They manage the day-to-day operations, including maintenance, credit sales to resellers, power management, and repairs. They operate a 5000KW microhydro and solar installation which powers the network as well as the school's lighting and computers. Unfortunately, the microgrid does not have enough reserve power to operate 24 hours a day, so the network is shut down manually between 11pm-5am (extended to 12-4:30am part way through the trial) to conserve power.

Our research group has a nearly a decade of experience working with them to explore different approaches to rural connectivity. The current iteration is a rural-optimized LTE network designed and installed in 2018. Demand has fluctuated, serving between 40 and 80 users when operational. Its topology is relatively simple, with Radio Access (RAN) provided by a commodity eNodeB installed on the top of the school's gymnasium, connecting existing user handsets. An x86 mini-computer hosts an Open Source EPC to terminate all LTE signaling from the eNodeB, connected to a generic IP router with NAT, and ultimately a consumer-grade very small aperture satellite terminal (VSAT) providing Internet backhaul. The site originally connected to another community via wireless relay to share a 3Mbps/1Mbps 8:1 (downlink/uplink & contention ratio) VSAT, but the relay was destroyed by lightning and a dedicated 1Mbps/256K 8:1 connection was temporarily established. This dedicated connection was later upgraded to 3Mbps/1Mbps 4:1.

Despite being an LTE network, no voice or SMS services are provided. The network operator encourages users to employ "over the top" (OTT) services like WhatsApp, Facebook Messenger, or Viber which are already extremely popular. Most users have dual-sim phones, and register themselves via SMS on the national carrier 2G network. Providing only generic IP data greatly simplifies the network architecture and reduces associated costs and liabilities from interconnect and identity (phone number) management.

3.2.1 Credit Model. The operator uses a prepaid model, where users pay cash to a reseller in 1:1 exchange for "credit" on their account denominated in the local currency. Users later use a locally hosted web application to convert their credit into "data," denominated in bytes, allowing Internet access until the corresponding amount of data has been transferred and the balance falls to zero. The country's main commercial carriers also use prepaid models, and they are well-understood by local users. For distribution, the operator first generates credit via an admin interface and sells it to three different resellers in the community at a wholesale rate. The resellers own small stores, selling basic goods like rice, oil, and candy, and are open for most of the day. They pass credit onto end-users with a small margin using a locally hosted mobile web-interface developed by the researchers for the network.

Users can purchase arbitrary amounts of credit from resellers, and transfer it between users accounts. After loading their credit, there are three data packages available: 10MB, 100MB, and 1GB. Data pricing is flat, at Rp250(~$0.01USD)/MB. Local services are zero-rated, and all external traffic is billed equally.

Table 1: Dataset summary statistics

	Log Count	%	GB	%
Internet Flows	56,001,999	74.5	1,324.9	98.9
Intranet Flows	19,179,804	25.5	15.1	1.1
Internet DNS Mapped	53,755,278	96.0	1214.6	91.7
Internet Org. Assigned	47,077,375	84.1	1250.7	94.4
Internet Categorized	46,826,941	83.6	1219.9	92.1
Transactions	40,450	100	-	-

3.2.2 Utilization. The Bokondini community network had ~50 users active at any point across the study period, where active means traffic to or from the user was measured in the network or the user made a credit transaction. Figure 1 shows the amount of traffic each day during the study period, the number of users active active at different aggregation levels, and the cumulative number of SIM cards registered with the network. We expected some attrition as SIMs are lost and replaced or users leave town. The combination of school break and several community members streaming a religious conference led to particularly high usage in late December and early January. During the last month of the study, a national operator began offering LTE service nearby, at a lower price point than the community network. This is correlated with the drop in traffic and active users observed in late February and early March. Despite the drop in activity, the network remained profitable.

4 DATASET & METHODOLOGY

We leverage two collected datasets: 1) network credit transaction logs, and 2) fused logs of IP and DNS metadata. We tightly integrated our instrumentation with the operator's existing systems to avoid overhead and gain access to ground-truth per-device information, but this required accepting the availability limitations of these existing systems detailed in section 3.2.

4.1 Data Collection

4.1.1 Credit Transactions. The credit transaction log records (1) when an administrator adds credit to the system, (2) when credit is transferred between users, and (3) when a user ultimately purchases data with their credit. Each entry contains a timestamp, user ID,

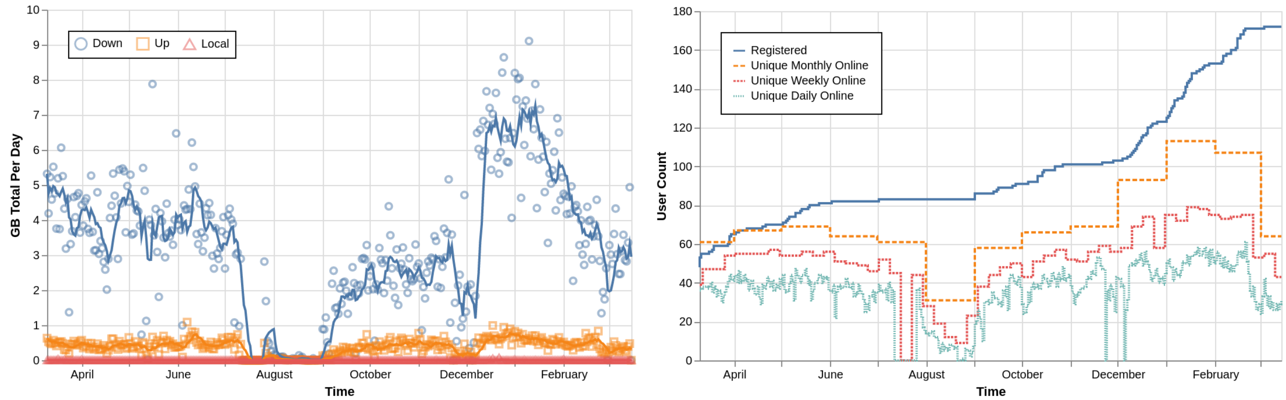

Figure 1: Network Activity vs. Time. The amount of traffic per day(Left) is highly variable and enveloped by the operational events detailed in table 2. Internet traffic eclipses local traffic by two orders of magnitude. The monthly online user count holds steady at around 60 users except for an extended outage in the summer and a peak in early 2020 as many new users join the network. Which users are online on particular days is variable and intermittent.

transaction type (see section 3.2.1), amount, and for transfers, a destination user ID. It is implemented as an asynchronous Javascript module extending the network's existing admin application. This analysis contains 40,450 transaction logs (see table 1).

4.1.2 Data Flows. The data flow log records an entry for each flow in the community cellular network. Flows are defined by their "five-tuple", consisting of the IP source and destination address, the transport layer protocol, and transport layer port numbers if applicable. The data was collected IPv4/IPv6 agnostically, and we use the term "IP" here to refer to either IPv4 or IPv6. The system aggregated individual packets from each flow into 20 minute intervals, and at the end of the interval recorded the five-tuple, the interval start and stop timestamps, and the total number of bytes transferred in uplink and downlink. A shorter interval was not used due to anonymization concerns (see section 4.3.1) and limited local storage. By integrating with the network's policy control and billing system, our instrumentation could map each flow to a SIM card and user account. A post-processing step replaced the flow's local address with a scrambled user ID, coded with the same key as the credit transaction log to associate flows and transactions with the same entity. The raw data contains 75,181,803 flow logs.

4.1.3 Intermediate DNS Responses. DNS establishes a likely mapping between observed destination IP addresses and the domain name a user is contacting. To augment the flow data with domain information, we collected an intermediate dataset of response timestamps, requesting scrambled user ID, domain requested, dns server response code, list of IP addresses returned from the DNS server, and list of address TTLs.

In post-processing we reconstructed the likely DNS state at each user's client over time, modeling each user as a single device with a shared DNS state, and iterating through the combined raw DNS and flow logs by time for each user. Each DNS response updates the client state to map the response IP addresses to the requested domain name. Since multiple DNS entries may point to the same IP address, the mapping can be ambiguous. We record the set of possibly ambiguous names and select the most recent for annotation. For each flow encountered, the current client DNS state is consulted.

Table 2: Notable events impacting network operation.

July 2017	Initial site visit and surveys
October 21, 2018	Initial launch of network with pilot users
October 31, 2018	Scaling of network by adding 10 new users
Feb 18, 2019	Transition to open network
March 10, 2019	**Beginning of dataset**
July 12-26, 2019	Extended outage due to relay lightning strike
July 26, 2019	Reconnected directly to school's VSAT
July 26-Sept 1, 2019	No credit sold while working with school
November 22, 2019	Operation extended to 4:30am to Midnight
December 1, 2019	VSAT Upgrade to 3/1 Mbps at 4:1 contention
February 20, 2020	National carrier begins operating 4G nearby
March 15, 2020	**End of dataset**

If the IP address has a known mapping, it is assigned from the DNS state. If the IP is not in the client state, we attempt a reverse DNS lookup for the IP address. Only if the reverse DNS lookup fails, we mark the host as unknown. For each flow we record the name, type of mapping (Client DNS or reverse DNS), and the count of possible ambiguities. While simplistic, we find this model mostly sufficient for this dataset, with 50,676,332 of flows covering 90.5 % of bytes coming from observed client DNS responses, and 41.1% of flows and 31.7% of bytes having an unambiguous mapping.

4.2 Data Processing

Once collected and anonymized, the data was uploaded to a central server for analysis. We periodically uploaded subsets of the data during off-peak hours to minimize the impact on user traffic. Before analysis, we removed users who join the network less than one week before the end of the data collection period (N=0) or who were active for less than one day (N=3). After filtering the final dataset contains 168 users and 72,278,238 flows.

4.2.1 Classifying Domain Names. Through manual inspection we hand-built rules to classify domain names and tag them with an "organization" and "category." We built the classifiers by iteratively grouping flows by domain, sorting by the total bytes transferred, and looking for patterns in the top domains. As passive observers in the network, our methodology provides no ability to accurately determine the ground truth contents of encrypted flows. We consulted the domains themselves, whois data, publicly available documentation in the case of APIs, and sites hosted at the domain to determine the parent organization responsible for each domain and its contents. We classified all domains and IP addresses with 200MB total transferred in the network or more, assigning an organization and category to over 83 % of flows covering over 92 % of the total traffic.

We have made a best-faith effort to categorize traffic as thoroughly and faithfully as possible, preferring more detailed categories like "video" or "messaging" over generic ones like "social media" or "software and updates" where possible. We fully acknowledge the limitations of this approach, discussing them in more detail in section 4.4. Ultimately we mapped 98 organizations and 20 categories. We provide these artisanal classification rules and processing scripts for independent scrutiny and reuse, see section 4.3.1.

4.2.2 Detecting Peer-to-Peer Flows. We found peer-to-peer flows facilitated by ICE(Interactive Connectivity Establishment) account for a notable fraction of traffic, particularly in the uplink. We separate these flows into their own category by reconstructing the ICE state at each client and in the network NAT/firewall. ICE works by having each peer open a connection from the client to a common server, establishing an open port in each peer's NAT at the NAT's public IP address, which is visible to the server. The server then shares the each peer's public port and address with their counterpart, which the peers attempt to re-use to establish direct connections if their NATs/firewalls allow it. If a direct connection cannot be established, the server falls back to relaying connection packets on behalf of the peers.

We reconstruct the ICE state by iterating through the flow logs for each user, and tagging any flows to the well-known STUN/TURN listening ports (UDP:3478 or TCP:5349) as likely ICE. After an ICE flow is detected, if a new flow starts within 1 minute to an unknown IP address but with the same client ephemeral port, we reclassify the flow as "Peer to Peer" instead of "Unknown." Due to port reuse, there is a low but nonzero probability of false-positive detection, so we do not reclassify flows already classified by domain.

4.3 Ethics

This work utilizes anonymized per-user flow metadata and network transaction history to better understand the dynamics and economic sustainability of Community Cellular Networks. In consultation with our institution's IRB and per locally applicable Indonesian data protection law, we determined that this work did not need explicit end-user consent since users were aware that this information was accessible to the network operator, there is low risk of harm, and a consent process would require us to collect identifiable user information which we would not gain access to otherwise. Insights from this analysis have been shared with the network operator to improve the quality of service in the community.

4.3.1 Anonymization and Data Access. During data collection the operator scrambled all network IDs and transaction IDs with a key, which remained in the community and was destroyed after data collection. The key scheme allows correlations between the flow and transaction logs over the study period only. In generating flow

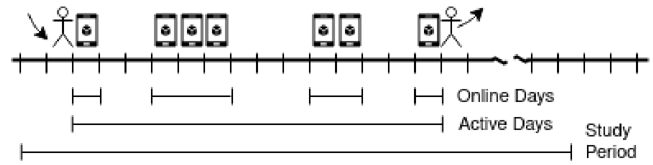

Figure 2: For each user, we count an "Online" day when the user transferred at least one byte that day. We define the number of "Active" days as the span between their first and last interaction with the network.

metadata, only L3 addresses, the L4 protocol number, L4 ports, payload size, and unencrypted DNS responses were collected. Packets were binned into 20 minute flow chunks and aggregated before storage, preventing fine-grained timing analysis. All traffic to organizations with N<5 unique users was grouped and references to the original domains were dropped from downstream analysis. The dataset is available open-access at https://doi.org/10.6084/m9.figshare.13116740.

4.4 Limitations

4.4.1 Flow aggregation. Using aggregated logs instead of per-packet traces limits the analysis resolution and obfuscates protocol-level performance. In particular, we cannot comment on the efficiency of individual flows at the transport level, leverage deeper packet inspection to verify which higher-level protocols are in use, or use ML-based timing analysis to predict the flow contents.

4.4.2 Domain and IP-Based Classification. The process of content categorization is subjective by its nature, but essential for high-level analysis, since modern CDNs and distributed edge infrastructures mean that large numbers of unique domains map to the same organization and service. Our dataset's limitation to aggregated flows only adds uncertainty to this process. As an example, we categorize general infrastructure from traditional social media providers like Facebook or Twitter as general "social media," except if the subdomain explicitly indicates it hosts video (`video.*.fbcdn.net`), messaging traffic (`mqtt-edge-*.facebook.com`), or advertising (`lithium.facebook.com`). While some organizations have an infrastructure more amenable to categorization, which uses different domains for different types of content, others do not. For example, all TikTok content appears to come from one set of converged media servers, even though the platform supports both video and messaging. We classified converged services by their predominant category ("video" in the case of TikTok), or a mixed category if there is no dominant content type. In the cases of Google and Facebook, it is difficult to distinguish traffic from different user-facing applications but that are part of the same corporate conglomerate. For example, it appears that WhatsApp and Instagram use media CDNs hosted at `fbcn.net` subdomains, and YouTube pulls content from `video.google.com`. All users connected to IP addresses in Google-owned blocks that had no publicly queryable DNS information. We attempted to break applications into their own classes when possible, but were not able to in all cases.

4.4.3 Generalizability. We partnered with the network in Bokondini due to its extreme remoteness and its uniqueness as a standalone satellite-backed LTE network with an approachable operator.

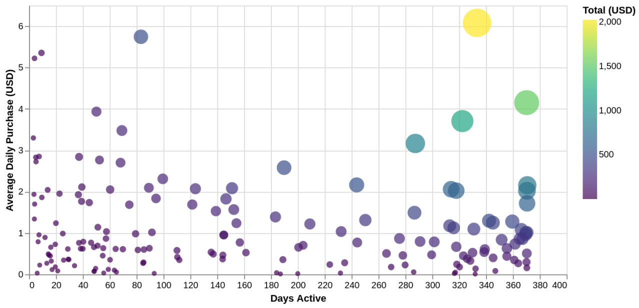

Figure 3: Daily Purchase vs. Days Active. Total purchased amount is encoded in the size and color of each point. Most users make small purchases independently of how long they have been active, contributing little to overall revenue.

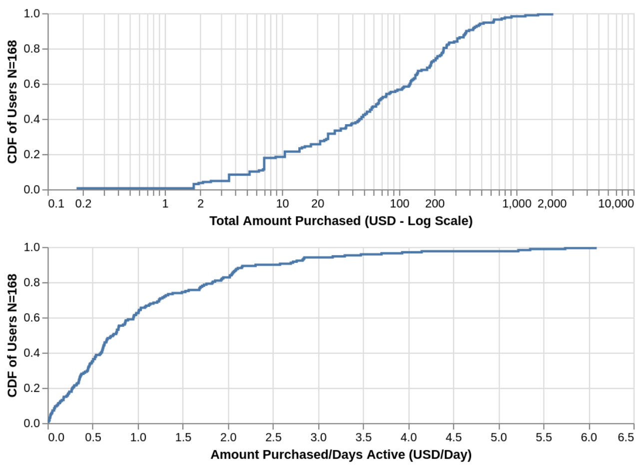

Figure 4: Distribution of spending across users. Consumption was highly unequal, with over half of users spending less than $100 USD while some users spent over $1000 USD.

All places are unique, but Bokondini is geographically and culturally similar to others in the remote highlands of Indonesian Papua. We don't claim to show generalizability due to the limitations of our dataset, but see no reason why our findings should be site specific.

5 RESULTS

In this section we detail the major results of our analysis. Contrary to our expectations given the network's low throughput (3/1Mbps D/U shared across all users), we find highly unequal consumption between users, that many users consume with intermittent access rather than frugal access, and that video and major platforms still play a large role in the network. Additional non-essential information and supplementary plots can be found at https://github.com/uw-ictd/ccn-traffic-analysis-2020.

5.1 Inequality & Sustainability

5.1.1 A Wide Range of Spending. Network use was highly skewed, with some users spending 5.5x the average amount, and 8.3x the median. Figure 4 shows the distribution of total network revenues from each user. While over 20% of users spent less than $10USD and 50% less than $100USD, three spent over $1000USD equivalent, generating an outsized portion of total revenue. Figure 3 normalizes spending by the amount of time users are active. The three heavy spenders are visible as outliers in the top right of the chart, spending a large amount per day on average and connecting consistently.

The heaviest users average over 300MB per day online, while the median consumes only 76.6MB. This disparity surprised us, since we hypothesized the network was bandwidth constrained, would not meet demand, and would have many users at the ceiling of available capacity. This has implications for the planning and operation of remote networks, further discussed in section 6.3.

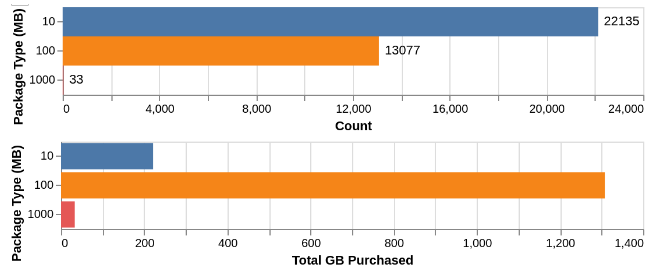

Figure 5: Relative use of data packages. The small (10MB) was most popular, but most bytes were purchased with the medium (100MB). The large (1GB) package was used rarely.

5.1.2 Frequent Data Topups, But Sporadic Credit. The network billing system offers 3 data packages: 10MB, 100MB, and 1GB to purchase with account credit at a uniform price per MB (see 3.2.1). Examining the transaction logs, we found the 10MB package is the most popular, while most "bytes" are purchased in the 100MB sized package. The 1GB package is rarely used (see figure 5). The network's flat pricing schedule does not incentivize purchasing large packages, and users will often quickly purchase several small packages to synthesize the exact amount they want. We grouped the chains of purchases which occur within one minute from one to the next, and plot them in figure 6. Synthesized packages are still often small, with the bulk of packages coming in at 200MB or below. The network's interface is designed to minimize friction for this workflow, requiring only two taps to make a data purchase.

Users tended to make frequent small data purchases through the web portal multiple times a day, even after accounting for purchase chaining. Figure 7 shows a CDF of the time between data purchase chains across users. 95% of users make a purchase every 10 hours or less on average, and over 90% of users have a 90th percentile inter purchase time of 10 hours or less as well. Frequent purchasing could help manage overall consumption and provide a sense of control over spending and background processes.

We hypothesized there would be a clear distinction observable between users who maintain a store of credit in the network for on-demand data purchasing and users who do not, but we found the reality to be much more ambiguous. 23.8% of users maintain a positive credit balance more than 95% of the time, but they tend to be new to the network. Of users active for more than 30 days, only 19.3 % have a positive balance more than 95% of the time. Figure 8 plots the fraction of time that each user had no credit while active versus the number of days they were active. We observe a wide variety of ratios at all time scales, indicating that the amount of

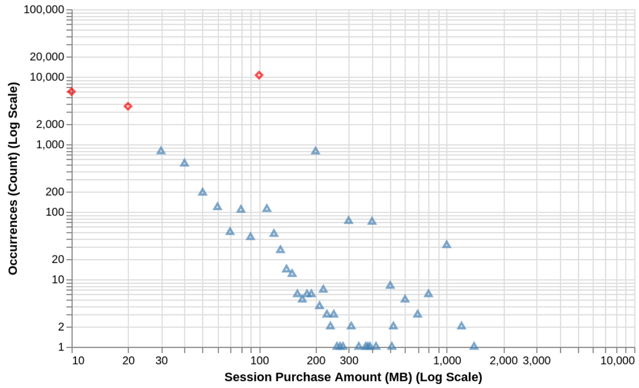

Figure 6: Amounts of data purchased in a single "chain". Amounts which occurred more than 1000 times are marked with red diamonds. The predefined 10MB and 100MB amounts were most popular, but users frequently synthesized non-standard packages to better suit their needs.

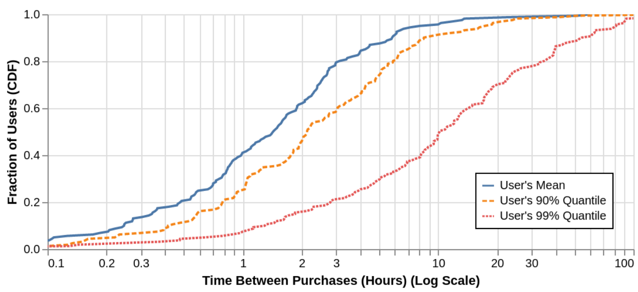

Figure 7: Time between user data purchase "chains". Most users commonly purchase data multiple times a day.

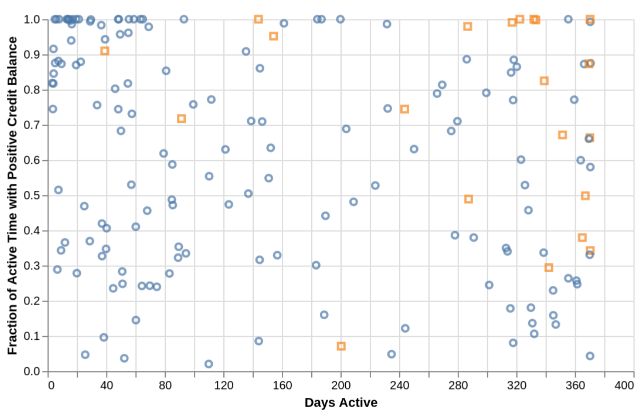

Figure 8: A cluster of new users tends to have nonzero balance most of the time, but a wide variety of ratios can be seen across both long-time and relatively new users. Orange square users were already members at the beginning of logging and have a manually adjusted start balance.

time a user has zero balance may be more random and situational rather than a strategic choice.

Having a 0 balance means that the user does *not* have data available on demand (in case of emergency or otherwise), and would need to visit a physical reseller to get access. In other contexts

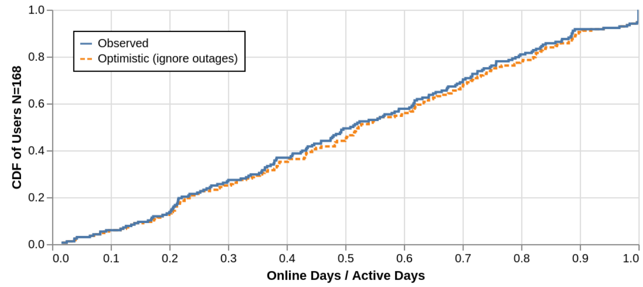

Figure 9: CDF of Online ratios. The fraction of days users connect is highly variable. ~8% connect most days, but the remaining ~90% are distributed across a range of online ratios, even conservatively ignoring days with unplanned outages.

researchers have noted that users may prefer to not carry a digital balance to avoid pressure from friends and family to loan them data [56], or to keep reserves in more flexible cash with different affordances for bargaining and negotiation [34].

5.1.3 Intermittent Use. Network use, even among the heaviest users, is highly intermittent. Comparing the number of days a user is online to the number of days they are active (see figure 2), we find that the median user is offline 53% of the days in their active time window. Only 7.7% of users access the network more than 95% of the days they are active. Figure 9 shows a CDF of the user Online Day/Active Day ratio, showing a roughly uniform distribution of the fraction of time online after accounting for the small number of users always online.

This itermittency impacts the operator's network planning and business operations, and is reflected in top-level statistics about the network. Figure 1 demonstrates this, where the count of unique daily, weekly, and monthly users differs substantially.

This intermittency combined with the varied amounts of time users spend without data-on-demand reflects on the core use-cases of the network. For the majority of users, their connection is *not* an always-available lifeline, but rather more sporadic and asynchronous. This may be partly due to the pace of life in general in the community, where residents are used to tasks taking days or weeks due to infrastructural limitations, or the availability of the national carrier's 2G network for small urgent messages, making the LTE network less essential.

5.1.4 The Network Is Financially Sustainable. Despite intermittent use and a relatively small number of users, the Community Cellular network in Bokondini is financially sustainable *without an external subsidy*. Re-use of local infrastructure and local installation labor kept the install capital expenses below Rp150M (~$10,000USD). Regular maintenance, mostly related to the power system, averages Rp1.3M (~$95USD)/month, and the satellite subscription costs $300USD/month (~Rp4.3M). Repairing the backhaul after the lightning strike cost two months of lost revenue and $1000USD (~Rp14M) in repair costs, but was covered by existing backup funds. Even with the downturn in use in February, revenues exceed costs, and are being invested in expansion to surrounding communities.

Figure 10 visualizes the network's cumulative revenue over the study period (ignoring early revenue from the pilot). Calling attention to the role of anchor users, we plot the revenue curves

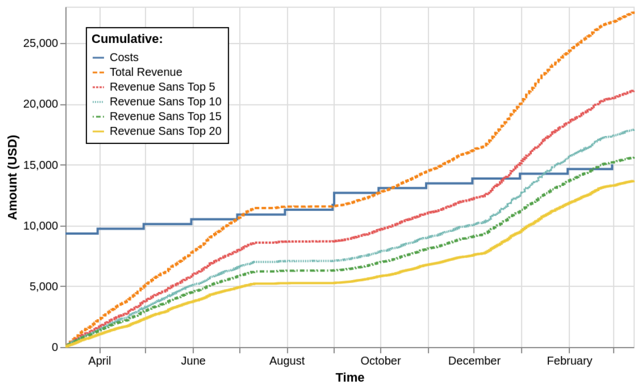

Figure 10: Revenue and Costs vs Time. The Bokondini network is financially sustainable, with top users contributing a large share of revenue. The costs line includes the upfront capital expense to deploy the network, regular operational expenses of monthly maintenance and backhaul, and the incidental operational expense to repair the network after lightning damage. User support is handled informally at the network's small scale and not accounted.

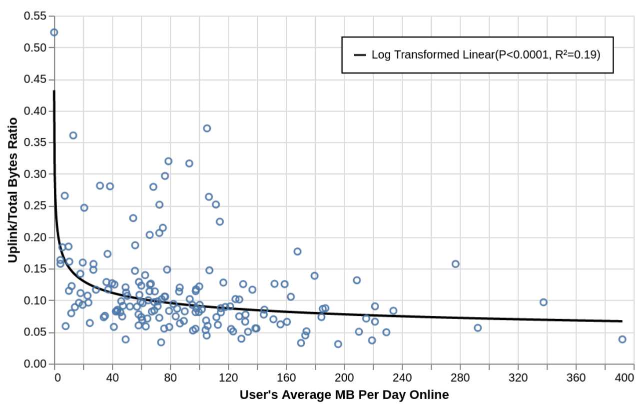

Figure 11: Uplink/Downlink Ratio vs. Consumption. Many light users have relatively more uplink traffic than heavy users, placing a different workload on the network.

excluding the top 5, 10, 15, and 20 overall users. The commonly quoted "ARPU (Average Revenue Per User)" metric does not capture the diversity of the underlying user population, and the impact that losing even a handful of these anchor users would have on the network. Without the top 15 users the network would still be sustainable after a year of operation, but a typical rural small cell site, with capex costs on the order of $40,000 USD [50], would not be. We discuss sustainability further in section 6.3.

5.2 Whales Engage Different Parts of the Web

We expected users would be limited similarly by the tight constraints of the Bokondini network on the modern web, but we find structural differences in the traffic of heavy and light users. Light users tend to have more balanced uplink/downlink ratios than heavy users, less video traffic, and also to abstain from games, content uploading, and dedicated antivirus. This indicates that rather than just using the network less, lightweight users are actually using the network differently, consuming a different mix of content, likely encountering different performance constraints, and placing a different burden on the network. Figure 11 plots the uplink/downlink ratio of each user vs. their average consumption. There is a weak but significant overall correlation ($P < 0.0001$, $R^2 = 0.19$).

5.2.1 Video. Examining the categories of flows attributable to heavy and light users, we find that video traffic makes up a disproportionate fraction of content from heavy users, while other categories stay relatively constant. All users have at least some traffic in the video category, but figure 12 shows the explosive growth of the video category between light and heavy users, concentrated in the top 10% of users overall. General social media use increases for the top 50% of users, but does not see the dramatic increase video does. We note that video from mainstream applications like Youtube, Facebook, and TikTok significantly outweighed adult video sites (~9:1).

In the network as a whole, video (both adult and non-adult) only consumes 37% of the download bandwidth, compared to the global mobile Internet market where video makes up 65% of mobile download traffic [14]. Examining only the 10 heaviest users, video still only makes up 49% of download bytes. All users in Bokondini consume less video than the average global user, and the median user consumes significantly less (~1/3) as a share of her total consumption. The under-representation of video overall compared to global trends could indicate that prices are too high to support carefree video streaming and the media rich Web, or the network may not have sufficient capacity to meet demand.

5.2.2 Hotspotting: NBD. While anecdotally most users connect to the network via a smartphone, we observe traffic to domains commonly associated with computers, such as `update.microsoft.com` (Users=5), `download.adobe.com` (Users=7), and `cdn.mozilla.net` (Users=16). Although we expected PC users to consume significant traffic, and these users all fell into the top 50%, they were not the heaviest users in the network. Any educational outreach and/or tool development to manage network traffic will need to focus on mobile media consumption for the greatest impact.

5.3 Platforms, Reach & Utilization

Breaking traffic down by organization, we see that some organizations interact with almost all users, while others communicate with only a small subset. Unsurprisingly Facebook and Google receive traffic from all users, and WhatsApp from almost all, but TikTok, QQ Messenger, Twitter, ShareIt Games, W Share, and UShareIt are also popular, even though they account for a smaller share of overall traffic. Compressed web content, consisting of AMP pages or sites served through the UC browser, was also relatively common. All users interacted with local services to purchase more Internet data from their credit balance.

5.3.1 Large Platform Dominance. Traffic to and from Facebook owned properties (Facebook, WhatsApp, Instagram) made up 39.9 % of bytes, exceeding the Asia-Pacific regional average of 35 % [14]. Google and google affiliated services account for 31.5 %. Taken together, these two platforms alone account for 71.4 % of traffic in the

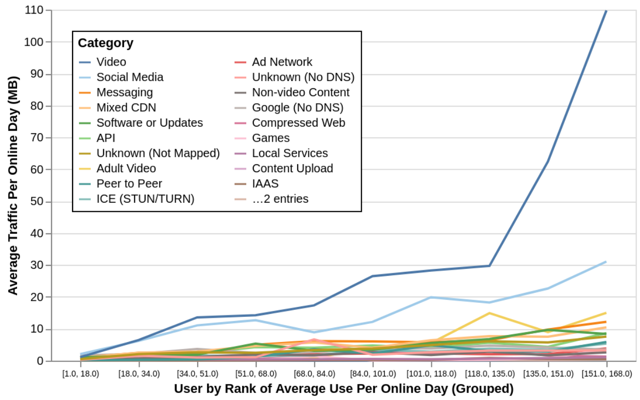

Figure 12: Bytes per Category vs. Decile. Video and Social media grows disproportionately between light and heavy users. The top 20% consumes a large share of total video traffic.

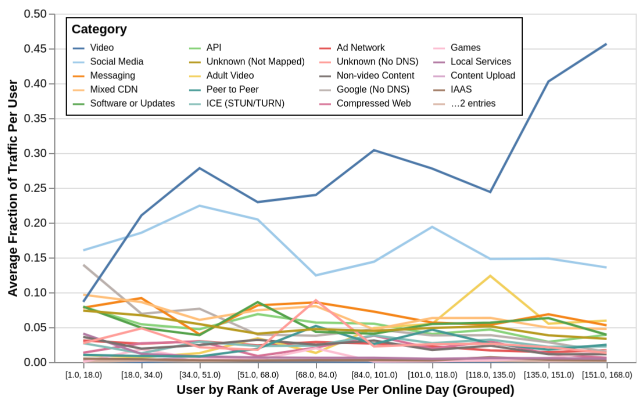

Figure 13: Share per Category vs. Decile. Video use as a share of total traffic increases sharply for the top 20% while most other categories remain flat.

network, supporting observations that large platforms (Facebook in particular) have wide reach in remote contexts.

5.3.2 Messaging Universality. While video and content consumption account for the most bytes on the network, we found that a significant amount of network resources go to messaging traffic and realtime UDP flows, particularly in the uplink. 90 % of users have some UDP uplink traffic facilitated by ICE (see section 4.2.2), and ICE or peer to peer bytes make up 6.1 % overall and 26 % of the uplink. Widespread messaging and communication via the Bokondini community network is surprising given how intermittently most users are connected and the availability of competing 2G voice and SMS services. Intermittent messaging use offers an example of how technology can be adapted to the constraints of remote edge networks in distinct ways from how it is used in well-connected areas. We discuss OTT applications and peer-to-peer communication further in section 6.2.2.

5.3.3 Local Trends. Due to the small number of users, local phenomena can cause large operational impacts to the network. In December and January we observe 22 users start streaming sessions from a month-long conference, consuming ~20 % of the network's

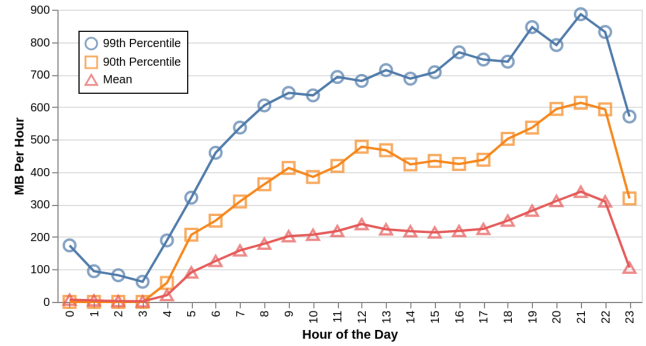

Figure 14: Bytes per time of day. The network is usually powered down at night, but it is a manual process and is skipped some days if there is power to spare.

resources for most of the month. The conference site was the 8th heaviest destination for the year overall, even though it was only visited for a little over a month by a small set of users. Ways to cache or "re-stream" content on the local side of the satellite link could greatly mitigate the impact of similar local phenomena.

5.3.4 Resource Utilization. The total data transferred per day had high variance, with a mean of 3.82GB and standard deviation of 2.15GB. There are some notable outlier days, where the network saw more than twice as much traffic as usual. Within each day, visualised in figure 14, there tends to be a slight bump in utilization around 12 noon (lunch time) and a stronger increase in use in the early evening from 6-10, peaking at around 9.

Although the operator maintains a symmetric infrastructure internally, their satellite backhaul is asymmetric with a 3/1 mbps down/up ratio. Observed traffic actually exceeds this ratio, with an overall down/up ratio of 8.5:1 (but which varies per user, see Figure 11). This indicates downlink is the likely bottleneck for the system workload, and the uplink could be better utilized or reduced.

5.3.5 Local Services and Local Only Traffic. The network has two zero-rated local services: one is a portal where members can view their balance, purchase more data with credit, or transfer credit to another user, and the other is a local media server stocked with material from the school hosting the network. All users interact with the portal, but it is a basic web application and does not contribute much to the total traffic. The media server is heavier, serving video and other rich content, but sees much less regular demand.

While the network allows peer-to-peer communication within the community, there is only very sparse peer-to-peer traffic. Almost all local interactions are between a network user and one of the two provided services. This may be due to a lack of knowledge that peer-to-peer is available, or assumptions built into the wider ecosystem that most LTE networks do not allow local connectivity. The widespread use of sharing apps indicates that there is demand for sharing, but almost all sharing activity in our dataset is mediated by Internet services rather than relying on direct discovery in the local network. This is supported by prior work [25] that found that community cellular networks are used disproportionately for external communications.

6 DISCUSSION

6.1 Practical Challenges Measuring Small Nets

This study presents a longitudinal deep-dive into the operations of a small Community Cellular Network providing sustainable data services in an area only recently reached by other operators. Conducting this study required significant multi-year effort to coordinate and integrate new systems, which does not scale down proportionately with network size. By web standards our dataset is miniscule, and a similar level of integration effort with a larger network would have yielded a much more statistically powerful result, just due to the scale of the targeted network. Yet small operators play a large role in rural access [24], and face unique constraints and challenges. Tools and techniques, both technical and organizational, that lower the burden of conducting research and developing technologies for these environments will likely play an important role in expanding Web access over the next decade.

6.2 Network Design for Content

6.2.1 Video and Social Media in Constrained Networks. Despite operating at the extreme edge, we saw video and social media weigh heavily in the composition of the network's traffic. Video is delay tolerant and resource intensive, but currently peaks at the same time as real-time communication flows. Prior studies have found that media often circulates locally in close-knit communities, and is delay tolerant by nature [54]. Traffic shaping tools to identify heavy flows and prioritize realtime traffic on the satellite backhaul downlink could improve the experience of a large number of light users at the expense of only a minor delay in a video download. Incentives to demand-shift video consumption (by marking videos to download in the early morning), or encourage local peer-to-peer sharing could also be explored.

Policy around video opens interesting ethical questions about the values embedded in the network's operations. Is video more or less important than other types of traffic, and is it acceptable that video consumes such a large fraction of the link?

6.2.2 Messaging Applications & Peer To Peer. A key differentiator between the network profiled in this research and prior work is that it is a **data only** LTE network. The Bokondini network operates neutrally, charging a price per byte independently of the type of traffic. This forefronts the primacy of flexible over-the-top messaging services, such as WhatsApp, Messenger, or Viber, over in-network protocols like Voice-over-LTE (VoLTE). While an all-data approach eschews traditional network-based quality of service differentiation, our experience in Bokondini shows IP-based designs to be a massive success in the context of rural networks with extremely limited backhaul, even for voice.

Anecdotally the network operator reports users think the call quality of WhatsApp over the community network substantially exceeds the quality of calls via the national operator's existing 2G network and new 4G network, yet it is not immediately clear why this is the case. Telecom standards by their nature lack flexibility to adopt new technologies, and can lag behind state-of-the-art techniques than be quickly pushed out to OTT applications. OTT services, unlike centralized telecom voice services, also naturally support peer-to-peer communication, and transparently establish low-latency connections within the community when possible. WebRTC extends this capability to browser-based Web applications as well [55]. While local call routing is possible in LTE, it is inconsistently implemented since it adds signalling complexity and limits the telecom's ability to correctly track calls. Peer-to-peer may be a nice-to-have feature in well-connected areas, but it can be the difference between usability and frustration on the extreme edge.

In future work, we hope to rigorously measure the impact of OTT app design on service quality, digging into the anectdotal evidence that OTT apps are both less expensive and more performant in extreme-edge conditions despite the lack of network integration.

6.3 Sustainability & Whales

The sustainability of rural networking solutions is a hot topic in many policy circles. A host of models exist, leveraging measurements such as the network ARPU (average revenue per user) to show where networks are and are not viable. Our analysis provides a more thorough look into user behavior, specifically the *distribution* of subscribers in an area. Figure 3, in particular, shows that these networks have a wide range of types of subscribers, perhaps more akin to the "whales" present in the freemium games literature [49]. In our case, three users dominate revenue generation, pushing the network ARPU (Rp190,308/mo) significantly above the country average reported by Telkomsel (Rp47,000/mo in April 2019 [11]). Coarse metrics like ARPU, and even average installation costs, obscure the reality that each community is a unique location with unique citizens. While the network is still sustainable without these three individuals, we believe it is important to understand that "there is no average user" [10].

In terms of building models for future deployments, a survey of the community may not find these individuals or may find just them. This variance, inherent in operating where the overall number of residents is low when compared to dense urban situations, makes predicting the sustainability of these networks more difficult. Because of this, we argue that bottom-up decision making, likely via *local* entrepreneurs interested in providing connectivity to their communities, is a more efficient and sustainable way to allocate resources on the edge.

6.3.1 Future Backhaul. Several companies (OneWeb, Amazon, and SpaceX, among others) are actively working to send thousands of satellites into low-earth orbit (LEO) with the promise of high-quality, low-latency global connectivity. As noted in Kleanse et al. [33] "Constellations of hundreds to thousands of satellites promise to offer low-latency Internet to even the most remote areas."

Given how much this paper focuses on the challenges of operation behind a backhaul-limited satellite link, these are exciting times. We wonder though what the ultimate long-term impact of widely available high-performance satellite networks will be. Will any of the current community network anchor tenants prefer to purchase their own private terminal and undermine the sustainability of the community network? Will rural site economics change such that a national operator's network will cover the entire region? We are cautiously optimistic that these networks will *eventually* deliver service in remote areas with fiber-like latency and an order of magnitude more throughput than existing geostationary VSATs.

While this will be a welcome development in the Bokondini network, the per-terminal costs, real-world performance, reliability, and market impact of these networks remain to be determined.

7 FUTURE WORK

Edge Measurement. The practical difficulty in gathering measurements from the network was surprising. Part of this was due to the fact that we were not colocated with the network, but with nearly 30GB of data to analyze, being present in Bokondini would have stripped us of the ability to investigate the data at scale using cloud resources. Our group is currently running or assisting in the operation of networks in the Philippines, Mexico, Hawa'ii, and the Arctic, so we expect this issue will continue to manifest.

We hope to develop a network-side application which could perform an efficient, potentially streaming, first-pass analysis and transfer only compressed results to a longitudinal telemetry service. This approach could draw from advances in federated learning [9], to keep sensitive data in the community while improving the quality of high-level analysis. This application could take into account network behavior and patterns, optimizing compute and network utilization against demand in the network. Optimizing the size and structure of the produced model given the expectation of future bandwidth availability will be a focus of our research.

Edge Caching. One result from this work is the importance of reducing backhaul reliance through caching and local loops. This is not a new topic in community networks; Guifi.net has done extensive work in local services [45] (though due more to political desires than backhaul limitations) and others have explored novel caching schemes in the developing world (notably Raza et al. [40]). Siskin [51], from Google, is a similar initiative to enable peer-to-peer connectivity in disconnected environments.

While these efforts bring novel ideas, there is no satisfactory answer for edge caching yet in an HTTPS world [18]. We do want to call attention to WPack [35], one particular effort that we find intriguing. WPack is a proposed standard for downloading and signing web content explicitly providing support for redistribution and caching. WPack seems well-suited to the remote community network environment and could re-enable network caching, while also supporting secure service utilization, if included in browsers. We are tracking WPack development closely and hope to eventually use it to implement a media cache for remote networks.

COVID-19 Analysis. Due to limited healthcare capacity, the government took aggressive action to quarantining area communities beginning on March 25, 2020. This began with closing schools, then all roads and markets the following week. The network ecosystem in the community changed drastically; the operators of two competing hotspot installations left town without providing infrastructure (e.g., network credits) for continued operation, and person to person contact within Bokondini was extremely limited.

It is difficult to draw meaningful conclusions from traffic data without a clear picture of how the situation evolved on the ground. We are exploring the impact of the pandemic and its response in-depth with a wider array of qualitative and technical data.

8 CONCLUSION

Through integration with a local operator's infrastructure, we gathered a unique dataset to characterize and report a year of finances and utilization in a, remote, data-only Community LTE Network. With visibility of single users, we found use highly unbalanced and the network supported by only a handful of relatively heavy consumers ("whales"). 45% of users were offline more days than online, and the median user consumed only 36 MB per day on average, making frequent purchases in small amounts. We showed that Internet-only Community Cellular Networks can be profitable despite most users spending less than $1 USD/day, and provided a characterization of unique properties of the network.

9 ACKNOWLEDGMENTS

We would like to thank David Haag, Ibel Itlai, Marmulia Siahaan, Scotty Wisley, and the Ob Anggen School for their support in operating and measuring the Ungu network. We also thank the UW-ICTD lab, Erin Wilson, our reviewers, and the WWW committee for their ideas, constructive feedback, and support. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1762114.

REFERENCES

[1] G. Aceto et al. 2016. Analyzing Internet Censorship in Pakistan. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI)*. IEEE, Bologna, Italy, 1–6. https://doi.org/10.1109/RTSI.2016.7740626

[2] Michael Adeyeye and Paul Gardner-Stephen. 2011. The Village Telco project: a reliable and practical wireless mesh telephony infrastructure. *EURASIP Journal on Wireless Communications and Networking* 2011, 1 (2011), 78.

[3] Mukhtiar Ahmad et al. 2016. A View from the Other Side: Understanding Mobile Phone Characteristics in the Developing World. In *Proceedings of the 2016 Internet Measurement Conference* (Santa Monica, California, USA) *(IMC '16)*. ACM, New York, 319–325. https://doi.org/10.1145/2987443.2987470

[4] Tayyab Arshad et al. 2016. Performance Evaluation of Mobile Broadband Cellular Networks in Pakistan. In *2016 IEEE 41st Conference on Local Computer Networks Workshops (LCN Workshops)*. IEEE, Dubai, United Arab Emirates, 104–111. https://doi.org/10.1109/LCN.2016.035

[5] GSM Association. 2017. Connected Society: Unlocking Rural Coverage: Enablers for commercially sustainable mobile network expansion.

[6] Muhammad Faheem Awan et al. 2015. Measuring Broadband Access Network Performance in Pakistan: A Comparative Study. In *2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*. IEEE, Clearwater Beach, FL, USA, 595–602. https://doi.org/10.1109/LCNW.2015.7365903

[7] Roger Baig et al. 2016. Making Community Networks Economically Sustainable, the Guifi.Net Experience. In *Proceedings of the 2016 Workshop on Global Access to the Internet for All (GAIA '16)*. ACM, New York, 31–36. https://doi.org/10.1145/2940157.2940163

[8] Roger Baig, Ramon Roca, Felix Freitag, and Leandro Navarro. 2015. Guifi. net, a crowdsourced network infrastructure held in common. *Computer Networks* 90 (2015), 150–165.

[9] K.A. Bonawitz et al. 2019. Towards Federated Learning at Scale: System Design. In *SysML 2019*. Google Research, Stanford, CA, 1–15.

[10] Neil F Budde. 2004. There's no such thing as an" average" user. *interactions* 11, 2 (2004), 54.

[11] CEIC Data. 2019. Indonesia Telecommunication Statistics: Monthly Average Revenue per Unit. https://www.ceicdata.com/en/indonesia/telecommunication-statistics-monthly-average-revenue-per-unit.

[12] Llorenç Cerdà-Alabern, Roger Baig, and Leandro Navarro. 2020. On the Guifi. net community network economics. *Computer Networks* 168 (2020), 107067.

[13] Marshini Chetty et al. 2013. Measuring Broadband Performance in South Africa. In *Proceedings of the 4th Annual Symposium on Computing for Development (ACM DEV-4 '13)*. ACM, Cape Town, South Africa, 1–10. https://doi.org/10.1145/2537052.2537053

[14] Cam Cullen. 2020. *The Mobile Internet Phenomena Report*. Technical Report. Sandvine, Plano, TX. 14 pages.

[15] Jonathan Donner. 2015. *After Access: Inclusion, Development, and a More Mobile Internet*. The MIT Press, Cambridge, Massachusetts ; London, England.

[16] N Echániz. 2017. The right to co-create the internet. Altermundi.

[17] Roderick Fanou. 2014. *On the State of Interdomain Routing in Africa*. Ph.D. Dissertation. Universidad Carlos III de Madrid, Spain.

[18] Adrienne Porter Felt et al. 2017. Measuring HTTPS Adoption on the Web. In *USENIX Security Symposium*. USENIX, Vancouver, BC, Canada, 17.

[19] A. Formoso et al. 2018. Deep Diving into Africa's Inter-Country Latencies. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*. IEEE, Honolulu, HI, USA, 2231–2239. https://doi.org/10.1109/INFOCOM.2018.8486024

[20] Pantelis A Frangoudis, George C Polyzos, and Vasileios P Kemerlis. 2011. Wireless community networks: an alternative approach for nomadic broadband network access. *IEEE Communications Magazine* 49, 5 (2011), 206–213.

[21] Oana Goga and Renata Teixeira. 2012. Speed Measurements of Residential Internet Access. In *Passive and Active Measurement*, Nina Taft and Fabio Ricciato (Eds.). Springer, Berlin, Heidelberg, 168–178. https://doi.org/10.1007/978-3-642-28537-0_17

[22] Sarthak Grover et al. 2013. Peeking behind the NAT: An Empirical Study of Home Networks. In *Proceedings of the 2013 Conference on Internet Measurement Conference (IMC '13)*. ACM, New York, 377–390. https://doi.org/10.1145/2504730.2504736

[23] Shaddi Hasan et al. 2019. Scaling Community Cellular Networks with CommunityCellularManager. In *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19) (NSDI'19)*. USENIX Association, Boston, MA, USA, 735–750. https://doi.org/10.5555/3323234.3323294

[24] Shaddi Hasan, Yahel Ben-David, Max Bittman, and Barath Raghavan. 2015. The Challenges of Scaling WISPs. In *Proceedings of the 2015 Annual Symposium on Computing for Development (DEV '15)*. ACM, New York, 3–11. https://doi.org/10.1145/2830629.2830637

[25] Kurtis Heimerl et al. 2013. Local, Sustainable, Small-Scale Cellular Networks. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers - Volume 1 (ICTD '13)*. ACM, New York, 2–12. https://doi.org/10.1145/2516604.2516616

[26] Kurtis Heimerl et al. 2015. Analysis of Smartphone Adoption and Usage in a Rural Community Cellular Network. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development (ICTD '15)*. ACM, New York, 1–4. https://doi.org/10.1145/2737856.2737880

[27] ITU Publications. 2020. *Measuring Digital Development: Facts and Figures 2020*. Technical Report. International Telecommunication Union, Geneva, Switzerland. 18 pages.

[28] Esther Jang et al. 2018. Crowdsourcing Rural Network Maintenance and Repair via Network Messaging. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, Montreal QC, Canada, 1–12. https://doi.org/10.1145/3173574.3173641

[29] D. L. Johnson, Elizabeth M. Belding, Kevin Almeroth, and Gertjan van Stam. 2010. Internet Usage and Performance Analysis of a Rural Wireless Network in Macha, Zambia. In *Proceedings of the 4th ACM Workshop on Networked Systems for Developing Regions (NSDR '10)*. ACM, New York, 1–6. https://doi.org/10.1145/1836001.1836008

[30] Matthew Johnson, Spencer Sevilla, Esther Jang, and Kurtis Heimerl. 2018. dLTE: Building a More WiFi-like Cellular Network: (Instead of the Other Way Around). In *Proceedings of the 17th ACM Workshop on Hot Topics in Networks (HotNets '18)*. ACM, New York, 8–14. https://doi.org/10.1145/3286062.3286064

[31] Kanchana Kanchanasut et al. 2007. DUMBONET: a multimedia communication system for collaborative emergency response operations in disaster-affected areas. *International Journal of Emergency Management* 4, 4 (2007), 670–681.

[32] Nabin Kumar Karn, Zhang Hongli, and Muhammad Shafiq. 2017. Measuring broadband internet performance in Nepal: A comparative study. *Procedia Computer Science* 107 (2017), 64–69.

[33] Tobias Klenze et al. 2018. Networking in Heaven as on Earth. In *Proceedings of the 17th ACM Workshop on Hot Topics in Networks (HotNets '18)*. ACM, New York, 22–28. https://doi.org/10.1145/3286062.3286066

[34] Deepti Kumar, David Martin, and Jacki O'Neill. 2011. The Times They Are A-Changin': Mobile Payments in India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, 1413–1422. https://doi.org/10.1145/1978942.1979150

[35] David Lawrence and Sean Turner. 2021. Web Packaging (Wpack) -. https://datatracker.ietf.org/wg/wpack/about/.

[36] Adisorn Lertsinsrubtavee et al. 2015. Understanding Internet Usage and Network Locality in a Rural Community Wireless Mesh Network. In *Proceedings of the Asian Internet Engineering Conference (AINTEC '15)*. ACM, Bangkok, Thailand, 17–24. https://doi.org/10.1145/2837030.2837033

[37] Ioana Livadariu, Ahmed Elmokashfi, and Amogh Dhamdhere. 2017. Measuring IPv6 Adoption in Africa. In *E-Infrastructure and e-Services for Developing Countries*. Springer International Publishing, Dakar, Senegal, 345–351. https://doi.org/10.1007/978-3-319-98827-6_32

[38] Andres Martinez-Fernandez et al. 2016. The TUCAN3G project: wireless technologies for isolated rural communities in developing countries based on 3G small cell deployments. *IEEE communications magazine* 54, 7 (2016), 36–43.

[39] Gregers Petersen. 2014. Freifunk: When Technology and Politics Assemble into Subversion. In *Subversion, Conversion, Development: Cross-Cultural Knowledge Exchange and the Politics of Design*. The MIT Press, Boston, MA, USA, 39–56.

[40] Ali Raza et al. 2017. xCache: Rethinking Edge Caching for Developing Regions. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development (ICTD '17)*. ACM, New York, 1–11. https://doi.org/10.1145/3136560.3136577

[41] Carlos Rey-Moreno, Amalia Sabiescu, and Masbulele Siya. 2014. Towards Self-Sustaining Community Networks in Rural Areas of Developing Countries: Understanding Local Ownership. In *ICTs for Inclusive Communities in Developing Societies*. International Development Informatics Association, Port Elizabeth, South Africa, 63–77.

[42] C. Rey-Moreno, W. D. Tucker, D. Cull, and R. Blom. 2015. Making a Community Network Legal Within the South African Regulatory Framework. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development (ICTD '15)*. ACM, New York, 57:1–57:4. https://doi.org/10.1145/2737856.2737867

[43] Brandon Schlinker et al. 2019. Internet Performance from Facebook's Edge. In *Proceedings of the Internet Measurement Conference* (Amsterdam, Netherlands) *(IMC '19)*. ACM, New York, 179–194. https://doi.org/10.1145/3355369.3355567

[44] Paul Schmitt et al. 2016. Community-Level Access Divides: A Refugee Camp Case Study. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development (ICTD '16)*. ACM, New York, 1–11. https://doi.org/10.1145/2909609.2909668

[45] Mennan Selimi et al. 2015. Cloud services in the guifi. net community network. *Computer Networks* 93 (2015), 373–388.

[46] Spencer Sevilla et al. 2019. Experiences: Design, Implementation, and Deployment of CoLTE, a Community LTE Solution. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19)*. ACM, Los Cabos, Mexico, 1–16. https://doi.org/10.1145/3300061.3345446

[47] Kushal Shah et al. 2017. An Investigation of Phone Upgrades in Remote Community Cellular Networks. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development (ICTD '17)*. ACM, New York, 1–12. https://doi.org/10.1145/3136560.3136569

[48] Asheesh Sharma et al. 2015. Revisiting the State of Cellular Data Connectivity in India. In *Proceedings of the 2015 Annual Symposium on Computing for Development (DEV '15)*. ACM, New York, 149–157. https://doi.org/10.1145/2830629.2830649

[49] Savannah Wei Shi, Mu Xia, and Yun Huang. 2015. From minnows to whales: An empirical study of purchase behavior in freemium social games. *International Journal of Electronic Commerce* 20, 2 (2015), 177–207.

[50] Javier Simo-Reigadas et al. 2015. Sharing Low-Cost Wireless Infrastructures with Telecommunications Operators to Bring 3G Services to Rural Communities. *Computer Networks* 93 (Dec. 2015), 245–259. https://doi.org/10.1016/j.comnet.2015.09.006

[51] Samuel Sudar, Matt Welsh, and Richard Anderson. 2018. Siskin: Leveraging the Browser to Share Web Content in Disconnected Environments. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18)*. ACM, New York, 1–7. https://doi.org/10.1145/3209811.3209820

[52] Srikanth Sundaresan et al. 2011. Broadband internet performance: a view from the gateway. *ACM SIGCOMM CCR* 41, 4 (2011), 134–145.

[53] D. Vega, L. Cerdà-Alabern, L. Navarro, and R. Meseguer. 2012. Topology Patterns of a Community Network: Guifi.Net. In *2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, Barcelona, Spain, 612–619. https://doi.org/10.1109/WiMOB.2012.6379139

[54] Morgan Vigil, Matthew Rantanen, and Elizabeth Belding. 2015. A First Look at Tribal Web Traffic. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1155–1165.

[55] Christian Vogt, Max Jonas Werner, and Thomas C. Schmidt. 2013. Leveraging WebRTC for P2P Content Distribution in Web Browsers. In *2013 21st IEEE International Conference on Network Protocols (ICNP)*. IEEE, Goettingen, Germany, 1–2. https://doi.org/10.1109/ICNP.2013.6733637

[56] Sarah Yu and Samia Ibtasam. 2018. A Qualitative Exploration of Mobile Money in Ghana. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18)*. ACM, New York, 1–10. https://doi.org/10.1145/3209811.3209863

[57] Yasir Zaki et al. 2014. Dissecting Web Latency in Ghana. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, New York, 241–248. https://doi.org/10.1145/2663716.2663748