

University of California
Santa Barbara

Network Measurement and Systems for Resource-Constrained Environments

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Computer Science

by

Paul Richard Schmitt

Committee in charge:

Professor Elizabeth Belding, Chair
Professor Ben Zhao
Professor Heather Zheng

September 2017

ProQuest Number: 10607215

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10607215

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

The Dissertation of Paul Richard Schmitt is approved.

Professor Ben Zhao

Professor Heather Zheng

Professor Elizabeth Belding, Committee Chair

July 2017

Network Measurement and Systems for Resource-Constrained Environments

Copyright © 2017

by

Paul Richard Schmitt

Acknowledgements

The material in this dissertation is based on joint work with Elizabeth Belding, Daniel Iland, Morgan Vigil, Mariya Zheleva, Ramya Raghavendra, Thomas Pötsch, Jay Chen, Barath Raghavan, Brian Tomaszewski, Ying Xu, Carleen Maitland, Petko Bogdanov, and Timothy Larock. Work was funded through NSF Network Science and Engineering (NetSE) Award CNS-1064821 and NSF Catalyzing New International Collaborations (CNIC) Award IIA-1427873. This dissertation was also supported by the UC President's Dissertation Year Fellowship and the UCSB Dean's Fellowship.

This PhD would not have been possible without the support I had from many people. First, I would like to thank my life partner Jess. Thank you for your boundless enthusiasm and for standing by my side throughout this adventure. You inspire me to keep an open mind and take risks that have universally turned out to be the best decisions I've made.

Thank you to my research advisor, Elizabeth Belding. Your advice and wisdom have been invaluable. Your willingness to listen to and critique my half-baked research ideas have provided me with an excellent example of how I hope to be as a research mentor in the future. Thank you for the opportunities to explore networks in far-flung locales that I could not have imagined studying before being a part of the Moment lab.

Thank you to my research committee members Ben Zhao and Heather Zheng for valuable feedback and interesting classes.

Thank you to my collaborators: Danny Iland, Morgan Vigil-Hayes, Mariya Zheleva, Jay Chen, Barath Raghavan, Ramya Raghavendra, and Carleen Maitland. Thank you particularly to Danny, Morgan, and Mariya for being great travel and research companions as we worked together all over the world in environments where things never go as planned.

Lastly, thank you to my family and friends. My parents have given endless support

and have been exemplary role models for how to be a decent human. Thank you for instilling in me my curiosity and dedication to teaching. Thank you to my sisters for helping me find my voice by creating an environment where one has to be confident to get a word in edgewise.

Curriculum Vitæ

Paul Richard Schmitt

Education

2017	<i>Doctor of Philosophy (Ph.D.)</i> Computer Science University of California, Santa Barbara Santa Barbara, CA
2011	<i>Master of Science (M.S.)</i> Software Engineering University of St. Thomas Saint Paul, MN
2009	<i>Bachelor of Arts (B.A.)</i> Computer Science University of St. Thomas Saint Paul, MN

Experience

2012 – 2017	<i>Research/Teaching Assistant</i> Dept. of Computer Science University of California, Santa Barbara Santa Barbara, CA.
2016	<i>Research Intern</i> International Computer Science Institute (ICSI) Berkeley, CA
2013	<i>Research Intern</i> IBM Research Yorktown Heights, NY
2006 – 2012	<i>Network Administrator</i> University of St. Thomas, Saint Paul, MN
2002 – 2006	<i>Network Administrator</i> Blue Earth County Mankato, MN

Awards

2016	<i>2016-17 UCSB President's Dissertation Year Fellowship</i>
2014	<i>2014-15 UCSB Dean's Fellowship</i>

Publications

- 2017 **P. Schmitt** and E. Belding, *Low On Air: Inherent Wireless Channel Capacity Limitations* Paul Schmitt, Elizabeth Belding, in the Third Workshop on Computing within Limits (**LIMITS 2017**), June 2017, Santa Barbara, CA
- 2016 T. Pötsch, **P. Schmitt**, J. Chen, and B. Raghavan, *Helping the Lone Operator in the Vast Frontier*, in the ACM Workshop on Hot Topics in Networks (**HotNets 2016**), November 2016, Atlanta, GA
- 2016 **P. Schmitt**, D. Iland, E. Belding, and M. Zheleva, *PhoneHome: Robust Extension of Cellular Coverage* in the International Conference on Computer Communication and Networks (**ICCCN 2016**), August 2016, Waikoloa, HI
- 2016 **P. Schmitt**, D. Iland, and E. Belding, “*SmartCell: Small-Scale Mobile Congestion Awareness*” in **IEEE Communications Magazine**, July 2016
- 2016 **P. Schmitt** and E. Belding, “*Navigating Connectivity in Reduced Infrastructure Environments*” in the second workshop on Computing within Limits (**LIMITS 2016**), June 2016, Irvine, CA
- 2016 **P. Schmitt**, D. Iland, E. Belding, B. Tomaszewski, Y. Xu, and C. Maitland, “*Community-Level Access Divides: A Refugee Camp Case Study*” in The International Conference on Information and Communication Technologies and Development (**ICTD**), June 2016, Ann Arbor, MI
- 2016 **P. Schmitt**, D. Iland, M. Zheleva, and E. Belding, “*HybridCell: Cellular connectivity on the fringes with demand-driven local cells*” **IEEE INFOCOM 2016**, April 2016, San Francisco, CA
- 2016 **P. Schmitt**, M. Vigil, and E. Belding, “*A Study of MVNO Data Paths and Performance*” in Passive and Active Measurements Conference (**PAM 2016**), March 2016, Heraklion, Crete, Greece
- 2015 **P. Schmitt**, R. Raghavendra, and E. Belding, “*Internet Media Upload Caching for Poorly-Connected Regions*” in The sixth annual Symposium on Computing for Development (**ACM DEV 2015**), December 2015, London, United Kingdom
- 2015 C. Maitland, B. Tomaszewski, E. Belding, K. Fisher, Y. Xu, D. Iland, **P. Schmitt**, and A. Majid, “*Youth Mobile Phone and Internet Use January 2015 Za’atari Camp, Mafraq, Jordan*” Penn State Univ., State College, PA, Tech. Rep., October 19 2015
- 2015 M. Zheleva, **P. Schmitt**, M. Vigil, and E. Belding, “*Internet Bandwidth Upgrade: Implications on Performance and Usage in Rural*

- Zambia*”, in Information Technologies and International Development Journal (**ITID**), v. 11, Issue 2, Summer 2015
- 2014 M. Vigil and **P. Schmitt**, “*Conferences: HotMobile 2014*”, in **IEEE Pervasive Computing**, v. 13, p. 84. July-September 2014.
- 2013 M. Zheleva, **P. Schmitt**, M. Vigil, E. Belding, “The Increased Bandwidth Fallacy: Performance and Usage in Rural Zambia”, in *The Fourth Annual Symposium on Computing for Development (ACM DEV), Cape Town, South Africa, December 2013*
- 2013 M. Zheleva, **P. Schmitt**, M. Vigil, E. Belding, “Bringing Visibility to Rural Users in Cote DIvoire”, in *The International Conference on Information and Communication Technologies and Development (ICTD), Cape Town, South Africa, December 2013*
- 2013 M. Zheleva, **P. Schmitt**, M. Vigil, E. Belding, “Community Detection in Cellular Network Traces”, in *The International Conference on Information and Communication Technologies and Development (ICTD), Cape Town, South Africa, December 2013*.
- 2013 **P. Schmitt**, M. Vigil, M. Zheleva, E. Belding, “Communication Flow Patterns in the Orange Telecom D4D Dataset”, in *Third conference on the Analysis of Mobile Phone Datasets (NetMob), Boston, MA, May 2013*

Abstract

Network Measurement and Systems for Resource-Constrained Environments

by

Paul Richard Schmitt

As mobile and wireless networks become the dominant user access technology in well-connected regions, they face an explosion in demand for high-bandwidth Internet and cellular connectivity. Additionally, wireless, both cellular and broadband, has become the de facto access technology at the Internet's frontier in developing contexts and other resource-limited environments due to underlying economic factors. In order for wireless and mobile networks to maintain pace with demand, and achieve global connectivity, we must empirically study existing networks and use those findings to inform new system designs across the telecommunications hierarchy.

Before we can build and deploy networks at the Internet's frontier to bring the next billion users online, we must understand the unique challenges currently faced by networks in resource-limited environments. In this thesis, we measure and characterize real-world operational networks, both cellular and wireless broadband, so that we can later use insights gleaned from this analysis to inform system designs aimed at ameliorating the performance inadequacies we witness. Our work includes a one-of-a-kind glimpse into cellular infrastructure performance in an operational refugee camp from the user's perspective. Our analysis shows that the networks we study often provide users with diminished performance due to factors such as oversubscription and relatively small backhaul links.

Whereas network systems are often designed for highly-connected, resource-rich contexts, this thesis strives to consider holistically (i.e. across multiple tiers of the telecomm-

munications hierarchy) the challenges to providing robust connectivity in limited settings. Based on our comprehensive analysis of real-world conditions, we produce cellular and broadband systems aimed at bridging the technological gap that exists for people in challenging network environments. While the focus of our work is in specific environments, developing world contexts share many similarities, enabling the generalization of the systems we create to other resource-limited scenarios.

Contents

Curriculum Vitae	vi
Abstract	ix
List of Figures	xiv
List of Tables	xix
1 Introduction	1
1.1 Thesis Statement and Contributions	4
1.2 Dissertation Outline	9
2 Research Background	10
2.1 Field Work	11
2.2 Resource-Limited Network Challenges	16
2.3 Discussion and Conclusion	18
Part I Characterization and Analysis of Resource-Limited Networks	19
3 Analysis: Cellular Performance Inference from Control Channel Messages	22
3.1 Introduction	22
3.2 Methods	24
3.3 Cellular Measurement Analysis	26
3.4 Discussion	37
3.5 Related Work	38
4 Analysis: MVNO Cellular Network Performance	40
4.1 Introduction	40
4.2 Background	42

4.3	Data Collection	43
4.4	Network Analysis	45
4.5	Related Work	53
4.6	Discussion and Conclusion	54
5	Analysis: Cellular Usage in Côte d'Ivoire	57
5.1	Introduction	57
5.2	Methodology	60
5.3	Network Analysis	65
5.4	Related work	79
5.5	Discussion and conclusion	81
6	Analysis: Internet Performance and Usage in Rural Zambia	85
6.1	Introduction	85
6.2	Network Analysis	87
6.3	Related Work	112
6.4	Discussion	113
6.5	Conclusion	114

Part II Network Systems Design for Resource-Limited Environments **116**

7	System Design: Cellular Systems for Resource-Limited Environments	119
7.1	Introduction	119
7.2	HybridCell	127
7.3	SmartCell	156
7.4	PhoneHome	162
7.5	Related Work	174
7.6	Conclusion	176
8	System Design: Automated WISP Network Topology Planning	177
8.1	Introduction	177
8.2	Planning Frontier Networks	179
8.3	Toward Automated Planning	186
8.4	Prototype Implementation	188
8.5	Discussion	190
8.6	Related Work	191
9	System Design: Unsupervised Transmitter Detection for Spectrum Sensing	193
9.1	Introduction	193
9.2	Related Work	197

9.3	Methodology	199
9.4	Evaluation	213
9.5	Wideband characterization with AirVIEW	223
9.6	Discussion and conclusion	227
10	System Design: Internet Media Upload Caching for Poorly-Connected Regions	230
10.1	Introduction	230
10.2	Traffic analysis	233
10.3	VillageCache design	243
10.4	Evaluation	250
10.5	Related work	253
10.6	Discussion and conclusion	255
11	Conclusion and Summary	257
11.1	Conclusion	257
11.2	Summary	263
Bibliography		264

List of Figures

1.1	Dissertation overview.	5
2.1	Za'atari camp map.	12
2.2	A map of Southern Africa highlighting the location of Macha.	15
3.1	GSM network immediate assignment procedure. We capture messages broadcast from the BTS to MS over the GSM air interface.	25
3.2	Percent of immediate assignment messages that were rejections in five minute bins on January 6 th 2015 in Za'atari.	29
3.3	Observed backoff wait values.	30
3.4	Calculated distances between phones and base stations.	31
3.5	Zain RSSI.	32
3.6	Orange RSSI.	32
3.7	CRO artificially increases the calculated signal strength when the phone determines which cell to use.	34
3.8	Data coverage. Color indicates connectivity: 3G HSPA (Green), 2G EDGE (Blue), No connectivity (Black).	36
4.1	RTT measurements for mobile carriers to non-location specific servers. .	46
4.2	RTT comparison for specific CDN locations identified by airport codes: Atlanta (ATL), Chicago (ORD), Dallas (DFW), Los Angeles (LAX), Miami (MIA), New York (LGA), San Jose (SJC), Seattle (SEA), Washington DC (IAD). Each carrier's top three preferred locations are indicated. .	48
4.3	Path metrics for mobile networks. MNOs and MVNOs exhibit similar AS path lengths but differ in the number of hops taken to reach the destination.	50
4.4	Paths taken for each carrier to reach servers located in Los Angeles (LAX) from clients in Santa Barbara (SBA). Arc weight indicates the number of times a particular hop was taken. Clockwise arcs indicate direction of traffic between adjacent hops.	52
5.1	The effect of removing the ego (depicted with a square) from the egocentric social graph.	63

5.2	Building a persistence graph.	64
5.3	Cellular Antennas in Côte d'Ivoire.	66
5.4	Antenna activity.	67
5.5	Mobile network activity over time.	68
5.6	Weekly patterns.	68
5.7	Communication patterns as a function of the distance between antennas.	71
5.8	Classification of communication between antenna pairs.	72
5.9	Mean call duration and mean call distance for connections of different types.	72
5.10	Weighted connectivity map with detected communities.	75
5.11	(a) The number of connected components (CCs) per ego and (b) the standard deviation of the number of connected components per ego over the observed period.	76
5.12	(a) The in- and out-degree of nodes in all persistence graphs and (b) the average Jaccard similarity for each persistence graph.	77
5.13	Number of occurrences of the first, second and tenth most frequent neighbor.	79
6.1	Network architecture and traffic monitoring.	88
6.2	Traffic load over time.	90
6.3	RTT.	91
6.4	Payload size.	91
6.5	Bytes by day.	91
6.6	Bytes in flight.	92
6.7	TCP flow success and failure in uplink and downlink direction.	94
6.8	Comparison of TCP performance in Windows and Linux.	97
6.9	Tstat analysis of service types.	98
6.10	Popular URI Requests.	100
6.11	TCP flow success and failure for URIs of interest.	101
6.12	Flow distance from Macha; CDF.	103
6.13	Flow distance from center mass; CDF.	105
6.14	Center-mass points with radii of gyration.	106
7.1	HybridCell consists of a Network Analyzer, a Local Cell and our network switching application CellSwitch. (1) The Network Analyzer gathers performance data from the commercial cell; (2) the Network Analyzer algorithm determines commercial cell congestion and failure; (3) the Local Cell is reconfigured based on the Analyzer's measurements; and (4) CellSwitch shifts users between Local and Commercial cells for seamless connectivity.	121
7.2	MTN Syria coverage map with camp locations. [131]	126
7.3	Cellular coverage map two weeks after Typhoon Haiyan [72].	126
7.4	Guatemala population density. Dark areas indicate higher density.	126
7.5	Movistar coverage in Guatemala.	126
7.6	Peaks with fixed periodicity are observed in the cyclic autocorrelation function, corresponding to timeslots used by cellular synchronization channels.	129

7.7	Comparison of local and commercial cellular network architectures. Key advantages of local cells include their highly-distributed nature, and use of free software, open hardware and generic backbone for interconnects.	134
7.8	Example of cellular-network-in-a-box extended with modems for active testing and additional storage, used for the authors deployment in [206].	136
7.9	Frame structure of CellSwitch.	137
7.10	Energy consumption increases during transitions, but peaks immediately following movement between cellular networks.	139
7.11	Time spent without service when moving to the local network from commercial networks.	140
7.12	Time spent without service when moving to commercial networks from a local cell.	140
7.13	As the duty cycle period P is increased, the energy burden and the time a phone spends disconnected from both networks is decreased.	142
7.14	Immediate assignment rejection percentage per minute, and successful and failed immediate assignment messages collected from Zain over 1 day in Za'atari.	145
7.15	Estimated availability for Zain with different α values.	146
7.16	Estimated availability for all carriers ($\alpha = 0.8$).	147
7.17	Simulation results with statically set estimated availability.	149
7.18	L-C Latency for callers assigned to different carriers.	152
7.19	L-C Latency with caller-callee designations.	153
7.20	C-L Latency with caller-callee designations.	154
7.21	BTS Availability (BTSA) for carriers in Za'atari on the morning of January 6 th 2015.	160
7.22	PhoneHome system diagram. Users connect their mobile devices to the local cellular network. PhoneHome uses a long-distance wireless link to a remote relay node in order to bridge local cellular services onto commercial carriers near the remote node.	163
7.23	PhoneHome remote node equipment. The remote node includes a PC, GSM devices to bridge traffic onto a nearby commercial carrier, and a long-distance Wi-Fi antenna to connect to the local node within the camp.	166
7.24	Long-distance Wi-Fi link locations used for experiments.	169
7.25	Call setup latency components.	170
7.26	Latency comparison for call setup.	171
7.27	Long-distance link QoS measurements.	172
8.1	Map of 50 km \times 20 km frontier network region and locations of a subset of (potential) customers.	178
8.2	Finding the best relay tower locations (blue markers) and links to inter-connect 8 sites (red markers).	190

9.1	(a) Transmitter characterization metrics; (b) Illustration of a transmitter comprised of multiple transmissions.	199
9.2	AirVIEW workflow.	201
9.3	Example of wavelet decomposition with illustration of key features employed by AirVIEW.	202
9.4	(a) TPR and (b) FPR of two instantiations of AirVIEW ($s = 4/\beta = 2$ and $s = 4/\beta = 4$) and the competing techniques Naive and Denoised at varying signal levels with noise floor fixed at -110dBm. (c) an example single-time detection of a transmitter at -85dBm.	215
9.5	(a),(b): Average accuracy of detection in synthetic data using various parameter setting combinations of s and β . Over 1000 runs, 2 transmitters with mean power of -88 were randomly placed in the data and detection was applied using 4 different scales. (c),(d): Similar evaluation in real-world sensed TV band spectrum for constantly transmitting stations of known bands.	216
9.6	Accuracy and alignment at mean transmit power of -103dBm for various (β, s) combinations. We vary scale from 8-9 in (a) to 5-6 in (d). On each graph we vary β from -1 to 4 in increments of 0.2. Across all (β, s) combinations the accuracy and alignment follow a similar trend, which indicates that alignment is a good proxy for accuracy.	219
9.7	Maximum and achieved accuracy with decreasing transmitter PSD. Alignment is a good proxy for accuracy across all PSD regimes. Achieved accuracy is high, persistent and close to the max accuracy.	220
9.8	β at max accuracy and β at max alignment for increasing scales. As s grows, max accuracy and max alignment correspond to decreasing β s. This underlines the importance of adaptive parameter estimation.	220
9.9	Stability of parameter estimation as a function of the number of spectrum sweeps used for training. AirVIEW learns the optimal parameters within 40 sweeps. Stability improves with more training.	220
9.10	Examples of transmission detection with varying smoothing parameters: (a) $\lambda = \infty$, (b) $\lambda = 0.5$ and (c) $\lambda = 0.0$	223
9.11	Occupied bands	225
9.12	Number of Transmitters	225
9.13	Temporal transmitter characteristics of a wideband spectrum scan. The average (black) and standard deviation (green envelope) of active time, cycle and gap illustrate that some intermittently-occupied bands have predictable incumbent patterns while in others the incumbent activity is very unpredictable.	226
10.1	Observed RTT in Macha.	235
10.2	User-interrupted HTTP POSTs in Macha.	236
10.3	Local vs non-local Facebook image interactions in Macha.	236

10.4 Local vs non-local Instagram media interactions in TDV network.	236
10.5 Lifetime for Instagram media in TDV network.	238
10.6 Macha DNS latency comparison.	238
10.7 Macha YouTube QoE.	238
10.8 TDV network Instagram uploads.	243
10.9 System architecture.	244
10.10 Average file download times.	251
10.11 Cached YouTube QoE analysis.	251
10.12 Cached video bandwidth savings.	251

List of Tables

2.1	Data bundles available on Jordanian carriers.	13
3.1	Overall mobile network statistics.	27
3.2	Android signal strength RSSI value to level map.	35
5.1	Antenna Density Classifications	66
6.1	General TCP statistics averaged over each time period.	87
6.2	TCP flow analysis.	93
6.3	TCP flow uplink (UL) and downlink (DL) characteristics.	94
6.4	HTTP Response Codes	102
6.5	Measured radius of gyration.	104
6.6	TCP flow uplink (UL) and downlink (DL) characteristics.	109
7.1	System messages used by HybridCell.	132
7.2	Trace-driven simulation call totals.	151
10.1	Macha network traffic characteristics.	239

Chapter 1

Introduction

Modern networks face an unrelenting demand for high-bandwidth Internet and cellular connectivity. The challenges of providing such connectivity are exacerbated in resource-limited environments, where infrastructure improvement or augmentation is often economically infeasible. Without access to adequate connectivity, the information and technology digital divide between the ‘haves’ and ‘have-nots’ widens. In order to achieve necessary performance and reliability, a concerted breadth of solutions must address challenges present across the telecommunications hierarchy in order to maintain pace with demand and provide global connectivity.

In recent years, great strides have been made to bring connectivity to areas where it previously was not. For example, in 2016 the number of cellular subscriptions in the world is roughly equal to the world population¹. While this statistic appears to provide cause for celebration, upon deeper inspection we see that the number of subscriptions is artificially bolstered by people carrying multiple SIM cards in order to obtain service on multiple service providers depending on time and place. Taking this into account, the number of unique cellular *subscribers* is roughly 5 billion. Overall, we observe that

¹<https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf>

networks serving rural areas and developing regions continue to significantly lag behind the rest of the world. Further, networks that do offer coverage in such resource-limited areas are often poor-quality and offer spotty, unpredictable coverage.

In this thesis we focus on *characterization* of real-world networks and subsequent *system designs* based on our analysis to improve connectivity in the measured contexts. Commercial cellular network infrastructure is built using expensive, highly-specialized equipment and relies on robust backhaul links in order to connect to the global telephony network. These factors make it challenging to deploy cellular infrastructure to previously unconnected areas when demand changes rapidly (i.e. formation of a refugee camp or unplanned settlement) and economically infeasible in rural areas with sparse user populations. In this work, we characterize the “health” of commercial cellular networks by listening to plaintext cellular control channel messages in multiple locations representing differing levels of economic development. Key findings include: chronic congestion in some of the networks, particularly in the refugee camp context, and varying levels of congestion and failure between multiple carriers serving the same areas. Given our observations we design a local cellular network dubbed HybridCell [165], which dynamically shifts users, based on the health of the commercial network, to the local cellular network in order to provide for local communications during times of commercial congestion and failure. HybridCell includes mechanisms for characterizing nearby commercial cellular performance as well as techniques for dynamic spectrum coexistence. Additionally, we design SmartCell [162], an Android application meant to provide users with insight into congestion on the cellular infrastructure they use, and PhoneHome [164], a system that extends existing cellular coverage to areas with no or damaged cellular infrastructure, or infrastructure that is otherwise performing poorly.

Mobile Virtual Network Operators (MVNOs) are cellular providers that do not own all of the infrastructure on which they operate; rather, they rent the use of infrastructure

from traditional carriers, known as Mobile Network Operators (MNOs). MVNOs have increased in popularity in recent years, particularly with economically disadvantaged user populations as they often have relaxed contract terms and lower cost rates. We examine data traffic performance on MVNO carriers and compare it with traffic on the underlying MNO carriers to uncover differences in performance and routing behaviors. Key findings include: a “light” MVNO carrier essentially mimics the underlying MNO, whereas “full” MVNO carriers differ greatly both from the MNO carriers as well as between other full MVNOs that are sold under the same brand.

In addition to cellular networks, we study broadband wireless connectivity in resource-limited environments and design systems based on our findings. We contribute a new dataset captured at the Internet gateway of a rural network in Macha, Zambia and analysis of the trace before and after an upgrade of the gateway link from a 256kbps satellite link to a 2Mbps terrestrial link. We find that the performance gains and user experience are not proportional to the bandwidth increase [210, 209]. The increase in network capability led to more ambitious use of the still-limited connectivity. As a result user traffic showed a marked *decrease* in performance. We find high locality of interest in social media content in the Macha network, leading us to design VillageCache [166], a system which allows for appropriate local transformation and redistribution of media *uploaded* through an Internet cache. Importantly, VillageCache is able to redistribute locally-generated content without requiring a single download of that content from the cloud.

Wireless Internet Service Providers (WISPs) represent a model for delivering connectivity to rural areas for relatively little cost. Unfortunately, WISP networks are challenging to design and are often created and maintained by individuals with little networking expertise, leading to brittle topologies and poor performance. We design a prototype system for automated planning of WISP network topologies using topographic

data [145].

Wireless spectrum is a scarce, finite resource that is a fundamental requirement for wireless communications. Recent advances in agile radio technologies have enabled the promise of dynamic spectrum access models, where unoccupied spectrum can be shared by primary and secondary users. In order to realize the promise of DSA, we must first have knowledge of the spectrum occupancy in a given location. With this goal in mind we design AirVIEW, a one-pass transmitter detection algorithm that is resistant to noise.

1.1 Thesis Statement and Contributions

This dissertation shows that:

The empirical study of existing resource-constrained wireless and mobile networks and the resulting findings are critical for the design of systems that can effectively meet modern connectivity needs.

In line with this statement, we measure and analyze multiple real-world, resource-limited networks and design new systems intended to alleviate the challenges we witness.

The overview of this dissertation is depicted in Figure 1.1. We measure and characterize cellular and wireless broadband networks in order to understand network capabilities and performance from the user perspective. We leverage our findings to inform system designs aimed at meeting actual needs.

Characterization and analysis. We perform measurement and analysis on two wireless access technologies: (i) cellular and (ii) broadband.

- **Cellular.** We collect cellular network measurements in order to characterize performance and identify under-performing aspects of the infrastructure under study.

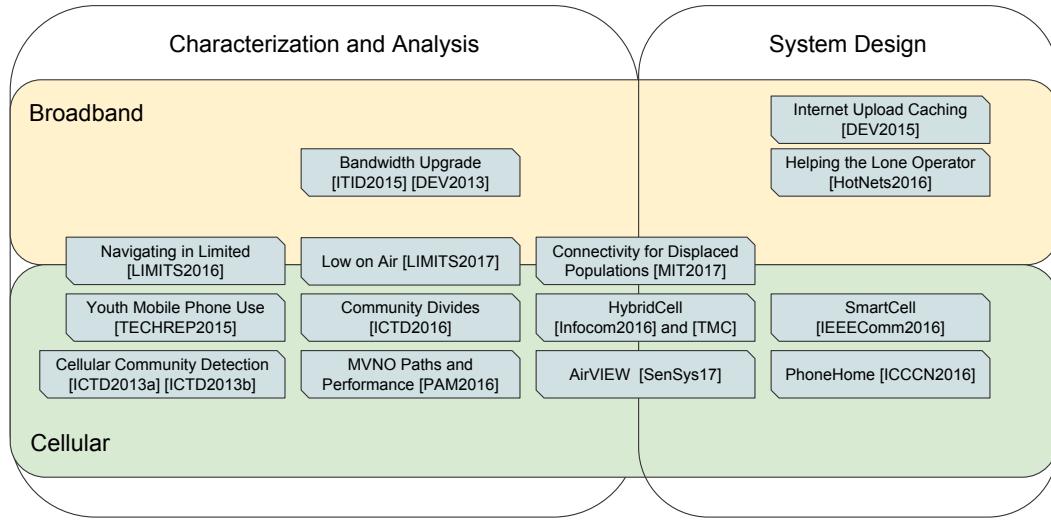


Figure 1.1: Dissertation overview.

Contributions: To understand cellular infrastructure performance and user needs in resource-limited environments, we utilize three different datasets. Our first dataset was collected during field visits to the Za’atari refugee camp in Jordan and the rural city of San Cristóbal Verapaz in Guatemala in early 2015. We collected GSM control messages that are sent over the broadcast channel in order to gain insight into the operating state of the cellular infrastructure in each setting. We study the performance of the networks for all of the service providers in each location. We find that, in some cases, different carrier infrastructure in the same location can exhibit vastly different performance characteristics.

Our second dataset focuses on the performance and routing behaviors for data traffic on mobile virtual network operator (MVNO) networks as well as the base carrier networks over which they operate. We examine differences in network performance and behavior. Ultimately, we find that traffic on MVNO networks takes more circuitous, less efficient paths to reach content servers compared to base operators.

We also analyze a large-scale trace of call data records provided by Orange from

their network in Côte D'Ivoire. By combining underlying population density datasets, we are able to differentiate between cellular base stations in urban and rural areas. Unsurprisingly, infrastructure is scant in rural areas, whereas it is relatively well-provisioned in urban centers. Importantly, we find high locality of interest in rural communication patterns (i.e. 70% of all rural-to-rural calls used a single cellular base station, meaning the users were located in the same geographic area).

- **Broadband.** We investigate traces collected in broadband networks to examine both user behaviors as well as topology characteristics.

Contributions: In order to understand user Internet behavior and network performance in a resource-limited environment, we study a network traffic trace collected at the Internet gateway in the rural village of Macha, Zambia. Our measurement window spans the time before and after a significant bandwidth increase for the backhaul link. We discover that the increased capacity was quickly consumed by users, ultimately leading to surprising *decrease* in performance as user began attempting to use more resource-heavy online services. Resultingly, user behavior changed as they were discouraged by the poor network performance. We also find high locality of interest in locally generated social media content. Network users were more likely to download media content that had been uploaded by other community members versus content from the broader Internet.

System design. We design system solutions tailored for the challenges identified in our measurement and analysis to improve connectivity in settings such as those we have studied.

- **Cellular.** Through our cellular measurement campaigns we find that infrastructure serving the Za'atari refugee camp and rural Guatemala offers poorer performance compared with infrastructure in an economically advantaged area. We

design systems aimed at improving user connectivity in such situations and for extending commercial network coverage to nearby areas.

Contributions: Given that carriers in the same location suffer from significantly different levels of control channel congestion and that the congestion differs over time, we design HybridCell, a system for dynamically offloading users to a local cellular network based on the health of the commercial network. We use software-defined radios to monitor the control channel of nearby commercial carrier infrastructure and devise a congestion estimation calculation based on the observable broadcast messages. Using the congestion estimation, we adjust the amount of time users spend on the commercial cell and the local cell provided by HybridCell with an aim for maximizing global connectivity while providing local-only connectivity during times of commercial network failure.

In some areas, deploying a local cellular network is not a viable solution (e.g. due to licensing challenges or lack of local expertise). We design SmartCell, an Android application, to inform users of the estimated congestion levels of the commercial cellular infrastructure they use. Given the information we are able to collect, users are provided a more complete picture of the infrastructure they are using and are empowered by information to seek better connectivity by physically changing locations or by switching SIM cards to a different provider.

Ubiquitous cellular coverage is often taken for granted, yet numerous people live outside, or at the fringes, of commercial cellular coverage. Further, natural disasters and human rights violations cause the displacement of millions of people annually worldwide, with many of these people relocating to shelters and camps in areas at or just beyond the margins of existing cellular infrastructure. We design Phone-Home, a system prototype that extends existing cellular coverage to areas with

no or damaged cellular infrastructure, or infrastructure that is otherwise poorly performing.

- **Broadband.** Our analysis of broadband traffic as well as our experience in deploying networks in resource-constrained environments enable us to identify unique challenges facing such networks and inform broadband-based system designs.

Contributions: WISP networks represent an increasingly common model for delivering connectivity in rural areas. However, the rural nature of WISPs means that they are often planned by lone operators with limited expertise and resources. The key challenge is the mismatch between the resources of these small operators and the complexity of their network planning. To aid in network planning, we build a prototype system with the goal of reducing the complexity for WISP operators and identify several open problems in the space.

The current paradigm of exclusive radio spectrum assignment and allocation creates an artificial spectrum scarcity that limits options for wireless connectivity. To this end, governments, industry and academia have endeavored to design mechanisms for spectrum sensing and management that allow shared access to radio spectrum. A key requirement for the realization of next-generation spectrum management is deep understanding of spectrum utilization in time, frequency and space. We propose AirVIEW, a one-pass, unsupervised spectrum characterization approach for rapid transmitter detection with high tolerance to noise. AirVIEW employs wavelet decomposition in order to amplify and reliably detect transmissions at a given time instant. It further reconciles transmissions in time using time-decay smoothing. As a result, AirVIEW identifies transmitters, even if their power level is very close to the noise floor, exhibiting better accuracy than simpler thresholding and wavelet-based alternatives.

We find in the traces from Macha, Zambia a high locality of interest for locally uploaded social media content. That is, village network users are much more likely to view and download content that other users of the network have uploaded, compared with content uploaded by users outside of the village network. Given the low gateway link capacity and generally poor performance, large media file uploads and downloads are prone to failure and we find that users often cancel large uploads. This leads to the inability for local residents to fully participate in the modern Internet. We propose VillageCache, a system which allows for local transformation and redistribution of media *uploaded* through an Internet cache. Importantly, VillageCache does not require even a single download of content from the cloud that has passed through the system.

1.2 Dissertation Outline

The remainder of this dissertation is organized as follows. In Chapter 2 we present the background of our research including a description of our field studies as well as challenges facing networks in resource-limited environments. In Chapters 3 - 5 we characterize cellular network performance and user behaviors in a variety of settings. Chapter 6 includes an in-depth look at a broadband network in a developing context. Chapter 7 discusses three cellular-based systems we have designed based on our prior observations. In Chapter 8 we present our design of a WISP network topology planning tool. Chapter 9 presents AirVIEW, our transmitter detection algorithm for spectrum occupancy information. Chapter 10 discusses VillageCache, our web media cache system that scrapes locally uploaded social media content and redistributes those media in subsequent local requests without the need for cloud retrieval. Lastly, Chapter 11 discusses our findings and concludes this dissertation.

Chapter 2

Research Background

It is difficult to underestimate the need for communications systems. Connectivity has become so fundamental that the United Nations stated in 2011 that Internet access is a basic human right [116]:

Given that the Internet has become an indispensable tool for realizing a range of human rights, combating inequality, and accelerating development and human progress, ensuring universal access to the Internet should be a priority for all States. Each State should thus develop a concrete and effective policy, in consultation with individuals from all sections of society, including the private sector and relevant Government ministries, to make the Internet widely available, accessible and affordable to all segments of population.

Mobile connectivity is often touted as the most plausible solution for providing connectivity in less developed regions due to its relatively cheaper capital investment cost compared to that of traditional wired infrastructure. Unfortunately, despite the lower investment cost, mobile infrastructure is still not economically viable in many areas due to low subscriber density and lack of purchasing power. Mobile providers simply lack the business motivation to provide connectivity to users in such environments. The stalemate

between business interests of providers and the goal of universal connectivity has come to, in many ways, define the digital divide.

In this dissertation, we study the usage and performance of various networking solutions in resource-limited environments. We leverage our trace captures to identify network usage and challenges related to user experience and design systems based on our findings. In the remainder of this chapter we give background of our field work and a brief description of the characteristics and challenges in working in resource-limited environments.

2.1 Field Work

We conduct field work in multiple locations around the world representing varying levels of economic development and network usage. We capture traces of cellular control channel messages, mobile data network path and performance probe traffic, and data packet captures for a rural broadband network.

Za’atari Refugee Camp, Jordan. Za’atari is one of four refugee camps in Jordan, and is the oldest and largest. Located in a desert region near the border of Syria and Jordan shown in Figure 2.1, it is one of the largest camps in the world. The camp was established when, as the influx of refugees increased, voluntary sheltering of refugees by Jordanian families could no longer absorb the flow. As a result, in July 2012, the United Nations High Commissioner for Refugees (UNHCR), together with the Jordanian Hashemite Charity Organization (JHCO), opened the camp. Within the first three months the camp was home to 30,000 residents who brought with them \sim 10,000 mobile phones.

Since then the camp population swelled to 120,000 residents, and is now stably hosting roughly 80,000 residents in roughly 6 square kilometers. As the camp grew, the Jord-

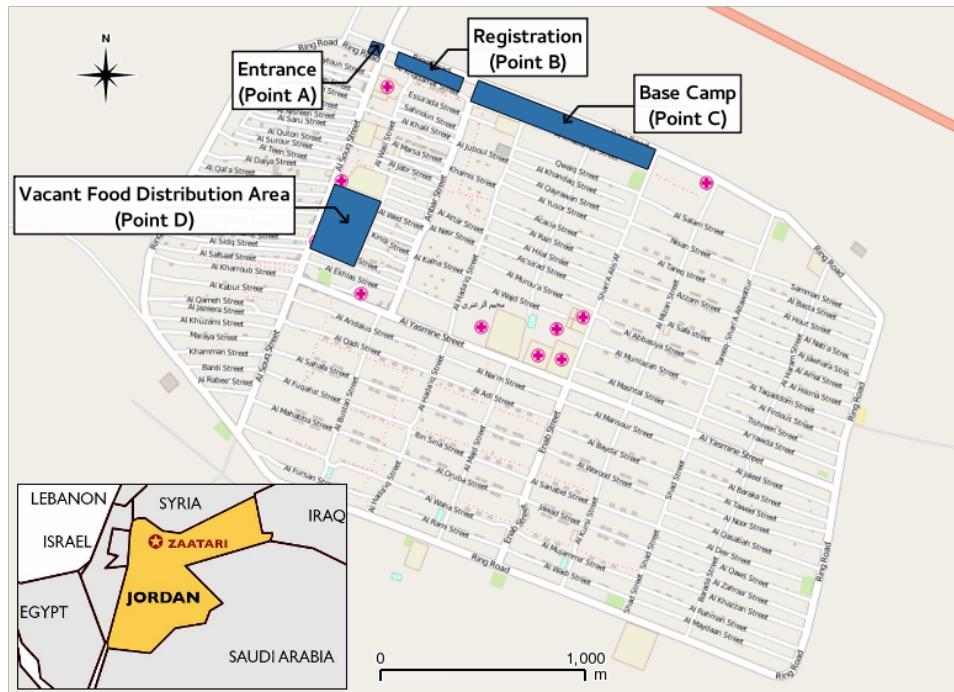


Figure 2.1: Za'atari camp map.

nian government became increasingly involved in its operations, particularly as security concerns and the need for educational services increased. Camp growth also required expanding functional boundaries and changes in land use itself. The increasing density of housing in the ‘old’ part of the camp, led UNHCR to expand housing areas to be places further afield, which were unfortunately distant from many service locations and economic activity in the camp. Also, changes in camp food management processes led to the abandonment of a food distribution location near the camp’s entrance in favor of modern grocery stores further from the entrance.

Mobile service in the camp is offered by three carriers: Zain, Umniah, and Orange. Much of their infrastructure is located outside the camp and was initially built to serve the surrounding rural community that existed prior to the camp's development. We discussed carriers with camp residents and were told that Orange is popular in the more urban areas of Amman, while Zain is popular for rural customers. Umniah was generally

not used by people with whom we spoke. Zain is the most popular carrier in the camp as SIM cards for the network are given to residents by camp administration as they first arrive and register at the camp.

As shown in Table 2.1, the cost of pre-paid data bundles vary amongst the three mobile carriers, with Umniah generally offering the cheapest rates and Orange offering the most expensive rates. All of the carriers offer reduced pricing for calls and SMS messages between same-network customers. Jordan’s nationwide cellular data connectivity is largely based on 3G HSPA technology as well as 2G EDGE technology. Carriers began to construct 4G LTE-capable infrastructure in 2015; it is generally not available outside of Amman.

		Data Bundle Size		
		200MB*	500MB†	1GB†
Carrier	Zain	2.00 JD	3.50 JD	5.00 JD
	Orange	1.99 JD	4.99 JD	6.99 JD
	Umniah	1.00 JD	3.00 JD	5.00 JD

* Weekly

† Monthly

Table 2.1: Data bundles available on Jordanian carriers.

We collected cellular measurements in Za’atari over the course of three days in January 2015. Our data collection focused on cellular control messages as well as received signal strength indicator (RSSI). We used eight Android phones with radio debug mode enabled, allowing us to log all cellular communications to a computer via USB. We used `xgoldmon`, an open source tool that converts debug logs into easily parsable and human readable packet capture (pcap) files. Each phone recorded all of its own uplink traffic as well as all broadcast traffic sent by cellular base stations. Using the eight Android handsets, we were able to log more than 95,000 cellular control messages. Phones also ran an application that logged latitude and longitude coordinates and cellular network details (e.g. RSSI, cell-ID, and connection state).

San Cristóbal Verapaz, Guatemala. Guatemala is one of the poorest countries in Latin America with a GDP per capita of around \$3,512. Despite this, mobile carriers within Guatemala have invested heavily in cellular infrastructure, providing connectivity to a large portion of the country. We collected measurements in the rural city of San Cristóbal Verapaz over eight days in January 2015 using the same methods as the Za’atari collection. Radio debug-enabled phones record broadcasts to pcap files for later analysis.

Three GSM carriers offer service in Guatemala: Tigo, Claro, and Movistar. All three are available in San Cristóbal Verapaz. Tigo is Guatemala’s most popular carrier and has built the most extensive cellular infrastructure network. Claro is the least popular carrier in San Cristóbal Verapaz and coverage is more focused on 3G in urban areas. Movistar is the second most popular and its network consists of a higher percentage of 2G base stations compared to the other carriers. It is difficult to compare prices for the carriers as there is no equivalent package offered on all three. Generally, Tigo is slightly more expensive than the others (e.g. 30GTQ for 500MB versus 25GTQ for 500MB from Movistar). Movistar is the most affordable carrier and prepaid scratch cards used for credit are widely available.

United States-based MVNO networks. We studied the routing behaviors and network performance of cellular mobile virtual network operators (MVNOs). MVNOs represent a different type of resource constraint compared with developing regions or displacement scenarios. MVNOs are virtual in the sense that they offer telecommunications services without owning all of the infrastructure used by clients. Instead, MVNOs pay mobile network operators (MNOs) for the right to service user traffic using the underlying base carrier network. MVNO networks are resource-limited in that their traffic is typically given lower priority compared with MNO user traffic. We performed active measurements of Internet traffic on four MVNO networks in the United States as well as the four underlying MNO networks in an effort to uncover potential factors that impact



Figure 2.2: A map of Southern Africa highlighting the location of Macha.

user performance on MVNO networks.

Macha, Zambia. We study Internet connectivity through a shared community network in Macha, Zambia. The location, shown in Figure 2.2, is a typical rural village with scant infrastructure; most residents are maize farmers. Macha’s local population is roughly 135,000, however it is rather sparse as the village has a radius of around 35 km, making the population density 25 per km^2 . The consequence of such sparse populations is a general lack of infrastructure as well as a lack of motivation for service providers to build infrastructure as revenues from such a population would likely be too low to justify the expenditure. The average estimated income is \$1/person/day – 5 times less than the round-trip cost to the closest town.

While Macha is connected to the national power grid, electricity is rarely available in individual households. The lack of electricity coupled with the high prices for user equipment and Internet provisioning makes it virtually impossible for Machans to use Internet at home. Internet users in Macha typically access the Internet from work, from

an Internet café or at school.

At the time of our study, Internet access in Macha was available through a microwave terrestrial link shared over a local wireless network connecting approximately 300 community members. The local network in Macha was delivered using 802.11 devices acting as point-to-point links, hotspots, and layer-2 bridges. Public Internet access was gained through an Internet café. To monitor traffic, we captured at the network gateway using tcpdump. With users' consent, we collect packet headers on the public-facing interfaces. We save network traces to an external hard drive and physically retrieved them for offline analysis to avoid affecting the traces or burdening the network with our traffic uploads. The Internet connectivity is distributed over an area of 6 km^2 , including schools, a hospital, a research institute and residential areas. We refer the interested reader to [123] for more details about the wireless network in Macha. Between 2008 and April 2011, the village was connected to the Internet through a satellite connection that cost \$1200/month and provided 256kbps downlink bursting to 1Mbps, and 64kbps uplink bursting to 256kbps. In April 2011, the village Internet access was upgraded to a higher quality microwave terrestrial link with speeds up to 2Mbps costing \$3600/month. At the time of the Internet link upgrade, approximately 300 residents were regular users of the Internet connectivity.

2.2 Resource-Limited Network Challenges

This section presents our experience in analyzing cellular and Internet connectivity and designing systems based on our analysis. This summary is based on traffic analysis and our anecdotal experience while spending time in the communities we have studied.

In refugee and displacement scenarios, network connectivity is of paramount importance for both displaced people and the aid agencies that serve them. Not only do

displaced people need to remain in contact with family they have left behind, they must reconnect with missing kin who may have been lost during relocation. While under normal circumstances, network connectivity is provisioned through a combination of fixed, wireless and cellular networks, in displacement scenarios, connectivity can be most plausibly resolved through robust and reliable cellular service.

Unfortunately, cellular services in refugee camps are often either unavailable, spotty, or of poor quality at best. The conflicts that caused the displacement often result in the destruction of any existing telecommunication infrastructure. Further, both planned and unplanned camps typically are established in undeveloped areas, leading to a concentration of people relying on rural cellular infrastructure that was never provisioned to support the sheer number of requests posed by a nearby camp, if that infrastructure exists at all. Such changed utilization is likely to challenge the capacity of any remaining communication equipment due to traffic overloads. Further complicating matters, displaced people often have little money, either because circumstances required they leave it behind, or because the infrastructure to access their money is not available. The lack of and poor quality technological infrastructure, coupled with the limited access to finances, present seemingly insurmountable obstacles to communication. Yet it is in exactly these circumstances that communication is most critical. Further, due to the limited financial resources of displaced persons, and the desire for refugee camps to remain “temporary”, existing cellular infrastructure is unlikely to be upgraded to handle the increase in load created by the new camp residents.

In rural, developing regions, the underlying economics and purchasing power of a sparse user base is traditionally the limiting factor in obtaining robust connectivity. From the service provider perspective there is little incentive to deploy costly infrastructure that will not likely lead to revenue that outweighs the capital and operating expenditures involved. Further exacerbating the challenges, in many developing regions there are

additional environmental challenges that networks must overcome such as poor quality of electricity. In our experiences in Macha, Zambia, the grid electricity we used was quite “dirty”, with large voltage swings being common which led to premature equipment failure.

2.3 Discussion and Conclusion

In many ways, connectivity is a necessity in our modern world. If we accept this premise, we must solve the unique challenges that arise when pushing the boundaries of the Internet’s frontier. While the global rollout of cellular and Internet connectivity in recent years has been one of mankind’s great feats of engineering, there remains a sizable percentage of the world population that have yet to be reached. This thesis examines the unique challenges faced by communications infrastructure in resource-limited environments, both rural developing regions and displacement scenarios. We identify opportunities for improvement and build systems for the settings we study.

Our work strives to bring connectivity to those that lack it, and where connectivity is of essential importance. Through our field studies of networks in Jordan, Guatemala, the United States, and Zambia, we are able to better understand inadequacies that network users in these locations face every time they attempt to connect. Using our analysis, we design network systems to help bring the next billion people online and to improve the experience for people that are currently limited to poor connectivity.

Part I

Characterization and Analysis of Resource-Limited Networks

In order to design effective connectivity solutions we must first understand the realities faced by networks in resource-limited environments by collecting data in-situ. We begin by performing measurements and analysis of cellular and broadband traffic in a variety of environments to identify opportunities for improvement.

We examine real-world cellular infrastructure performance using traces collected in an operational refugee camp and in rural Guatemala. We find that the commercial cellular infrastructure can suffer from radio resource congestion, which manifests in the inability for users to obtain voice/SMS or data service in a timely manner. We design a congestion metric that mimics a telephony industry key performance metric to better understand the level of congestion at the radio resource level on cellular base stations. Importantly, our metric is available to third party observers of networks and does not require our trace collector to be “inside” the cellular network.

We conduct a study of data traffic on MVNO networks in comparison with the underlying MNO networks on which the MVNOs operate. We find that performance can significantly differ between the MVNO and the MNO, as well as between MVNOs that are all sold as the same brand yet utilize different MNOs depending on the SIM card in use. We also find differences when using a ‘light’ MVNO versus a ‘full’ MVNO. Our findings motivate further research toward understanding the behaviors and performance of MVNO cellular carriers.

In many developing regions, cellular networks offer the only means of access to global connectivity. We analyze cellular network traces provided by Orange from their network in Côte d’Ivoire and find communication patterns correlated with population density where rural users are more likely to call other rural users. This locality of interest in communications means that local cellular network technologies could provide a great benefit to rural communities without relying on global connectivity for a large portion of the cellular traffic. We also examine social graph persistence. We find that urban areas,

although geographically distant, are strongly connected in terms of cellphone network activity.

Broadband Internet has become a crucial enabler of prosperity; however, broadband is not available in most rural developing regions. In such areas access is often provided through low-bandwidth, high latency solutions such as satellite if it is available at all. To understand the challenges faced by users of such networks, we study the performance and usage implications of an Internet access link upgrade, from a 256kbps satellite link to a 2Mbps terrestrial wireless link in rural Zambia. While performance improved in the near term after the upgrade, we found very poor network performance after three months. We found that the users, apparently encouraged by the increased link capacity, began using more bandwidth-intensive, and latency-sensitive content such as video streaming. Thus, while one would expect an Internet link upgrade ought to translate to improved performance and user experience, in our observed rural environment the demand outpaced the capacity of the link and brought unexpected consequences.

Chapter 3

Analysis: Cellular Performance Inference from Control Channel Messages

3.1 Introduction

A look at the worldwide mobile cellular subscription rate indicates that by the end of 2015, there was more than 7 billion mobile cellular subscriptions, corresponding to a penetration rate of 97%¹. While this seems like an immediate cause for celebration, a deeper look is required to more fully understand what this number represents. If we parse the numbers by region, we find that in 2015, 78% of residents of Europe and 77% in the Americas had mobile broadband subscriptions, compared with only 42.3% in Asia and the Pacific and 17.4% in Africa. Breaking coverage into geographic region begins to give us an understanding of access differences; however, stopping there masks critical disparities in the *quality* of available cellular connectivity. In particular, the cellular

¹<http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf>

technology (LTE vs 3G vs 2G), the data rate, the number of users per base station, and the cost of the subscription, are all critical factors that play into the persisting inequities between cellular users in developed and developing regions. Simply looking at who has access does not give an indication of whether the quality of that access is equivalent for different users in different locations.

To dive deeper into these disparities, we conduct a cellular infrastructure measurement campaign in three locations of diverse population characteristics. Specifically, we passively observe, collect and analyze messages broadcast by cellular base stations in order to assess coverage quality and usability. We collect measurements in San Cristóbal Verapaz, Guatemala, a city of roughly 20,000 in one of the poorest countries in Latin America; in the Za’atari refugee camp in Jordan, the oldest and largest refugee camp in Jordan with a population of roughly 80,000 in only 6 square kilometers; and, for comparison, in Santa Barbara, California, a community of about 90,000 residents in 42 square miles. Through our measurements, we reveal a number of interesting and important anomalies about coverage in these different locations that begin to paint a clearer picture of cellular quality divides. Most importantly, we discover chronic network congestion in some of the networks, which leads to a consistently poor quality of experience for associated users. Interestingly, we find that in our measurement locations, there are multiple providers available, which have varying levels of traffic load, ranging from minimal load (completely uncongested) to very heavy loads (chronically congested).

In the refugee camp context specifically, our novel use of detailed community-level cellular network analyses identifies unique divides and their relation to the camp’s social systems of refugee service providers. Service providers offer food, shelter, and clothing, among other services, and include agencies of the host country government, and the mobile network carriers. The technical systems are characterized by the properties of the camp’s wireless and cellular networks as well as the social and technical systems that

influence the availability of SIM cards, mobile phones, laptops and desktop equipment in the camp. With this background, our research identifies digital divides in the Za’atari refugee camp in Jordan, explains how these divides emerged, and makes recommendations for future infrastructure development and related policies.

3.2 Methods

Our measurement studies are focused on GSM networks as that is the most prevalent cellular technology in our selected communities. In this section we provide an overview of the GSM system and control messages we use for infrastructure characterization. We refer the reader to Chapter 2.1 for descriptions of the Za’atari refugee camp and San Cristóbal Verapaz, the locations we use for our case study. For comparison purposes, we capture messages on two cellular carriers in and around Santa Barbara, CA ($\sim 90,000$ residents) including a mix of urban and suburban base stations. We focus on the two major GSM carriers in the U.S.: AT&T and T-Mobile. We collected traces over roughly ten days for AT&T and five days for T-Mobile in October 2015.

3.2.1 GSM technology

The GSM core network is a complex hierarchy composed of many logical entities. The pertinent objects for our study are mobile stations (MS) and base station transceivers (BTS). The MS is the user device, commonly a mobile phone equipped with a SIM card. BTSs are the components that communicate with the user device over the air interface. MSs communicate with only one BTS at a time.

GSM control messages enable the inference of BTS control channel congestion serving the measurement locations. Figure 3.1 displays the message sequences that take place when an MS requests a private communication channel from the BTS, which is necessary

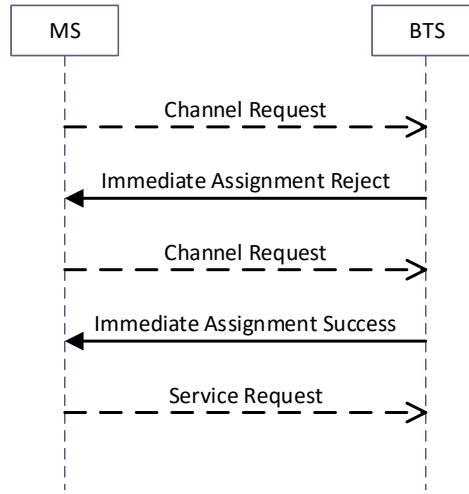


Figure 3.1: GSM network immediate assignment procedure. We capture messages broadcast from the BTS to MS over the GSM air interface.

for voice/SMS or data. We capture messages broadcast from the BTS to the MS (solid arrows in Figure 3.1) on the common control channel. When an MS needs to use the network for a voice/SMS or data session, it issues a channel request to the base station. If a channel is available, the base station responds with an *immediate assignment success* message that provides information about the available channel. A BTS operating at full capacity such that it is unable to allocate a channel will issue an *immediate assignment reject* message, in accordance to the 3GPP 04.08 specification², indicating no channel is available. Because these control messages are broadcast to all MSs connected to the BTS, we can leverage the immediate assignment success rate to approximate stand-alone dedicated control channel (SDCCH) blocking [114, 82], a GSM industry-standard key performance indicator (KPI) used to measure BTS congestion and to inform capacity expansion. We calculate the immediate assignment success rate by dividing the number of observed successful immediate assignment messages by the total number of immediate assignment (success and reject) messages.

²<http://www.3gpp.org/DynaReport/0408.htm>

Message capture

We capture broadcast messages using a total of eight cellular phones, which consist of Samsung Galaxy Nexus, S2 and Galaxy S4 handsets with radio debug mode enabled, which logs all cellular communications to a computer via USB. We use `xgoldmon`, an open source tool that converts debug logs into packet capture files using the GSMTAP pseudo-header. Each phone records all of its own uplink traffic as well as all broadcast traffic sent by BTSSs. Critically, our message capture is **non-invasive** and **non-intrusive**. Messages broadcast over common control channels are received in plaintext and intended for all MSs connected to a BTS. The GSMTAP pseudo-header does not include private user data. In all locations we capture using two phones per carrier, one set to prefer 2G and another set to prefer 3G. We do not capture 4G LTE as LTE is only available in the U.S. location.

3.3 Cellular Measurement Analysis

We investigate network performance and congestion via immediate assignment success and rejection messages. As described in Section 3.2, an MS issues a request to the BTS for a communication channel at the time that the MS needs to use the network for voice/SMS or data. In response, the BTS broadcasts an immediate assignment message to inform the MS whether or not its request is granted. We also examine signal strength and connectivity logs gathered during our visit to Za’atari.

3.3.1 BTS messaging load

We examine ‘busyness’ for our observed BTSSs on each of the carrier networks by calculating the number of immediate assignment messages per second. Table 3.1 summarizes

	Capture Duration (s)	Mean Immediate Assignments per second	Total Immediate Assignment Success Messages	Total Immediate Assignment Rejection Messages	Immediate Assignment Reject Percentage
AT&T (US)	840,114	0.04	34,539	9	0.03%
T-Mobile (US)	379,897	0.05	19,112	2	0.01%
Claro (GT)	625,051	0.01	7,661	0	0.00%
Movistar (GT)	625,073	0.21	127,996	1,055	0.82%
Tigo (GT)	622,340	0.21	129,837	846	0.65%
Orange (JO)	243,678	0.09	21,675	1,355	5.88%
Umniah (JO)	243,678	0.12	27,385	1,550	5.36%
Zain (JO)	243,678	0.12	19,804	8,800	30.76%

Table 3.1: Overall mobile network statistics.

our results across the measurement locations. It is important to note that immediate assignment messages are BTS-specific and they do not indicate the service footprint or population served by a BTS. As shown in the table, AT&T and T-Mobile in the U.S. have relatively lightly-used BTSSs, with 0.04 and 0.05 immediate assignments per second, respectively. These results could indicate networks comprised of a larger number of BTSSs, where each BTS serves a smaller area and hence a smaller number of users. Claro also exhibited very few messages. We believe this may be caused by Claro's lack of popularity in our measurement location. Tigo and Movistar exhibit similar message loads and have the highest rates of immediate assignment messages, roughly four times that of the U.S. networks. The three carriers in Jordan have similar messaging rates compared with one another and fall in between the observed load on the U.S. networks and those in Guatemala.

3.3.2 BTS congestion

In terms of congestion, we find qualitative performance divides between locations. Predictably, we capture almost no immediate assignment rejection messages on the U.S. provider networks. This indicates adequately provisioned BTSSs for the given client load. Apart from Claro, which has a very low load, the Guatemalan networks experience higher

rejection rates (minimum 21 times higher) than AT&T, the U.S. carrier with the highest rejection rate. Turning to the refugee camp network we see an even greater divide. The Jordanian carriers all have vastly higher rejection percentages than in the other locations, with Orange and Umniah both experiencing more than 150 times higher rejection rates than AT&T. Zain, the most popular carrier in the camp, has an extremely high rejection message rate (more than 1,000 times higher than AT&T). For perspective, prior work considers a 5% SDCCH blocking rate to be ‘high congestion’ [115].

While the rejection percentages may seem low, their impact can be drastic. Each rejection includes a value for a backoff timer that forces the MS to wait up to 255 seconds before resubmitting a resource request. In terms of user quality of experience, such backoffs result in very poor performance. For instance, the GSM specification classifies call setup times, the time between a user pressing the call button and the call connection, into three categories³. ‘Fast’ call setup is defined as 1 - 2 seconds; ‘normal’ 2 - 5; and ‘slow’ 5 - 10. We analyze rejection messages in the traces and find the Jordanian networks frequently send backoffs of up to 128 seconds during congestion events. Clearly, the additional time attributable to backoffs we observe extends call setup times far beyond the classification of ‘slow.’

3.3.3 Za’atari temporal congestion characteristics

Because the carriers serving Za’atari experienced the greatest congestion, we take a deeper look into carrier-specific performance in the traces to investigate any cross-network traffic load relationships. Figure 3.2 shows the percent of immediate assignment messages that were rejections in five minute bins over the course of a single day during our measurement window⁴.

³http://www.etsi.org/deliver/etsi_gts/02/0267/05.00.01_60/gsmts_0267v050001p.pdf

⁴We choose to examine January 6th as it was the day we were able to capture messages without interruption. Additionally, a winter storm on January 7th caused the camp to be closed early. The

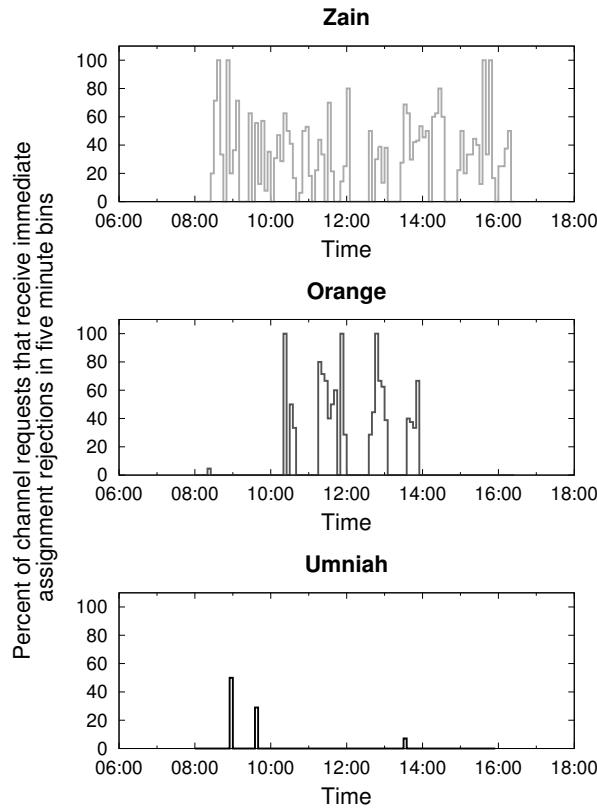


Figure 3.2: Percent of immediate assignment messages that were rejections in five minute bins on January 6th 2015 in Za’atari.

We see that the carriers’ rejection patterns do not resemble one another. The Orange network was congested in short, severe bursts. It also appears as though congestion on Orange was higher at particular times of the day that correspond with workday schedules (i.e. before and after lunch). Umniah, on the other hand, exhibited almost no evidence of congestion throughout the day. The Zain network experienced sustained congestion, occurring throughout the day and frequently reaching rejection percentages above 50%.

The asynchronous nature of congestion across providers implies that users can potentially achieve improved connectivity and performance by switching points of attachment when the BTS or carrier on which they are connected experiences congestion.

Immediate assignment rejection messages include a *backoff wait value* indicating how impending storm may have affected network usage as residents were making preparations.

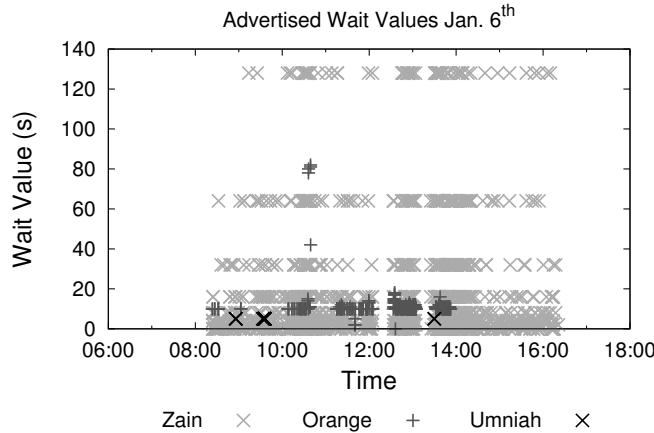


Figure 3.3: Observed backoff wait values.

long a phone must wait until it repeats its request for a resource. Emergency calls are not subject to the wait value restriction. Cellular networks use the backoff wait value to ease congestion on the control channel caused by phones repeatedly requesting unavailable resources in quick succession. The value ranges from 0 - 255 seconds and the advertised value is determined by the level of overload on the base station. We can leverage this value to indicate the *severity* of congestion. Figure 3.3 shows the observed wait values for the three carriers over the course of the day. It is clear that Zain has the most rejection messages as well as the highest severity of congestion as we see many 128-second waits. Such underprovisioned infrastructure causes a negative impact on user experience, evidenced by two-minute waits to place a voice call, send an SMS, or use data services. We believe that the congestion of Zain may be in part due to the policy of giving newly arriving residents SIM cards for Zain.

Timing advance

As discussed above, immediate assignment messages are broadcast to all phones connected to a cellular base station. These messages also include *timing advance* information for the phone for which the message is intended. The value indicates, in steps of roughly

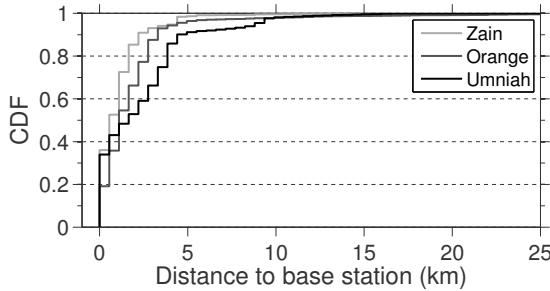


Figure 3.4: Calculated distances between phones and base stations.

$3.69\mu\text{s}$, how distant a phone is from the base station. Because radio waves travel at the speed of light, we know the timing advance value changes for every 550 meter change in the distance between a phone and the base station. Using the observed values we can therefore estimate distances between all the phones that receive immediate assignment messages and the base station on which our measurement phone is connected. Though this is a rough estimate of distance, we can use it to generally infer the coverage footprint of the camp-serving cellular infrastructure. Cellular carriers configure the coverage area of base stations depending on location and nearby population in order to balance infrastructure costs and ability to serve users. Base stations in densely populated urban areas are typically set to serve small coverage areas, while rural-serving infrastructure has much larger coverage footprints.

Figure 3.4 displays a cumulative distribution function (CDF) of the calculated distances for each of the carriers. While the majority of distances are within a few kilometers, we see that all three carriers have base stations where users are connected from longer distances. Umniah has the most long-distance users while Zain users experience the lowest distances. Our observations indicate that some of the camp-serving cellular base stations are configured to cover large areas, as in rural configurations. We posit that this may be caused by infrastructure that was installed prior to camp construction as the area was previously rural. The maximum data throughput for a phone is negotiated

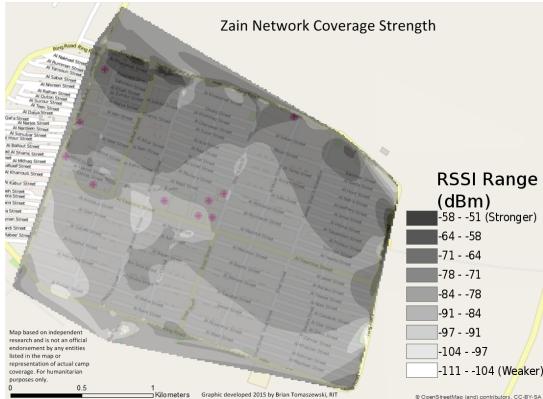


Figure 3.5: Zain RSSI.

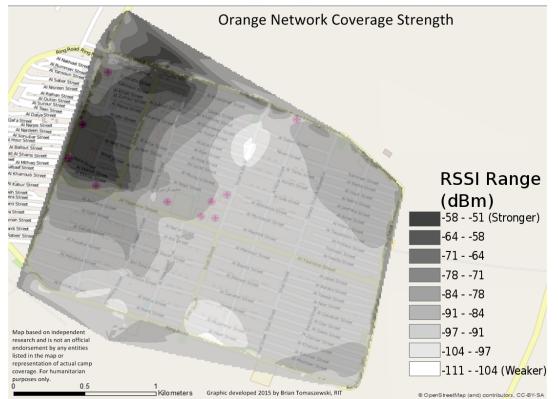


Figure 3.6: Orange RSSI.

based on the wireless channel quality between the phone and the base station, which is heavily impacted by distance. Given our observations, camp residents are likely limited to poorer data rates than that which could be attained via physically closer base stations at least some portion of the time. These results could help to partially explain the poor network performance reported in [143].

3.3.4 Za’atari geographic analyses

We measure the cellular infrastructure signal strength and data connectivity type with respect to location in order to explore community-level divides of the cellular infrastructure serving Za’atari.

Received signal strength

We first investigate the cellular signal strength our measurement phones observed during our time in the camp. Figures 3.5 and 3.6 show the collected RSSI values for two of the three carriers, interpolated using natural neighbor interpolation [60]. We do not include an RSSI map for Umniah as our measurement phones failed while we were in the middle of the camp. The highest signal strength values for Zain and Orange are

near the entrance gate and base camp, located on the northern edge of camp just east of the entrance. Both carriers generally have lower signal strength coverage in the more residential areas of the camp.

The carriers clearly have different coverage profiles throughout the rest of the camp. Zain, the most popular carrier according to residents, has more locations with high signal strength, particularly focused along the northern edge of the camp. Orange, on the other hand, has very high signal strength near base camp as well as the market area in the northwest area of camp. Interestingly, the vacant food distribution location, depicted as Point D in Figure 2.1, has extremely high signal strength values. At the time of our visit this location was an abandoned area that had previously been used by the World Food Programme and was now a fenced-off, unused area. Inside of the vacant area was a cellular base station, presumably owned by Orange given our signal strength readings. It is unclear why this location was chosen for an installation, whether it was simply due to convenience, availability of space, or that previous use of the area led to a high density of users, and hence, coverage was needed in this location. The highly variable nature of refugee camps, as evidenced by the changing spatial configurations (food distribution centers convert to unused spaces when people are shifted to western-style supermarkets) makes it difficult to anticipate coverage areas and more important to have flexible network connectivity arrangements. Phones on Orange also collected some very low RSSI measurements in the south central and north central residential parts of the camp. This could be due to shadowing caused by a small hill in the middle of the camp.

Cell reselection offset

Cell reselection is the process phones use to choose which base station to utilize. Phones accomplish this by constantly monitoring received signal strength for the base

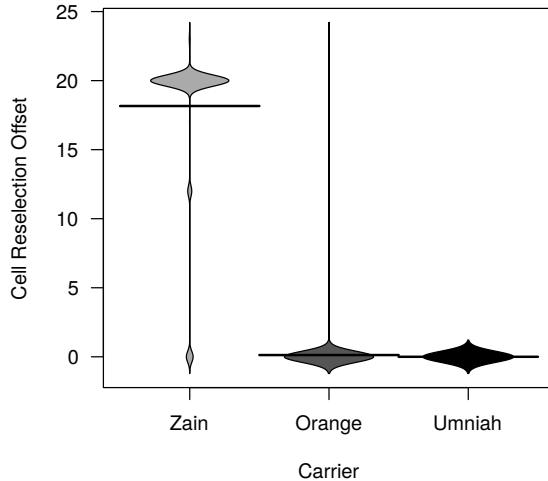


Figure 3.7: CRO artificially increases the calculated signal strength when the phone determines which cell to use.

station they are connected to as well as for a list of neighboring base stations provided to the phone by the current base station. In addition to the received signal strength indicator (RSSI), the 3GPP specification allows for carriers to advertise cell reselection offset (CRO) values, which effectively increase the signal strength of a cell in the phone's internal RSSI calculation. The CRO values can range from 0 - 63 with each number representing an increase of 2 dBm, meaning base stations can cause phones to artificially increase the calculated RSSI of a base station up to 126 dBm. Carriers can use this to shape and load balance infrastructure usage in areas by artificially increasing or reducing the 'attractiveness' of base stations for phones.

Figure 3.7 shows a violin plot with the distribution of offset values that were observed in the camp for each of the carriers, with horizontal lines indicating the mean. Zain, the most popular carrier, makes extensive use of CRO with 87% of messages advertising a CRO of 20, resulting in an artificial increase of 40 dBm. The effect of such an increase is illustrated in Table 3.2, which displays the mapping Android uses to convert RSSI values to 'levels' that define how many signal strength bars are displayed on a phone. For example, a 40 dBm increase would cause a phone to view a 'poor' signal of -100 dBm

RSSI Range	Signal Level
>-51 dBm through -65 dBm	Great
-67 dBm through -81 dBm	Good
-83 dBm through -97 dBm	Moderate
-99 dBm through -111 dBm	Poor
<-113 dBm	None

Table 3.2: Android signal strength RSSI value to level map.

as a ‘great’ value of -60 dBm. Thus, while the phone may indicate to the user strong connectivity to the base station, in reality it may have a very weak connection. Traces from the other two carriers included very few non-zero CRO values.

We inspect the cellular IDs of the Zain base stations that advertise non-zero CRO values. We find that the data connectivity on all of those base stations was limited to either 2G EDGE or no data connectivity (voice/SMS functions without data). We did not observe any 3G Zain base stations in the camp utilizing CRO. While the precise reasons behind such a configuration are unknown and could be attributable to Zain purposely engineering capacity planning into the deployed infrastructure, the overall effect of the configuration is to steer mobile devices on the Zain network toward 2G base stations serving the camp. This can be detrimental for some users as phones ‘prefer’ the artificially-enhanced base stations even when they may be better served by a base station with a higher real-world RSSI value. Such a configuration could be desirable in areas where a carrier aims to reduce the service footprint of 3G base stations not using CRO, as phones will only connect to the CRO-disabled 3G base station if they detect very high RSSI values (i.e. they are physically near the base station). This can advantage users in close physical proximity to a base station, such as those allowed inside base camp.

Our measurements indicate that base stations using CRO often failed to provide any data connectivity. This is a sign of congestion, which is a logical outcome of a base station appearing to client phones as having very high signal strength, thus increasing

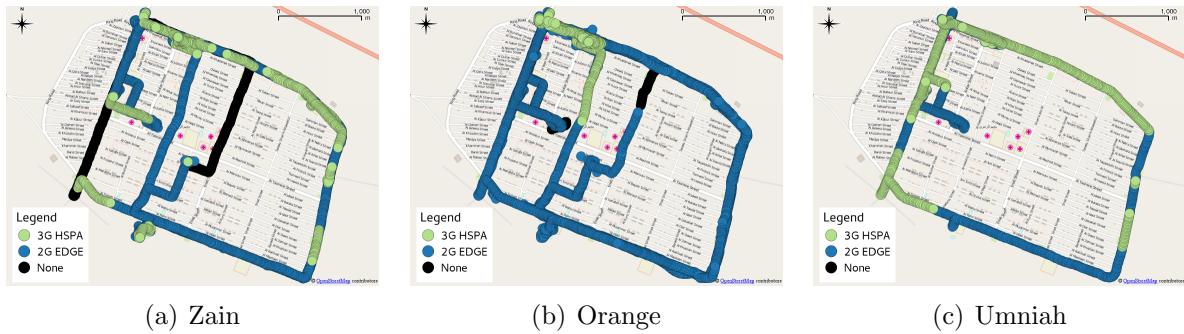


Figure 3.8: Data coverage. Color indicates connectivity: 3G HSPA (Green), 2G EDGE (Blue), No connectivity (Black).

the number of phones that attempt to use it. Essentially, the base stations advertising high CRO values worsen user connectivity in two ways: they attract devices away from nearby 3G base stations onto 2G base stations, and this attraction consequently causes the farther-away 2G base stations using CRO to be more heavily used and thus more highly congested.

Data connectivity

Beyond signal strength, we collect the data connection *types* during our measurement for the carriers. All three carriers have 3G HSPA and 2G EDGE connectivity in some portions of the camp. We configure our measurement phones to prefer 3G connectivity and record the actual connectivity as we travel around the camp. Figure 3.8 shows the attained data connectivity types for each of the carriers. Similar to the signal strength maps, we observe that the best data connectivity speeds are attainable in the northern areas of the camp and near base camp for the carriers. Zain's 3G connectivity (Figure 3.8(a)) is mostly focused around the northern road along with a few small areas around the perimeter of camp. Both Zain and Orange (Figure 3.8(b)) have areas with no data connectivity in the north central area of camp; Zain also lacks data in the southwest area of camp. Orange has the smallest area of 3G coverage, with the majority of

the camp covered by 2G EDGE connectivity. Umniah (Figure 3.8(c)), the least popular carrier according to residents surveyed, has the highest number of points indicating 3G coverage.

One reason for the lack of widespread 3G connectivity is low antenna height. While in base camp, we observed a base station with antennas placed roughly 5 meters above the ground, a height that results in a small service footprint. In this case the base station's location within base camp and the height is likely chosen to purposely limit the service area. Our general observations of data connectivity are that 3G coverage is focused around administrative and NGO locations, while residential areas are largely covered by 2G EDGE connectivity.

Generally, the southern areas of the camp have slower data speeds compared to the northern areas. This is important as UNHCR has expressed the intent to increase spatial distribution of both services and residents in the southern and eastern areas of camp; these areas are quite sparsely populated while northern and western are dense. Some residents have resisted these efforts of relocation. While certainly not the sole reason, data connectivity speeds could impact the desirability and livability of locations in the camp.

3.4 Discussion

Our work examines cellular infrastructure in resource-limited environments at the community-level. We find that infrastructure serving these environments often struggles to provide connectivity guarantees to users when demand outstrips network capacity. Given the highly dynamic environment of refugee camps and the lack of permanence, rapidly-deployable cellular infrastructures present an interesting potential direction. Recent work has deployed local cellular networks [88, 87, 206] in remote areas without

existing service. Successful deployments by nonprofit organizations show that it is technologically and economically feasible for agencies like UNHCR to deploy and operate their own ‘pop-up’ cellular networks, which can quickly expand and shrink in concert with the demands of a refugee camp. The challenge posed by Za’atari is the presence of existing cellular service, as any local cellular network would be required to coexist with incumbent carriers. Such a scenario was explored in [165] using an application designed for rooted Android smartphones to programmatically duty cycle between commercial and local cellular networks without requiring user intervention.

Our investigation characterizes community-level divides so that they may be narrowed. We observe inter-carrier divides in the form of congestion and a spatial distribution divide throughout the camp manifested in different levels of connectivity. Our hope is that by using information gleaned from our analysis, network operators and users in resource-constrained environments can more effectively leverage the limited available technology.

3.5 Related Work

Our measurement of the cellular infrastructure focuses on messages broadcast over the GSM air interface rather than application-level performance. Two common approaches to understanding cellular performance are application-level studies that look at end-to-end characteristics and radio-level studies that focus on the access link. Our measurement of the cellular infrastructure is a radio-level study, that specifically utilizes messages broadcast over the GSM air interface. Recent work has illustrated the potential impact that cellular radio state has on end user experience [187, 154, 120]; and how air interface messages can be used to infer cellular user activity [113, 23]. Our work demonstrates how detailed, small-scale analysis can support research on local-level infrastructure. This ap-

proach is in contrast to efforts to measure network performance on a global or nationwide scale, such as the FCC's Measuring Broadband America⁵, and enables us to separate performance attributable to client-facing infrastructure serving the measurement locations from that which is related to carrier's core networks.

Our research contributes to the digital divide literature in three ways. First, we provide a middle ground for physical access research [189], offering a more granular perspective as compared with those conducted at broader geographic scales [33] or at the level of individual users [83]. Second, we also provide middle ground by highlighting the role of organizations as both potential providers and influencers of network infrastructure, that at the same time are consumers of bandwidth at the community level. As such, our research fills the gap between studies seeking to explain divides based on analyses of government policies at national and sub-national (state, province) levels, as well as those that seek to shed light on policies and norms affecting individuals' cognitive and social access [189]. Finally, our community level analysis complements the primary scholarship at this level, namely community technology centers. By highlighting the broader technical aspects of access, particularly in the very popular mobile and wireless realms, our research expands notions of community level digital divides.

⁵<https://www.fcc.gov/measuring-broadband-america>

Chapter 4

Analysis: MVNO Cellular Network Performance

4.1 Introduction

What factors cause one mobile Internet provider to be faster than another, even if they share some common core infrastructure? Traditional metrics chosen to represent speed may not perfectly correlate with end-user performance and are heavily influenced by the design and behavior of the underlying mobile data network. The challenge of mobile network characterization is further extended with the rise in popularity of mobile network virtual operators (MVNOs). In this chapter, we shed light on observable traffic behaviors exhibited by mobile networks that affect performance metrics and user experience. We examine mobile data network behavior when connecting to popular content delivery networks used to serve media. We are particularly interested in performance comparisons between the four major mobile carriers in the United States and MVNOs that license use of the underlying base carrier infrastructure. Ultimately, we want to explore network topology factors that affect traffic in mobile data networks.

Increasingly popular due to relaxed contract terms, MVNOs have quickly grown their market share in recent years [108, 62, 170]. They operate by leasing access to base mobile network operator (MNO) infrastructure, thus avoiding the high cost of building their own networks or licensing spectrum. Performance of MVNO data networks is often assumed to be inferior, but ultimately at least somewhat attributable to the underlying base carrier network. Previous work [203] has shown that is indeed the case; application performance suffers when using MVNO networks compared to MNOs. We investigate MVNOs and MNOs, searching for potential causes of degraded performance such as server resolution location and inefficient (e.g. excess hop counts and geographically indirect) paths.

We focus on traffic to content delivery networks (CDNs), which improve performance for end users by replicating identical content across geographically diverse locations [186]. CDNs are important factors in the user experience as they are typically responsible for delivering large web objects. The exact CDN server chosen by the client when browsing is typically dependent on DNS resolution with the expectation that the client is ‘near’ the DNS resolver. Unfortunately for most users, mobile data networks are strongly hierarchical and it has been shown that accurately localizing mobile users is a difficult challenge [156]. The localization problem illuminates a critical issue for mobile networks: the closest or best server depends on the mobile network core topology as well as peering arrangements between the content providers and mobile carriers. We study geographic paths taken by traffic on all of the mobile networks in order to better understand the obtained performance and routing behavior of the networks. Specifically, we are interested in the following questions:

- *Can we identify reasons behind MVNOs performing worse than MNOs?* We characterize network performance for all four major U.S. carriers as well as a single MVNO for each, discovering that performance appears to be dramatically affected by destination server location.

- *Can we find potential areas for improvement in order to reduce performance gaps between mobile carriers?* We find that MVNOs have more intermediate hops, which are also geographically inefficient in the case of full MVNOs we study. From our study we believe there is room for improvement with regards to mobile network topology.
- *Do we observe marked difference between full and light MVNOs?* We observe that a light MVNO closely resembles the underlying MNO, while traffic on full MVNOs differs, often exceptionally, compared to respective MNOs.

4.2 Background

CDNs and DNS. The use of CDNs to deliver content from distributed replica servers is commonplace in order to improve performance as Internet content has become increasingly heavy and media-rich. Client DNS requests resolve to particular replica server IP addresses when the clients browse the Internet. Ideally, the resolved servers are ‘near’ (e.g. lowest round trip time) the client relative to other potential servers in order to maximize application performance [147, 27]. A challenge for mobile data networks is that the limited number of public-facing gateways in the cellular core network, as well as the location of cellular network DNS resolvers, make localizing clients from an outside perspective difficult. Peering arrangements, or lack thereof, between mobile providers and content providers also leads to inefficient traffic routes even with the presence of a nearby replica [202].

MVNOs. Recently, MVNOs have increased in popularity worldwide. MVNOs are virtual in the sense that they offer telecommunications services without owning all of the mobile infrastructure used by clients. Instead, MVNOs pay MNOs for the right to service user traffic using the underlying base carrier network. The rise in popularity of MVNOs

is often attributed to relaxed contract terms such as pay-as-you-go and pre-paid plans compared to traditional base carriers in the U.S. which have traditionally operated using post-paid plans. MVNOs can be classified in one of two ways: **full** or **light**. Full MVNOs are carriers that license only the radio network of the base carrier. They implement their own core, including authentication and billing services (i.e. they distribute their own SIM cards). Light MVNOs, also called *resellers*, are re-branded versions of the base carrier, which means they can fully use the base carrier infrastructure. Mobile operators often create light MVNOs to target specific demographics or to lower consumer cost by cutting back on support services.

4.3 Data Collection

We collect data from eight mobile devices running on eight different carriers between March 6 and March 20, 2015. We conduct the experiment over two weeks to account for performance differences attributable to time-of-day patterns. All measurement phones are located in Santa Barbara, CA and left in a static location. All phones report ‘good’ or ‘great’ signal strength via the Android telephony API throughout the experiment. For simplicity, we focus on routes and performance associated with the popular social media sites Facebook and Instagram. These services are responsible for huge amounts of mobile Internet traffic, 19.43% and 4.27% respectively in North America [158], and are widely replicated across many well-known CDN data centers, which allows us to explore geographic differences between carriers. Measurements gathered across additional locations, carriers, and sites would be ideal; however, this study is an initial look at potential factors impacting MVNO network performance and we hope to motivate further, more in-depth research. The list of CDNs that we use as measurement points can be accessed on our project repository at <https://github.com/schmittpaul/mobileCDNs>. The list

includes 108 servers: 72 associated with Facebook and 36 associated with Instagram. Some servers are location-specific, identified by location clues in the name. We include international servers in our study as through initial work we find that mobile traffic surprisingly resolves to such servers a significant portion of the time ($>5\%$) for multiple carriers.

4.3.1 Carriers and phones

We collect data on all four of the major base carriers in the United States. We identify base carriers as A, B, C, and D. Carriers A and C are GSM networks while carriers B and D use CDMA technology. MVNO carriers are identified as A-1, B-1, C-1, and D-1, with their letters indicating the underlying base carrier. MVNO B-1 is a light MVNO, which means that it has full access to the infrastructure of carrier B. Carriers A-1, C-1, and D-1 are all branded as the same full MVNO with different SIM cards and contracts specifying the base carrier used. All phones run Android 4.4 and we leave them in a high-power state to avoid latency due to radios entering low-power states. All phones are attached to their carrier (i.e. not roaming). We choose to run all experiments while connected via 3G rather than 4G due to uneven 4G LTE coverage in our area between carriers. Recent work [202] has found that 3G and 4G mobile networks in the U.S. have few Internet ingress points, meaning 4G networks will exhibit similar behavior in terms of routes and CDN resolution as 3G networks.

4.3.2 Traceroute and location data

Each hour of the testing period, each device records a `traceroute` to each of the servers in the CDN server list, resulting in $14 \times 24 \times 8 \times 108 = 290,304$ records. We then use multiple techniques to estimate the location of each IP address in the traceroute records. We first employ the IP2Location DB5 database in order to map the traceroute

IP addresses to latitude and longitude coordinates. Unfortunately, prior work has established that IP geolocation databases are often rather inaccurate [144]. We also verify through a manual sanity check of the IP-location mapping, where we find improbable location mappings. To fix inaccuracies we use two other sources to manually estimate location for 5,172 unique, routable IP addresses observed over the course of the experiment. We use `nslookup` to resolve the human-readable name of the IP address if it exists. We do this because routers and servers often include three or four character location clues in their names. Next, we use Internet looking glass servers, available through `traceroute.org` from multiple cities around the U.S., to traceroute to each IP address. Observing the paths taken and RTT values from geographically diverse vantage points enables us to further estimate location (e.g. RTT of a few milliseconds from a particular looking glass server and intermediate hops containing location identifying names). Overall, we find that out of the 5,172 unique IP addresses, we override 1,988 addresses (36.4%) from the IP2Location database with our manual location estimate.

We run `whois` on each observed IP address to determine the associated Autonomous System (AS) number. With this information, we create a data set corresponding to each attempted traceroute that includes: the number of hops, the IP address associated with each hop, the geographic coordinates associated with each hop, the autonomous system number for each hop, and the observed RTTs associated with three traceroute probes.

4.4 Network Analysis

We measure traffic on the four major mobile network operator networks in the United States as well as MVNO carriers operating on each of them. We first look at network performance using standard metrics such as round trip times, hop counts, and autonomous system paths. We then combine geographic information and traceroute records to explore

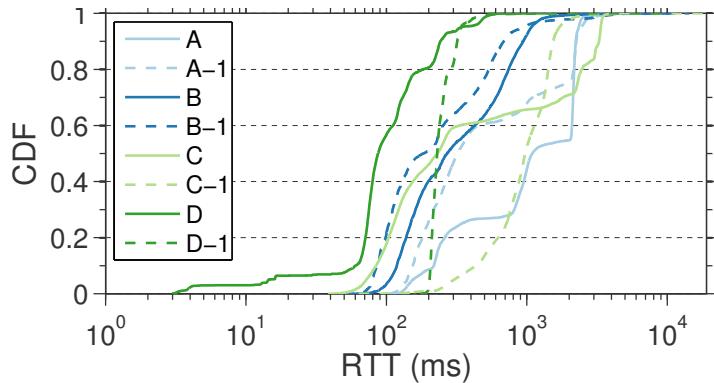


Figure 4.1: RTT measurements for mobile carriers to non-location specific servers.

traffic route path characteristics.

4.4.1 Round trip times (RTT)

We begin by investigating RTTs for packets traversing the mobile networks to the 73 **non-location specific** servers specified in our CDN server list. RTT is a critical metric in network performance as the majority of TCP variants rely on RTT to determine throughput [103]. Figure 4.1 shows a cumulative distribution function (CDF) plot of measured RTT values for all carriers in our study. We see considerable performance variance between the networks despite all measurements originating from the same location.

We also observe significant performance differences between base carriers and MVNO carriers operating on the corresponding base carrier infrastructure. For instance, in Figure 4.1 we see a 772.03 ms difference between the median RTT values for carrier A and the MVNO carrier A-1. However, the most surprising results are that MVNO carriers A-1 and B-1 outperform their respective base carriers in terms of achieved RTTs, with the aforementioned 772 ms lower median value for A-1 and a 87.24 ms median difference between B and B-1. These results contradict the expectation that MVNOs universally offer inferior performance. Previous work has established the widespread use of transparent middleboxes on mobile networks [188], which could help explain why

networks with better RTT performance do not necessarily outperform others as such middleboxes likely ignore our measurement traffic. In order to understand round trip performance more fully we must also consider the *locations* of servers to which client traffic is resolved, explored in the next section.

4.4.2 Location-specific RTTs

We study performance by examining the data center locations to which carriers are most likely to resolve. We record the geographic location for the destination server in all of the traceroutes corresponding to non-location specific requests in the previous experiment and find that the vast majority of requests resolve to data centers in nine US cities and most carriers heavily favor relatively few server locations. We then measure RTT performance to all locations using our list of location-specific servers, which are identified using 3-character airport codes in server names (e.g. `scontent-a-lax.cdninstagram.com` corresponds to an Instagram server in Los Angeles). Figure 4.2 shows RTT CDFs for each of the data center locations and highlights each carrier's top three ‘preferred’ locations. We find location preference by calculating the percent of ‘hits’ at each location for all non-location specific requests.

The figure illustrates large performance differences and unique behaviors between carriers. Some MVNOs appear to mimic the underlying base carrier, while others behave in drastically different ways. Perhaps the most interesting performance is seen on carriers A and A-1. Carrier A experiences vastly different round trip times between different CDN locations. Additionally, carrier A favors CDN sites (Dallas, Seattle, Los Angeles) that have the slowest median RTT compared to the other locations. The latency to Los Angeles servers is the second longest, despite Los Angeles being the data center nearest our measurement location of Santa Barbara. Carrier A-1 (Figure 4.2(b))

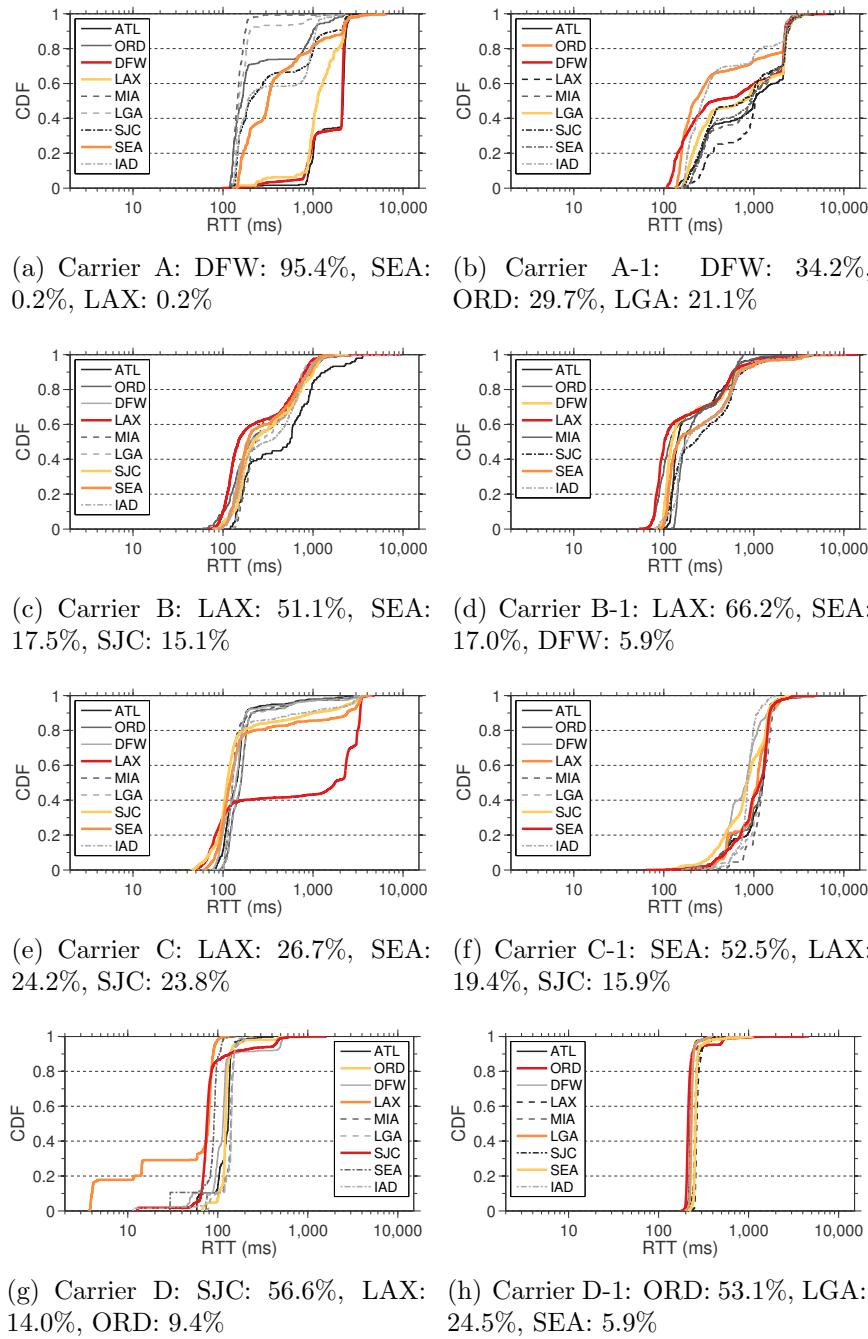
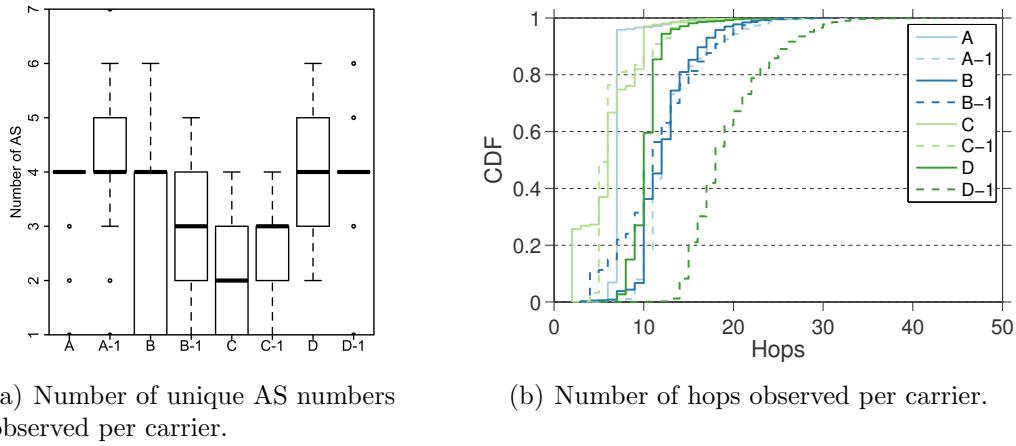


Figure 4.2: RTT comparison for specific CDN locations identified by airport codes: Atlanta (ATL), Chicago (ORD), Dallas (DFW), Los Angeles (LAX), Miami (MIA), New York (LGA), San Jose (SJC), Seattle (SEA), Washington DC (IAD). Each carrier's top three preferred locations are indicated.

displays the broadest range of RTT values across all CDN sites, and also favors servers located in Dallas, TX. However, carrier A-1's second and third most popular locations are Chicago and New York, respectively. We believe these results are due to A-1 being a full MVNO; thus, they employ their own core infrastructure and have service and peering arrangements independent from the base carrier A.

Carriers B and B-1 (Figures 4.2(c) and 4.2(d)), on the other hand, perform more similarly in both latency measurements and preferred destinations. In fact, MVNO B-1 slightly outperforms the base MNO in terms of RTT in our experiments. Both carriers tend to route traffic toward Los Angeles. Los Angeles also tends to correspond to the lowest RTT values for both carriers. The striking similarity can be explained as carrier B-1 is a ‘light MVNO,’ thus B and B-1 use the same infrastructure to handle client traffic. In this regard, it stands to argue that customers considering carriers B and B-1 are essentially choosing between the same service when it comes to connecting to our specified CDN sites.

Carriers C and C-1 are quite different from one another in terms of performance even though they favor the same three data center locations. Interestingly, carrier C (Figure 4.2(e)) routes the highest percentage of its traffic to servers in Los Angeles, which achieve highly variable RTT values (seemingly bimodal). We speculate that this result is due to the carrier load-balancing flows across dissimilar paths. Latency values on carrier C-1 (Figure 4.2(f)) are rather consistent to all CDN locations, with higher RTTs overall compared with carrier C. Carrier D (Figure 4.2(g)) experiences the lowest network latencies overall. This carrier tends to favor CDN servers located in San Jose, CA, which also has the lowest median RTT value for carrier D. MVNO carrier D-1 (Figure 4.2(h)) shows the most consistent latency across all data center locations, but interestingly favors CDN servers in Chicago, 2,961 km away from San Jose. Similar to A-1, we believe this is likely due to D-1 being a full MVNO, with traffic traversing a different core network



(a) Number of unique AS numbers observed per carrier.

(b) Number of hops observed per carrier.

Figure 4.3: Path metrics for mobile networks. MNOs and MVNOs exhibit similar AS path lengths but differ in the number of hops taken to reach the destination.

than the base MNO.

Overall, we observe that some MVNOs exhibit drastically different RTT performance from their MNO counterparts, while others are similar. While it seems that the light MVNO can be characterized as simply a re-branded version of the base MNO, our experiments using full MVNOs show unique latencies and resolutions between them and their MNOs. Thus, these carriers do not appear to simply reflect the performance of the MNO network on which they reside. Such behaviors will clearly impact network latency and throughput and may help to explain why MVNOs networks generally perform worse than MNOs.

4.4.3 Autonomous system paths and hop counts

We next investigate traffic routes with respect to autonomous system (AS) paths in the traceroutes for each carrier to non-location specific servers. We use `whois` queries to map all IP addresses seen in carrier traceroutes to AS numbers. Figure 4.3(a) shows the number of unique AS numbers observed across the carriers, with the dark line indicating the mean. As shown, it appears as though MVNO behavior overall is similar to MNO networks. This result illustrates that MVNO networks are bound to some degree to

the MNO network configuration. We study the actual AS numbers traversed by traffic between MNOs and MVNOs and find that they generally match, and as such omit this analysis for brevity. Interestingly, although carriers A-1, C-1, and D-1 all fall under the ‘same’ MVNO brand, traffic for each client behaves differently based on the underlying carrier. It appears that in terms of AS paths, MVNOs closely reflect the underlying MNO. These results lead us to investigate hop counts to help explain performance differences between MVNOs and MNOs.

We consider the total number of hops in traceroute records for traffic on each mobile network to non-location specific servers. We only consider records that reach the destination server. Figure 4.3(b) shows the results. As with other metrics, we observe considerable variability between carriers. Carriers A and D both use dramatically fewer hops to reach the server compared to their respective MVNOs, while carriers B and C closely resemble their respective MVNOs. The path length inflation seen on carriers A-1 and D-1 could help explain poorer RTT performance compared to the base carrier. We also observe that MVNOs A-1, C-1, and D-1, which are all marketed as the same carrier, experience very different length paths to reach the destination servers. We believe the variance may be attributable to the different preferred locations observed in Figure 4.2.

4.4.4 Geographic path analysis

Lastly, we study the geographic paths taken by traffic on each mobile provider going to specific CDN locations. This analysis provides us visual insight into the carrier network behavior and performance. Figure 4.4 illustrates hops taken by traceroutes over our two-week experiment where the destination servers are located in Los Angeles, CA (140 km away from our location in Santa Barbara, CA). Due to space limitations, we only include plots for Los Angeles and four carriers, as it is representative for all locations observed and illustrates the contrast between full and light MVNOs. The figures indicate

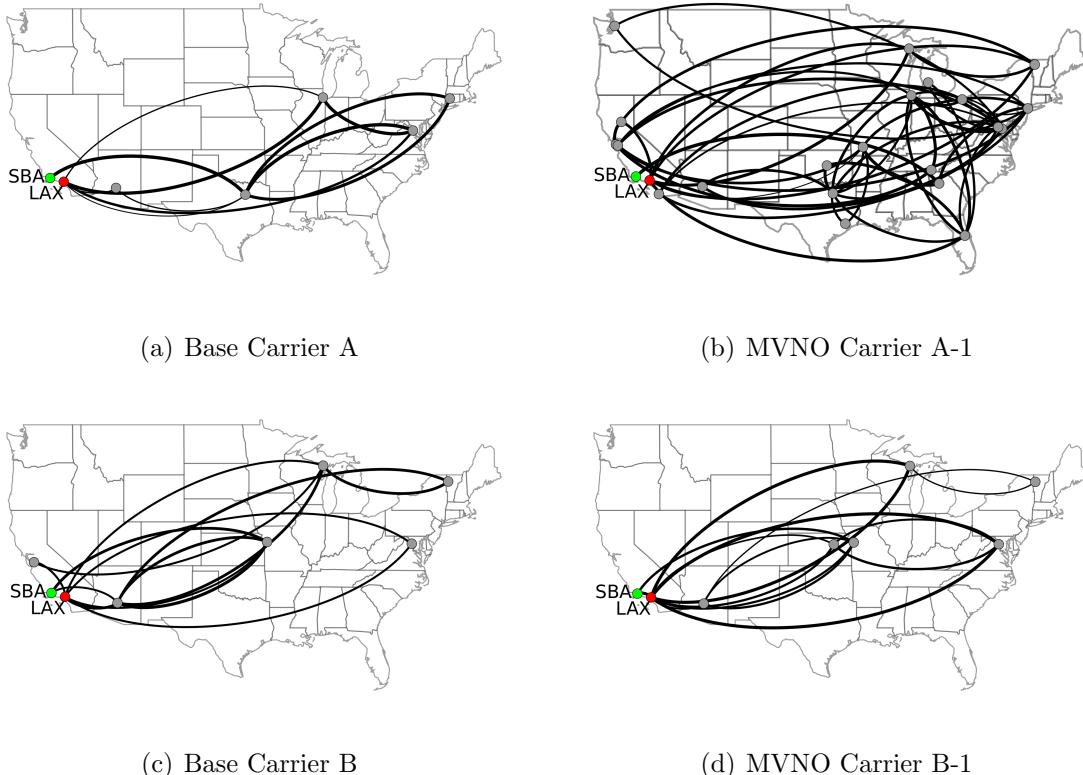


Figure 4.4: Paths taken for each carrier to reach servers located in Los Angeles (LAX) from clients in Santa Barbara (SBA). Arc weight indicates the number of times a particular hop was taken. Clockwise arcs indicate direction of traffic between adjacent hops.

the differences in behavior between MVNOs and their base carriers. For instance, carrier A (Figure 4.4(a)) clearly operates over different, more stable routes compared with its corresponding MVNO (Figure 4.4(b)). This helps explain why there is such a marked difference in performance between the two when connecting to various data centers. Carriers B and B-1 (Figures 4.4(c) and 4.4(d)), on the other hand, are quite similar to one another. This result depicts the difference between light and full MVNOs, where the light MVNO (B-1) routes traffic in the same way as the base MNO, while full MVNOs that implement their own core are influenced by outside factors and differ from their respective base carriers.

A curious finding is that many of the carriers, particularly the MVNOs, contain paths

that pass through Los Angeles only to continue with subsequent hops in distant locations before finally returning to Los Angeles. This seems to indicate the lack of peering between the network that the earlier Los Angeles hop is within and the content provider located in Los Angeles. This behavior is interesting given that [202] found three of the four major US MNOs have peering arrangements with Google servers in Los Angeles. The propagation delay introduced by such scenarios can be considerable, without accounting for additional potential for congestion or queuing delays. Bottlenecks such as these must be removed in order for mobile data to shrink the performance gap between mobile and traditional wired connectivity.

Overall, path visualization gives us an increased understanding of how carriers differ. The full MVNOs that we measure share many locations with their underlying MNO, but their routes are more frenetic. This could be due to different peering arrangements versus the base carrier or simply due to different overall Internet connectivity. We also see that the light MVNO in our study closely resembles its base carrier. Given all that we have observed it seems clear that light MVNOs are, at their foundation, re-branded base carriers.

4.5 Related Work

There has been significant effort towards measuring, characterizing, and improving the performance of cellular network infrastructure with respect to the user experience [173, 134, 202]. Sommers et al. [173] compare the performance of cellular and WiFi networks using a crowdsourced approach for measuring network throughput. Nikrashev et al. [134] measure longitudinal performance from end-devices to uncover the prevalence of middleboxes in mobile networks. Zarifis et al. [202] use end-devices to identify latency caused by inflated routes and the relationships between user performance, Internet

ingress points, and peering agreements. Similar to previous work, we use measurements from the end-user perspective to understand the impacts of network infrastructure on user experience.

Zarinni et al. [203] compare application performance over two major carriers and three MVNOs per carrier. Our work focuses on performance with respect to underlying network layers (e.g. latency and route paths) and considers all four major U.S. carriers and MVNOs operating on top of each.

As cellular networks become the primary mode of Internet connectivity, research efforts have focused on the analysis of the impact of content placement and network configuration on end-user experience [202, 156]. Zarifis et al. [202] find that route inflation leads to increased RTT experienced by end users connecting from locations with limited infrastructure. Rula et al. [156] explores the relationship between cellular DNS infrastructure and the location of selected content replicas, finding that instability of cellular DNS resolvers significantly degrades the experience of mobile users. We find that locations of resolved content servers are not universally attributable to one single factor.

4.6 Discussion and Conclusion

Given the results of our measurement study, what are the overriding lessons?

Round trip times. We observe that round trip times vary significantly between MNOs as well as MVNOs. Additionally, we see that location of destination servers drastically affects RTTs, and resolved server locations do not appear to be logical in that they are often physically distant from the client location. Such behavior could be the result of mobile carrier peering arrangements, DNS infrastructure, and Internet ingress points. Future work should focus on making more efficient network topologies in order to close the performance gap between mobile carriers.

Route paths. We find that MVNOs typically traverse the same autonomous systems as their MNO counterparts in their paths to reach servers. However, we often observe a higher number of hops on the MVNOs. The root cause of such path inflation needs more thorough investigation, as it could be attributable to multiple factors such as: Internet ingress points or middleboxes used for accounting or traffic shaping in the mobile core network. Given our geographic analysis, we believe that full MVNOs, which operate their own core networks, route traffic through seemingly inefficient paths. Perhaps increasing the number of ingress / egress points as well as replicating middlebox functionality across more geographic locations could improve the directness of mobile traffic on such networks.

MNOs vs MVNOs. With the exception of carrier B-1, we observe marked performance differences on MVNO networks compared with their underlying MNO networks. As carrier B-1 is a light MVNO, while the others are all full MVNOs, we can argue that consumers should expect a different user experience when connecting via full MVNOs compared with base carriers. The observed light MVNO leads us to conclude that its use is in essence the same as the base carrier. It remains to be seen whether the same is true for all light MVNOs. We find that full MVNOs tend to share some infrastructure with the MNO, but that they are less predictable in terms of routing paths. Latency differences are also considerable between MNOs and full MVNOs and some variability can be attributed to destination server location. It seems likely that MVNOs may have fewer peering agreements with content providers, evidenced by considerably longer, more circuitous paths taken.

We do not believe that MVNOs, by their nature, are bound to offer inferior performance compared to MNO carriers. There appears to be multiple avenues available to explore for MVNO carriers in order to maximize traffic efficiency. For researchers, this subject deserves more in-depth, longitudinal studies from many locations to fully

understand performance of these networks. For consumers considering which MVNO or plan is the best option, there is currently no clear answer. Additionally, the ‘best’ carrier will likely vary based on what content the user intends to consume on the Internet. The inherent tradeoffs between carriers are worthy of future exploration using real-world user traffic.

Limitations. Our measurement study provides only a limited glimpse into the performance of mobile data networks given a single measurement location and targeting a small set of servers. A longitudinal, in-depth measurement campaign is required to fully understand the tradeoffs between mobile carriers and content delivery networks. Measurements also rely on the efficacy of the tools we use, such as `traceroute`, and the equal treatment of measurement traffic by the carrier core networks. A larger study must include more real world traffic.

Chapter 5

Analysis: Cellular Usage in Côte d'Ivoire

5.1 Introduction

The availability of mobile networks has revolutionized the way people communicate in the developing world. Our first hand experience in rural Macha, Zambia indicates that access to cellular services is of critical importance to residents. While the reasons for adoption of cellphone technology in developing communities are not drastically different than those of the Western world, the benefits for people in these remote communities without infrastructure or other means of telecommunications is much more pronounced. Obtaining information via cell phone, as opposed to in person after travel, saves both critical time and money.

A plethora of applications that improve the wellbeing of people in remote communities, leverage cellular networks. Such applications span from health care [30, 53, 112, 136] and education [152, 8, 139] to agriculture [64, 138, 137] and mobile banking [124]. Multiple successful projects in Africa have originated from observing user behavior in mobile

or social networks. As a result of Facebook traffic analysis, Johnson et al. designed a system to facilitate local content sharing within remote rural communities [105]. Mbiti et al. described a system called mPesa [124] that enables transfer of money in the form of airtime in rural Kenya. The design of this system was inspired by analysis of mobile network usage in Kenya, which indicates that people tend to transfer airtime between one another as a means for payment or financial support. Follow up studies on the adoption of mPesa in Kenya show that theft decreased, as users no longer needed to carry cash.

Such projects are of critical importance to introducing new services and enhancing the wellbeing of people in under-serviced areas. At the same time, special attention should be paid in the design process of these systems to make sure that they meet an actual need in the community. Analysis of large scale datasets generated by the targeted communities naturally facilitates the identification of actual community needs.

We approach a cellular network dataset from Côte d'Ivoire with this end in mind. The dataset provides information on an hourly basis for pairwise antenna communication over a period of five months. The dataset also features information for the personal network of 5,000 randomly selected individuals; these personal networks are called *egocentric social graphs*. Our hope is to identify unique usage patterns based on population density. We correlate cellphone usage patterns with population density and focus on aspects such as inter-antenna distance and call duration to reason about connection strength and locality of interest. Further, we analyze the egocentric social graphs hoping to identify community persistence in an attempt to motivate feasibility of information relays in user-centered cellular communication.

Previous work on geography of mobile communications focuses on traces from European countries [135, 48, 37]. In contrast, we analyze cellular network activity from a predominantly rural sub-Saharan country, where communication patterns could be different than those in Western countries due to individuals economic power, rate of adoption

of cellular services and population sparsity. Other work that employs rural and urban classification of cellular subscriptions focuses on extraction of behavior patterns of individuals living in cities vs. those in rural areas [68]. Instead, we study urban and rural area mobile usage to identify differences in call duration, distance and temporal patterns. Social network analyses using mobile traces focus on implications of network diversity [69], extracting relations [70] and community formation [135]. These works, however, are not concerned with temporal aspects of individuals' communication networks. This chapter makes several contributions:

- we evaluate temporal aspects of cellular communication in Côte d'Ivoire and identify differences with respect to mobile network utilization in the Western world;
- we identify stark differences in cellphone usage patterns in rural and urban areas;
- we find that the communication patterns between antennas in similar population densities are largely different than between antennas in different population densities;
- we design a model based on *persistence graphs* to study temporal persistence of social groups in egocentric graphs; and
- we discover that while there is a weak community persistence in egocentric graphs, there are individuals in an egocentric network that are highly persistent over time.

We start by describing our methodology of correlating population density with cellular network activity. We then describe our model for evaluation of community persistence in egocentric social graphs. We continue with extensive network analysis in section 5.3. In section 5.4 we talk about related work. Based on findings from our analysis we provide discussion of future directions in section 5.5.

5.2 Methodology

We explore the differences in communication patterns between three categories of antennas: Urban, Suburban and Rural. Apart from these typological communication patterns, we also analyze single user communication patterns over time. In particular, we look at community persistence in egocentric social graphs, whereby a subscriber of interest is centered in a graph and the periphery nodes of this graph are other subscribers with whom the central node communicates.

In this section we start by describing the datasets as provided by the telecom. We then give details about our antenna categorization. Finally, we talk about our model for egocentric social graph analysis.

5.2.1 Datasets

Our analysis is based on cellular network traces provided by one of the major mobile operators in Côte d'Ivoire. The datasets were collected over the course of 150 days between December 1, 2011 and April 28, 2012. To assure homogeneity, the data features records only for users who were subscribed with the network for the entire capture period. Incoming and outgoing calls associated with the same session have been combined and counted as a single call. We complement this information with a dataset from AfriPop that provides high resolution population density information for Côte d'Ivoire. A detailed description of each of these datasets follows.

Antenna-to-antenna. This dataset provides information, aggregated on an hourly basis, for the number of calls and call duration between every pair of communicating antennas in the network. Each antenna has assigned location information in the form of latitude and longitude.

Ego dataset. In this dataset the entire capture period is divided into ten equal sub-

periods. The dataset contains the personal communication networks of 5,000 randomly selected subscribers (egos); one network per period per ego. These personal communication networks include up to second degree neighbors of an individual and are called egocentric graphs. An edge between neighbors in these ego-graphs indicates that there was at least one call between these two users; no information for number of calls, call duration or direction is provided. Edges are drawn between (i) the ego and its first order neighbors, (ii) between two first order neighbors or (ii) between first and second order neighbors.

Population density. We use a dataset provided by AfriPop¹ that contains detailed spatial-geographic data for population density in Côte d'Ivoire. Population density information is provided in ESRI Float format and can be extracted with various resolutions, the highest of which is $100\ m^2$.

5.2.2 Antenna classification

We categorize the antennas in the *Antenna-to-Antenna* dataset in three categories based on population density: Urban, Suburban and Rural. To do this, we employ the new European Union typology of “predominantly Rural”, “Intermediate”, and “predominantly Urban” areas. This typology is a modification of the Organisation for Economic Co-operation and Development (OECD) methodology that seeks to minimize distortions caused by large variations in the area of local administrative units[9]. Using the new OECD method, rural local administrative units are defined as areas with a population density below 150 inhabitants per km^2 , urban are above 300 people per km^2 and suburban (or intermediate) are with population density between 150 and 300 people per km^2 .

We utilize the population density information contained in the AfriPop data set and

¹<http://www.afripop.org/>

use Quantum GIS² to project this data as a raster layer. The AfriPop data includes population density information formatted as the number of people per 100 square meters. Since $100\ m^2$ is too high of a resolution with respect to typical cellular antenna coverage, we re-sample this density data at a lower resolution creating a grid of 2 km squares where the population density assigned to each square is the mean density value of the AfriPop data bounded by this square. Each square is then assigned one of the population density categories using the OECD typology. Our grid assigns population density at a resolution suitable for associating antennas with the underlying population statistics.

The majority of land area in Côte d'Ivoire – 96.98% – is classified as Rural; 1.22% is Urban and 1.79% is Suburban. At the same time, according to the World Bank [16], urban areas account for 48.8% of the total population in the country, while the remaining 51.2% is classified as rural. Thus, almost half of the population in the country is concentrated in the few major cities, constituting a small geographic area.

5.2.3 Egocentric graphs analysis

We examine the egocentric social graphs dataset to determine persistence of social groups for each ego over time. We also analyze the likelihood that one or few nodes (users with which an ego communicates) persist over time in an egocentric graph. We hope to see persistence in both communities as well as individual subscribers. We hypothesize that such continuously-present entities can be used as information relays to strengthen information distribution amongst community members. Our analysis indicates that while community persistence is relatively low, persistent nodes indeed exist.

In order to extract the separate social groups of an ego, we remove the ego node from each egocentric social graph (Figure 5.1) and analyze the connected components that remain. Each connected component corresponds to one social group. Note that in the

²<http://www.qgis.org/>

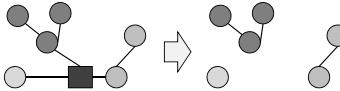


Figure 5.1: The effect of removing the ego (depicted with a square) from the egocentric social graph.

text we use the terms *connected component* and *social group* interchangeably.

After extracting the connected components we evaluate the persistence of these components over time. A connected component is 100% persistent over two consecutive periods if the nodes in this connected component are identical in the two periods. For this evaluation we define a *persistence graph* $G = (N, E, W)$ with N nodes, E edges and W weights assigned to each edge. Each node in G is a connected component labeled with the period to which it belongs. An edge exists between two connected components if they overlap in consecutive periods. The weight assigned to each edge is the *Jaccard similarity*, J , between the connected components [178]. For two sets A and B , the Jaccard similarity J can be calculated as follows:

$$J = \frac{A \cap B}{A \cup B} \quad (5.1)$$

The Jaccard similarity ranges between 0 and 1, where 0 indicates no overlap and 1 indicates full overlap.

Figure 5.2 presents an example of building the persistence graph for a single ego over three consecutive periods. The left-hand side of the picture presents the set of neighbors in each of the three periods. The social groups comprised by these neighbors are color-coded. The right-hand side of the picture presents the resulting persistence graph. Each node corresponds to a connected component (CC) in a given period. In the figure nodes' labels are of the format CCID-PeriodID. Edges exist only between connected components that overlap fully or partially in consecutive periods. There is no edge between connected

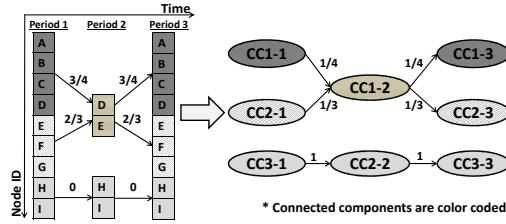


Figure 5.2: Building a persistence graph.

components that persist over non-consecutive periods (e.g. there is no edge between node “CC1-1” and node “CC1-3”).

Our persistence analysis is based on the described persistence graphs and consists of two parts. First, we analyze the in- and out-degree distribution of the nodes in the persistence graph. We note that if the social groups of an ego persist over time, all the nodes in the persistence graph should have in- and out-degrees of either 0 if the node belongs to the first or last period, or 1 if the node is in the intermediate periods. In cases where social groups do not persist, nodes can have a degree of 0 if the corresponding social group does not re-appear in following periods. Nodes can also have in- and out-degrees larger than 1 if social groups merge or split in consecutive periods.

We attempt to quantify the level to which social groups overlap by considering the weights of the edges in the persistence graphs. As detailed earlier, edges are drawn between nodes that overlap fully or partially in consecutive time periods. The weights assigned to these edges are the Jaccard similarity between the nodes connected by these edges. For each transition between period t and period $t+1$ we find the normalized Jaccard similarity $\hat{JS}^{(t,t+1)}$ between these periods: that is the sum of edge weights $W_i^{(t,t+1)}$ divided by the number of edges $|E^{(t,t+1)}|$ between the two periods.

$$\hat{JS}^{(t,t+1)} = \frac{\sum_{i=1}^{|E^{(t,t+1)}|} W_i^{(t,t+1)}}{|E^{(t,t+1)}|} \quad (5.2)$$

We then find the average Jaccard similarity for the entire persistence graph by summing the normalized Jaccard similarities and dividing this sum by the number of period transitions K .

$$\bar{JS} = \frac{\sum_{j=1}^K \hat{JS}_j^{(t,t+1)}}{K} \quad (5.3)$$

Informally, the higher the average Jaccard similarity, the more persistent the social graphs of an ego are over time.

We present our results for social group persistence in Section 5.3.7.

5.3 Network Analysis

We begin our analysis by mapping antennas to population density and discussing general trends in antenna utilization. We then investigate temporal trends in mobile communication in general and across areas of different population density types. We expect to observe regular temporal trends along weekly and monthly intervals, with Rural areas having distinctive patterns from those of Urban areas. We also explore trends related to population density; we expect to see differences in call duration and call frequency based on the population density of the sender's geographical area. Next, we analyze how the inter-antenna distance impacts the call duration and call frequency. Finally, we examine patterns in social groups, hoping to observe persistence in social groups over time.

5.3.1 Antenna activity

We study patterns of mobile communication in Côte d'Ivoire by associating antennas with their geographic location and population density. The resultant mapping of antennas to location is shown in Figure 5.3. The figure presents average population density per

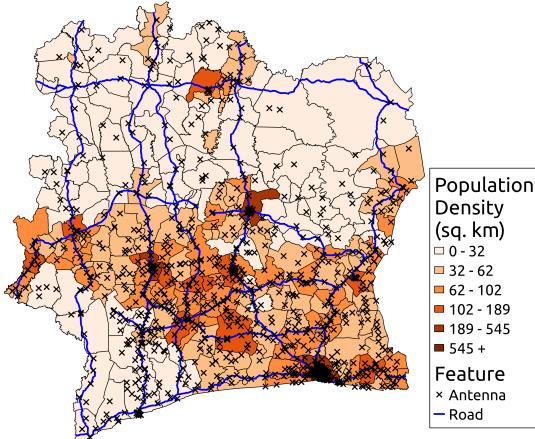


Figure 5.3: Cellular Antennas in Côte d'Ivoire.

Table 5.1: Antenna Density Classifications

Classification	Antenna Count	Source Calls
Rural	528	146,481,488
Suburban	90	21,529,115
Urban	598	331,630,147
Unknown	15	65,393,926

sub-prefecture with overlaid antennas. It is evident that antennas are densely clustered in urban locations while more sparsely located in predominantly rural regions. We also find that high activity antennas are often located along major transportation corridors.

We explore the relationship between the population density of a sending antenna and the average number of outbound calls associated with the antenna. Because of the predominant use of “Calling Party Pays” (CPP) policy in sub-Saharan Africa, we focus on the number of outbound calls rather than incoming calls [129, 66]. Due to the CPP policy, we anticipate a larger mean number of outbound calls from antennas in high population density areas, which coincide with Côte d’Ivoire’s financial district and center of commerce. In Table 5.1 we show the number of antennas that fall into each of the classifications as well as the total number of calls originated from each antenna type. As expected, while the number of Rural and Urban antennas is almost the same, the amount

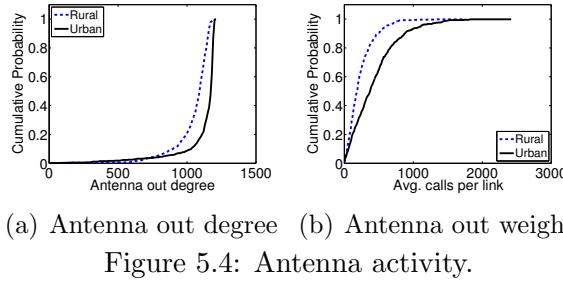


Figure 5.4: Antenna activity.

of calls originated by Urban antennas is more than twice as large as those originated by Rural antennas. Very few antennas and calls are classified as Suburban. This is consistent with the fact that the subset of Suburban population is very small in comparison with Urban and Rural. In the rest of this chapter we will focus our evaluation on activity associated with Rural and Urban antennas.

We finalize our antenna activity evaluation by examining the outbound communication trends per antenna pair in Urban and Rural areas. Again, we focus on call originators. In particular, we evaluate over the entire period, (i) the outbound degree of each antenna, meaning the number of connections each antenna establishes with other antennas, and (ii) the sum outbound weight of each antenna expressed as the sum of the number of calls originated on each outbound link. We present our results in Figure 5.4. Figure 5.4(a) plots the CDF of out degree of antennas in Urban and Rural areas. As we can see, the mean outbound degree for Rural antennas is lower than that for Urban. This indicates that Rural antennas tend to communicate with fewer antennas than Urban. We then evaluate the average strength of antenna to antenna links by examining the average weight of the outgoing links associated with a given antenna; link weights are assigned according to the number of originated calls. Figure 5.4(b) plots a CDF of average link weight for Rural and Urban antennas. The average weight for Rural is twice as small as for Urban, which implies that on average more calls originate from Urban than from Rural antennas.

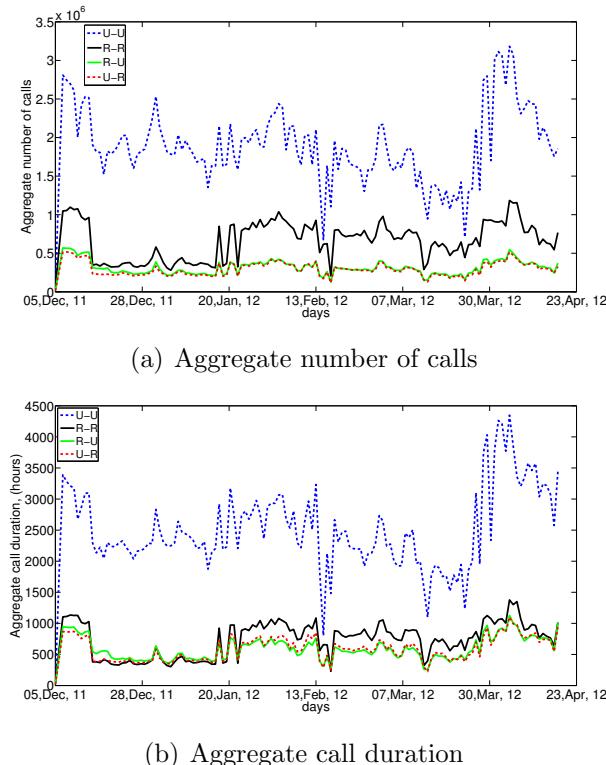


Figure 5.5: Mobile network activity over time.

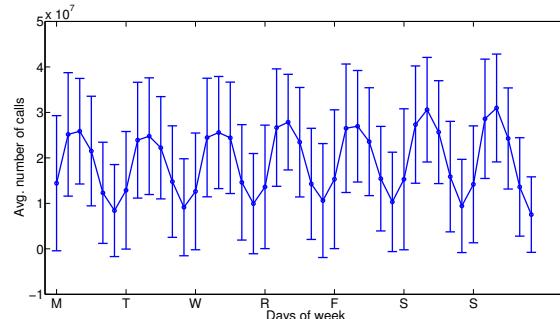


Figure 5.6: Weekly patterns.

5.3.2 Temporal usage patterns

We evaluate the cellular network activity patterns over the entire capture period.

We categorize calls by originating and receiving antenna type based on the antenna categorization from section 5.2.2. In Figure 5.5 we plot aggregate number of calls and call duration per day for four categories of antenna pairs. In the legend, U and R stand for Urban and Rural, respectively. As the figure shows, there is no distinctive call pattern

on a weekly or monthly basis; instead, subscriber activity seems to be widely correlated with events in the country. We hypothesize that the peak near the beginning of the period coincide with the weeks before and after the parliamentary elections on December 11th, 2011, while the second peak is most likely traffic related to New Years. The increased utilization from the end of March through April is likely associated with the military coup in Mali and the ECOWAS³ summit that took place in Abijan, Côte d'Ivoire. Such irregular usage pattern is very different than what had been observed in cellular network traces from the Western world [111].

In Figure 5.5 we also illustrate that the calling patterns in Rural areas differ from these in Urban areas. As the figure shows, calling patterns for all four categories follow similar trends, where the number of calls and the aggregate call duration between Urban antennas is about three times higher than between Rural antennas. We note that while the number of Rural to Rural calls is larger than the number of Rural to Urban and Urban to Rural, the aggregate call duration for these three categories is the same (Figure 5.5(b)). This result indicates that while calls between Rural residents occur more often, they are shorter in comparison to calls between Urban and Rural residents.

Next we study weekly communication patterns. To extract weekly behavior, we average the number of calls and call duration over the entire capture period for each week day. Figure 5.6 presents our results. Each point on the plot presents an average over four hours over all occurrences of each day of the week (that is the first data point from the graph presents the average number of calls for the hours from Midnight to 4 AM for all Mondays in the capture period). The figure clearly presents diurnal pattern of network activity with slight increase over the weekend. However, the standard deviation of this graph is very high, indicating that the network activity varies dramatically over the observed period.

³<http://www.ecowas.int/>

5.3.3 Antenna distance

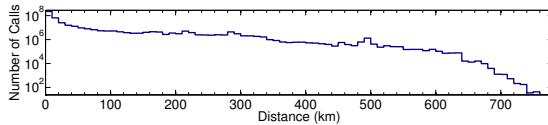
We investigate the relationship between call distance and the average duration of calls. We calculate the distance in kilometers between all pairs of antennas with known geographic location using the Haversine formula [153] with mean Earth radius of 6,372.80 km. We group connection distances into the nearest 10 km in order to calculate aggregate statistics for each group.

First we examine trends in frequency of communication as a function of distance between calling parties, shown in Figure 5.7(a). The figure plots number of calls in log-scale on the y-axis and distance on the x-axis. We see a long tail distribution of number of calls over distance, whereby antennas in close proximity tend to have many more calls between one another in comparison with antennas that are further apart.

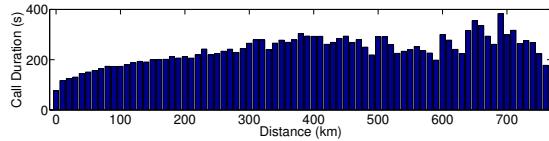
Next we evaluate the impact of distance between antennas on the mean call duration. We find the mean call duration by grouping all antennas within a certain distance from one another and dividing the total call duration by the number of calls. The impact of distance between source and destination antenna on mean call duration is shown in Figure 5.7(b). While the number of calls in close proximity is much higher, the average call duration is lower when communicating antennas are nearby, and increases as connection distance grows. We hypothesize that the reason for such increase in call duration is that calling parties who are further apart have fewer opportunities for in-person interactions, thus they tend to talk longer over the phone. Lastly, note that with relatively few call records for distances greater than 500 km, more noise is introduced into the graph.

5.3.4 Call typology classifications

We investigate the potential correlation between population density and calling patterns by associating antennas with the corresponding local population density. This pro-



(a) Number of calls vs. antenna distance



(b) Mean call duration vs. antenna distance

Figure 5.7: Communication patterns as a function of the distance between antennas.

cess yields antennas denoted as Rural, Suburban, Urban, or Unknown (for the antennas which have no geographic location). We process the *Antenna-to-Antenna* set to classify call records by each typology source and destination pair in order to investigate potential communication patterns. In this analysis we do not consider records for antennas with no geographic data or records without valid antenna IDs.

We start by analyzing the distribution of antenna pairs in four categories: U-U, R-R, U-R and Other. Note that in this classification we do not consider directionality. The Other category contains antenna pairs featuring Suburban antennas as well as such that are unclassified. As seen in Figure 5.8, the majority of connections are classified as U-U connections. This is followed by 20% of connections classified as R-R. Mixed links of R-U account for 18% of the total. The relatively small fraction of calls between mixed antennas can be explained with the high locality of calls observed in Figure 5.7(a), according to which the majority of calls occurs between antennas which are a few tens of kilometers apart.

Next, we search for differences in mean call duration across the connection classifications and show results in Figure 5.9(a). We find that the two call classifications with the longest mean call duration are Urban to Rural and Rural to Urban. An observable

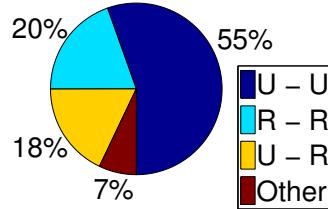


Figure 5.8: Classification of communication between antenna pairs.

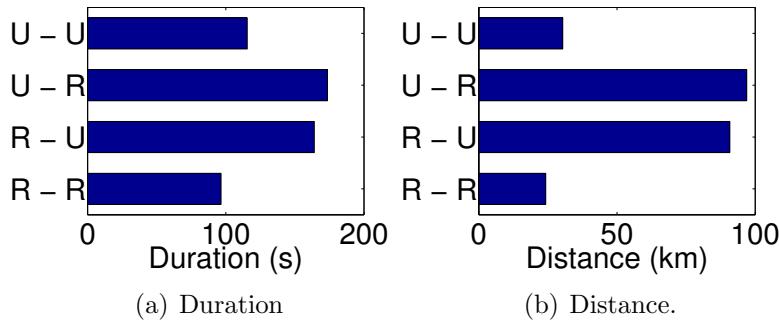


Figure 5.9: Mean call duration and mean call distance for connections of different types.

phenomenon is that calls confined to the same source and destination density type are noticeably shorter on average compared to calls between mixed types. Given our prior finding of the relationship between call distance and average duration, we posit that the majority of calls that do not cross classification boundaries are confined to a smaller geographic region. For instance, we believe Urban to Urban calls are more likely to be sourced from and destined for the same urban area. Lastly, an interesting observation is that calls originating from Urban antennas generally have a longer duration for any destination type. This is likely due to the common policy of “Calling Party Pays” and higher buying power of individuals who reside in urban areas.

This trend leads us to look at the average distance between connecting antennas associated with each connection density classification type, shown in Figure 5.9(b). The longest average distance between connecting antennas occurs in the Rural to Urban and Urban to Rural cases. The shortest average distance occurs between similar source/destination connections. In the case of Urban to Urban this is likely due to the

fact that there are only two Urban areas, which are possibly closer to one another in comparison with an arbitrary Urban to Rural case. The case for Rural to Rural, however, is more interesting. While the vast majority of the country is Rural, thus statistically Rural to Rural communications can cover long distances, the Rural to Rural patterns presented in Figure 5.9(b) indicate that people in Rural areas who call one another tend to be in close proximity. This indicates high locality of interest in Rural to Rural cellular communications in Côte d'Ivoire.

To further explore this locality of interest, we investigate associating call patterns and population density for calls that have the same source and destination antenna ID. We find that in 57% of all Rural to Rural calls are sourced from and destined to the same antenna. We posit that this is due to fewer available antennas in predominantly rural areas. Furthermore, the coverage area of a single antenna in rural settings is typically larger (up to 35 km), which means that a higher proportion of local users are associated with the same antenna. Nevertheless, this high percentage of same antenna calls confirms that cellular communications in Rural areas are very local. Interestingly, Urban connections sourced from and destined to the same antenna represent 23% of all Urban to Urban calls. We believe that the higher density and smaller cell range of Urban antennas provides more diverse antenna association possibilities for users.

5.3.5 Transportation infrastructure

A close analysis of Figure 5.3 shows that a large fraction of antennas in areas of low population density are situated in close proximity to major transportation corridors in Côte d'Ivoire. In our current antenna typology, these antennas close to roads are categorized as Rural antennas. However, we expect that the usage patterns of antennas associated with transportation corridors will differ significantly than those located

in Rural residential areas. Thus, we divide the Rural category in two subcategories: Transportation and Rural-residential, where an antenna is labeled Transportation if it is within 5 km of a highway. As a result we find that 51.7% of the antennas the were originally classified as Rural are associated with road infrastructure. Of note is that a Transportation antenna can be used by both travelers as well as Rural residents, which is why all results except for the ones presented in this section, feature both Transportation and Rural-residential antennas in the Rural antenna type.

Based on our new classification, we evaluate communication patterns in terms of three indicators: (i) call duration, (ii) number of calls and (iii) percentage of calls where both the originator and the terminator are associated with the same antenna. We find that the average call duration for Rural-residential to Rural-residential calls slightly decreased from 95.5 seconds to 84.8 seconds. More drastic, however, was the change in the same antenna calls and the average distance between calling parties. While previously the same antenna Rural calls were 57%, in the case of Rural-residential communication the same antenna calls increases to 70%. Accordingly, the average distance between Rural to Rural calls decreased from 24.2 km to 13.4 km after the Transportation antennas were removed from the Rural typology. This result once again confirms the strong locality of interest in cellular communications in rural Côte d'Ivoire.

5.3.6 Community discovery

We examine the antenna connectivity data as a weighted graph in order to seek out underlying community structures using the modularity algorithm proposed in [38]. Communities are defined as subsets of highly connected nodes in a graph [133]. Nodes in our graph are represented by sub-prefectures with edges indicating connectivity between antennas lying within the two sub-prefectures. Edge weights are represented by the

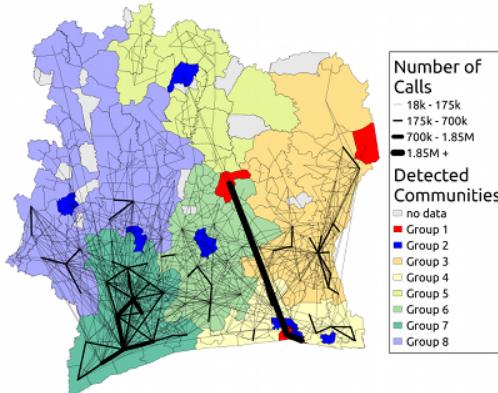


Figure 5.10: Weighted connectivity map with detected communities.

total number of calls between a given sub-prefecture pair. Figure 5.10 shows detected communities using a resolution value of 1.15 for the modularity algorithm. The resolution value is data-dependent and determines the number of detected communities for a given data set. We find 1.15 to be an optimal value for our data set in order to avoid too many or too few communities. The figure edges are weighted based on the number of calls with only the top 2% of edges displayed for readability. Sub-prefectures are shaded to indicate community membership. Sub-prefectures labelled as “no data” are those in which an antenna does not exist, therefore they are not considered a part of the graph.

After applying the community detection algorithm we find that with a few exceptions, communities are largely based on close proximity within geographic area. As the figure shows, “Group 1” includes two of the largest cities (Abidjan and Bouaké) despite the fact that they are geographically non-contiguous. We believe the easternmost sub-prefecture included in “Group 1” is an anomaly given that only two antennas exist in the region. Perhaps the sub-prefecture includes a facility in which workers often contact major cities. Most interestingly, “Group 2” includes multiple geographically dispersed medium-sized cities as well as the area surrounding Abidjan. “Groups 1” and “2” indicate that as population density increases, geographic distance becomes less of a factor in relative

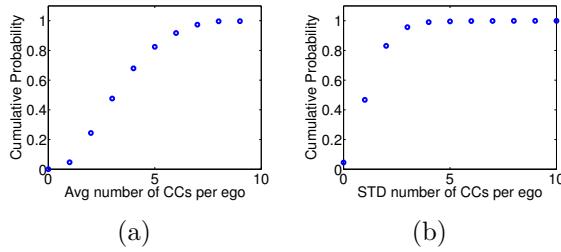


Figure 5.11: (a) The number of connected components (CCs) per ego and (b) the standard deviation of the number of connected components per ego over the observed period.

“connectedness” in the network. The remaining groups largely consist of rural regions that are geographically similar.

The results support our observation of a distinct divide between rural and urban regions. Rural communication patterns are largely regional and tightly connected with the surrounding area. On the other hand, urban areas are tightly connected with other urban areas and not necessarily with nearby rural areas. As predominantly disconnected rural areas are often the target of connectivity research, this insight can be used to inform the design of any solution. While global connectivity is the ultimate goal, local and regional telecommunications connectivity can clearly benefit users in a localized rural area.

5.3.7 Egocentric graphs

Now that we understand regional connectivity patterns, we take a deeper look at egocentric social graphs provided in the *Ego* dataset. Our analysis focuses on persistence of social groups with which individual egos communicate. We regard this analysis as preliminary work on identifying persistent neighbors within one’s social network who can serve as reliable information relays.

First we provide high level analysis of the average number of social groups with which

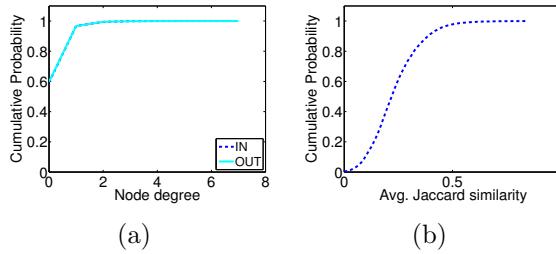


Figure 5.12: (a) The in- and out-degree of nodes in all persistence graphs and (b) the average Jaccard similarity for each persistence graph.

each ego communicates over the entire capture period from December 2011 to April 2012. For this analysis we sum the number of connected components that appear in each two-week period and divide this sum by the number of capture periods. Figure 5.11(a) plots a CDF of the average number of connected components for each ego. While the average number of components across egos spans from 1 to 10, the majority of egos – 68%, have between 2 and 5 connected components on average. Further, we examine how the number of connected components deviates for each ego. Figure 5.11(b) plots a CDF of the standard deviation of the number of connected components per ego over the observed period. Almost half of the egos (47%) have standard deviation of less than 1, while 96% of all the egos have standard deviation of less than 4. This indicates that the number of connected components in an egocentric graph remains relatively constant over time.

Next we analyze the persistence of these social groups over time. First, we look at the in- and out-degree distribution of nodes in the persistence graphs. As detailed in Section 5.2.3, a node in period t has in- or out-degree of 0 if it belongs to the first or last observed period or if it does not overlap with any node from the preceding ($t - 1$) or the following ($t + 1$) period. Nodes have in- and out-degree of exactly 1 if they persist over time, and degree larger than 1 if they split or merge over consecutive periods.

We calculate that out of all the nodes in all persistence graphs, 9.49% belong to the first period (i.e. have in-degree of 0) and 8.93% belong to the last period (i.e. have out-

degree of 0). At the same time Figure 5.12(a) indicates that in nearly 60% of the cases nodes have in- or out-degree of 0. This means that about 50% of all the social groups that we observe, and which were not associated with the first or last period, did not occur in the preceding and following periods. 40% of the nodes have in- or out-degree of 1, indicating that 40% of the social groups persisted in consecutive periods. Only about 3% of the cases have in- or out-degree larger than 1; social groups rarely split or merge over consecutive periods.

This result indicates an important quality of the observed egocentric social graphs: there are two distinctive types of social groups with which an ego communicates – (i) those that likely occur only once (in- and out-degree is 0), and (ii) those that likely persist over time and strictly correspond to one social group from the preceding and one social group from the following period. The former group can be associated with one-time calls, for example calling to schedule a doctors appointment, while the latter can be associated with calls recurring over time, such as these between relatives and friends who stay in touch.

We continue our evaluation of social group persistence by analyzing the weight of edges (representing the similarity) of social groups in consecutive periods. We leverage the average Jaccard similarity metric as defined in Section 5.2.3; the closer this similarity is to 1, the larger the overlap between social groups in consecutive periods. Figure 5.12(b) plots a CDF of the average Jaccard similarity for the 5,000 ego-centric graphs. The median of this CDF is only 0.22, which means that on average the overlap of social groups over time is relatively small – about 22%.

Finally, we evaluate the frequency of occurrence of the neighbor that appears most often in the social network of an ego. For this evaluation we count in how many of the ten observed periods each neighbor appears. We then sort the neighbors in decreasing order

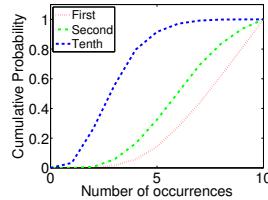


Figure 5.13: Number of occurrences of the first, second and tenth most frequent neighbor.

of appearance frequency. We compare the first, second and tenth most frequent neighbors to determine whether there are groups of neighbors that appear more often and what is a typical size of such groups.

Figure 5.13 presents our results. The median value for the first top neighbor is 8, while for the second and the tenth top neighbor it decreases to 6 and 3, respectively. This means that in 50% of the cases, the most frequently occurring neighbor exists in 8 out of the 10 observed periods. These results indicate high persistence of at least one neighbor in the social graph. At the same time, a group of two most persistent neighbors would appear ten times in only 6.8% of the cases, which indicates that a group of most persistent neighbors would typically have very few members.

5.4 Related work

Analysis of mobile network traces provides a unique opportunity for large-scale verification of socio-economical models that were previously derived and studied on much smaller scales. The correlation of mobile network traces with other data provides for realistic analysis of problems spanning from epidemiology [194] to individuals' economic prosperity [69], relational dynamics [70] and gender divides [39]. Previous work focuses on epidemiology informed by human mobility patterns in cellular networks. For example in malaria tracking, mobile network traces with subscriber location information are correlated with malaria prevalence information. This correlation enables quantification of

the impact of human mobility on the spread of malaria: see [194] and references therein.

Previous research related to our work can be divided in three categories: (i) geography of mobile communications, (ii) mobile communication patterns informed by population density and (iii) social interactions analysis.

Geography of mobile communications. Previous work explores geographical aspects of human social interactions in mobile networks. Onnela et al. analyze one month of mobile traces from a European country to discover relations between community formation and geographical distance between community members [135]. Other work based on European traces [48] analyzes patterns in pairwise communication to determine whether call duration and frequency change based on the physical distances and opportunities for in-person communication between individuals. Blondel et al. apply a grouping algorithm on a cellular network trace from Belgium to extract the geography of mobile communication in the country [37]. The presented results indicate that grouping of mobile call interactions is both region and language dependent. In contrast our work focuses on mobile traces from sub-Saharan Africa, where communication patterns could be different than those in Western countries due to specifics related to population sparsity, adoption of cellular services and individuals purchasing power.

Population density. Previous work on mobile network analysis informed by population density focuses on studying behavioral patterns in rural and urban areas [47, 68, 77]. Eagle et al. employ mobile data to study differences in human behavior related to personal network topologies and travel patterns of individuals living in cities as opposed to those living in rural areas [68]. In contrast, we study urban and rural mobile usage to identify differences in call distance, duration and temporal usage. Our findings make a strong case for locality of interest in rural cellular communications in Côte d'Ivoire.

Social analysis. Social network analyses based on cellular traces focus on implications of network diversity [69], extracting relations [70] and community forma-

tion [135]. Studies demonstrate that diversity of one's mobile social network influences socio-economical prosperity [69]. Other work extracts biases in self-reported friendships by comparing characteristics of self-reported relationships with those extracted from cellular traces [70]. These studies, however, are not concerned with variability of social networks over time. In contrast our analysis explores temporal trends of cellular communication in individual subscribers' communication networks and provides insights on community persistence in egocentric social graphs.

Dynamic graph mining. In the area of dynamic graph mining, research has focused on evolutionary community detection [109], conserved relational states [24] and high-scoring dynamic subgraphs [40]. Bogdanov et al. propose a method to identify the highest-scoring temporal subgraph (e.g. most congested road segment) in a dynamic network [40]. Our analysis is different, as we seek to summarize the persistence in different egocentric networks without observing the whole graph at a time. Other work mines relational patterns in a dynamic network [24] in order to detect maximal evolution paths in time-evolving networks. While this work utilizes a model for tracking similarity that is similar to our persistence graphs, the proposed scheme is only concerned with full overlap of graph entities over time. In contrast, our method captures partial overlaps and allows for fine-grained analysis of community persistence.

5.5 Discussion and conclusion

Our analysis of cellphone traces from Côte d'Ivoire indicates that temporal usage patterns differ than those in European countries [111]: while studies in Europe show persistent diurnal and weekly patterns over time, such patterns are not typical to Côte d'Ivoire. The reason for such difference could be related to several factors. First of all, while in Europe access to cellular communication is a regular part of daily life,

cellphone services in developing areas are still being deployed at large and the number of subscriptions is actively growing. In this sense the cellphone user population and usage habits are still being established, which could lead to highly-dynamic utilization trends. Secondly, because urban areas, which are the center of governance and commerce, are the main contributor to cellphone activity in Côte d'Ivoire, the cellular network utilization is likely largely dependent on socio-political or cultural events in the country.

Urban areas are the main contributor to cellular activity in Côte d'Ivoire: 55% of all calls occur within urban areas and another 9% of the activity is initiated by urban areas. We speculate that this could be due to several reasons. First, people in urban areas have more coverage than those in rural areas: while only 1.22% of the territory of Côte d'Ivoire is urban, 48.8% of the population and half of the cellular antennas reside in urban areas. Furthermore, it is possible that people in urban areas have higher buying power than those in rural areas. In line with this assumption and the CPP policy in sub-Saharan Africa, we find that Urban-to-Urban calls and mixed calls initiated by urban residents are longer on average than those originated by rural residents. Of note is that the paying ability is not the only factor that influences cellular usage; factors such as adoption, needs and communication habits can also influence usage patterns. Thus, further analysis that involves fine-grained per-subscriber call information and air time purchasing information would help clarify calling trends.

We show that urban and rural usage patterns in Côte d'Ivoire differ significantly. For instance, while the total number of rural and urban antennas is comparable, rural antennas tend to connect with fewer antennas than do urban antennas. Furthermore, a rural antenna originates fewer calls on average than an urban antenna. At the same time the locality of interest in rural communication is much higher than in urban communications: 70% of the Rural-residential calls happen in the vicinity of the same antenna. This means that people in rural residential areas tend to call other rural residents in close physical

proximity. While global connectivity is the ultimate goal, this finding makes a strong case for the feasibility of local cellular communication solutions such as the ones discussed in [29, 86, 206]. The actual applicability of such solutions, however, depends not only on availability of low-cost power-saving technology but also on licensing. The current licensing schemes are extremely conservative, costly and oblivious to return of investment. We argue that these licensing aspects are the main factor preventing entrepreneurs from deployment in rural remote areas.

Lastly, we evaluate the persistence of social groups in egocentric social graphs. With this analysis we provide preliminary insights to inform future work on extraction of information relays from egocentric cellular social networks. Such information relays can be used to strengthen information channels between educators and students or health workers and patients. Our first-hand experience in rural health care indicates that improvement of information channels is indeed needed since health care services such as immunizations are often available; however, it is difficult to bring information about availability to the interested patients. Analyzing 5,000 random users from Côte d'Ivoire, we find that on average an egocentric network has four social groups; this number is stable over time. We also find that 50% of the observed social groups did not occur in the corresponding preceding and following periods. At the same time, less than 1% of the communities split or merge over time; thus communities that do persist tend to be independent of one another. This means that multiple information relays can be selected from each independent community, to increase the chance that information will flow to the ego. Finally, we observe that on average there is a 22% overlap of social groups over time. Persistence of a subscriber in one's social group likely means that there is a stronger personal connection between the ego and the corresponding subscriber. Thus, such subscriber persistence needs to be taken into account in the process of relay selection. In order to devise models for information relay extraction, however, more detailed

information is needed that contains not only location but also frequency and duration of interaction. Availability of such information will enable true extraction of individuals that are strongly connected to an ego and can serve as reliable information relays.

Chapter 6

Analysis: Internet Performance and Usage in Rural Zambia

6.1 Introduction

Access to the Internet is critical for improving the wealth of nations and promoting freedom. Bright examples of advancements facilitated by Internet access span from democratic changes [26], to government [132], e-learning [171] and health care [74]. Broadband Internet access, however, is still largely unavailable in developing countries with only 6% of the population having broadband connectivity [99], the majority of which is in urban areas.

Recent efforts to bring connectivity to rural areas of the developing world utilize asymmetric satellite or other low-bandwidth wireless links [123, 176]. At the same time the bandwidth demand of online applications is increasing; for example the average web-page size has grown 110 times since 1995 [110]. As a result, residents of developing rural regions access the web with inadequate connectivity for the bandwidth requirements of modern content. These opposing trends in content growth and limited capacity render

Internet access frustrating or even unusable [67, 106] in many developing areas.

Previous work on traffic analysis shows a “strong feedback loop between network performance and user behavior” [106], whereby residents in bandwidth-constrained environments tend to focus more on bandwidth-light applications such as web-browsing, as opposed to those in bandwidth-rich environments which enable multimedia streaming, content upload, and real-time user interaction. In the face of limited bandwidth, the failure rate of uploads is high [105], discouraging rural residents from contributing to the Internet content and resulting in consumption of largely Western content [191]. Thus, while recognizing the potential benefits of the Internet, residents of developing regions express concerns that the flood of Western culture, coupled with decreased ability to document and transfer their own traditions, threatens the existence of local cultures [190].

The focus of our work in this chapter is in Africa, where the increased fiber-optic capacity [6], coupled with higher-bandwidth, lower-latency technologies such as terrestrial microwave wireless gives hope for improved Internet access in remote areas. In this chapter we study the implications of an Internet access upgrade from a satellite to a microwave terrestrial link on the performance and Internet usage in the rural community of Macha, Zambia, as described in 2.1. To the best of our knowledge, this is the first real-world comparative study of pre- and post-upgrade Internet usage and performance. As such, our dataset offers a unique opportunity to study the change in user behavior and Internet usage following an eight-fold increase in access bandwidth. We evaluate a total of three months of usage: one month before the upgrade, one month after the upgrade and one month three months later. Our results show that while usage did not change immediately, application performance improved. However, as time progressed subscribers began to change their Internet usage behavior, which ultimately resulted in network performance degradation and subsequent deterioration of user experience. The Internet access upgrade broadened users’ abilities to access con-

Table 6.1: General TCP statistics averaged over each time period.

	Total GB	Packets (x10 ⁶)	Control packets (%)	Avg. Window (kB)	Avg. RTT (s)	Retrans- missions (%)
Pre	123	373	56.59	38	0.1436	1.12
Post	163	338	47.69	52	0.1085	1.09
LT	210	432	49.72	62	0.3190	1.16

tent, use online applications, and express themselves on the Internet. At the same time our results make a strong case that one should not assume that advanced technologies and higher access speed grant better experience and increased adoption of the Internet in rural communities; rather one should carefully consider the evolution of usage and performance in order to assess the actual impact and adoption of Internet technologies.

6.2 Network Analysis

We evaluate the network performance and usage for three months. We select one month immediately before (which we call Pre-upgrade) and one month immediately after the upgrade (Post-upgrade) to measure the short term impact on the network usage and performance. We also evaluate one month of traffic approximately three months Post-upgrade to determine whether performance changed as time progressed. We call this time period Long-term.

We start by describing our traffic collection methodology as well as our approach to calculating evaluation metrics. We then continue with detailed results from our network analysis. We first focus on overall network performance analysis, which indicates that the majority of traffic traversing the network is TCP (93%). We, thus, focus our analysis on TCP performance following the increased bandwidth. We describe trends in uplink and downlink performance of TCP flows, and we identify the most popular applications

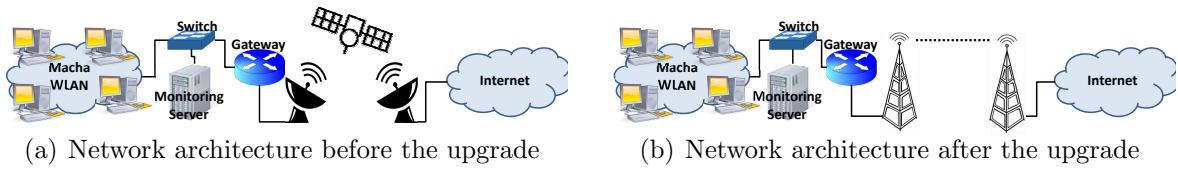


Figure 6.1: Network architecture and traffic monitoring.

based on TCP port number. We then assess the success and failure rate of TCP flows. We conclude our TCP analysis by outlining performance trends in Windows and Linux machines. We then switch to evaluation of network usage focusing on popular URIs. We conclude by analyzing the “worldliness” of network flows initiated in Macha in an effort to determine whether Machans started using more global services once they had better Internet access.

Pre-upgrade the network was typically saturated, resulting in high round trip time (RTT), congestion, and aborted sessions. Post-upgrade, we saw a decrease in the number of retransmissions and RTT due to improved network performance and movement away from the saturation point. By three months after the upgrade, the traffic had increased once again to saturation. Our analysis shows a difference in network performance and utilization Post-upgrade and Long-term: while Post-upgrade user behavior did not change, automatic programs, such as software updates, were suddenly able to complete, which resulted in an increase in traffic demand. In Long-term, subscribers utilized the faster Internet access for more bandwidth-hungry applications such as video streaming. Once the saturation point was reached in Long-term, network performance deteriorated, but was still better able to support bandwidth-intensive applications than Pre-upgrade. We describe these network usage and performance patterns in detail in the following sections.

6.2.1 Methodology

We capture traffic at the Internet gateway in Macha. As shown in Fig. 6.1, we connect a monitoring server to the switch that bridges the Internet gateway and Macha’s WLAN. We configure a mirror port at that switch, which allows us to capture all the traffic traversing the WLAN. With user consent, we capture packet headers and store traces on the monitoring server. During our last trip to Macha in Summer 2012, we offloaded the collected traces to an external hard drive and brought them to our research facility for offline analysis.

We now describe our methodology for extracting metrics from the collected network traces. In our evaluation we use metrics such as TCP window size, RTT, TTL and re-transmissions. We extract these metrics by running `tshark` in an offline mode on the collected traces. For our flow analysis we developed a tool that reassembles unidirectional flows from a list of packets based on packet signature (source IP, source PORT, destination IP, destination PORT, timestamp). In the process of flow reassembly we count the number of packets and bytes associated with this flow and calculate its duration. We calculate the packet Inter-Arrival time (IAT) as the difference in time of consecutive packets. In order to obtain bidirectional flows, we then combine the unidirectional flows based on flow signature and timestamp.

6.2.2 Overall network performance

Traffic load. We start with evaluation of the traffic load. We calculate the load as the aggregate number of bits that traverse the gateway each hour divided by the number of seconds in an hour; our results capture the average combined uplink and downlink rate. We find that the average traffic load Pre-upgrade is 367.3 $kbps$, Post-upgrade is 495.3 $kbps$ and Long-term is 648.1 $kbps$. Fig. 6.2 plots over time the traffic load averaged per hour

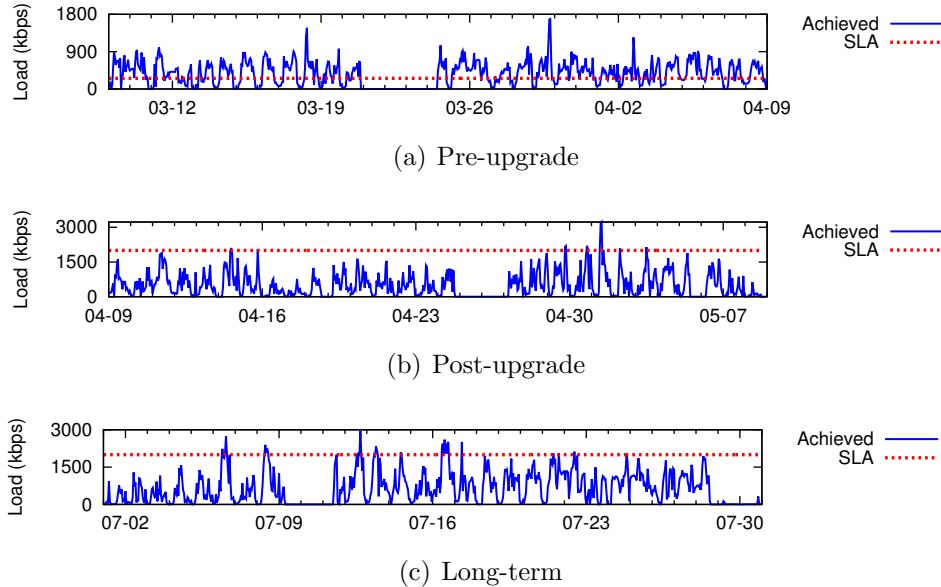


Figure 6.2: Traffic load over time.

in blue and the Service Level Agreement (SLA) with the Internet provider in red¹. In the period before the upgrade, the demand frequently exceeded the SLA of 256 kbps. This is less often the case for the period immediately after the upgrade, as users have not yet adapted to the increase in bandwidth. However, three months after the upgrade the demand often approaches the SLA. As detailed later in our analysis, this is likely due to changed usage patterns whereby users began to access more bandwidth-hungry applications once more bandwidth was available. The gaps in the plots correspond to time periods in which traffic captures were unavailable due to power or network outages.

General trends. We continue our evaluation by discussing general trends over the three observed periods. Table 6.1 presents a detailed look into performance. As we can see, the total bytes that traversed the gateway nearly doubled in the course of three months. The total number of packets dipped Post-upgrade, as the same traffic demand was first accommodated with fewer retransmissions. As time progressed usage

¹Note that while the guaranteed speed was 256 kbps, bursts of up to 1Mbps were possible depending on link utilization. This is why the actual traffic load Pre-upgrade consistently exceeds the SLA of 256 kbps.

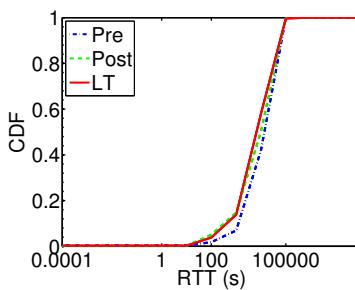


Figure 6.3: RTT.

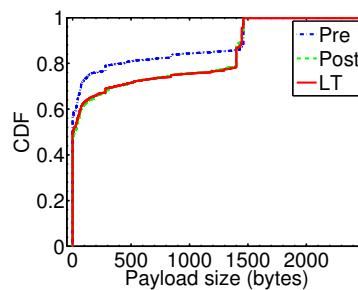


Figure 6.4: Payload size.

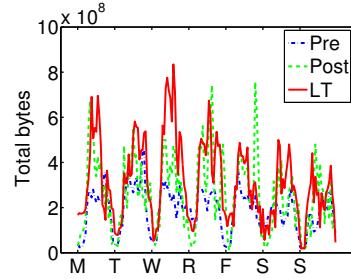


Figure 6.5: Bytes by day.

changed, which resulted in drastic increase in the total bytes traversing the gateway and a corresponding increase in the number of packets.

A similar trend is observed in RTT. While immediately after the upgrade the average RTT decreased by about $35ms$, it nearly tripled as time progressed. We explore the RTT dynamics in more detail in Fig. 6.3, which plots a CDF of RTT for the three periods. We observe a long-tail distribution of RTT in Post-upgrade and Long-term performance; however, the median values of RTT for those two periods are lower than those observed Pre-upgrade. As we will see later in our analysis (in section 6.2.5), the long-tail distribution of RTT after the upgrade is due to changed browsing habits and tendency to use services that are physically further away (such as streaming video from servers abroad). We provide in-depth discussion of usage patterns in section 6.2.4 to validate our hypothesis.

We analyze payload size in Fig. 6.4. We see a clear bi-modality [172] of payload size, which is due to the prevalence of either control packets with 0 bytes payload or data packets with payload of about 1500 bytes. Clearly, the percentage of large data packets Post-upgrade as well as Long-term increased. We also see an increase in the average TCP window size (Table 6.1), which allows more packets to be sent in the network before an acknowledgement is received. This increased TCP window size is critical to improved TCP performance as it translates to higher achieved throughput.

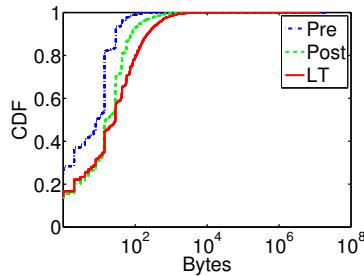


Figure 6.6: Bytes in flight.

We next measure the overhead, focusing on the percentage of the total packets that are due to retransmissions and control packets (e.g. TCP control packets are ACK, SYN, FIN). The payload of control packets is zero bytes. As Table 6.1 and Fig. 6.4 indicate, the fraction of control packets decreased after the link upgrade from 56.59% to 47.69% and then slightly increased in Long-term to 49.72%. The number of retransmissions follows a similar trend. This overall decrease in control overhead can be attributed to improved network performance, which resulted in less protocol overhead from retransmissions and repeated acknowledgements, as well as fewer attempts to re-establish failed TCP sessions. The uptick in retransmissions and control packets over the Long-term can be attributed to decrease in performance due to the increase in offered load to the new saturation point.

Temporal trends. We now discuss performance trends over time. We evaluate byte count in Fig. 6.5, which plots the average on a weekly basis aggregated per hour. For example, the first data point of Fig. 6.5 presents an average over all occurrences of the first hour of Monday, for each of the one month time windows. As the figure shows, there is a clear diurnal pattern in link utilization. Furthermore, the amount of traffic generated during weekdays differs from that on weekends, with weekday traffic loads typically being heavier. The increase in traffic after the network upgrade is also observable in the figures.

Table 6.2: TCP flow analysis.

Period	Total GBytes	Flow size (B)	IAT (s)
Pre-upgrade	105	3445	1.92
Post-upgrade	145	7708	1.49
Long-term	183	8103	1.91

6.2.3 TCP performance analysis

Our analysis shows that more than 93% of the traffic traversing the gateway in Macha is TCP. The performance of TCP improved significantly after the link upgrade. One factor indicative of this improvement is the *bytes in flight*, which is the fraction of sent data that has not yet been acknowledged. The bytes in flight is influenced by the TCP window size: the better the link performance, the larger the window size, which allows more data to be sent on the link before an acknowledgement is received. As indicated in Table 6.1, the TCP window size increased Post-upgrade and Long-term, allowing the amount of bytes in flight to ramp up. Fig. 6.6 presents a CDF of bytes in flight for the three periods. Immediately after the upgrade, the bytes in flight drastically increased and continued growing over the Long-term.

We continue our analysis by exploring TCP flow trends following this improved TCP performance. In order to extract uni-directional TCP flows from our `tshark` captures we develop a tool that examines packet signatures (sourceIP-sourcePORT-destinationIP-destinationPORT) and time-stamp and groups them in flows accordingly. We start by presenting general trends of TCP flows in Table 6.2. The total bytes associated with TCP flows increased after the upgrade and continued growing in Long-term. This increase in bytes is due to increased demand in browsing and streaming applications as well as increased rate of completion of larger TCP flows. We evaluate flow success and failure rates later in this section.

We next examine the average flow size across the three periods. As we can see in

Table 6.3: TCP flow uplink (UL) and downlink (DL) characteristics.

	Total GBytes		# of Flows ($\times 10^5$)		Pkt size (B)		Flow size (B)	
	UL	DL	UL	DL	UL	DL	UL	DL
Pre	18.65	85.9	189	194	132.7	616.0	988.9	4427
Post	19.26	125.7	114	116	158.6	877.0	1691	10856
LT	38.14	145	157	168	227.7	787.6	2422	8613

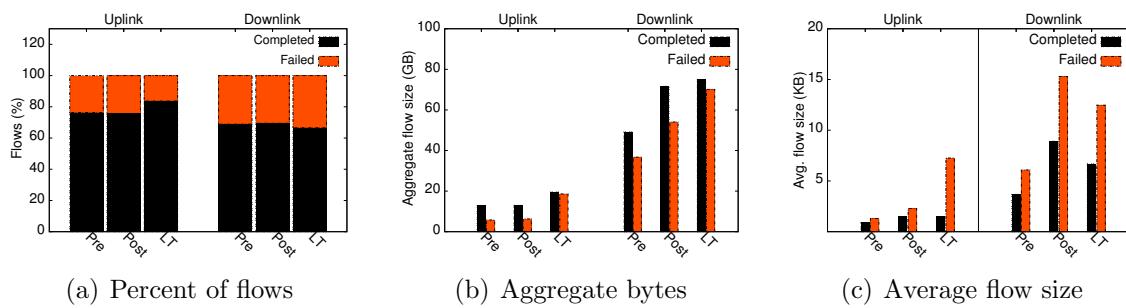


Figure 6.7: TCP flow success and failure in uplink and downlink direction.

Table 6.2, the flow size doubled Post-upgrade and then continued increasing in Long-term. The increase of flow size can be attributed to different applications utilizing the link immediately after the upgrade and in Long-term. Indeed, we see many software updates Post-upgrade, which are then replaced by other applications as we explore in section 6.2.4. The average packet inter-arrival time (IAT) decreased Post-upgrade and then increased Long-term.

Uplink and downlink flows. Next we differentiate flows into uplink and downlink to analyze direction-specific trends. In Table 6.3 we first present aggregate bytes in each direction. Both uplink and downlink bytes increased after the link upgrade. While downlink increased rapidly, uplink remained almost unchanged Post-upgrade but then increased drastically over the Long-term. Average uplink packet size and flow size increased Post-upgrade and in Long-term. At the same time, downlink packet and flow sizes increased Post-upgrade and then slightly decreased over the Long-term. These trends can be explained with differences in applications accessing the Internet, as well as with changes in network performance due to link saturation in Long-term. The rapid increase

in downlink activity Post-upgrade is due to an increase in automated activities such as software updates. The increase in uplink happens more gradually, which is attributed to a slower change in user behavior and, in particular, a gradual increase in content upload attempts.

Finally, we concentrate on the number of flows. As we can see in Table 6.3, the number of flows in both up- and downlink directions decreased dramatically Post-upgrade and then increased. The initial decrease can be attributed to a higher rate of successful flow completions, which directly results in fewer flow re-initializations. The subsequent increase in the Long-term is due to a combination of increased user activity as well as an increase in flow failure rate as user demand again reaches link capacity.

TCP flows success and failure. We now focus on flow completion and failure. In compliance with RFC 793 that mandates the operation of the TCP protocol, we accept that a FIN packet indicates a completed flow, while lack of a FIN packet or exchange of a RST (reset) packet indicates a failed flow. Fig. 6.7(a) presents the fraction of completed and failed flows in uplink and downlink in each period. The completion rate of uplink flows remained unchanged Post-upgrade and then slightly increased in Long-term. On the other hand, the downlink flow completion rate remained unchanged. In Fig. 6.7 we also analyze success and failure trends correlated with byte volume and flow size. Fig. 6.7(b) plots the aggregate flow size in bytes for each direction. The aggregate size of both completed and failed uplink flows remained the same Post-upgrade and then increased in the Long-term. Unfortunately, the amount of bytes in failed flows approaches the amount of bytes in completed flows, which indicates that, while users were likely more successful in uploading content, over the Long-term half of the total content that users generated failed to upload. Similarly, in terms of total size of downlink flows, we see a gradual increase in successful downloads; however, over the Long-term the aggregate size of download flows that failed also increased, nearly reaching the aggregate size of

successful downloads.

We evaluate average flow size of completed and failed flows in Fig. 6.7(c). We find the size of an individual flow by summing the packet sizes of all packets associated with the flow. In the uplink direction, the average size of failed flows over the Long-term is four times larger than the size of completed flows. This implies that smaller content uploads such as Facebook posts and small images are more likely to succeed, while larger uploads of videos or high quality pictures had a higher probability of failure. Similarly, the average size of failed downlink flows is persistently higher than that of completed flows. This points to the success of smaller flows, such as e-mail and web access, while the increase in downlink average flow size for failed flows is likely due to increased attempts to download larger files, such as video content.

Windows vs. Linux. Lastly, we evaluate the TCP performance of two of the most prevalent operating systems used in Macha: Windows and Linux. Using the observed TTL values, we were able to distinguish between the two operating systems [183]. Linux implements CUBIC TCP, which has optimized congestion control mechanisms for high bandwidth networks with high latency. This optimization occurs by calculating the window size according to the last congestion event. In this way, CUBIC TCP measures congestion independently from long RTTs [81]. This differs from Windows, which implements TCP Reno in Windows XP and Compound TCP in Windows Vista and subsequent Windows versions. TCP Reno and Compound TCP base window size on the RTT – window size increases with low RTT values and decreases with high RTT values. This method of congestion calculation causes Windows machines to interpret network latency as network congestion [128].

In Fig. 6.8 we plot frame size and retransmissions per hour for Windows (W) and Linux (L) machines. As we can see, Linux maintains higher mean frame size over all three periods. Thus, it is much more aggressive in pushing data onto the link. Naturally, this

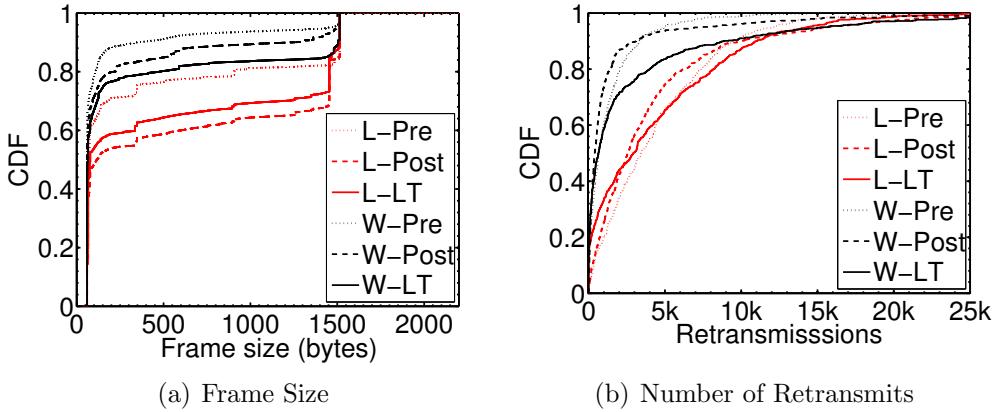


Figure 6.8: Comparison of TCP performance in Windows and Linux.

results in more retransmissions per hour in comparison with Windows. Linux's aggressive behavior, however, leads to higher achieved throughput of 487.6 Kbps in comparison with Windows, which only achieves 106.2 Kbps in the Long-term.

Most popular services. We analyze the most popular services accessed by Machans in the three periods. For this analysis we make use of a tool called Tstat², which can perform layer-7 packet inspection to determine service type.

Our initial results show that the most popular services across periods were HTTP, P2P and SSL/TLS. Thus, the remainder of this section focuses on those services. Fig. 6.9 presents our results for (a) percent of bi-directional flows to each service, (b) percent of total bytes, (c) uplink bytes and (d) downlink bytes. We examine trends across the three periods and look for correlations between different services in order to capture changes in user behaviour. A substantial amount of the flows could not be classified by the layer-7 inspection and were labelled as Unknown. As we can see on Fig. 6.9(a) and 6.9(b) this Unknown traffic constituted the majority of the flows observed across all periods; however, the fraction of bytes due to Unknown flows is minimal. We postulate that the Unknown traffic was due to malware such as port scans, which generated a

²<http://tstat.tlc.polito.it/>

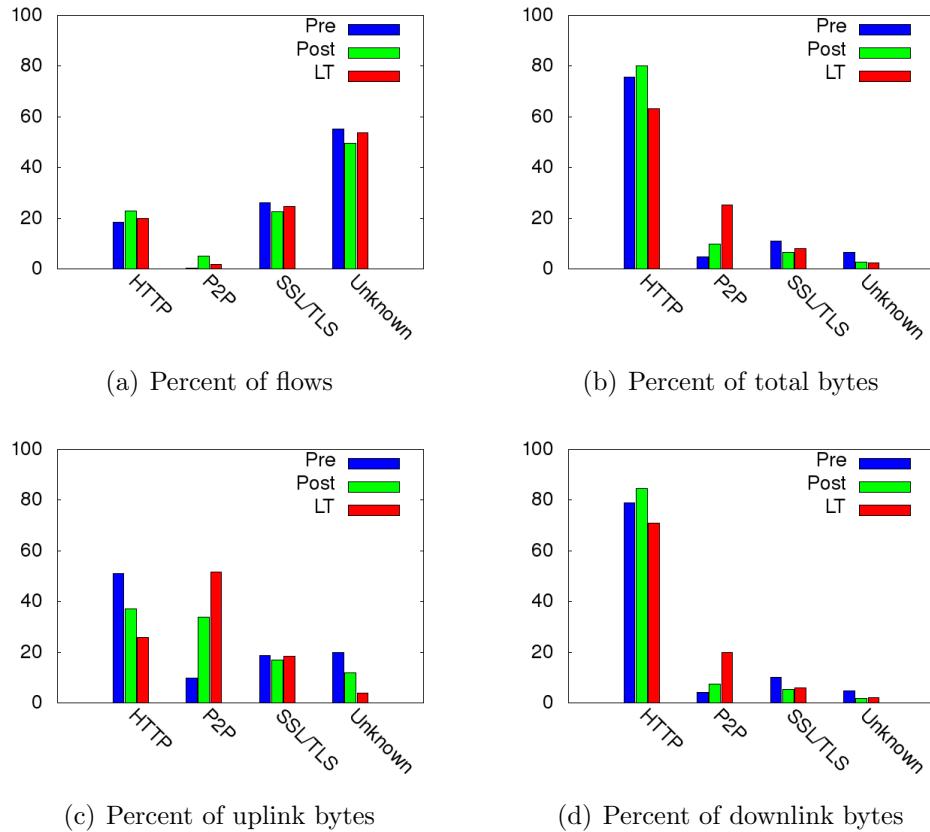


Figure 6.9: Tstat analysis of service types.

large volume of small flows. The fraction of bytes due to Unknown traffic (Fig. 6.9(b), 6.9(c) and 6.9(d)) decreased Post-upgrade and over the Long-term. This is likely due to the successful completion of software updates, which allowed computers to better defend against malware. SSL/TLS is the next most accessed service, followed by HTTP. On the other hand, in terms of generated bytes, HTTP is much more prevalent than SSL/TLS. Lastly, P2P is the least popular in terms of percent of flows; however the bytes due to P2P flows are substantial.

Exploring trends across the three analyzed periods we see a reverse correlation between the bytes due to HTTP and P2P flows over the Long-term (Fig. 6.9(b), 6.9(c) and 6.9(d)). As Fig. 6.9(b) shows, there was an increase in both services Post-upgrade. Over the Long-term, the bytes due to HTTP activity dropped and those due to P2P

flows increased significantly. This indicates a shift in user interest from web browsing to P2P file downloads. Further analysis of the upload and download bytes (Fig. 6.9(c) and 6.9(d)) confirms this trend. We see a gradual increase Post-upgrade in both uplink and downlink P2P bytes, while a more substantial increase can be observed over the Long-term. Facebook, Google and software updates were among the top applications accessed through HTTP. 40% of the P2P traffic was through BitTorrent applications and the remainder was other unclassified P2P traffic. The nature of BitTorrent applications provides a partial explanation of the increase in both uplink and downlink P2P bytes. Particularly, when a user downloads a torrent, this user is simultaneously a seeder; i.e., this user becomes a source of the file and uploads that file to other torrent clients. As a result, 28% of all the uploads were BitTorrent uploads, or in other words, (potentially unintentional) seeding activity. The remaining 72% of the uploads were user-initiated and consisted of HTTP (23%), SSL/TLS (19%), unclassified P2P (23%) and other (7%).

6.2.4 Network usage

The most prevalent application protocol used in Macha is web-browsing. 87% of the Pre-upgrade traffic in up- and downlink direction was a combination of HTTP and HTTPS. This number remained almost unchanged Post-upgrade, and dropped to 71% in the Long-term. At the same time, P2P and Unknown traffic, which includes services to unspecified ports (e.g. Skype and BitTorrent) increased in the Long-term. This is a strong indication of a shift of usage habits to more real-time services, which is typical for well-connected Internet users.

In this section we investigate web traffic to understand user behavior. We correlate our findings about popular applications with network performance and make inferences about the user experience based on this correlation.

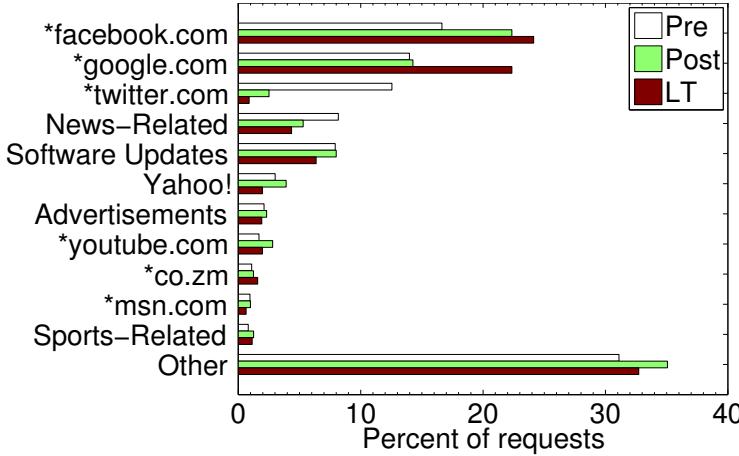


Figure 6.10: Popular URI Requests.

Popular URIs. We begin our analysis by evaluating popular web services. Fig. 6.10 shows web URI requests classified by the destination domains and includes the top 14 requested sites. For clarity of presentation we combine related sites (e.g. Facebook with the associated Content Delivery Networks). Facebook and Google are clearly the most popular sites. Both sites see a significant increase in the percentage of requests after the link upgrade, further extending their dominance. At the same time, access to Twitter, the third most popular domain Pre-upgrade, dropped significantly. The “News” classification includes *postzambia.com, *lusakatimes.com, and BBC news sites. The popularity of these websites is important as it shows user interest in local content, a pattern also seen in [104].

Software update sites such as those associated with Windows, Adobe, and Ubuntu remain relatively unchanged throughout the measurements; however, as shown later in this section, their completion rate significantly increased Post-upgrade. While requests for multimedia-rich sites or large binary downloads remained the same across periods, the actual traffic associated with such requests increased as more requests were successfully completed. We explore TCP sessions count and size later in this section.

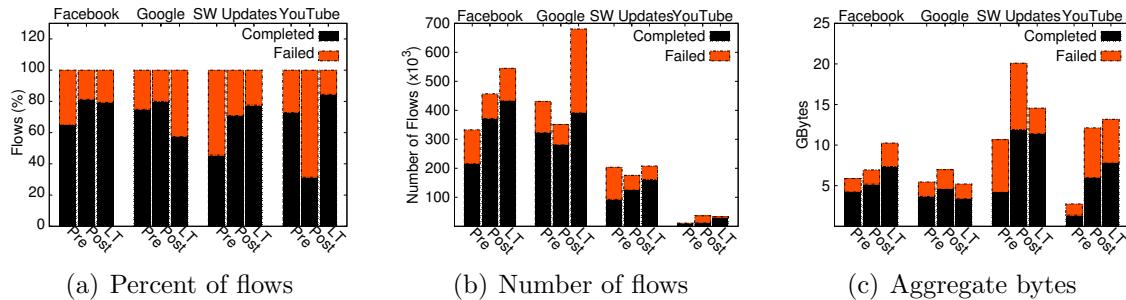


Figure 6.11: TCP flow success and failure for URIs of interest.

Advertisement-related sites are the seventh most popular request type, representing roughly 2% of all requests. Traffic generated by such requests is equivalent to wasted bandwidth as most advertisements are targeted at more affluent urban consumers and are likely of no interest to users in rural Zambia. As bandwidth is clearly a scarce resource in this network, such wasteful access to advertisements can lead to further deterioration of user experience.

Following our URI findings we evaluate TCP flow patterns associated with four of the most accessed online services: Facebook, Google, YouTube and Software updates. For this analysis we combine the previously extracted uni-directional flows into bi-directional sessions based on flow signature and timestamp. We then extract flows of interest based on the URIs that have been accessed in the corresponding session. Fig. 6.11 plots (a) the percentage and (b) the number of flows as well as (c) the aggregate bytes over each period for the four services. The results are divided in terms of flow completion and failure. Both the number and total bytes associated with Facebook flows increased over the three periods. This trend is different than the one followed by Google, which in terms of number of flows remained almost unchanged Post-upgrade, but increased over the Long-term. Similar to Facebook, YouTube also increased immediately after the upgrade both in terms of flows and aggregate bytes. Interestingly, the failure rate of YouTube flows was high Post-upgrade and then de-

Table 6.4: HTTP Response Codes

Response	Pre-upgrade	Post-upgrade	Long-Term
200	4,289,578	3,333,240	4,667,380
400	5,933,008	2,627,842	3,514,872
408	17,146	68	162
Total	12,638,744	7,507,975	10,186,110

creased. This might be due to software updates using a large fraction of the bandwidth Post-upgrade, which caused YouTube to fail more often. Of note, while only 16% of YouTube flows failed in the Long-term, those accounted for 40% of the YouTube flow bytes. This implies that large flows were most often the ones to fail. Due to the increased interest in access to real-time streaming services such as YouTube, the network quickly achieved its maximum capacity, inhibiting these services with substantial flow failures.

Lastly, we look at TCP flows from software updates. The number of such flows decreased slightly Post-upgrade and then increased in the Long-term. Our analysis indicates that the short-term decrease is due to improved network performance resulting in fewer TCP session re-initializations. Furthermore, the quantity of bytes associated with software updates doubled immediately after the link upgrade. This is likely due to long-postponed software updates finally being able to complete. We see a decrease in software update bytes in the Long-term due to successful completion of updates in the period Post-upgrade.

Next, we measure HTTP response codes in an effort to find discernible differences between observation periods. We find noticeable changes in three response types: 200, 400, and 408. 200 (OK) responses indicate a valid request for which an HTTP server can correctly craft a response. As shown in Table 6.4 the percentage of HTTP 200 responses increases more than 10% after the link upgrade. 400 (Bad Request) errors indicate a request that the web server does not understand. These errors typically are caused by

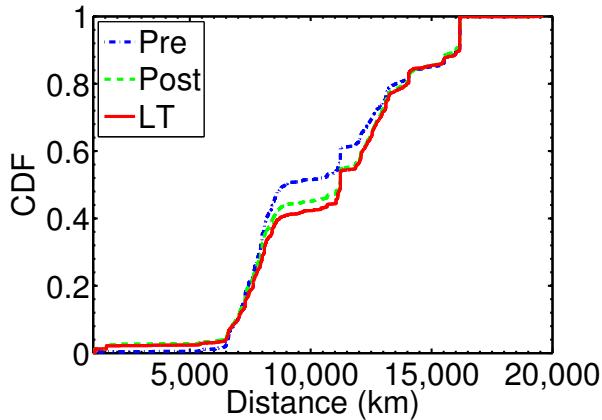


Figure 6.12: Flow distance from Macha; CDF.

bad syntax or potentially a host infected with malware that sends poorly defined HTTP requests. The table shows 400 errors decrease significantly after the link upgrade. We believe that this could be due to two changes. First, immediately after the upgrade hosts could have implemented overdue software updates which could rectify browser version issues associated with request format. Secondly, in a similar fashion to operating system software, anti-virus software was updated to newer versions which could potentially allow for the detection and removal of malware on hosts. The final response code we investigate is 408 (Request Timeout) which indicates that the server was expecting a request from the client in some amount of time and the client failed to produce such a request. Such errors occur in networks with very limited bandwidth or where multiple packets are dropped some point along the path. The number of 408 errors decreases dramatically after the link upgrade. This is an encouraging result, as it shows that even a small bandwidth increase can make a large difference in the user experience.

6.2.5 Flow Distance

We investigate the network traffic using geographical information to characterize usage. For each traffic flow we identify the external node IP address. Using these IP

Table 6.5: Measured radius of gyration.

Period	Distance (km)
Pre-upgrade	6,363.26
Post-upgrade	6,851.41
Long-term	7,096.86

addresses, we query the MaxMind GeoIP database [12] to correlate each flow with geographic coordinate information. Our preliminary investigation involves calculating the straight-line distance between Macha and the given coordinates for the other side of each connection using the Haversine formula [153]. Fig. 6.12 shows the CDF of the flow distances from Macha in each of the three observation periods. We find that flows generally occur over longer distances in the periods after the network upgrade. Of note is the large increase between the Pre period and the Post period in the roughly 8,000 to 12,000 km range. While Long-Term flows show even longer distances as compared to Post-upgrade, the increase is not as pronounced. We posit that a potential reason for the increase in distances from Macha is the result of a better user-experience after the network upgrade, which encouraged users to access such services that are physically further away.

We also use the GeoIP database to find the country code for each external node. We calculate the number of bytes associated with each country code and rank them. Interestingly, traffic to and from nodes in Zambia itself increased dramatically after the network upgrade. In the Pre period, Zambia ranks as the thirteenth most popular country in terms of bytes, representing 0.9% of all traffic. In the Post period, Zambia jumps up to rank second representing 23.4% of all bytes; in Long-Term it is ranked third with 12.1%.

Our initial distance findings lead us to investigate not only the distance from Macha that flows represent, but also the overall “worldliness” of the network flows. That is to say, we investigate the distribution of the geographic coordinates in order to further characterize network usage. We utilize the Radius of gyration metric to provide a value for the spread of the data. Radius of gyration has been used extensively to characterize

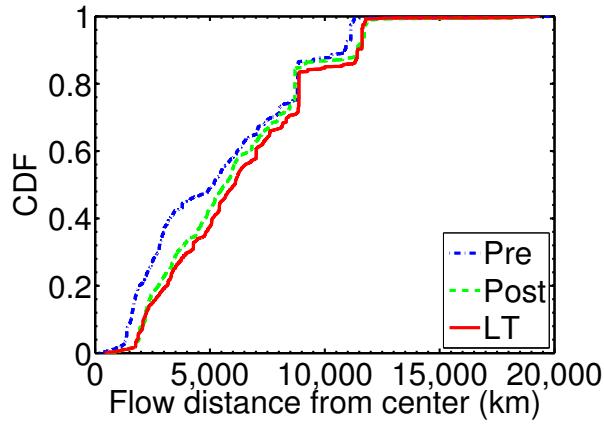


Figure 6.13: Flow distance from center mass; CDF.

user mobility in wireless networks [77] and provides a technique for measuring dispersion. It can be understood as the range of observed points up to time t and can be calculated by the formula:

$$r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)} \sum_{i=1}^{n_c^a} (\vec{r}_i^a - \vec{r}_{cm}^a)^2}$$

where \vec{r}_i^a represents the i^{th} coordinate for period a and \vec{r}_{cm}^a is the calculated center-mass for the period. $n_c^a(t)$ represents the number of points measured up to time t . Table 6.5 shows the results for the observation periods. We see that each successive period shows an increase in the radius compared to the prior periods. This means that not only are flows connecting to locations further away from Macha as seen in Fig. 6.12, they are actually spreading out further as well. Assuming users are behind the majority of network traffic we can argue that network users are connecting to content from a larger geographic variety of the world.

We verify these results in two ways. First, we find the distances between each flow and the center-mass and plot the CDF shown in Fig. 6.13. As expected, we see longer-distance flows in the periods after the network upgrade. Further, the increases seem to be incremental and uniform rather than drastic changes. We also investigate the center-

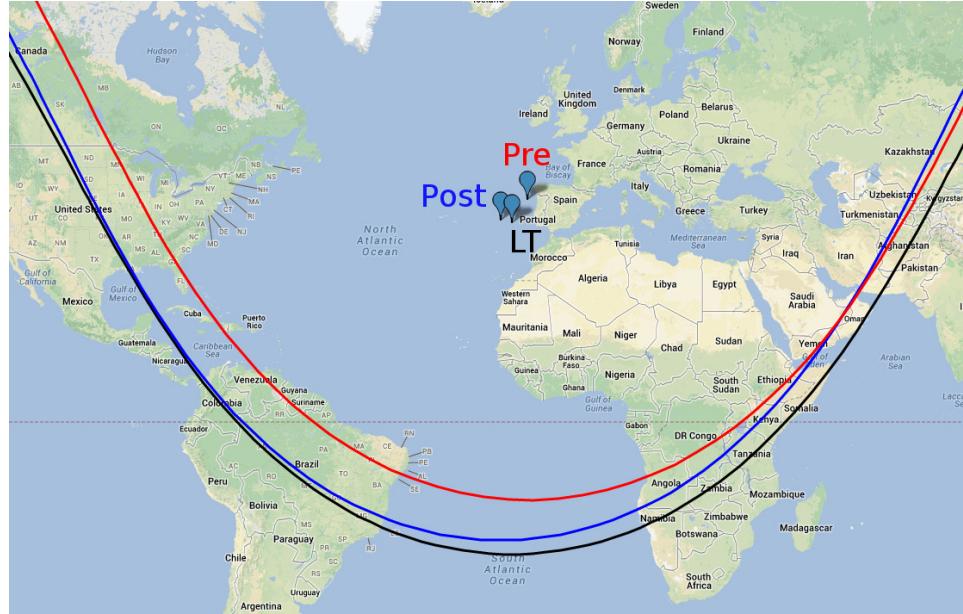


Figure 6.14: Center-mass points with radii of gyration.

mass values to determine whether the upstream provider (which changed along with the network upgrade) drastically altered the distribution of external nodes. While we expect the center mass values to be different for each period, we also expect them to be somewhat clustered. Should the upstream providers apply unexpected policy-based traffic routing (e.g. resolving all CDN queries to a particular location), we expect to see the center mass values vary dramatically between the Pre period and those after the upgrade. Fig. 6.14 shows the center-mass points for each period as well as each calculated radius of gyration. We find the three center-mass values are within a reasonable range of each other given the global scale and as such we do not credit the upstream provider with the radius of gyration increase. Given these results we are confident that the increase in spread can be credited to an increased geographic diversity of external nodes.

6.2.6 Long Term Trends Persistence

To verify whether the observed long term trends persist over time, we examine network performance and utilization in one additional time interval seven months post-upgrade. In order to assure that a change in user population would not skew our results, we focus on November, since this month has roughly the same number of users as the previous months of interest. We measure the same set of characteristics as for the previous three months and compare results from November and July (our Long-term month). Our results show that the overall network utilization in November was similar to that of July. The user behavior, however, changed, whereby users were no longer attempting to upload large files to the Internet but instead increased their download activity. As a result, the total uploaded traffic in November decreased and became comparable to that before the upgrade, while the total downloaded bytes increased significantly by 60

We analyze a total of two and a half weeks of November. The reason for not analyzing the full month is two-fold. First, the first four days of the month were not captured in our traces due to failure of the managed switch that was mirroring the network traffic for monitoring purposes. Second, the Android phone phenomenon that we observed in July was aggravated throughout November, whereby approximately 4 million out of a total of 5 million requests sent to Google were produced by this phone. In order to conduct an objective comparison of network performance and utilization, we discard the minute intervals throughout November where this phone was present in the network, as the presence of the phone significantly skews the results. For the remainder of this section we present results from the two and a half weeks analyzed; we refer to this traffic as the November traffic or simply November.

In November, a total of 158 GB traversed the network, which consisted of 373 million packets. This is about 1GB/day more than what we observed in July. This increase in

the total amount of data that traversed the network is due to a combination of improved TCP performance and larger download attempts. The improved TCP performance is indicated by factors such as control overhead and bytes in flight. The control overhead in November decreased to 45.2% (compared to 49.7% in July) and the bytes in flight increased from 10.5kB to 36.4kB. We postulate that the large value of bytes in flight is due to longer TCP flows, which allows enough time for TCP to ramp up the amount of bytes sent and achieve higher throughput. The latter results in increased bytes in flight. We also examine the possibility that this aggressive behavior is due to more Linux traffic traversing the network. Our verification, however, shows that the ratio of Linux and Windows packets that traversed the network is comparable to that in July. Thus the reason for this aggressive behavior is not related to increase in the amount of Linux traffic. Lastly, we examine retransmissions as an indicator of TCP performance. We observe an increase of retransmissions from 1.16% in July to 3.68% in November. While this increase contradicts the TCP performance improvement, a closer look at retransmissions over time shows that there were several short periods where the network was experiencing poor performance, which resulted in few tens of flows with millions of retransmissions. If we omit those flows, the retransmission performance is comparable to that in July.

We now examine uplink and downlink trends in an effort to assess whether user experience in sharing and accessing content persisted throughout November. The total uplink bytes in the November traffic is 11.7GB, which is approximately 650MB/day. This is a decrease in the uploaded traffic in comparison with July, where uploads were 1.3GB/day. In fact, the amount of uploaded bytes per day in November is comparable to that before the upgrade (620MB/day). As we saw in section 6.2.3 the performance of uplink flows deteriorated dramatically as users began to upload more content. Thus, the decreased amount of uploaded bytes in November can be explained by users having

Table 6.6: TCP flow uplink (UL) and downlink (DL) characteristics.

	Percent of Flows		Aggregate GB/day		Avg. Flow Size (kB)	
	July	November	July	November	July	November
Failed DL	38	42	2.33	2.08	12	6.2
Success DL	62	58	2.5	4.21	6.5	9
Failed UL	19	21	0.633	0.147	7	0.9
Success UL	81	79	0.667	0.504	1.5	0.8

become discouraged from sharing their content online by the poor upload performance. At the same time the amount of downloaded traffic increased from 4.8 GB/day in July to 6.3 GB/day in November, indicating a clear trend of more Internet traffic consumption than generation.

To get a better understanding of how success and failure of uploads and downloads influences users inclination to share and consume content, we separate the November traffic into completed and failed TCP flows and analyze these flows in the uplink and downlink directions. As Table 6.6 shows, there was a slight increase in the percentage of successful uplink and downlink flows. At the same time, the traffic volume measured in aggregate bytes changed more substantially in both directions. The aggregate bytes in failed uploads decreased by 77%, reflecting on peoples decrease in interest to upload content online. Similarly, the aggregate bytes of successful uploads decreased from 667MB to 504MB/day. In line with this trend, the average flow size of both completed and failed uplink flows decreased in November. The aggregate of successfully download bytes increased by 60% in comparison with July, indicating a rapid increase in the amount of consumed traffic.

To better understand this shift in usage characterized by rapid decrease in uploads and growth in downloads, we examine the most popular layer-7 services in November. We observe a notable change in usage as compared with July, whereby P2P traffic decreased and HTTP increased. The fraction of P2P bytes in November dropped to 9%, in

comparison with 25% in July. This decrease was shared between uploads and downloads but was much more pronounced in the upload direction: P2P bytes dropped from 52% to 7%. Because the P2P uploads constituted a large portion of the overall uploads in the network (51%) in July, their decrease had a significant impact on the overall upload volume observed in November. Simultaneously, the percentage of HTTP downloads increased from 63% to 78%.

The goal of the analysis presented in this section was to evaluate persistence of network performance and usage in the post-upgrade period. While network performance improved in comparison with July, the usage of network services changed. In particular, we observe a rapid decrease in the volume of upload flows, while more emphasis was put into downloads. We saw a shift of usage from HTTP immediately after the upgrade to P2P file transfers in July and back to HTTP in November, which indicates that users were eager to try new services but ultimately reverted back to HTTP, likely due to poor user experience.

6.2.7 Benchmark

In this section we answer the question “Given the limited gateway capacity in Macha, how well can the network perform?”. To this end we provide a benchmark of network performance in November and compare this benchmark performance with the actual observed network performance. Intuitively, the performance of services such as Facebook and Google is influenced by the specific service configuration. For example, whether services are centralized or use content distribution networks, and they handle poor network performance, can directly influence the user experience. In order to be able to objectively compare the performance of the Macha network with a benchmark, we need to ensure that the benchmark setup resembles the same service configuration as the one used by

Machans during the trace capture period. At the same time, service configurations often change to meet the ever growing user demand and application growth. Thus, we base our benchmark on the November traces we capture in Macha. In particular, we analyze parts of the trace when fewer users used the network in order to ensure there was minimal contention on the link. Our analysis focuses on TCP performance and shows that the benchmark traffic performs significantly better than the average performance in November.

Therefore, to generate the benchmark trace we divide the November capture into one minute chunks and consider the minutes with three or fewer users. This method generated a total of 2853 minutes (nearly 48 hours), which we use as our benchmark. The website access distribution during those hours is comparable to that during busy periods. The benchmark traffic amounted to 5.88GB. Out of a total of 9,556,817 packets, 16% were control packets, which is a significant reduction in the amount of protocol overhead as compared to the more general scenario. The average RTT decreased from 157ms to 119ms. Because the same websites were accessed, this likely indicates that there was less queuing delay, as the ISP gateway and/or core network was less overloaded.

With high-delay, low-speed connections, rural users are often limited in their ability to utilize the Internet. Furthermore, the benchmark presented in this section clearly shows that when multiple users share a limited Internet connection, the end user experience can be further deteriorated. In less overloaded scenarios, however, the TCP performance becomes comparable to that of the western world, whereby the RTT reduces and the amount of control overhead drops significantly. Thus, if a single 2MBps link such as the one in Macha was used by several users, there is a hope that the Internet experience can be on par with that in the western world. When a single 2MBps connection is shared between tens or hundreds of users, however, the network performance will continue to negatively impact the user experience.

6.3 Related Work

Our work builds upon earlier analysis of Internet access in Macha [106]. This prior work focuses on network performance and usage during a two week measurement period in 2010. Other work analyzing rural Internet usage patterns includes [67] and [96]. Web traffic from Internet cafés and kiosks in Cambodia and Ghana is analyzed in [67]. Here the focus is on the characterization of HTTP traffic to guide caching techniques for web users in developing regions. Ihm et al. [96] focus on understanding the network traffic in developing regions as compared to their OECD counterparts. This chapter characterizes national traffic patterns based on network usage, with the goal of improving caching techniques for developing regions. Anokwa et al. identify the impact of latency on network performance in developing regions and propose a flow-based prioritization scheme as a solution [32]. In contrast, our work focuses on a smaller scale and characterizes web traffic in order to ascertain the impact of a network upgrade on usage and performance.

Our analysis of TCP performance builds upon the measurements employed by Johnson et al. [106] by engaging a more in-depth analysis of TCP performance, including measurements of TCP windows, retransmissions, inter-packet arrival times, RTT, and packet sizes. Performance of CUBIC TCP, Compound TCP, and TCP Reno interactions is measured via simulation of high delay wireless networks in [22]. This work has an explicit interest in measuring goodput and TCP fairness. Our analysis is based on TCP performance in a low-bandwidth, high-latency real network; we measure TCP performance in aggregate and as separated by operating system network stacks, and draw conclusions about TCP fairness of different variants found on the network.

6.4 Discussion

Our analysis of a network upgrade in a rural community indicates that even a small increase in access bandwidth can improve the usability of the network: for example, successful software updates and updated anti-virus protection immediately after the upgrade grant better performance in HTTP request generation and, overall, decrease the traffic due to malware activity, resulting in possibilities for better performance. While these results are encouraging, incremental increase of available bandwidth can often bring only marginal improvement of user experience, as indicated by the large volume of failed requests in the case of Macha. In the face of such increased usability but still low quality of user experience, the need of systems such as VillageShare [105] (and others [67, 102, 193]), that can intelligently manage activities in the network, is even more pronounced.

One immediate need that arises from our analysis is the one of prioritizing bandwidth allocation to critical services. For example, as usage patterns in Macha did not immediately change post-upgrade, critical software updates were finally able to complete. This, in turn, resulted in rapid improvement in browsing experience (as indicated by the drop in HTTP Bad Request and HTTP Request Timeout messages in the network) on one hand and by the decrease of traffic associated with malware on another. This observation hints of a need for a system that can detect critical services and emulate a bandwidth increase for such services.

Such a system would be able to perform real-time detection of network traffic anomalies (for example increase in abnormal HTTP requests or traffic to ports associated with viruses) and would prioritize bandwidth assignment for software updates. Two major concerns arise with regards to such a system. First, to ensure that such bandwidth prioritization does not compromise the user experience, this functionality can be embedded in time-shifted proxies such as [67, 105, 193]. Time shifting to off-peak hours, however,

runs the risk of users turning off their computers, which brings us to the second challenge in such system design. To handle this, local caching techniques [102, 105, 193] can be employed which make particular content (e.g. software updates) available in the local network for use during peak hours.

6.5 Conclusion

We utilize a unique dataset from a rural sub-Saharan village that captures usage before and after an Internet access speed upgrade. We study the effects of this upgrade on the network performance and user behavior. We find that performance improved immediately after the upgrade, whereby automatic services that were previously failing due to slow access speed were finally able to complete. With improved network performance, subscribers were encouraged to use more bandwidth-demanding services such as YouTube video streaming. There also was a substantial increase in attempts for content sharing online, whereby the uplink byte volume doubled in the Long-term. Unfortunately, with the increase of upload attempts, the failure rate of uploads grew as well.

Another trend that stood out from our analysis is the stark difference in performance between Windows and Linux operating systems. As our results show, Linux outperforms Windows, achieving five times better throughput on average in the Long-term. This makes a case for careful selection of operating system and/or modifications of the network stack to facilitate better networking performance in bandwidth-constrained environments.

Internet access upgrade in the context of developing rural regions is not a trivial task. Although such upgrades are perceived to lead to overall better performance and user experience, this is not always the case for communities that are largely bandwidth-impaired. In such communities, an Internet upgrade can be just a small increment to

the more substantial access speed that is needed to accomodate modern web content and applications. Each such increment gives users the ability to more fully utilize the modern Internet with bandwidth-intensive applications; however, it is clear that in the developing regions case, even an eight-fold increase in network capacity can be not nearly enough. Many rural communities like Macha have a long way to go before their Internet experience parallels that of users in the Western world.

Part II

Network Systems Design for Resource-Limited Environments

Given our characterization and analysis presented in Part I, we now turn to designing systems specifically tailored to address some of the connectivity challenges we have witnessed.

We design three separate cellular-based systems to mitigate the challenges faced by users in fringe network coverage areas and where infrastructure is unable to meet user demand. We design HybridCell, a system that dynamically shifts usage onto a local cellular network in response to nearby commercial cellular congestion. For settings where a local cellular network cannot be deployed we offer SmartCell, an Android application that informs the user of commercial congestion conditions, empowering them to make choices such as physically relocating or changing to a different provider using another SIM card. Lastly, we create PhoneHome with the goal of extending commercial coverage and localizing traffic to ease the burden on a nearby commercial network.

WISP networks are often the most economical solution to providing access to rural communities. However, such networks are often planned by lone operators with little to no networking expertise. The key challenge is the mismatch between the resources of these small operators and the complexity of their network planning. We create a prototype system to aid in WISP network topology planning and detail the many challenges in this area.

Radio spectrum is a finite resource that is needed for wireless communication. Recent advances in agile radio technologies have led researchers and governments to begin thinking about allowing dynamic spectrum access (DSA) from multiple users in a given frequency / location. In order for DSA to be viable, we must have detailed knowledge of the spectrum usage. We present AirVIEW, a spectrum characterization approach for rapid, arbitrary transmitter detection that is resistant to noise. AirVIEW is designed to be single-pass, making it ideal for wide-band spectrum scans.

Given our analysis of the network performance in Macha, Zambia, we investigate

user content patterns and discover high locality of interest in social media content. We create VillageCache, a web cache that scrapes locally uploaded media content, locally preforms the necessary transformations and injects into the cache a copy of the media for subsequent local downloads. Importantly, VillageCache removes the requirement of even a single download of locally-create media to install that media in the cache, as even a single download can prove impossible in poorly connected networks such as the network we have studied.

Chapter 7

System Design: Cellular Systems for Resource-Limited Environments

7.1 Introduction

Millions of people throughout the world live in areas at the fringes of cellular connectivity, where the cost of increasing local infrastructure exceeds the expected return on investment. In India, for example, roughly 25% of the population (nearly 315M [7]) resides in areas without cellular coverage. Furthermore, coverage that does exist in fringe areas is often spotty and overburdened. As a result, residents typically carry multiple SIM cards in order to obtain connectivity in locations where one provider is present while another is not. Nevertheless, and the inability to obtain any service during busy times of day is common. Cellular networks in these areas are simply unable to service the demand placed on them, yet upgrading the infrastructure is infeasible due to cost. Further, despite continued expansion of commercial networks, hundreds of millions of people reside in areas without coverage. Clearly, ubiquitous coverage will not be achieved without exploring the use of more cost effective technologies.

Rural areas are just one example where infrastructure capacity is unable to meet the demand. When disasters strike, people often move to makeshift camps located in areas with available space, where existing cellular infrastructure is not provisioned for the increase in user load. Likewise, political conflicts lead to displacement of people to refugee camps. Such camps are often located in rural areas, on the fringes of infrastructure, in order to reduce disruption of the residents of the host territory. Even infrastructure in urban areas can experience sudden spikes in demand (e.g. 100,000 people gathered for a protest), where the expense of building the infrastructure necessary to meet peak demand makes little financial sense.

In this chapter, we take our findings from Chapter 3 to inform the design of three different systems with the overall goal of connecting displaced people or those living in rural areas with poor coverage.

7.1.1 HybridCell

Recently, local cellular networks have gained traction as a solution for providing cellular connectivity in remote areas that lack coverage [17, 85, 87, 206]. These installations use small-scale base stations running open source software such as OpenBTS [1], which translate GSM to voice over IP (VoIP), and offer free or low-cost cellular services. Unlike femtocell technology [52], local cellular networks can exist autonomously and do not require a reliable connection to a commercial provider in order to operate. When a reliable connection to a provider is unavailable, a local cellular network can still provide voice and SMS capability between users connected within that local network. The question arises, *can local cellular networks be leveraged to alleviate and bolster poor commercial connectivity in areas where providers are either unable or unwilling to augment their infrastructure?* Prior work on local cellular networks has focused on areas with *no existing*

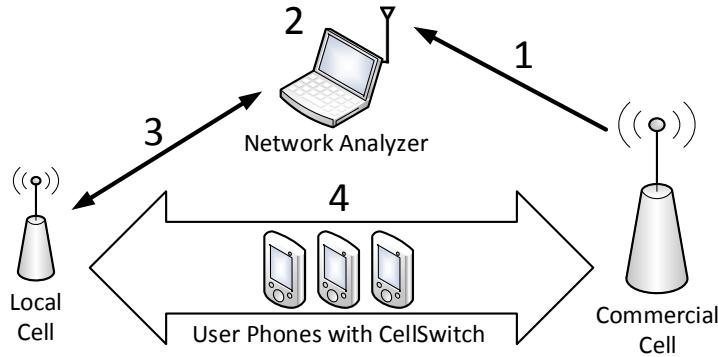


Figure 7.1: HybridCell consists of a Network Analyzer, a Local Cell and our network switching application CellSwitch. (1) The Network Analyzer gathers performance data from the commercial cell; (2) the Network Analyzer algorithm determines commercial cell congestion and failure; (3) the Local Cell is reconfigured based on the Analyzer's measurements; and (4) CellSwitch shifts users between Local and Commercial cells for seamless connectivity.

coverage [87, 206, 88, 2]. In contrast, we use local cellular technology to create HybridCell (Figure 7.1), a self-contained, community-owned and operated cellular network that intelligently *augments and coexists* with existing commercial coverage. HybridCell provides supplemental capacity when commercial networks become overloaded or non-operational by shifting a portion of local calls and SMS messages between HybridCell users to the local network. We design HybridCell to improve connectivity in areas with existing, yet under-performing cellular coverage rather than areas devoid of coverage.

The design of HybridCell raises several research challenges. For instance, how can we determine when a cellular network is overloaded without having access to that network, and without adding any traffic load? How can we define congestion given that base stations can have varying capacities? Can we create a solution that respects spectrum occupancy and does not cause interference?

HybridCell solves these challenges through three key components: (i) a commercial network analyzer, (ii) small local cells, and (iii) an Android application dubbed CellSwitch. HybridCell's analyzer performs non-intrusive, passive measurements to quantify

the level of congestion in the commercial network and thus determine the performance and health of the commercial network. When the commercial network is fully functional, HybridCell operates quietly in the background, simply monitoring network load. As congestion and failures increase, HybridCell adaptively shifts clients from the commercial to the local network in order to decrease the load on the commercial network. The Cell-Switch application controls each user’s network association programmatically, removing the need for users to explicitly choose networks via physically changing SIM cards or reconfiguring network association settings. HybridCell works alongside any commercial carrier, in contrast to the connectivity model used by ProjectFi [10], as described in Section 7.5.

To more specifically put the HybridCell system into context, we examine the communication needs due to the plight of Syrian refugees, and demonstrate how HybridCell could help improve cellular access in such circumstances. The UNHCR estimates that the conflict in Syria has caused over 4.9 million Syrians to leave the country as of March 15, 2017 [18]. The influx of refugees has led to the establishment of roughly 30 refugee camps in neighboring countries. The Za’atari refugee camp, located in Jordan, has become one of the largest refugee camps in the world. As described in Chapter 2.1, we visited Za’atari in January 2015 to evaluate current cellular network coverage and usage within the camp. Like many refugee camps, Za’atari quickly sprung into existence, forming virtually overnight. Within 9 months of its July 28, 2012 establishment, the population of the camp had increased to over 200,000 people [18]. As a result, infrastructure struggled to keep up with the rapidly growing population. We captured measurements of the existing cellular infrastructure serving the camp and conducted surveys and interviews of camp residents as well as administrative staff. In Section 7.2.2, we use our measurement data to both convey the dire, present need for a system such as HybridCell, as well as to evaluate the potential for HybridCell in such an environment.

Our work on HybridCell makes several key contributions:

- We design a first of its kind method for passively quantifying cellular congestion as third party observers without requiring access to the cellular network’s core traffic.
- We use our congestion detection mechanism to assess network load and performance in a real-world multi-carrier environment in Jordan’s Za’atari refugee camp.
- We implement an Android application, CellSwitch, that programmatically shifts phones between cellular networks, allowing phones to use both local and commercial cellular networks without manual intervention.
- We integrate our analyzer and CellSwitch into HybridCell; the first system to combine commercial and local cellular networks for seamless user connectivity.

7.1.2 SmartCell

Residents in developing countries are acutely aware of coverage disparities between carriers, and often use SIM cards from multiple carriers, switching between carriers depending on location. Our interviews with Za’atari residents indicate that they typically carry SIM cards for each available provider, but do not have any specific algorithm for selecting which to use at a given time. Reduced cost calling and texting between users on the same carrier creates a network effect, and may cause users to prefer a given SIM, simply because the provider is the most popular. This leads to uneven traffic distributions and load between carriers.

Additionally, in some locations deployment of a local cellular network as proposed by HybridCell is not feasible due to licensing or lack of local expertise. To account for such situations, we design SmartCell, an Android application that gives users awareness of quality of service by detecting congestion on their cellular base station. SmartCell informs

the user when congestion is detected, thereby empowering them with information about their local network. With this information, the user has multiple options for obtaining better quality of service. Most simply, the user could decide to use one of her alternate SIM cards, switch to an underloaded network, and obtain higher quality service. If she has a multi-SIM phone, SmartCell can automatically select a SIM from a less congested carrier to use for voice calls, SMS, and mobile data traffic. Without information about the quality of the given cellular network provided by SmartCell, the user is not informed that a better connectivity option is available, and is likely to continue to suffer with unacceptable performance.

7.1.3 PhoneHome

Through our experience in the field, we learned that there are often large groups of people that live *near* cellular coverage but for a number of reasons are at the edges of commercial coverage. We term these locations as ‘fringe’ areas. As an example, refugees who are displaced from their home countries due to civil unrest often resettle immediately across the border of their home country. In many, there are documented instances where refugees have been able to obtain low quality access to their home cellular providers, if there is, for instance, a hill nearby [143, 118]. As an example, Figure 7.2 illustrates the cellular coverage map of Syria. We see that cellular coverage easily “leaks” across political boundaries into neighboring countries. Refugees will often pursue these connections, where possible, to circumvent international calling rates to communicate with family back home [143, 118]. This “rogue access” also avoids restrictions from host country cellular providers, where these providers will not offer new refugees SIM cards and cellular access, due to their inability to obtain credit and provide a permanent address. In scenarios of relocation within the home country due to natural disaster,

such long-distance, low quality cellular access is also likely, as camps form in rural areas outside, but nearby, damaged cities. An example is shown in Figure 7.3, which maps cellular coverage and shelter locations two weeks after Typhoon Haiyan passed over the Philippines in November of 2013. As the figure shows, most of the sites are in areas with degraded service.

To illustrate the pressing need for connectivity in rural fringe areas, we investigate geographic coverage in Guatemala. We leverage population density information from [184] as well as the OpenCellID database [61] to map base station coordinates for the three major cellular carriers, Tigo, Claro, and Movistar. The population density is a 1-km grid shaded based on the number of inhabitants per square, as shown in Figure 7.4. Due to fundamental timing constraints, the maximum distance a GSM mobile device can be from a base station is 35km. We therefore buffer each base station by 35km to generate the map shown in Figure 7.5. Note that the coverage areas (white) depicted do not reflect real-world conditions and in fact represent best-case, maximum coverage for all base stations. Coverage is, in reality, diminished by terrain impacting line-of-site and user demand. To conserve space, we only include a coverage map for Movistar. As expected, all three carriers focus coverage on highly populated areas.

We can estimate the number of people living outside, but *near* cellular coverage. We buffer all base stations by an additional 25km and calculate the population of people living in the area between 35km and 60km from the base stations (grey shaded areas in Figure 7.5). We find that Tigo has a total of 17,438 people living in these ‘fringe’ areas, Claro has 35,294, and Movistar has 113,795. These areas represent opportunities for extending existing coverage using a system such as PhoneHome.

We leverage local cellular network technology to create *PhoneHome*, a system prototype designed to extend cellular coverage from a commercial carrier into an area where the carrier does not offer sufficient, quality coverage, or where coverage does not exist.

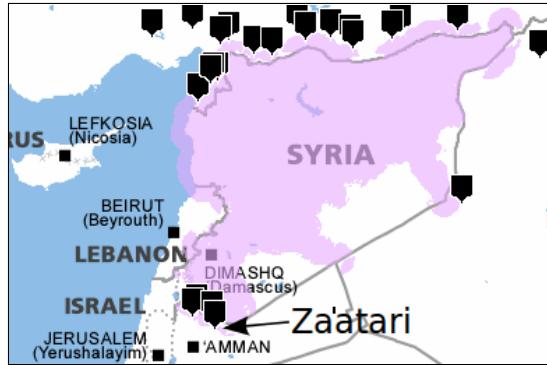


Figure 7.2: MTN Syria coverage map with camp locations. [131]

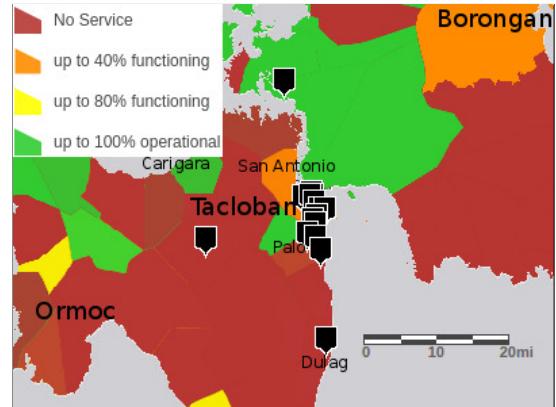


Figure 7.3: Cellular coverage map two weeks after Typhoon Haiyan [72].

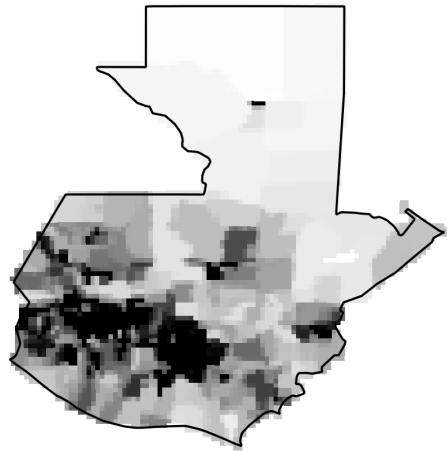


Figure 7.4: Guatemala population density. Dark areas indicate higher density.

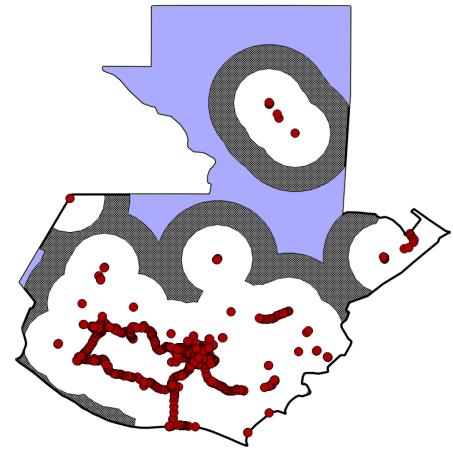


Figure 7.5: Movistar coverage in Guatemala.

PhoneHome introduces three critical functions: (i) it independently provides cellular coverage into poorly-connected areas; (ii) it localizes the cellular core, enabling local cellular traffic regardless of the presence or absence of global connectivity; and (iii) it virtually extends nearby commercial coverage without requiring commercial carrier involvement.

7.2 HybridCell

The main goal of HybridCell is to improve user connectivity when a nearby commercial network is unable to provide adequate service. We design HybridCell with the intention that phones should remain on commercial networks as much as possible since HybridCell is meant to be a secondary, rather than primary, means of connectivity. When users are connected via their commercial network, they are globally reachable and can receive incoming calls and text messages from users outside the local network. Reachability of users on the local cellular network is deployment-dependent: if the local cell does not permit incoming calls through a backhaul link using VoIP, then the only way to call someone on the local cell is for the caller and callee to both be on the local cell network. In such a case, the traffic HybridCell is able to offload is limited to that between local users. Prior work, however, shows that as much as 70% of rural networks traffic can be local [207]. Thus HybridCell has the potential for high impact in our target areas.

7.2.1 HybridCell System Design

HybridCell consists of three components as illustrated in Figure 7.1: (i) a network analyzer that detects active commercial cellular networks and characterizes their performance; (ii) a local cell that augments cellular services in the face of a failing commercial network; and (iii) the CellSwitch application that resides on users' phones and transparently migrates users between the commercial and the local network in order to assure seamless connectivity. In this section we provide a detailed description of HybridCell's components and operation.

Commercial Network Analyzer

Before taking any action, HybridCell first assesses the performance of the nearby commercial network using a network analyzer, which operates in two phases. The first *detects* available carriers and the second *characterizes* the performance of these carriers. The detection phase identifies all the Absolute Radio Frequency Channel Numbers (ARFCNs), i.e. the operating frequencies of individual cells, along with the technology they use (i.e. GSM, 3G or LTE). The characterization phase then taps into the control channels of each carrier and identifies the health status of the network based on the control messages. In what follows, we detail our carrier detection and characterization mechanisms.

Commercial carrier detection. HybridCell begins by detecting all active carriers in its vicinity. This detection is necessary for two reasons. First, it is needed in order to identify available ARFCNs that can be used by the local cell without interfering with the commercial carriers. Second, the detection determines the technologies (GSM, 3G or LTE) used by the commercial carriers, which in turn informs the network characterization. For the purpose of detection we implement a two-pronged approach: 1) *blind service identification*; and 2) *cellular-aware incumbent detection*. Our techniques provide information about all of the nearby base stations (including technology and ARFCN) that can be augmented by HybridCell should they become overloaded.

Blind service identification. Three commonly proposed spectrum sensing methods for blind service identification are energy detection, matched filter detection, and feature detection [75]. Energy detection cannot discern the underlying technology behind a signal and matched filter requires a priori knowledge of the signal for which we are searching. Hence, they are not good choices for our solution. On the other hand, feature detection is able to exploit the known periodicity (i.e. cyclostationary characteristics) of the target

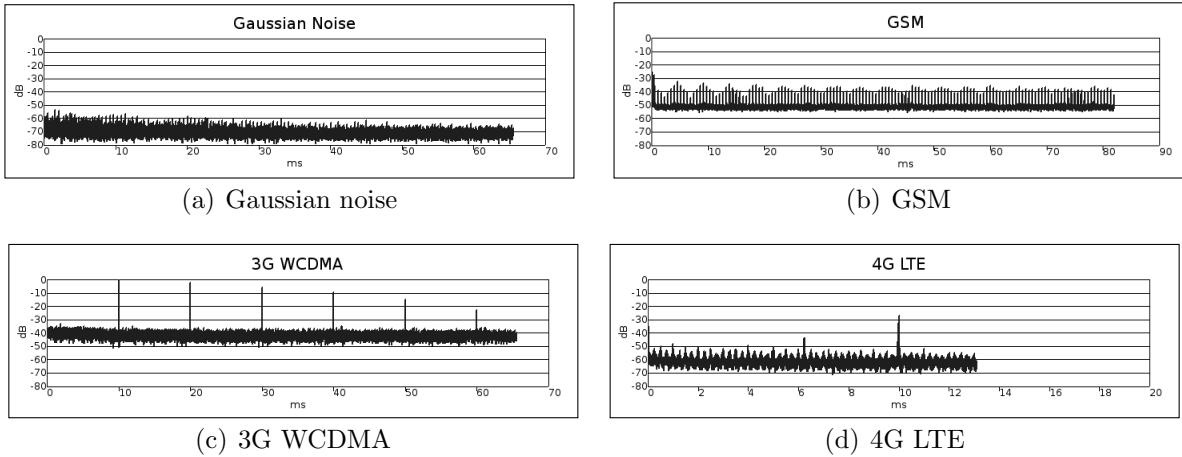


Figure 7.6: Peaks with fixed periodicity are observed in the cyclic autocorrelation function, corresponding to timeslots used by cellular synchronization channels.

signals. The pilot signals present in cellular standards result in our ability to use a single detection scheme for GSM, 3G, and 4G LTE. Furthermore, cyclostationary detection is more robust in noisy environments than simple energy detection, making it ideally suited for use in HybridCell.

We implement blind service identification using an Ettus Research USRP2 and Gnu-Radio to generate the cyclic autocorrelation function (CAF) values for each signal type. Figure 7.6 shows the CAF for Gaussian noise and different types of cellular signals. Peaks in the figures mean that the observed signal exhibits periodicity at the corresponding x-axis (time) value. The Gaussian noise plot (Figure 7.6(a)) demonstrates the case of a channel with no service provided. As expected, we do not see any clear peaks in the figure as there is no periodicity. In contrast, as shown in Figure 7.6(b), the CAF of a GSM signal includes many peaks at the TDMA time slot interval of $576.9\mu\text{s}$ [21]. We search for peaks corresponding to the synchronization channel (SCH), which is transmitted every 10th TDMA frame (8 TDMA slots), resulting in a periodicity of $576.9\mu\text{s} \cdot 8 \cdot 10 = 46.15\text{ms}$. We observe similar clear pilot signals in 3G and 4G LTE synchronization channels at 10ms

(Figures 7.6(c) and 7.6(d)). Since 3G and 4G LTE have the same pilot interval, we must apply further logic to the detector to distinguish between the two. We can differentiate 4G LTE signals from 3G by detecting the OFDM subcarriers spaced 15KHz apart within an LTE signal; this can easily be accomplished using a fast Fourier transform and a threshold.

We do not anticipate a high churn rate in terms of neighboring base stations. Therefore, blind service identification can be run with low periodicity, reducing processing burden and allowing the software defined radio to be utilized to observe and characterize commercial cells between scans. Once the analyzer has finalized its blind server identification, it connects to the identified cells in order to collect control channel messages for further characterization as detailed below.

Cellular-aware incumbent detection. Blind identification suffers from the hidden terminal problem, whereby base stations that are beyond the sensing range of the network analyzer's radio may not be detected. While those farther base stations will not be augmented by HybridCell, we must ensure that HybridCell does not interfere with their frequencies. Cellular-aware incumbent detection is thus used to determine the ARFCNs of those farther base stations. HybridCell observes System Information Types 1 and 2 (Table 7.1) messages broadcast in the beginning of every multiframe by nearby, detected commercial cells, and extracts *cell channel descriptions* and *neighbor cell descriptions* from those messages. Neighbor cell descriptions include ARFCNs of nearby base stations in order to enable mobility between base stations; phones constantly monitor the list of frequencies and select the base station with the highest signal strength. By combining the list of ARFCNs generated by our blind service identification scans with ARFCN lists observed in commercial base station system information messages, HybridCell avoids frequencies used by detectable commercial cellular base stations *and* all frequencies used by their neighbors. In this way, HybridCell avoids all frequencies that a carrier *announces*.

they are using in the local area, rather than just all frequencies HybridCell *observes* a carrier using. This mitigates the hidden terminal problem for HybridCell, making it less likely HybridCell will interfere with base stations it cannot directly detect.

By combining blind service identification and cellular-aware incumbent detection, HybridCell can utilize information from a software defined radio and commercial cellular base station broadcasts to quickly and efficiently identify incumbents. Due to the large amount of channel occupancy information periodically broadcast in GSM System Information messages, HybridCell can operate in licensed spectrum without causing interference, and without relying on a centralized spectrum occupancy database. A feasible path to deploying HybridCell is to use licensed spectrum with the permission of the incumbent(s), who would benefit from reduced congestion due to a HybridCell deployment. As spectrum policies evolve, HybridCell may be permitted to operate as a secondary user of licensed spectrum, or to operate in a ‘general authorized access’ tier such as the FCC has recently proposed for the 3.55-3.7 GHz range [11].

Commercial carrier characterization. We base the characterization of commercial carrier performance on messages exchanged on the Broadcast Control Channel (BCCH) and Common Control Channel (CCCH). In this section we first discuss trade-offs of different carrier characterization approaches. We then give a brief overview of the BCCH and CCCH messages. Finally, we detail how we use these messages to characterize active carriers’ performance. While carrier characterization can also be performed at the end user device, we choose a dedicated device approach (i.e. at the network analyzer) due to the specific system requirements of carrier characterization, as described in our implementation (Sec. 7.2.1). We note that our analyzer-based approach may lead to inaccuracies if the analyzers’ view of the network is different than that of the end users’ phones. For the envisioned use cases, however, we anticipate that both the analyzer and the end users will be residing in the same commercial cell, thus the analyzer-based

Broadcast Message	Information used by HybridCell
System Information Type 1	Cell Channel Description: List of ARFCNs Band Indicator Access Control Classes (ACC)
System Information Type 2	Neighbor cell descriptions: List of ARFCNs Access Control Class (ACC)
System Information Type 3 and Type 4	Mobile Country Code (MCC) Mobile Network Code (MNC) Location Area Code (LAC) Cell Identity (CI) Access Control Class (ACC)
Immediate Assignment	Timeslot Single channel ARFCN
Immediate Assignment Reject	Wait Indication Request Reference

Table 7.1: System messages used by HybridCell.

characterization will be accurate from the end user's view point.

Control channel overview. Cellular phones or software defined radios can be used to collect System Information Messages broadcast by nearby cellular networks. This message collection is non-intrusive and non-invasive: all captured messages are broadcast by base stations in plain text on control channels and are intended to be received and processed by all phones associated with a given base station.

HybridCell makes use of the messages listed in Table 7.1. System Information Messages types 1-4 are broadcast on the Broadcast Control Channel (BCCH) and are akin to Wi-Fi beacons, providing basic information about the base station configuration and the services it offers. The messages include frequencies occupied by the base stations as well as those of neighboring cells, access classes to notify the handset of the level of service available, and the mobile network and country codes. HybridCell uses these messages to identify and characterize the nearby base stations. Immediate assignment messages, sent on the Common Control Channel (CCCH), indicate whether the base station is able to reserve resources for a mobile device (e.g. assignment of a voice traffic channel for a

call). We leverage these messages to estimate the health of the observed base station by recording the number of successful immediate assignments as well as rejection messages.

Congested and failed cell detection. A critical function of the system is the ability to characterize service availability and quality of service of nearby commercial base stations through passive observation. To this end, HybridCell uses observations of channel assignment broadcasts to determine the quality and reliability of provided cellular service. We estimate the operational state of commercial networks and configure local cells accordingly, shifting traffic to local cells when commercial networks are congested.

Several messages are indicative of network overload. For example, overloaded base stations may attempt to reduce load on the system by preventing users from using the network, by barring one or more *access classes*. As shown in Table 7.1, access class settings are broadcast in system information messages. Each of the bits in the Access Control Class (ACC) field of a System Information message represents a class of users. Base stations can block one or more Access Classes from connecting by modifying the ACC value broadcast in System Information messages [20]. When HybridCell detects access class restrictions, it learns that the observed cell is overloaded.

Another way to discern an overloaded network is through monitoring *Immediate Assignment Reject* messages. A cellular base station operating at full capacity that cannot allocate radio resources to serve a user will issue Immediate Assignment Reject messages. In the GSM 04.08 specification, Immediate Assignment Reject messages are defined to indicate that no channel is available for assignment [20]. The link between available channels and Immediate Assignment Reject messages makes this message an excellent indication that a base station is overloaded.

We use observed radio resource management messages to infer congestion. Our system estimates congestion every minute, based on the exponentially weighted moving average of channel assignment success rate. Our goal is to estimate the likelihood of a user

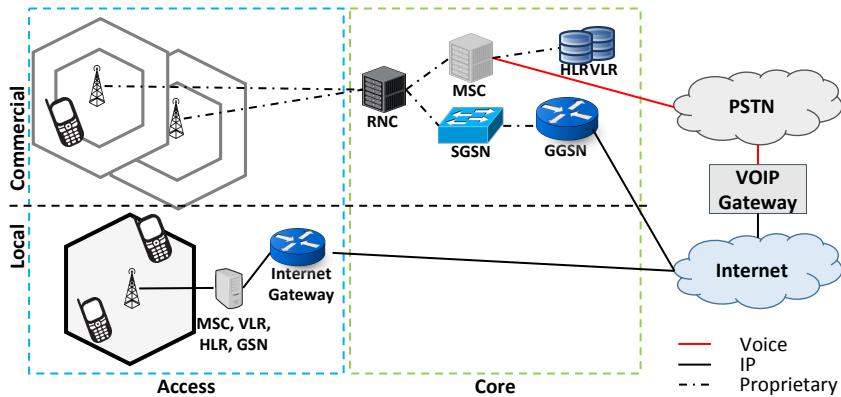


Figure 7.7: Comparison of local and commercial cellular network architectures. Key advantages of local cells include their highly-distributed nature, and use of free software, open hardware and generic backbone for interconnects.

being allocated a channel when they request one. Specifically, we define a *Channel Availability* metric based on the ratio of observed Immediate Assignment messages and Immediate Assignment Reject messages. Let $\alpha \in [0, 1]$ be a weighting factor for previous measurements where higher α means that past observations will be given more weight in estimating channel availability; let ρ be the number of Immediate Assignment Reject messages observed since the last calculation; let χ be the number of successful Immediate Assignments observed since the last calculation; and let Ψ_m be the estimated availability at time m which is defined as:

$$\Psi_m = 1 - ((\alpha \times \Psi_{m-1}) + \left((1 - \alpha) \times \frac{\rho}{\rho + \chi} \right))$$

Intuitively, Ψ_m varies between 0 and 1, where $\Psi_m = 0$ indicates a dysfunctional network, where each attempt to connect to the base station receives an Immediate Assignment Reject, while $\Psi_m = 1$ characterizes a fully-functional network with no occurring Immediate Assignment Rejects.

We analyze the Channel Availability of all three Jordanian cellular network operators by using our Za'atari traces to compute and evaluate this metric in Section 7.2.2.

Network Analyzer implementation. We design the network analyzer to perform blind service identification as well as commercial carrier characterization without the use of carrier SIM cards. We use affordable software-defined radios, including Nuand BladeRFs [3]. We leverage the existing open-source tool `gr-gsm` [4] to tune the SDR to the commercial base station frequency, decode GSM control channel messages, and output the messages using the GSMTAP [5] pseudoheader format which is easily parseable by network protocol analyzers.

Local Cells

Local cellular networks, based on open source software and open hardware, have gained traction as affordable means to provide cellular connectivity in infrastructure-challenged environments [206]. Recent deployments [88, 2] operate in areas without, or with limited existing commercial coverage in order to satisfy residents' communication needs. Key benefits of these local cells, as depicted in Figure 7.7, are their highly-distributed nature, and use of free software, open hardware and generic IP backbone for interconnects. As shown in the figure, local cells push mobile core functionality to the network edge by running local software processes that perform the functions of corresponding GSM architecture entities (e.g. Mobile services Switching Center (MSC), Home Location Register (HLR), etc.), whereas commercial networks' architecture is highly-centralized and relies on proprietary backhaul links to connect the edge with such GSM entities residing in the carrier core. The commercial network backhaul requirements, along with RF frontend investments, often render commercial deployments in infrastructure-challenged environments economically-infeasible [14]. Local cellular networks, in turn, are often realized as a cellular-network-in-a-box (Figure 7.8). As such, both the radio-frontend and the core components reside in the access network areas, making the design highly-distributed and more feasible for remote deployment.

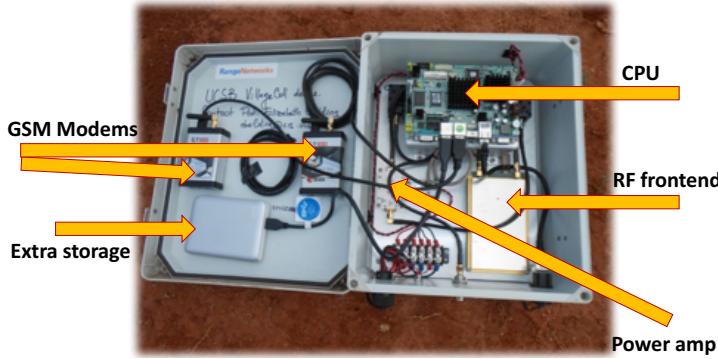


Figure 7.8: Example of cellular-network-in-a-box extended with modems for active testing and additional storage, used for the authors deployment in [206].

HybridCell is intended for scenarios where global connectivity is unreliable due to user demand or lack of sufficient commercial infrastructure. Therefore, we leverage local cellular networks as they do not rely on always-on global connectivity. In areas with mixed local and commercial networks, HybridCell augments cellular coverage by moving users between the commercial network and the local network when it detects congestion or failure of the commercial network. This behavior eases congestion by shifting some of the commercial network load to the local network. In the case of complete failure, the local network provides an alternate means of connectivity. We leverage existing local cellular network technologies in order to provide local cellular connectivity without requiring a reliable Internet connection, a key limitation of femtocells.

Depending on antenna location and height, local base stations with as little as 1 watt of amplification can provide coverage with a radius of a few kilometers. In densely populated areas, a number of local cells with reduced coverage can be deployed to provide additional capacity. The local network can easily be extended by interconnecting base stations using point-to-point Wi-Fi infrastructure, as voice and SMS traffic is encapsulated by standard UDP packets.

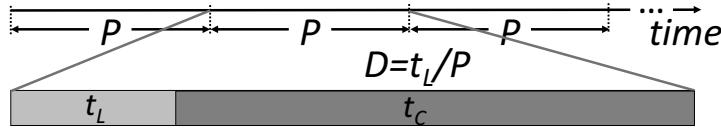


Figure 7.9: Frame structure of CellSwitch.

CellSwitch Android Application

The final component of HybridCell is CellSwitch, an Android application that allows a user's mobile phone to switch to specific nearby cellular networks without user intervention. We target Android, as affordable Android-based smartphones are widely available and popular throughout the world, including, as we observed, in the Za'atari camp. Our application uses an Android system function in the telephony framework's GSMPhone class to instruct the baseband processor to register on a specific network. This function allows the application to programmatically *duty cycle* between cellular networks without requiring user involvement. Tradeoffs inherent in configuring the behavior of CellSwitch are based on balancing user reachability with power consumption, as detailed below.

CellSwitch network switching protocol. CellSwitch employs duty cycling to define the amount of time users spend connected to the commercial and local networks. As illustrated in Figure 7.9, our application divides the time into equal periods P ; within each period it switches between the commercial and the local cell according to its duty cycle D . The period P is equal to the sum of the *time spent on local network* (t_L) and the *time spent on commercial network* (t_C), and defines how often the cycle repeats. The duty cycle D , in turn, is defined as the ratio of time spent on the local network over the entire period: $D = t_L/P$.

Duty cycle percentage. HybridCell adaptively adjusts the duty cycle percentage based on estimated channel availability (Ψ) observations made by the network analyzer detailed in Section 7.2.1. For instance, a Ψ value of 0.2 will result in users spending 20%

of P connected to the commercial network and 80% of P connected to the local network. The analyzer communicates the congestion level to the CellSwitch users via the local network. Thus, CellSwitch requires each phone to spend a minimum amount of time t_{Lmin} on the local network within each cycle for the congestion updates to be received. As the congestion updates are a single floating point value, the minimum time for transfer is a fraction of a second, even assuming relatively slow GPRS throughputs of roughly 35Kbps. We test the time to authenticate and join a local network in Section 7.2.1. More users connected to the local network will lead to an increased proportion of calls and SMS messages between pairs of users that happen on the local network, offloading a portion of the load away from the commercial network. It is possible to overload a local network given enough users. This can be avoided by controlling the number of HybridCell users or supplementing local network capacity by increasing the number of local cells.

In the beginning of each period, the phone connects to the local network and later switches to the commercial network, as defined by its duty cycle. This operation requires two network switches per cycle, which incur overhead in call establishment delay and power consumption. Thus in selecting our duty cycle, we need to balance the tradeoff between network availability on one hand, and user reachability and power consumption on another. In the remainder of this section, we provide measurement-driven quantification of these two overheads and present evaluation to justify our duty cycle selection.

Key tradeoffs. Power consumption and disconnected time are both important factors when configuring the duty cycle functionality of CellSwitch. To characterize the additional power burden imposed by switching networks, and to determine the time a phone takes to move between networks, we perform several experiments. We configure three different models of Samsung Android phones to programmatically switch between a commercial network and a local network once per minute. We measure the power consumed by the phone each second during the test using an in-line USB current and

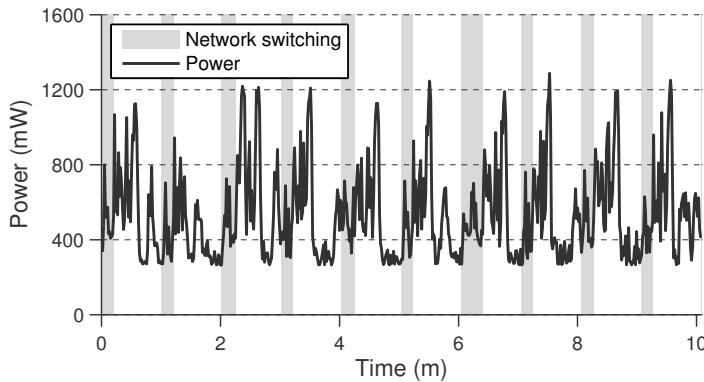


Figure 7.10: Energy consumption increases during transitions, but peaks immediately following movement between cellular networks.

voltage monitor based on the Texas Instruments INA219 DC current sensor. Our test devices were fully charged before testing to minimize the impact of charging the battery on power consumption. We create an Android test driver application that triggers transitions between networks. Our application registers a PhoneStateListener, allowing it to receive and log notifications when the phone goes in and out of service. Our test driver logs the current time in milliseconds when transitions are triggered and completed. The phones are synced to the same Network Time Protocol server as the computer recording power measurements to allow for accurate alignment of power logs with user transition logs. We force phones to connect to 2G and 3G commercial networks, allowing us to examine the impact of commercial cell technology on transition time and power.

Impact of network switching on power consumption. We first explore how switching networks affects phone power consumption. As shown in Figure 7.10, our experiments measured an increase in power consumption during the transition between networks. The power consumed during a migration is roughly double the idle power consumption, but, interestingly, consumption remains high for up to 30 seconds after joining a network, with peaks at almost four times idle power consumption. This result echoes prior findings that mobile radios remain in high power states for some time after

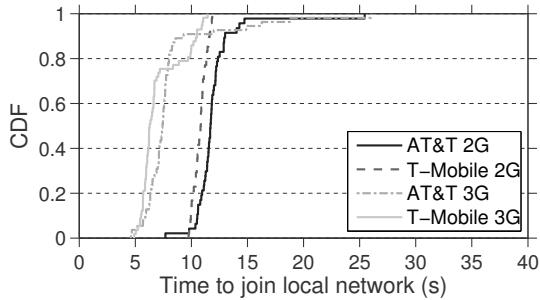


Figure 7.11: Time spent without service when moving to the local network from commercial networks.

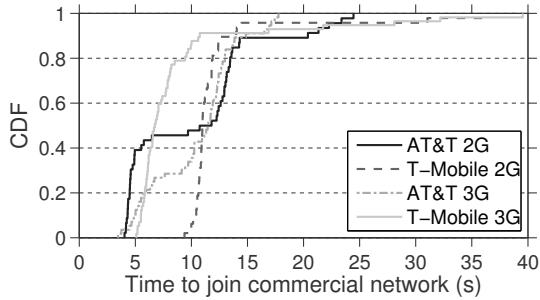


Figure 7.12: Time spent without service when moving to commercial networks from a local cell.

usage [35, 146] in an effort to avoid the latency penalty incurred when forced to transition from idle to active. Such power behavior must be taken into account when configuring client duty cycling in HybridCell.

Impact of network switching on disconnected time. During the transition between networks, a phone is disconnected from both networks; hence the frequency, determined by the duty cycle period, of shifting between networks must be kept relatively low.

To understand how the cellular technology used by commercial carriers and ‘direction’ of transition impacts disconnected time, we perform migration experiments in both ingress and egress directions, with phones configured to prefer 2G or 3G networks. Figure 7.11 shows transition times for phones moving from the two major U.S. commercial GSM networks to a local GSM network. We observe distinct performance differences be-

tween 2G and 3G devices, with 3G transition times roughly half of 2G transition times. We also explore times for phone transitions in the opposite direction, from the local network to the commercial network. Figure 7.12 shows a cumulative distribution function (CDF) of time required to join the commercial networks from the local network. We observe slightly improved best case performance compared to the ingress case in Figure 7.11. Interestingly, for AT&T 2G we observe two distinct groupings at 4-5 and 11-12 seconds. We posit that this could indicate the connection attempt was delayed or experienced retransmissions in the longer cases. Overall, we see transition times are typically between 4 and 7 seconds for 3G users, while 2G users most often spend approximately 10 to 12 seconds disconnected during a transition.

These results must inform the chosen duty cycle period P , as shorter lengths will result in a larger percentage of time spent disconnected. An extremely short P (e.g. 60 seconds) will cause a user to be unreachable for up to 120 4-12 second intervals over the course of an hour, or up to 33% of the time. As P increases, disconnected time decreases. For a 3G device that averages a 5.5 second transition time, $P=120$ seconds results in 9.16% disconnected time, while $P=200$ seconds results in 5.5% disconnected time and $P=300$ s decreases disconnected time to 3.66%.

Duty cycle period. The length of the duty cycle period P dictates the maximum length of time between leaving a network and returning to it. That is, a handset will join both the local and commercial networks once in each period. User phones must connect to the local network at least briefly each period to check for queued SMS messages and to learn the most recent estimated availability in order to adjust their duty cycle percentage accordingly.

Therefore, the P value is an upper bound on call initialization and SMS delivery latency, as calls can begin when both users are on the same network and SMS messages are queued until the recipient connects to the network. An effective cycle period can be

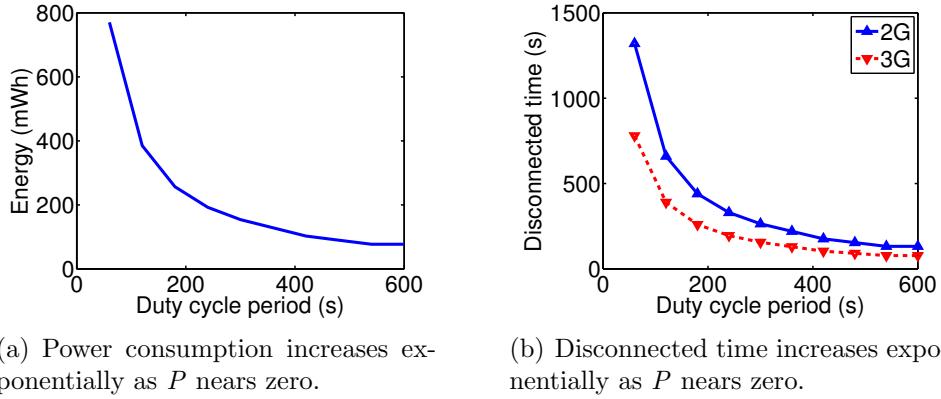


Figure 7.13: As the duty cycle period P is increased, the energy burden and the time a phone spends disconnected from both networks is decreased.

thought of as one in which the system minimizes call initialization or SMS delivery delays while balancing the constraints imposed by limited battery life offered by a phone.

We explore the relationships between the duty cycle period P and power consumption, as well as between P and disconnected time, in Figures 7.13(a) and 7.13(b), respectively. Based on our empirical findings in Figures 7.11 and 7.12, we calculate the power consumption and disconnected time for a phone for one hour as the period P grows from 60 to 600 seconds. We calculate the disconnected time as the sum of time intervals a phone spends not connected to any network, and the power consumption as the sum of energy drawn from the battery every time the phone switches networks. Our results show that shorter P values incur more power consumption and disconnected time within an hour, compared to longer P . Thus, a short P value (less than 200 seconds) may reduce the overall battery life of a device beyond what is likely to be acceptable by many users. As the period increases above 200 seconds, the steepness of the power consumption curve decreases, and variations in P length have less impact on the aggregate energy draw and, in turn, on the device battery life. Our disconnect time results (Figure 7.13(b)) indicate similar trends for power consumption. Thus, these combined results suggest that cycle lengths of 200-300 seconds provide satisfactory balance of power consumption, call

initialization delay, and disconnected time.

Detailed HybridCell operation

HybridCell reduces the load on nearby overburdened commercial networks by adaptively duty cycling users onto a local network. This dynamic leads to a number of potential use cases depending on the current home networks of the caller and the callee. A participating phone's SMS or voice call is attempted on the network that the phone is associated with at the time of the attempt. As a result, there are four possibilities at the moment a voice / SMS event is triggered:

1. a local-local (L-L) call, where both users are connected to the local network, in which a call or SMS proceeds immediately with no delay;
2. a commercial-commercial (C-C) call, where both users are connected to the commercial networks, and a call or SMS proceeds on the commercial network with no delay;
3. a commercial-local (C-L) call, where the caller is on the commercial network while the callee is on the local network. The call will be placed immediately but will result in a 'miss' due to the callee's local association; the caller will be sent to voicemail. An SMS will be sent and queued by the commercial carrier. The recipient will receive the queued message, or voicemail notification, when they connect to the commercial carrier during their duty cycle; and
4. a local-commercial (L-C) call, where the caller is on the local network and the callee is on the commercial network. When the call is placed the local network will place the caller on 'hold' until the callee associates with the local network. Alternatively, an SMS message sent by the local user will be queued on the local cell and will be

delivered when the callee associates with the local cell in their duty cycle.

HybridCell therefore introduces call initialization and SMS latencies as well as the possibility of missed calls for its users. We are particularly interested in L-C and C-L events as they result in latencies not present in traditional cellular systems. However, the system is designed for deployment in areas where calls and SMS messages are failing due to overburdened networks. Thus, though there may be a delay, HybridCell can provide successful communication during times when the commercial network cannot. When the commercial network is operating well, the duty cycle can be set such that users spend only a few seconds on the local cell each period to gather Ψ updates and queued SMS messages. This operation will result in minimal perceived delay, and in turn, seamless user interaction with the commercial network.

7.2.2 HybridCell Evaluation

In this section, we leverage our data collected in an active refugee camp to evaluate critical components of HybridCell. We combine in-situ measurements from Za'atari with a simulation environment to examine the HybridCell user experience. The simulation is required as an actual deployment in Za'atari is not feasible given regulatory and pragmatic constraints.

Data set: Za'atari refugee camp

For our evaluation we leverage the cellular message dataset that was collected in the Za'atari refugee camp in Jordan in January 2015 as detailed in Chapter 3. We used software-defined radios and mobile phones to record raw spectrum measurements as well as cellular broadcast messages for 2G and 3G service on all three cellular providers offering service in the camp. In addition, our team conducted interviews and administered

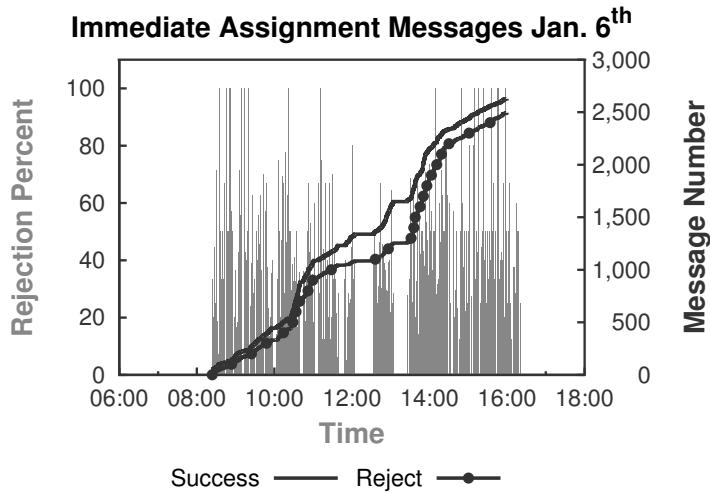


Figure 7.14: Immediate assignment rejection percentage per minute, and successful and failed immediate assignment messages collected from Zain over 1 day in Za’atari.

a survey on mobile phone and Internet use to camp residents.

Network analyzer congestion detection

We evaluate HybridCell’s network analyzer algorithm by computing the estimated availability (Ψ) for each of the three Jordanian cellular carriers using the data we collected in Za’atari. First, we explore how the selection of α values impacts our congestion metric using data from Zain, the most congested and most popular carrier. Then, we compare the results of our congestion detection algorithm for all three carriers using the same α value for each carrier. This demonstrates the efficacy of our metric in differentiating between heavily congested, sometimes congested, and rarely congested networks.

Using one minute bins, Figure 7.14 shows observed Immediate Assignment messages and Immediate Assignment Rejects collected over the course of one day for Zain, as well as the percent of immediate assignment messages that were rejections during our measurement window. We see different rates and percentages of rejection messages throughout the day. This result points out the need for flexible, adaptive offloading as congestion is

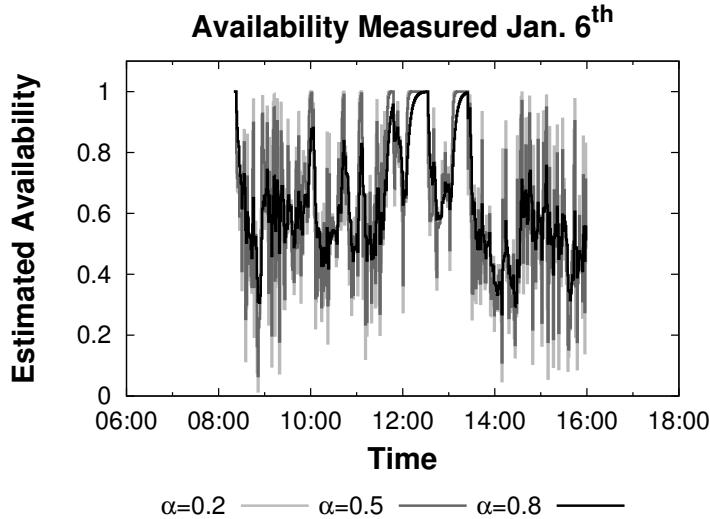


Figure 7.15: Estimated availability for Zain with different α values.

not constant. We calculate Ψ over time for the carrier using a range of α values, from 0.2-0.8. Figure 7.15 shows the computed Ψ over the course of one day for Zain with each α value. As expected, smaller α values result in higher estimated availability variance as less weight is assigned to history. Our metric succeeds in detecting congestion in the data presented in Figure 7.14; we see low availability values during time windows of high rejection percentage, and a return to high values during windows of few rejections (e.g. 12:00). Higher α values, such as 0.8, do not impact the detectability of congestion events, but reduce the variance of Ψ . A potential drawback of using small α values, thus resulting in rapid fluctuations, is that Ψ could increase rapidly after a congestion event ends, causing HybridCell to quickly shift many users to the recently congested network, potentially causing further congestion. This would result in unnecessary and frequent migration between networks, with recent estimations no longer accurately detecting congestion. Smoothing out large fluctuations in the availability metric is desirable for HybridCell, as channel availability is used to configure duty cycling. From these measurements, we determine that our metric for Estimated Availability with $\alpha = 0.8$ is satisfactory in

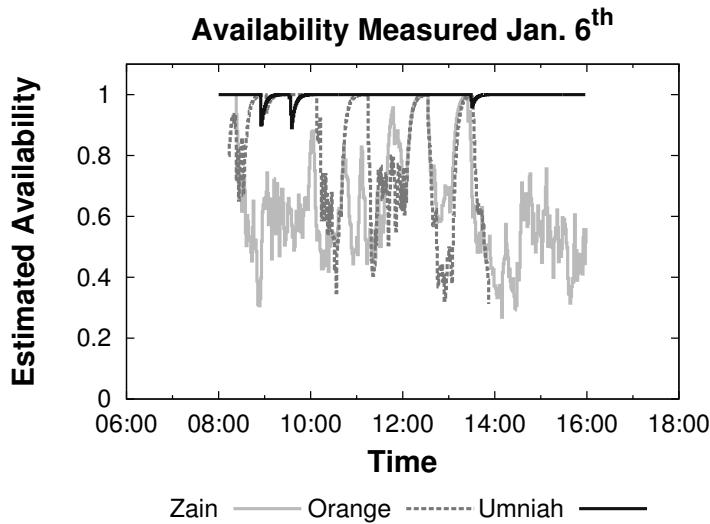


Figure 7.16: Estimated availability for all carriers ($\alpha = 0.8$).

balancing responsiveness with congestion detection.

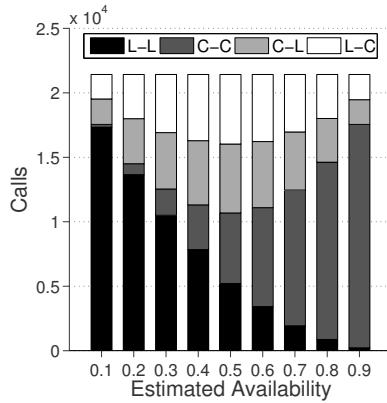
Using $\alpha = 0.8$ our algorithms detected several periods of congestion for each carrier. Zain suffered from the most congestion. This follows logically from our survey data, which indicated that Zain is by far the most popular carrier in the camp. Figure 7.14 shows that of five thousand channel assignment messages observed on Zain, successful channel assignments barely outnumbered rejected channel assignments collected during the same time period. As shown in Figure 7.16, our congestion metric identified continuous congestion throughout the course of one day in the camp for Zain. Orange suffered from both ephemeral and prolonged congestion events, but generally recovered between incidents. Figure 7.16 shows that Orange's Ψ fluctuates throughout the observation, again corresponding roughly to busy hours impacted by the workday, dropping as low as 0.4, but typically returning to 1.0 within an hour. Umniah is the least congested network based on the frequency of Immediate Assignment Reject messages in our traces, which may because it is the least popular carrier according to our survey. Over our entire collection period in Za'atari, Immediate Assignment Reject messages totaled 5.9%

of channel assignment messages for Umniah, compared to 33.7% for Zain and 15.2% for Orange. As shown in Figure 7.16, Channel Availability is near 1 for Umniah throughout the test period, and never drops below 0.8. A HybridCell user of Umniah therefore would spend the least time on the local cell, often only the minimum time required to receive updates and queued SMS messages.

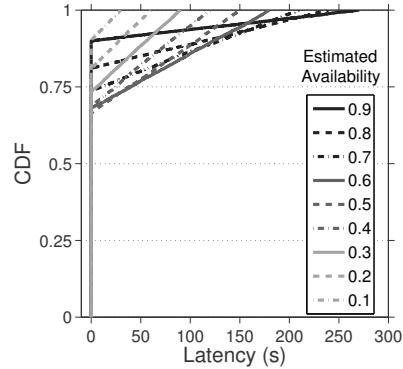
Simulation

Lastly, we evaluate the offloading behavior and expected user experience for HybridCell via simulation. We parse the cellular traces gathered in Za’atari to record timestamps for 21,426 Immediate Assignment success and reject events that represent a phone requesting a resource (e.g. to place a call or SMS). For simplicity, in our simulation we treat all events, both success and reject, as requests for voice calls. We run two simulations in order to understand: 1) the anticipated effect of different values calculated by the availability metric on system usage; and 2) the behavior of HyridCell in a real-world environment.

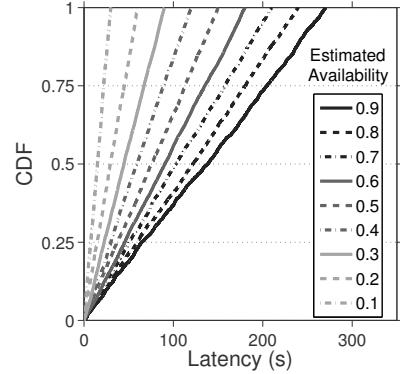
Effect of availability metric. We first seek to learn the impact of the availability metric on system usage and end-user experience. We are particularly interested in the locations of caller and callee at the time of a call as well as any latency introduced by HybridCell in L-C or C-L events. Our first simulation includes two pools of 500 users, callers and callees, with user duty cycle start and end times uniformly distributed across the duty cycle period P as CellSwitch timers are kept by each phone running CellSwitch and are not centrally synchronized. For simplicity, we assume that all users are HybridCell users, meaning they have access to the local network. For each immediate assignment event in the trace, we randomly select a caller-callee pair and categorize the pair based on the users’ associated networks at the time of the event as one of L-L, C-C, C-L, or L-C as described in Section 7.2.1.



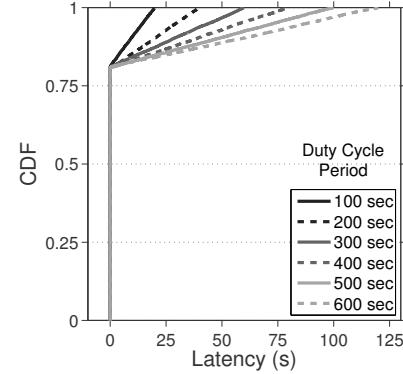
(a) Event designations with various estimated availability. As the commercial network is less available, more traffic is shifted to the local network.



(b) Latency values for non-miss events. Higher availability results in longer latencies.



(c) Latency values for L-C events. Higher availability results in longer latencies.



(d) Duty cycle period comparison for networks with 0.2 estimated availability.

Figure 7.17: Simulation results with statically set estimated availability.

Figure 7.17(a) shows the breakdown for simulated events with varying levels of estimated availability and a duty cycle period set to 300 seconds. As expected, decreased availability leads to increased local network usage. We also see that mixed events (L-C or C-L) are most prevalent when clients spend 50% of their time on both networks. ‘Misses’ account for roughly 25% of calls in the worst case; however, this case would be caused by an estimation of 50% availability on a commercial network. In other words, we would observe a rejection rate of 50% for all requests if our system did not exist.

HybridCell's presence in such a situation, on the other hand, would result in shifting 50% of call origination to the local cell (25% L-L plus 25% L-C calls), leaving 25% solely on the commercial network (C-C). Overall, we expect 75% of calls will succeed without a miss in this scenario, an overall increase of 25% compared with 50% failure.

Figure 7.17(b) shows a CDF of latencies for all ‘non-miss’ events (e.g. L-L, L-C, C-C). We see that a large percentage of events experience no latency, corresponding to L-L or C-C events. L-C events cause latencies greater than zero; different congestion estimates clearly affect user latency. We single out L-C events in Figure 7.17(c) and observe that as estimated availability decreases, users spend more time on the local cell, leading to shorter latencies. Lastly, we run the simulation with various duty cycle periods with a constant 0.2 estimated availability. Figure 7.17(d) shows the resulting CDF and the effect of cycle time on latencies. These results can be used to inform duty cycle configuration as each situation in which HybridCell can be deployed is unique. As noted previously, short cycle times will result in excessive power consumption. In areas where low latencies are most critical and device battery life may not be a concern, short P values can be selected. On the other hand, where latency is less important and battery life must be maximized, long P values work best.

HybridCell augmentation of real-world networks. Next, we simulate the presence of HybridCell in Za’atari in order to understand system usage and impact on end-user experience. We split a pool of 83,500 users (the estimated population of Za’atari) among the three commercial carriers based on the results of our user surveys. 74.1% (61873) are assigned to Zain, 17.1% (14279) to Orange, and 8.8% (7348) Umniah. We use 300 second duty cycles for all users. We use the immediate assignment events from our trace to calculate the availability for all three networks and adjust duty cycle percentages based on our findings. For this simulation we again assume that all immediate assignments, successes and rejections, are voice calls. We randomly select caller and

		Call Classification				
		L-L	C-C	C-L	L-C	Total
Caller-Callee	Z-Z	1,586	9,980	121	120	11,807
	Z-O	46	1,166	8	133	1,353
	Z-U	97	2,303	7	273	2,680
	O-Z	33	1,219	159	16	1,427
	O-O	2	156	1	1	160
	O-U	6	311	4	1	322
	U-Z	88	2,316	278	14	2,696
	U-O	9	332	1	3	345
	U-U	25	607	3	1	636

Table 7.2: Trace-driven simulation call totals.

callee pairs from the user pool and record the call designation (L-L, C-C, C-C, L-C) as well as carriers involved and any latencies introduced by L-C or C-L calls.

The resulting breakdown of calls, along with caller and callee designations is shown in Table 7.2. We observe that, as expected, the majority of calls involve Zain due to its popularity. We see that the majority (85.83%) of all calls are still placed solely on commercial networks. This is expected given the estimated availability shown in Figure 7.16. 8.83% of calls are L-L, meaning the burden of those calls has been entirely removed from the commercial networks. We also see a relatively small number of L-C and C-L events (2.62% and 2.72%, respectively).

Overall L-C latency. We plot L-C latency distributions for the three carriers in Figure 7.18 with the tops and bottoms of the boxes representing the 25th and 75th percentiles, respectively and the horizontal line the median for each distribution. Calls originated by Zain users have the highest median latency at roughly 205 seconds, while Orange callers have the lowest. We see that Zain and Umniah have wide distributions of latency, whereas Orange experienced a much smaller range. This can be partly attributed to small number of L-C calls in Table 7.2, as well as the availabilities shown for the carriers in Figure 7.16. Orange experiences brief, severe periods of roughly 50% availability at the same time as Zain, which should result in latencies of roughly half the duty cycle

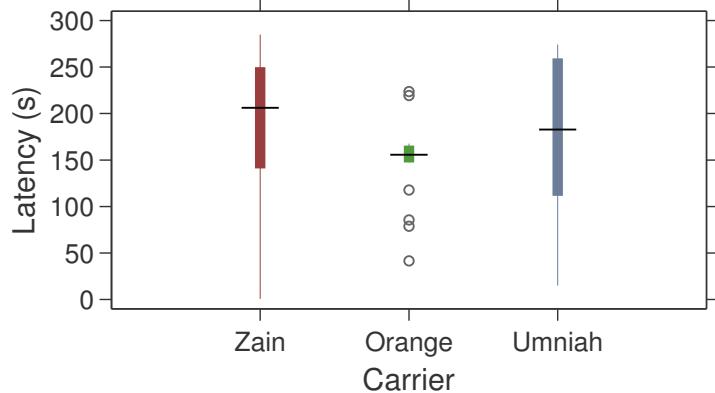


Figure 7.18: L-C Latency for callers assigned to different carriers.

period P , or 150 seconds. Availability for Umniah, on the other hand, remains rather high throughout the observation period and does not appear synchronous with the other carriers, leading to a broad range of potential latency values.

L-C latency with caller-callee pairs identified. The previous results, with Zain callers experiencing the highest median latency, are less than ideal. In order to understand the underlying cause, we plot L-C latency distributions for all caller-callee pairs in Figure 7.19. We observe that L-C latencies between Zain users are roughly 110 seconds, less than half of the duty-cycle period, while Zain-Orange and Zain-Umniah calls have much higher median latencies. This plot illustrates a challenge for HybridCell and an opportunity for improved design. When caller and callee are from the same carrier (e.g. Zain-Zain), the duty-cycles of that carrier will be the same for all users, resulting in low L-C latencies when the commercial network experiences low availability as predicted in Figure 7.17(c). Our selection of random caller-callee pairs results in a *worst-case scenario* for HybridCell, as the availability for the three carriers is independent (as shown in Figure 7.16). While one carrier may be suffering from low availability at a given point in time, resulting in a higher percentage of calls from that carrier's users placed originated on the local network, callees from the other carriers, which could have high availability,

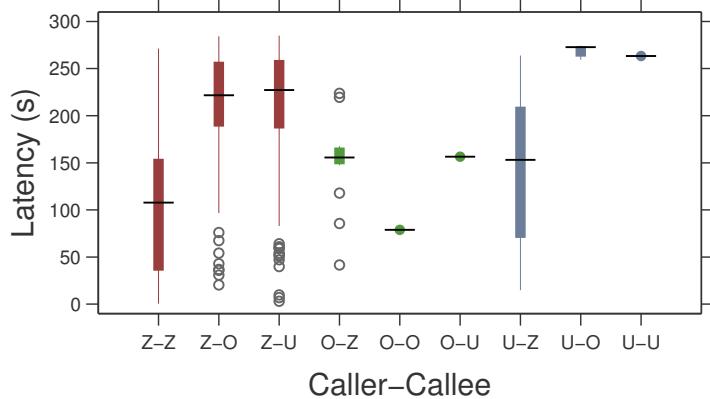


Figure 7.19: L-C Latency with caller-callee designations.

may be using duty cycles where they spend the majority of time on the commercial network. Intuitively, we anticipate that callers and callees are *not* random, families and social groups may tend to use a single carrier, thus avoiding the potential synchrony variance. This worst-case scenario is illustrated by L-C and C-L events with caller-callee pairs involving Zain and the other two carriers. For instance, the highest percentage of C-L calls are between Orange-Zain and Umniah-Zain. Because Zain experienced sustained congestion throughout our observation period, Zain users are likely to spend a higher proportion of their duty cycle connected to the local network. Conversely, users of the other carriers, which have generally higher estimated availabilities, are more likely to be connected to the commercial network.

This finding exposes an opportunity for design improvement in HybridCell. There are several potential avenues to pursue in order to solve this problem. Carrier-specific actions, such as CellSwitch maintaining availability state of all area carriers and triggering specific phone behaviors based on the callee carrier, may be warranted. We leave this for future exploration.

C-L latency with caller-callee pairs identified. Recall from Section 7.2.1 that C-L events will result in a queued SMS or a voicemail message as the callee is not globally

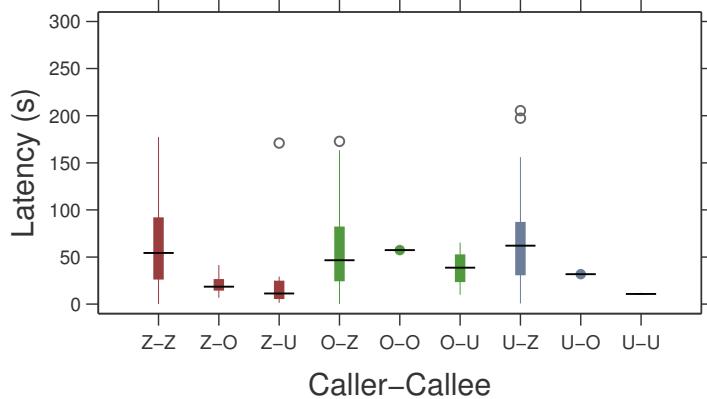


Figure 7.20: C-L Latency with caller-callee designations.

available at the time of the event. We plot the latencies for the callees to receive an SMS or voicemail notification in Figure 7.20. Across all three carriers, C-L latencies are lower than L-C latencies. This is expected, as the estimated availability of the commercial carriers tends to be above 0.5, meaning users generally spend more time on their commercial carriers than the local network. It is difficult to give general observations for calls that do *not* involve Zain, given the low count as shown in Table 7.2. For calls that do involve Zain, we see two trends. First, when the callee is a Zain user (e.g. Zain-Zain, Orange-Zain, Umniah-Zain), the distributions are similar and have median values of roughly 60 seconds. This is because Zain callees, no matter the caller, have the same likelihood to be connected to the local network when a call is initiated. Second, for C-L where the caller is a Zain user and the callees are from Orange or Umniah, the latency medians are quite low at roughly 25 seconds. We believe this is because Orange and Umniah were generally more available than Zain, so callees of those networks are likely to spend the majority of their time connected to the commercial network.

Overall, our simulations show that HybridCell successfully offloads a significant percentage of traffic from the nearby commercial networks, while not introducing an unacceptable amount of latency in mixed local and commercial call events. We anticipate

that a real-world deployment would result in tuning some of the assumed settings used in our simulation (e.g. duty-cycle period, α used for estimated availability, and frequency of availability calculation) for environment-specific characteristics.

7.2.3 HybridCell Discussion

In this section we design HybridCell, a system that makes use of independent local cells to augment cellular connectivity where commercial networks are overloaded or failing. HybridCell’s mixture of local networks with commercial networks and leveraging of cellular communication locality creates a new connectivity model, offering service to the millions of people currently on the fringes of cellular connectivity.

While HybridCell builds fundamental mechanisms for network characterization and switching, there are several questions that remain open related to the operation of the system. The current design incurs call completion latency that is manageable yet larger than that of a single-network communication. This aspect could be improved by the design of a smart duty-cycling mechanism that is also informed by social graph analysis to schedule duty cycles such that time on the local network coincides with the schedules of the user’s frequent contacts. While the system currently supports SMS and voice, we are working towards adding data offloading to local cells. This will require advanced per-service network characterization and an improved suite of network switching protocols that cater to data offload. Additionally, while the current prototype includes occupied channel avoidance mechanisms to ensure that the system does not interfere with existing commercial networks, we acknowledge that HybridCell operates in licensed frequencies. We believe that with licensed shared access regulations progressing in both Europe and the U.S., a system such as HybridCell will be feasible in the near future. Given a real-world deployment, we would also be able to incorporate call data record (CDR) analysis

to illuminate areas for further optimization such as intelligently scheduled duty cycles and user synchronization.

7.3 SmartCell

Informed by our analysis in Chapter 3, we propose SmartCell, an Android application that helps users avoid congested cellular base stations. SmartCell observes GSM control channels, allowing it to passively detect when the BTS serving a user is congested. Essentially, SmartCell exposes the SDCCH blocking KPI in near real-time to end-users; such metrics are typically only known to the cellular carriers themselves.

7.3.1 SmartCell System Design

The operation of SmartCell depends on a user’s preferences and the features of their mobile device. SmartCell periodically logs a timestamp, mobile network code, cellular base station ID, approximate location, and a BTS Availability (BTSA) value based on observed cellular network congestion. If permitted by the user, SmartCell may share these measurements with other SmartCell users, by periodically uploading this information to SmartCell’s servers. By combining measurements from multiple nearby users, SmartCell enables users to make an informed choice from among their possible providers, even if their phones do not support the diagnostic interface that permits SmartCell to directly observe the GSM air interface.

For users with multi-SIM phones, SmartCell may operate automatically, relocating users away from congested base stations using the phone’s observations and observations collected by other users nearby. SmartCell can select which SIM card should be the phone’s ‘primary’ SIM, which is used for outbound calling and mobile data. If a user’s phone only supports a single SIM card, SmartCell will notify the user if it determines

that they might benefit by switching mobile networks or cells. This determination begins with SmartCell observing congestion on the user's current network above a threshold.

As congestion approaches this threshold, SmartCell will provide context to the measurements by querying its servers to determine if a user is observing worse congestion than nearby users on other cells. If congestion on the user's current cell is above average for their area, SmartCell alerts the user which carriers offer superior service in their area, and of better performing cells of their own carrier. This may prompt users to switch SIMs to use a less congested carrier, or to relocate in order to switch to a less congested base station of the same carrier.

Hardware platform

During our visit, our team surveyed 228 residents of the Za'atari refugee camp. We found that Android phones are very popular in Za'atari: 64% reported owning an Android device, 22.4% owned a Nokia device, while 4% owned an iPhone. We believe the universal availability of Android handsets (e.g. 85% global market share), powerful features and APIs, and wide range of models at varying prices, make Android an ideal platform for SmartCell.

SmartCell relies on a diagnostic interface of the phone's cellular baseband to collect the cellular broadcast messages used to detect congestion. This data collection technique has been used to detect baseband attacks, and is supported on a wide range of Android handsets with popular baseband chipsets [187].

Use of multiple networks

SmartCell relies on the use of multiple SIM cards, which is common in developing regions. Unlike areas where 'good' cellular connectivity is ubiquitous, carriers in rural and developing regions often have vastly disparate coverage and quality of service. In contrast

to the multi-year cellular contracts common in countries like the US and Japan, the availability of low cost and contract-free prepaid SIMs enables users in developing regions to use multiple cellular networks. This leads to users carrying SIM cards from multiple carriers and switching as needed to obtain acceptable connectivity. For example, the number of cellular subscriptions per 100 residents in Guatemala is 106.6, and in Jordan it is 147.8, while in the United States it is 98.4¹. Respondents to the survey administered in Za’atari use three SIM cards on average, switching SIMs to take advantage of less congested networks, ‘same network’ discounts, and cheaper data-only plans. This level of comfort in switching carriers is promising for a system such as SmartCell.

Switching SIMs traditionally requires powering off the phone. However, phones that support multiple SIM cards are increasingly popular. OpenSignal reports 57% of Android users in Guatemala have multi-SIM phones, and more than 50% of Android users in several other developing countries own dual-SIM phones². When SmartCell is used on a multi-SIM phone, physically swapping SIM cards is unnecessary. Instead, SmartCell can take the user directly to the SIM settings activity, allowing the user to select which SIM to use for voice calls, SMS messaging, and data traffic. Alternatively, SmartCell can reconfigure the phone automatically, attempting to select the least congested mobile network in the area.

SmartCell operation

SmartCell observes cellular control messages, specifically immediate assignment and immediate assignment rejection messages, to understand network availability and load on a user’s current BTS. Recall that these messages are BTS-specific and allow an attached MS to infer SDCCH blocking affecting not only itself, but also other MSs connected to

¹<http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf>

²<http://opensignal.com/reports/2015/01/android-devices-dual-sim/>

the BTS. SmartCell detects these messages and uses them to compute a *BTS Availability* (BTSA) metric. When this metric surpasses a threshold, SmartCell notifies the user that the network is congested or automatically switches the priority of a user's SIM cards to an alternate network. We use 0.9 as the threshold based on doubling the classification of 5% blocking as high congestion [115].

Availability algorithm. SmartCell estimates BTSA periodically based on an exponential weighted moving average of immediate assignment success rate, as shown in Equation 7.1. Recall that we infer SDCCH blocking from the immediate assignment success rate.

$$BTSA_t = A_t \times \alpha + (1 - \alpha) \times BTSA_{t-1} \quad (7.1)$$

Using our algorithm, we plot the BTSA for each of the carriers serving Za'atari on the morning of January 6th 2015 in Figure 7.21. We use a period of one minute and an α of 0.25. Note that α , the periodicity, and the threshold are tunable values as system responsiveness requirements may differ depending on the environment. We see that the carriers have distinct availabilities that roughly correspond to the immediate assignment rejection percentage from Chapter 3 plotted in Figure 3.2. Importantly, congestion on one carrier does not necessarily correlate with congestion on another. If the value crosses below the threshold (indicated by the red line in Figure 7.21), SmartCell informs the user that their current BTS is congested, prompting the user to switch BTSs or networks, or, in the case of a multi-SIM phone, switches to a different network automatically.

An example scenario of SmartCell's operation can be observed in Figure 7.21. A Zain customer running our application would cross the availability threshold at around 8:30 (shaded region) and be advised they may wish to switch SIMs or move to a different location. At that time, Orange also experienced congestion, though to a lesser degree, while Umniah remained congestion-free. The Zain user could switch to either network

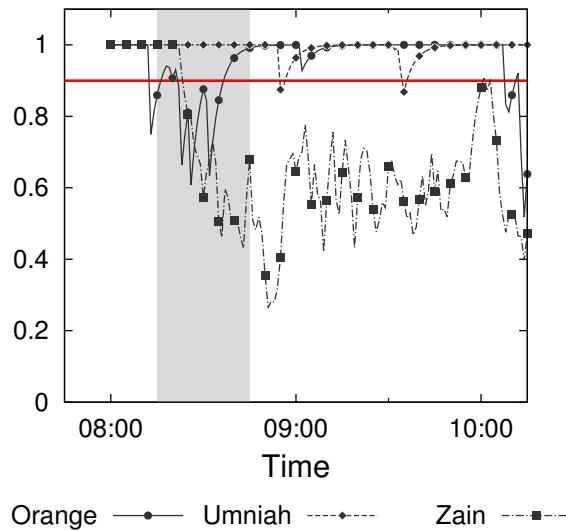


Figure 7.21: BTS Availability (BTSA) for carriers in Za’atari on the morning of January 6th 2015.

and obtain a BTS with higher availability, and hence receive better quality of service.

Reachability. The use of multiple SIM cards negatively impacts user reachability, which is the ability for a correspondent to connect with a user via one of the user’s phone numbers. Issues with reachability manifest as delays in correspondence, as calls and SMSs sent to an inactive SIM will not be received until the user inserts or re-activates the inactive SIM. SmartCell may exacerbate or improve this problem, depending on usage and the cellular networks in the area. In cases like Za’atari, where one carrier has significantly lower availability than others, users of SmartCell may visit the worst performing network less often, decreasing their reachability via their phone number from that network. However, as widespread use of SmartCell would likely result in loose load balancing across networks, this decrease in reachability is likely to be short-lived.

By contrast, in more balanced network environments, SmartCell may encourage users to switch between networks more frequently than they would have otherwise, reducing the average time their SIMs remain inactive and therefore reducing the delay before a user receives a missed call notification or SMS. We do not believe that attempting

to contact a user of SmartCell will be drastically different than contacting non-users, as users commonly switch between cellular networks currently. In contrast, SmartCell should *improve* reachability when a user is receiving a call on an active SIM, as SmartCell users prefer base stations with higher availability, which are more likely to have the capacity to provide service for the incoming call.

Global congestion. In the case of global congestion in an area, where all carriers are severely congested, SmartCell may cause users to switch between carriers when the switch will not actually improve connectivity. However, assuming a large population of users, SmartCell should lead to homeostasis where all carriers in an area have roughly the same availability, rather than drastically different levels as witnessed in Za'atari.

In cases of severe congestion on a single carrier, shifting many users onto a different carrier network may lead to simply migrating congestion from the egress network to the ingress network. As such, SmartCell can be tuned to avoid rapid, drastic changes in estimated availability by increasing the weight assigned to past measurements and/or setting the threshold at a lower value. By increasing the weight assigned to past measurements the availability metric is smoothed.

7.3.2 SmartCell Discussion

SmartCell opens multiple possibilities for further research. A current limitation of SmartCell is that each phone has limited a-priori information about whether selecting alternate carriers or BTSs will result in improved service. Each phone with SmartCell relies on its own measurements of the user's current cell, and recent measurements from nearby SmartCell users. This presents a challenge in areas with a low density of SmartCell users. Unless other nearby users have recently reported congestion information for multiple nearby carriers and cells, the selected network could be as congested or worse

post-switch. An algorithm to *predict* availability for carriers and BTSSs based on past data submitted to the SmartCell repository is one area for future improvement. The repository could serve tiles to clients in a similar manner to map applications, where each tile contains long-term congestion history of BTSSs in an area. This would reveal long-term trends, give users a better intuition of which carriers and cells are likely to be congested at any given time, and enable a more informed choice of networks and cells in the absence of real-time measurements.

The use of a central repository in congested environments can also be a limitation. If poor connectivity precludes the use of a central repository, cooperative peer measurements could be shared using direct channels such as local WiFi networks, WiFi direct, or Bluetooth. As each measurement is only a few bytes of information, a measurement can be encoded into short strings. For example, the string “1458049642:410:54012:34.4312:-119.7598:.8” indicates that at epoch time 1458049642, a cell on ATT (MNC 410) with cell ID 54012 was observed near 34.4312:-119.7598 with availability .8. These strings can be shared in a connectionless fashion, such as in Wi-Fi SSIDs or in the UUID field of Bluetooth Low Energy beacons.

Although cellular coverage is rapidly increasing throughout developing regions, it is clear that disparities in the quality of the coverage exist, and will likely continue to exist for the foreseeable future due to a variety of economic and geographic factors. Our work scratches the surface of understanding these inequities in two specific, very different regions. Clearly, more work remains to be done to fully characterize the problem.

7.4 PhoneHome

PhoneHome, as illustrated in Figure 7.22, consists of (i) a *local network* that operates in the outskirts of cellular coverage; (ii) a *PhoneHome bridge* comprised of two physical

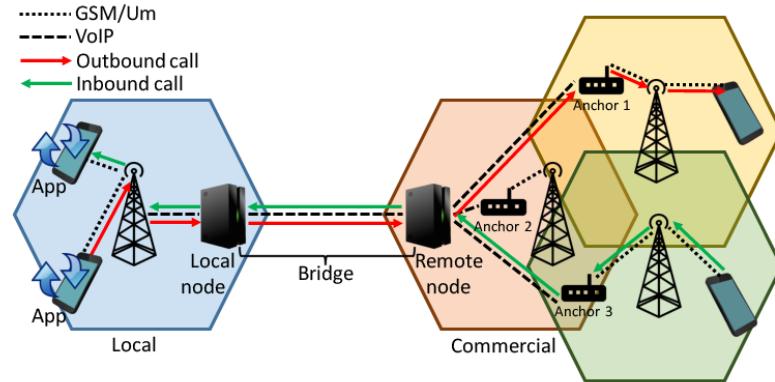


Figure 7.22: PhoneHome system diagram. Users connect their mobile devices to the local cellular network. PhoneHome uses a long-distance wireless link to a remote relay node in order to bridge local cellular services onto commercial carriers near the remote node.

nodes – a local and remote; and (iii) the *PhoneHome smartphone application*. The local network provides cellular connectivity to clients using their unmodified cellphones and SIM cards. Using the PhoneHome bridge, the local network also provides a virtual extension of the commercial network’s services to the outskirts. The bridge consists of two physical nodes that are connected via a long distance wireless link. The local node is situated in the desired coverage area (outskirts), while the remote node is positioned within the commercial provider’s coverage area. The link between the local and remote nodes serves as a bridge between the commercial carrier and the outskirts. Beyond physical connectivity, the PhoneHome bridge provides translation of GSM signals to VoIP and back to GSM in order to bridge local and commercial traffic while using low-cost IP backhaul. Finally, the PhoneHome application automatically switches users between the local and commercial networks based on their usage.

The remainder of this section details our system prototype implementation. We start with an overview of local cellular networks in Section 7.4. We then delve in the design and implementation of the PhoneHome bridge in Section 7.4. Finally, we describe the

PhoneHome application in Section 7.4.

Local cellular networks

Local cellular networks have been proposed in recent years to address the communication needs in disconnected or fringe areas [206, 88, 165]. These networks make use of free open source software, open hardware and generic IP backhaul in order to bring down the cost of cellular network deployments several orders of magnitude [206]. Since local cells are backward-compatible with commercial cellular standards, users can harness their existing phones and SIM cards to gain connectivity.

A typical local cellular network consists of a base station that runs open source implementation of the GSM stack such as OpenBTS. This base station communicates via the GSM protocol with user devices through a Um interface. Once GSM signals are at the base station, they are converted to a Voice over IP (VoIP) session and handed to the local network's core for authentication and switching. This transition from GSM to VoIP allows the use of inexpensive IP backhaul. The local network's core also makes use of open source software for traffic switching and authentication. While there are several options for such implementation, in our work we use a combination of FreeSwitch³, Sipauthserve and, Smqueue⁴. For more detailed description of local cellular networks' integration and operation we refer the interested reader to our previous work [206].

PhoneHome bridge

One of the key contributions of this work is the PhoneHome bridge. The main purpose of this bridge is to provide virtual extension of the commercial coverage to the local network. There are several challenges associated with this virtual extension concerned

³<https://freeswitch.org/>

⁴<https://github.com/RangeNetworks/smqueue>

with (i) the bridge integration, and the routing of (ii) outbound and (iii) inbound calls. We detail these in turn.

Bridge integration. The PhoneHome bridge is comprised of a local node and a remote node. As illustrated in Figure 7.22, the local node is connected to the local cellular network on one end and to the remote node via a long-distance wireless link on the other. The remote node, in turn, makes use of a set of anchor modems to interface with the commercial network(s) and bridge calls globally. Both the local and remote nodes run FreeSwitch and handle the voice/text traffic as VoIP. For this purpose, the nodes are configured to bridge VoIP sessions from the local network to the bridge (at the local node) and from the bridge through the anchors to the commercial network (at the remote node).

In terms of implementation, the local node is typically integrated with the local cell's core and is implemented on top of the local network's FreeSwitch instance. The local network is, in turn, equipped with a long-distance antenna to establish the wireless link for the PhoneHome bridge. The remote node is a Linux PC running FreeSwitch and equipped with several anchor modems and a long-range wireless antenna. The anchor modems can be implemented using any device that features cellular baseband, such as cellular USB dongles, GSM modems, or Android cellular phones. Figure 7.23 illustrates our remote node prototype that is comprised of a Linux PC running Ubuntu 14.04 and FreeSwitch. Our prototype makes use of Android phones as anchors running the built-in SIP client to interface with FreeSwitch in order to bridge calls between the local IP backhaul and the commercial GSM/Um interface.

One of the critical challenges for PhoneHome is to handle rendezvous for outbound and inbound calls. **Outbound calls** are established from a local user to a commercial user. After bridging the call to the remote node, our current prototype triggers the appropriate anchor to initiate a call via GSM/Um to the destination commercial node and



Figure 7.23: PhoneHome remote node equipment. The remote node includes a PC, GSM devices to bridge traffic onto a nearby commercial carrier, and a long-distance Wi-Fi antenna to connect to the local node within the camp.

then bridges this call with the corresponding VoIP session in order to connect the local and the commercial users. The key point to note here is that the anchor establishes the final stretch of call to the commercial user. This approach results in a major drawback: that is, the callee sees a call from an unknown number and may choose to ignore this call or wrongly search for the local user using the anchor's number.

Our future work will resolve this problem through the employment of electronically programmable SIMs (eSIMs) at the anchors. A key advantage of eSIMs is that their identity can be reprogrammed on the fly, which will allow the anchors to impersonate the local user before bridging the call to the commercial user. The latter will completely resolve our previously outlined drawback of the current approach as callees will see calls from a known number.

Inbound calls, on the other hand, are initiated by a commercial user and are destined for a local network user. In order to establish such calls, the commercial network needs to know that a particular user is reachable in the local network. This is unfortunately impossible, since commercial networks are unaware of local networks. This requires the local network to be proactive in receiving incoming calls. To this end we employ

added functionality to the anchors, that enables them to continuously monitor *pages* on the commercial network. Pages are broadcast messages that inform mobile devices of incoming calls or SMS, triggering the target device to request a private communication channel. PhoneHome inspects pages and determines whether the desired mobile device is in fact connected to the local network.

In our current implementation, the PhoneHome application will trigger a PHY-switch causing the local user to disconnect from the local cell and connect on the commercial in order to receive the call. There are two major problems with this solution. First, in cases where the commercial network is congested or failing, inbound calls will fail. Second, the PHY-switch may take up to 40 seconds, depending on carrier and technology. Taking into account that pages typically timeout within 10 seconds and carriers can configure whether or not they are repeated, it is possible that despite a successful PHY-switch, the call will still fail due to page timeout.

These problems will be resolved in the next version of our anchors, which will use eSIMs. The anchors will continuously monitor the broadcast channel of their respective commercial network for pages whose destination is in the local cell. Whenever such page is intercepted, the anchor will perform e-switch, which will change the identity of the anchor to that of the paged local user. The anchor will then be able to accept the incoming call on behalf of the local user and bridge that call to the local user. We anticipate electronically reprogramming SIMs will result in shorter switch times compared to PHY-switching.

PhoneHome application

The final component of PhoneHome is our smartphone application that performs two key tasks. First, it provisions users in the local network by reading and submitting their IMSI to the local network via SMS. Users with the PhoneHome application

can use any GSM SIM card with PhoneHome, including valid or expired SIMs from any provider. This functionality practically allows system operation without the need of customer service and support. While traditionally phones will only connect to the cellular base stations associated with their SIM card provider, our application uses an internal Android API to switch between available cellular networks at will, without user interaction, regardless of the issuer of the user's SIM. The second task the application performs is intelligently switching between the local and commercial networks based on user behaviors. As discussed in Section 7.4, user phones must shift between networks to receive inbound calls from the global network. Our application automates network switching without requiring the user to manually change settings.

Illustrative examples

To illustrate the operation of PhoneHome, let us consider two example scenarios of operation: one with an outbound and one with an inbound call. These two scenarios are illustrated in Figure 7.22, where the red path presents an outbound call, while the green path presents an inbound call.

First, in an **outbound scenario** (designated in green in Figure 7.22), the local user dials a destination number. The call request travels from the phone to the local base station via the GSM/Um interface. Once at the base station, OpenBTS translates the GSM/Um signal to a VoIP session and forwards the call to FreeSwitch over the IP backhaul. FreeSwitch, in turn, checks the destination number and determines that this call is to be connected globally via the bridge. At the local node, FreeSwitch then bridges the call to FreeSwitch at the remote node via the long-distance wireless link. The remote node bridges the VoIP session to the corresponding anchor. Once at the anchor node, the call is once again translated from VoIP to GSM/Um in order to enter the commercial network and be connected to the global user. We note that with our future eSIM-based

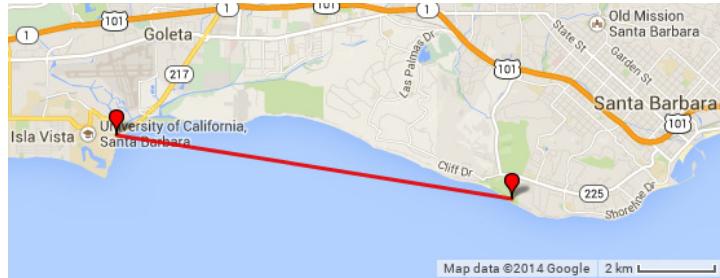


Figure 7.24: Long-distance Wi-Fi link locations used for experiments.

design the line of action will be the same, except that the anchor will impersonate the callee before completing the call.

For PhoneHome to receive an **inbound call** (designated in red on Figure 7.22), the orange anchor continuously monitors pages on the shared control channel to determine whether a local user is being called. When the anchor receives a page, it notifies the local network that a user is being called. The local network, in turn, triggers a PHY-switch by querying the PhoneHome app on the respective phone. If commercial coverage is currently available, the local user will migrate on the commercial network and will complete the call. We note that this chain of events will be very different in our future eSIM implementation of the anchors. With an eSIM, as soon as the orange anchor receives a page, it impersonates the local user who is being called and receives the call on this user's behalf. The anchor then bridges the call through a VoIP session with the local cell. Ultimately, the local cell completes the call by establishing a GSM session between OpenBTS and the local user.

7.4.1 PhoneHome Evaluation

We evaluate PhoneHome in a local testbed to appraise its efficacy in extending cellular coverage. For the purposes of our experiments, the local cellular network consists of a single Range Networks [150] Snap Unit GSM base station, which utilizes a Range

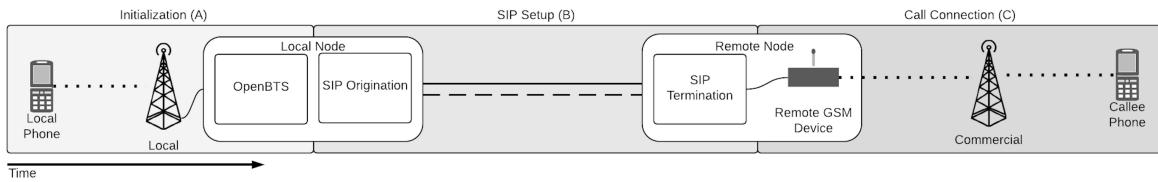


Figure 7.25: Call setup latency components.

Networks RAD-1 GSM transceiver, a 1W amplifier, and the OpenBTS 4.0 software stack. The local side of the bridge uses a Ubiquiti Rocket M5 with a sector antenna to operate a point to multipoint 5GHz Wi-Fi network, enabling one or more remote nodes to connect.

To evaluate the network performance with the presence of a remote node, we build a long-distance testbed link. We connect a remote relay node to the local node in our testbed using a roughly 10km 5GHz Wi-Fi link. The link locations are shown in Figure 7.24. Our remote node consists of a Lenovo X61 laptop running FreeSwitch, two Galaxy Nexus Android devices, and a Ubiquiti AirGrid 23dBi antenna. The remote node computer connects to commercial carriers using the two Android phones as anchors, one with a T-Mobile SIM and one with an AT&T SIM. Each phone is connected to the remote node via USB to provide power and signaling, and via 3.5mm cables to bridge call audio between the remote node and the phone.

Call setup latency

We examine latencies introduced by each of the call setup components in PhoneHome with and without the presence of a remote node. We define ***initialization*** as the time between a phone on the local cell dialing a number and the local base station establishing a session. ***SIP setup*** is the time taken by FreeSwitch, the IP-based private branch exchange (PBX), to set up a call. The final measured component is ***call connection***, the time between call initialization on the local or remote node and the callee's phone

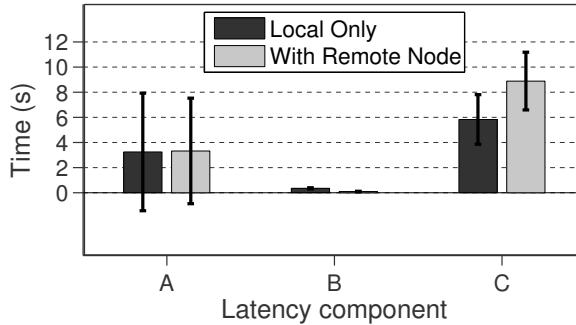


Figure 7.26: Latency comparison for call setup.

ringing. The three components with a remote node present are illustrated in Figure 7.25.

We evaluate the latency for each component by placing 30 consecutive calls from a phone on the local network to a commercial carrier using a remote node, then placing 30 calls from one phone on the local network to another local phone. The results are displayed in Figure 7.26. The bar values correspond to the mean time and the error bars indicate standard deviation. We observe that initialization (A) and SIP setup (B) components are similar in both scenarios. Call connection (C) is the only stage where we see significant latency added when a remote node is present; roughly three seconds is added to the total call setup time. We believe this 30.66% additional latency is within tolerable limits for users. It is important to note that the latencies we measure in this section correspond to call setup only, not in-call latency. In-call latency is explored in the quality of service (QoS) section.

QoS

We investigate voice session quality with the presence of a long-distance wireless link. We increase the number of voice sessions running across the 10km wireless link from 1 to 125 simultaneous calls with 700Kbps of background traffic. We choose 700Kbps as the maximum GPRS bandwidth per OpenBTS instance is 140Kbps with the highest coding

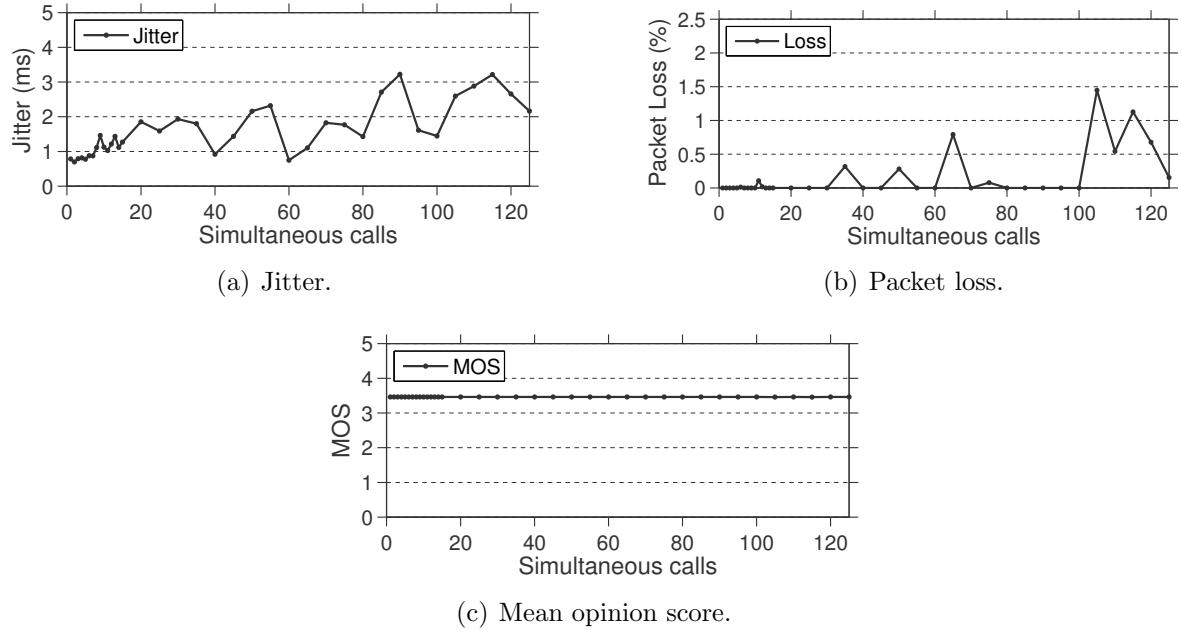


Figure 7.27: Long-distance link QoS measurements.

scheme and the maximum time slots dedicated to data traffic. Given this maximum, we inject traffic representing five local nodes using the remote node.

Latency, jitter, and packet loss are crucial metrics to determine voice quality. We first measure latency across the 10km wireless link. The ITU G.114 recommendation [98] specifies an upper threshold for one-way latency of 150 ms as satisfactory performance. The mean one-way latency for our 10km link is 28.09 ms and the median is 27.54 ms, well within the tolerable limit. We next investigate jitter and find that as simultaneous call load increases, the mean jitter value only slightly increases, as shown in Figure 7.27(a). The measured values are clearly within acceptable limits. Packet loss is displayed in Figure 7.27(b). As with jitter, we see a slight decrease in performance as the number of simultaneous calls surpasses 100. However, the overall loss observed remains acceptable.

Mean Opinion Score (MOS) is a metric used to describe perceived call quality, with a score between 1 and 5 where 5 is excellent and 1 is very poor. The maximum achievable MOS is dependent on the codec used to digitize the audio. In our case, the codec used

is GSM-FR (6.10), which corresponds with a maximum MOS of 3.46. We calculate the expected MOS using the E-model [100], which is dependent on both packet loss and the audio codec. Figure 7.27(c) shows that the calculated MOS values are stable and near the maximum achievable score for all evaluated loads. Overall, we see that the transmission of voice traffic across our 10km link does not unacceptably degrade the QoS of calls.

7.4.2 PhoneHome Discussion

We envision PhoneHome as a starting point for exploring user-extensible, bottom-up communications infrastructure in locations where connectivity is unlikely to be improved by commercial carriers. Many open questions remain for the idea of third-party network extension, pointing to important directions for future research. For instance, how many GSM devices are required at the remote gateway in order to provide adequate scaling given some number of users on the local network? How can we leverage electronically reprogrammable SIMs to provide global connectivity without forcing the user phones to switch to commercial carriers? How can we ensure users' SIM identities are secure and private in an eSIM environment? Can such a system provide data services in the same manner as voice and SMS? This section has focused on the initial proof of concept system-building aspect of providing a solution. We leave the open questions for future work.

An outstanding challenge relates to the coverage we extend. PhoneHome relies on a remote node located within the coverage area of a carrier. The placement of the node is likely to be chosen based on ease of access to the area and the constraints of building a long-distance link to the node. What we do not account for, however, is the capacity of the infrastructure our remote node utilizes. That is, we may be introducing a load that the existing infrastructure was not designed to handle. We are exploring load-balancing

across multiple commercial base stations to address this issue.

We have presented PhoneHome, a prototype system that extends existing cellular coverage on a community scale without requiring traditional wireless infrastructure investment. Our system is specifically designed for areas where existing coverage is inadequate for the local population, such as refugee camps or post-disaster shelters. Millions of people now live at, or beyond, the margins of traditional cellular coverage because they were forced to leave their homes due to man-made or natural disasters. PhoneHome provides a critical solution for facilitating communications for such people in a time when they need it most.

7.5 Related Work

Our work on HybridCell and PhoneHome utilizes recent research on local cellular networks. OpenBTS [1] is an open-source GSM base station that has been used to provide community-scale cellular coverage in rural and underdeveloped areas. Prior works use OpenBTS to provide coverage where no commercial carriers exist [85, 87, 206]. In contrast, our focus is on using local cellular networks to *characterize* and *augment* the coverage of incumbent wireless carriers in areas where commercial coverage *does* exist, but does not provide acceptable quality of service.

HybridCell explores moving users between independent cellular providers based on observed quality of service. In April 2015 Google announced Project Fi, which moves users between T-Mobile and Sprint base stations based on expected mobile data speeds [10]. The goals of Project Fi and HybridCell are related, but the implementations are distinct due to Google’s integration with two existing commercial providers. While HybridCell learns about nearby cellular networks through passive and independent observation, we expect Google has access to carrier metrics for base station performance. Additionally,

HybridCell is designed to be backwards compatible with existing SIM cards and Android devices, while Project Fi requires a particular phone model with a special SIM card. The always-best-connected concept [80] also touches on the use of multiple networks, however it assumes business relationships exist between providers, whereas HybridCell includes a completely independent local network.

Nomadic GSM also addresses non-interfering frequency selection for base stations [85]. However, this work relies on user handsets to scan the GSM frequency range, requiring active local cell users to discover incumbents. In contrast, our system monitors incumbent control channels to determine frequencies used by incumbent carriers without relying on local user handsets. Each System Information message may reveal up to 16 frequencies in use by commercial carriers, and each System Information message we use is broadcast multiple times per second. This allows our system to more quickly identify incumbents, and to identify them before transmitting.

To provide *global reachability* while a user is on the local network, prior work detailed integration of a local cellular network with Skype [97]. This enables users of the local network to make and receive Skype audio calls from any GSM handset, and to send and receive chat messages via SMS. However, this work was focused entirely on the reliability, rapid deployability, and VoIP gateway aspects of the system and did not address incumbent detection, user migration, or utilization of multiple cellular networks.

Two common approaches to understanding cellular performance as we aim to do with SmartCell are application-level studies that look at end-to-end characteristics and radio-level studies that focus on the access link. Our measurement of the cellular infrastructure is a radio-level study, that specifically utilizes messages broadcast over the GSM air interface. Recent work has illustrated the potential impact that cellular radio state has on end user experience [187, 154, 120]; and how air interface messages can be used to infer cellular user activity [113, 23]. Our work demonstrates how detailed, small-scale analysis

can support research on local-level infrastructure. This approach is in contrast to efforts to measure network performance on a global or nationwide scale, such as the FCC’s Measuring Broadband America⁵, and enables us to separate performance attributable to client-facing infrastructure serving the measurement locations from that which is related to carrier’s core networks.

7.6 Conclusion

In this chapter we have presented three different systems for improving connectivity in challenging cellular environments. Our system designs were informed by real-world analysis as well as experience in the field at multiple locations representing varying levels of economic development and cellular infrastructure usage. While these systems were tailored with specific situations in mind, they are generalizable to other settings. It is our hope that by demonstrating the potential using prototypes such as those we have presented, regulators and humanitarian agencies will be more likely to pilot systems similar in spirit to our work in the future.

⁵<https://www.fcc.gov/measuring-broadband-america>

Chapter 8

System Design: Automated WISP Network Topology Planning

8.1 Introduction

A number of years ago we set out on a mission to identify and eliminate barriers to the universal adoption of Internet access. As is still true today, we knew that Internet access is as much about economics as it is about technology. Therefore, despite ISPs being unwilling to build expensive infrastructure to serve regions with low user densities, we were certain of our inevitable success so long as cheaper, faster, longer range, and more rugged wireless equipment continued to become available [175].

Today, commodity wireless equipment is cheap, Internet is a basic human right, and major companies have joined the effort. However, despite buzz about high-cost, high-complexity, high-tech solutions to the problem, we have made only slow progress toward universal access. Near highly-connected cities there are communities connecting via dialup and their connections are getting *slower*—now crawling along at 9600 bps. Such neglected rural areas are home to 45% of the world’s population.

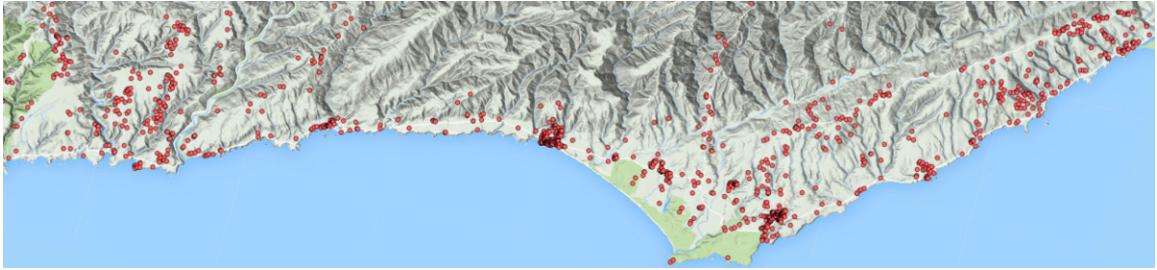


Figure 8.1: Map of $50\text{ km} \times 20\text{ km}$ frontier network region and locations of a subset of (potential) customers.

In this chapter we explore how to meet the challenges faced by the lone operator in the vast unconnected frontier. Building basic infrastructure in this frontier, even in wealthy nations, is an enormous endeavor. If universal connectivity is to be achieved, it will be not through the few, large operators connecting the last billions. Instead, connectivity will flow through the thousands of lone operators, often one-person outfits, who have a stake in bringing access to their own communities.

The key challenge for these operators is not one of hardware—commodity hardware is widely available and easy to set up—nor is it of management, as there are a number of free systems to aid them once up and running. Instead, it is a mismatch between the skills of to-be operators and the task at hand: planning a frontier network often requires a combination of extensive knowledge and practical expertise seldom found in one individual. As we discuss in Section 8.2, even for our expert team it was difficult to build such a network quickly, at low cost, and with few missteps; for unskilled lone operators who do not have our resources the difficulty is far greater.

Our remedy is a not purely technological, but a mix of a technological approach—automating the planning of frontier networks—and a social approach—identifying lone operators as the key enablers of universal access, and meeting their needs. The task of network planning typically falls to large carriers (in the case of backbones) and cloud providers (in the case of datacenters), both of which have the financial and political

resources to overcome physical obstacles (e.g. dig trenches, acquire spectrum, build large towers, buy land). In contrast, frontier network operators must plan and operate *within* existing constraints and cope with the complex myriad of network planning tasks.

No one task dominates others in importance when planning frontier networks, but the accumulation of poorly made decisions can easily bring down a network, leaving users in the dark once again as is the history of numerous rural operators. In this chapter we motivate the problem by describing one of our own experiences planning and deploying a frontier network. We highlight the practicalities of building such networks, what distinguishes them from other types of networks, what can and cannot be automated, our initial work on this challenge, and what remains to be done.

8.2 Planning Frontier Networks

To make the challenges of planning frontier networks apparent, we present a shortened case study on a network we deployed in a previously-unconnected region in rural Northern California [84]. Our experience illustrates some of the problems faced by a team of networking professionals when designing and deploying a frontier network infrastructure. While all networks are unique to their circumstances, the challenges are broadly similar, as found when we spoke with many dozens of rural network operators in North America, Asia, and Africa and given our firsthand knowledge of many of these networks. Our aim is to give the reader a sense of the challenges faced in frontier network design and the operational issues that should be taken into account.

8.2.1 Case Study Context

We learned of a region, about 50 km by 20 km, that was without broadband Internet connectivity. Local users who wanted Internet access either used a small regional dialup

Internet provider or used slow satellite Internet. Figure 8.1 is a rotated map of the region populated with data of a subset of (potential) customer sites.¹

The region had no coaxial infrastructure and poorly-distributed twisted-pair copper infrastructure, and thus no cable or DSL service. Cellular coverage was spotty, with no 4G service and unreliable 3G service; incumbent telcos had expressed no interest in improving service to the region, and even left backup generators in disrepair, resulting in frequent outages due to unreliable grid power. Several rivers and creeks cross the main road, which flood frequently cutting off road access. The region as a whole was economically depressed, including a local tribal community, with about a quarter of households living in poverty; however, there were pockets of affluence. Over the past decade at least three other operators have provided service to the region for a time, only to fail due to poor network planning and infrastructure and other challenges, resulting in poor network reliability and performance and leading to eventual business failure. Given this context, our challenge was this: how do we build a cost-effective, performant network to provide connectivity to the population depicted in Figure 8.1?

8.2.2 Deployment Process

Our deployment team consisted of several engineers and technicians. Despite being a skilled and experienced group we encountered numerous complex issues in planning, deploying, and managing the network and had to grapple with these issues with few tools at our disposal. As a result, while we were careful to weigh the decisions we made in designing and deploying the network, many decisions were still ad-hoc, and some decisions we made turned out to be mistakes that took time and money to undo.

Our initial task was to identify a source of upstream bandwidth. No universal map of

¹This data is a subset because these sites are only those who directly reported a desire to be connected, and therefore does not represent all households, businesses, civic buildings, and other locations that might need connectivity.

this information exists, and large telecoms (that are the usual providers of such service) do not publicize locations of their fiber facilities in such regions. After hearing local reports of a facility in the region, we contacted a large provider who, after months of our effort following up with them, confirmed for us that they would be able to sell us upstream bandwidth.

The lack of wireline infrastructure and the cost of building cell infrastructure and acquiring spectrum made microwave links (e.g., directional WiFi) a natural choice [140]. This hardware is cheap, low power, and easy to set up. However, such links require line-of-sight, have distance limitations, and can struggle with reflections, intermittent obstructions (i.e., severe weather), and is spectrum constrained.

Our first challenge was to determine how to distribute connectivity from the upstream gateway site. The telecom rejected our proposal to mount gear at their facility at low cost, leaving us with no option but to trench fiber from their site to another location nearby where this could be distributed. An ideal nearby site was a large, empty hillside near the facility. After another two months of tracking down and negotiating with the reclusive, elderly owner of the empty land, we were told that we could use the hill only for an exorbitant monthly fee. In parallel we considered several other neighboring sites, all of which were further away and none of which had any elevation. After the hillside was eliminated from consideration, we opted to trench fiber further to an alternative, low-lying location, from which we then had to set up backhaul links to a more distant hilltop location we secured, which would serve as a major distribution hub. In each area that we aimed to expand connectivity, we first spent many weeks using existing rudimentary planning tools [89] to manually identify multiple locations in concert that had line-of-sight and were located with good proximity to user populations. This was ultimately a guess-and-check approach. Once we had narrowed the list of sites, we then spent additional time to negotiate with land owners, businesses, and civic institutions.

The topography of the region—a narrow stretch of land between ocean and mountains that rise 1,000 m—dictated where we could place relay sites. Our constraints were further modulated by additional factors: where we could get power², where line-of-sight existed, where we had access to sites, and where potential users were situated. Existing tools only serve to compute line-of-sight between pairs of nodes, something available in many GIS planning tools. Since such networks have been built for a number of years, we expected that existing tools might be capable of doing semi-automated planning, but we found that the state of the art has scarcely advanced over the last decade.

The choice of frequencies for our radios at our sites was also decided manually and after many considerations. Spectrum contention was commonplace; despite our heavy use of unlicensed 5 GHz spectrum, in which there are numerous non-overlapping channels, we were forced to use other unlicensed bands as well due to contention at major sites. One other network began serving the region after us, in primarily different areas and providing significantly lower speeds.

After over six months of extensive planning, negotiation, and rollout efforts, our modest network consisted of about six sites and provided coverage to perhaps fifty users; it eventually took *years* for our network to expand to serve the majority of the region’s userbase. When unthrottled, many subscribers could receive 30-60 Mbps symmetric throughput to the Internet with less than 5ms latency within our network. At major infrastructure sites we also deploy batteries and networked power monitors, and power all key network devices using Power-over-Ethernet (PoE).

Several aspects of our network still did not perform well, leading us to attempt to use certain sites only to give up later. We deployed multiple parallel high speed backhaul links between major sites using different bands. However these different bands exhibited dramatically different performance characteristics during different types of severe

²The devices also tend to require DC power as they are sensitive to fluctuations in AC power.

weather. Another link in our network was from a peninsula to a point up the coast and had poor link quality in certain seasons of the year due to the sea.

8.2.3 Other Networks

As another example, we worked with one operator in North America, ‘Bob’, who had built his network over the course of over a decade and serves a few hundred users. Bob is a one-man operation who serves a region of about $60\text{ km} \times 50\text{ km}$. His network has been built in an entirely ad-hoc fashion, and to date he has no definitive map of all his sites and hardware, of spectrum use, of planning constraints, of capacity/traffic engineering considerations, of address allocation, or of other important information. Consequently, Bob’s network suffers frequent outages that can last for days at a time, but due to the lack of alternatives, customers continue to subscribe to his service.

Bob is a diligent and sincere individual, and the longevity of his network where others have failed is evidence of this fact, but he is not a networking expert. Bob’s network is constantly on the brink of failure. He requires money and time to repair and upgrade his network, which suffers frequent outages, but has little of either. There is a pressing need to design a network planning tool for non-experts like Bob. Such a tool would allow unskilled network operators to run reliable, high-performance frontier networks, potentially providing thousands of communities and millions of users in these communities with network access.

8.2.4 Complexity Challenges

Frontier networks, unlike ad-hoc networks, require complex planning to ensure high performance, robustness, and cost efficiency. As such, network operators must invest significant effort in planning, or deal with the consequences later. Almost all networking

research, however, focuses on network complexity problems *internal* to the network (e.g., wiring, cooling, protocols, management, etc.); networking researchers and engineers are typically insulated from the many *external* planning problems (e.g., facility siting, tower siting, fiber path planning, power management, etc.) that other well-resourced teams are responsible for handling in most large organizations. In frontier networks, when all of these problems are borne by a single individual or a very small team, the task becomes overwhelming. Furthermore, frontier networks do not have the financial or political capital to mitigate the sources of complexity and therefore must address them directly. Here we briefly highlight the differences in external complexity between a few network types.

Datacenter Networks. In a datacenter network, the number of possible options for deployment is limited by physical constraints within the region being considered, including where the planning department will permit construction and where sufficient power is available from the grid. While significant effort must go into obtaining a list of these potential sites, there are few degrees of freedom and only one location needs to be selected after evaluation. Within a conventional datacenter, considering a typical topology such as a Clos, there are a key design choices but they are *internal* (e.g., m and n for the Clos and its number of stages, rack design, transport protocols).

Enterprise Wireless Networks. In an enterprise wireless network that must provide wireless service across a number of buildings, we might consider each site to be a floor of a building, all of which must have service. Across a company’s campus, for example, this may be on the order of 100 sites. Thus the number of potential sites is equal to the number of sites for the network deployment. Within each site—each floor of a building—there are only a modest number of choices for equipment locations. While coverage must be established despite physical obstructions and hardware choices, the complexity of these considerations is still relatively low as the antennas are omnidirectional.

Regional Wireline ISP Networks. In a regional wireline ISP network, which is most comparable to a frontier network in terms of the number of network nodes and in terms of its service goals, much of the network’s design is predetermined by existing wireline copper infrastructure (twisted-pair for DSL and coaxial for cable). Network hardware, such as DSLAMs, is relegated to sites where existing head-end equipment can be placed, such as the regional telco’s central office or existing neighborhood-level cabinets.

Frontier Networks. In contrast to the above categories of networks, a typical frontier network is deployed across a large and topographically diverse area (e.g., $50\text{ km} \times 20\text{ km}$, or 1000 square km). In such an area, considering, crudely, that sites are typically parcels of rural land on the order of a couple of hectares each, there are about 50,000 potential sites. Even if we immediately aggregate or discard as non-viable 80% of these potential sites using various heuristics, some 10,000 possible site options remain. At each site, the number of constraints to be considered for placement of devices (which are directional, not omnidirectional) is on the order of twenty, including power availability, tree cover, slope of terrain, orientation, type of radio, type of antenna, type of tower or mast, type of hardware, and more.

Across these potential sites, the network only needs on the order of a dozen sites to serve the area, and such sites must be selected jointly, as the best set of sites (and their configurations) out of the thousands of options. This selection of the best small set from a large set of options results in combinatorial explosion, yielding many orders of magnitude greater design complexity than in other types of network design. It is the inability to cope with the combinatorics of the problem that frequently pushes network operators to make many ad-hoc design decisions that result in networks that are unreliable and slow—and thus expensive and short-lived.

8.3 Toward Automated Planning

Our goal is to enable the semi-automated design of a frontier network. We envision the process involving a would-be network operator (who may or may not have any network design or management expertise) articulating the geographic locations to be served, policy aims, and other limitations or criteria, and being given a fully-specified network design by the system, including the relay/backhaul locations and the network hardware to deploy, device configurations including spectrum allocation, and physical deployment specifics including elevation and power considerations. Such a design could then be improved through iteration with the design system—for example, as land use is negotiated—and a final design could be used as a blueprint for deployment.

To enable this, the system we aim to build must translate constraints *from the wild* along with user-specified policies into a cohesive model that then enables the construction of a network design by a solver. Eventually, the design generated by the system’s solver must be re-represented to enable the operator to refine and converge upon a network design. At the core of this iterative design process is a solver that combines the operator’s design specifications and physical models to produce a network design. Thus, we must design an appropriate representation for the solver, write translation tools from the varied and large datasets of constraints, and design a specification language for the user. Here we take the first steps toward automated planning by modeling and representing constraints.

8.3.1 Modeling

The interactions between elements that compose frontier networks make the problem of automated design particularly complex. There are numerous design elements that must be considered in a frontier network design, including site topography, upstream bandwidth availability, line-of-sight occlusion, spectrum, hardware choice, cost, tower/mount

choice, power availability, and more. It is these physical elements we aim to model.

To build the models we require, we must first begin with two sources: raw data (e.g., land topography datasets) and existing research on the behavior of specific elements (e.g., on wireless signal propagation). We must identify high-quality data sources for all of the elements we consider. Many of these data sources are incomplete, vary in granularity in different regions, and can disagree with each other. Since such variability in the data sources can confuse the solver, we must clean these sources.³ We must then express uncertainty explicitly in the models we build for each of these elements.

Each element itself has numerous important properties, and some are inherently difficult to model accurately. However, unlike work that is fundamentally about modeling, our aim in modeling them is not for their own sake, but towards the goal of automated design, and incomplete but practical models can be valuable to this end. Thus we must identify those properties of each design element most salient to the design task at hand, and focus our modeling work on those.

8.3.2 Solver Representation

Design specifications include three key components: 1) operator goals (e.g., bisection bandwidth, reliability, served geographic areas), 2) operator constraints (e.g., financial limits, deployment time, spectrum/hardware limits, known land availability), and 3) physical models.

The network planning problem naturally lends itself to a graph optimization representation where the vertices represent potential sites and edges represent connectivity (based on line-of-sight and transmission range). Within this optimization problem further specification by the network operator or dictated our physical models may be incorporated

³Ordinarily this is not a research challenge, but no existing stakeholder has both the incentive and means to do so.

into edge weights. The remaining optimization problem of connecting the vertices with the minimum total weight edges can be performed by a solver.

We do not view the theoretical development of a solver in this context as a novel research contribution, nor do we expect to definitively solve this problem from a theoretical perspective. Given that multiple aspects of the design problem are NP-Hard, such as site selection and spectrum allocation [142], the underlying constraint datasets are rife with error, and the policies expressed by the operator are ambiguous, there is limited room for or value in developing an “optimal” algorithm. Instead, it is sufficient to develop a practical system that is capable of producing a result that is substantially better than the status quo today. The resulting decrease in life-cycle cost over the course of the network’s design, deployment, and management will make all the difference for the viability and longevity of frontier networks.

8.4 Prototype Implementation

Our prototype planning system is built on top of the GRASS GIS library [78], which provides routines to access, modify and analyse geospatial data. Furthermore, the library provides useful functions for frontier network planning such as terrain analysis, network analysis, data visualization, etc. We use elevation data from the latest SRTM dataset [71] post-processed using GRASS GIS. We model the planning problem as a graph problem that can be solved by the calculation of a Steiner Tree; since this is NP-Hard, we leverage an approximation algorithm. Our system currently consists of the following stages:

Rasterize possible tower locations. Since the number of potential relay tower locations within the region of interest is large, we define a fixed raster of possible locations on the elevation map. That is, the raster defines possible tower locations so that they are equidistant and narrows the number of possible locations to a finite set (vertices).

Although these locations are an approximation and have a lower resolution compared to actual possible locations, this is a reasonable initial approximation before the actual mounting point of the antenna is known.

Determine visibility to other locations. For every possible site location on the pre-defined raster and also given their surrounding terrain, the geographical area that is visible from every location, or *viewsheds*, can be calculated to determine visibility. These viewsheds include all surrounding locations that are within line-of-sight and exclude points that are beyond the transmission range. If two nodes are visible to each other, the potential network link is represented by the addition of an edge between the two vertices. We apply this process pairwise to test the visibility of all tower locations with each other for all combinations of equipment transmission ranges to produce a basic visibility network graph for all possible equipment combinations.

Weight Edges with Costs. The edges in the visibility network are assigned weights that represent the cost associated with the link such as hardware costs, etc. In our current implementation, we use uniform weights for simplicity.

Find optimal tower locations. Given the visibility network graph $G = (V, E, w)$ and desired service locations $S \subseteq V$, the Steiner Tree calculation aims to produce the minimum cost tree that spans the vertices of S where extra vertices from V may be included in the tree so as to reduce the total weight. We leverage a heuristic Steiner Tree solver that produces an approximate minimum-weight Steiner Tree for S .

Figure 8.2 shows a sequence of steps from our implementation finding the best interconnect among several, arbitrarily chosen customer sites (indicated by red markers). Figure 8.2(a) shows the elevation data as colored regions where darker indicates a higher altitude of the terrain. The pre-defined possible tower locations (vertices) are projected on the elevation map as orange crosses. For illustration purposes, the horizontal and

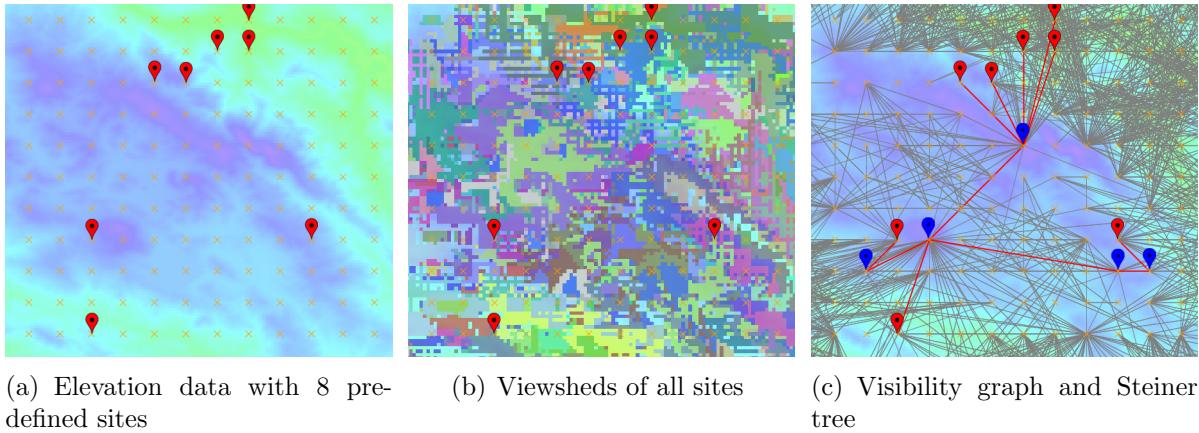


Figure 8.2: Finding the best relay tower locations (blue markers) and links to inter-connect 8 sites (red markers).

vertical distances between two locations are chosen to be ~ 750 m and only a restricted region (~ 170 km 2) of the map is shown in this example. Figure 8.2(b) depicts the visibility range of each location in a different color and represents the viewsheds of each location. The viewsheds are further processed and locations that are visible to each other are connected by edges. The resulting visibility network is shown in Figure 8.2(c) and indicated by gray lines. Finally, intermediate nodes (blue markers) are added to construct a Steiner Tree (red lines) from the 8 initial customer sites.

8.5 Discussion

Many obvious improvements can be made to our current system including: reducing the raster size, incorporating variable mounting height, antenna sizes, power output, tree cover, Fresnel zones, spectrum, equipment costs, reliability metrics, bandwidth, etc. The incorporation of these constraints (at the resolutions we require) will increase the problem size by at least two orders of magnitude. Currently, our implementation—not considering the additional variables above—takes approximately 40 minutes to compute the visibility graph for a region on a single standard machine. Given that the compute time of our

algorithms increase quadratically with the problem size, a high fidelity optimization will take about five orders of magnitude more computation, and adding the above variables will increase the complexity by another four orders of magnitude.

We presented a very clean representation of the problem as a set of fixed specifications to optimize against known constraints. However, the real world is messy and there are many unpredictable events. Consider what happens when an operator builds two sites of a five site backhaul network segment, only to lose land rights to the third site, or to find that the third site is in fact blocked by a new obstruction. While the planner's constraints can be changed to reflect this, this change might otherwise cause our system to produce a wholly different design without the first two sites that were already built. In this situation, the operator needs the ability to express the cost of abandoning work that has already been done (i.e., if there is truly no other way to build the network without abandoning the already-built sites, then our system should find those options, but should prefer leveraging already-completed work).

Another problem is that the Steiner Tree formulation (and other off-the-shelf solvers) may not capture the goals of the network designer. In this initial implementation we used the Steiner Tree optimization because it is a natural fit for the basic planning task, but we did so knowing it would prove inadequate. For example, the output of the algorithm produces a tree, which by definition has no redundant links. Given the general unreliability of wireless links in these settings, it is important to consider backup and parallel links.

8.6 Related Work

Networking in developing and rural regions is a topic that has attracted attention over the last decade [130, 42]. This focus has intensified in recent years and researchers

have recognized that the challenges presented in these networking regimes are substantially different and require different solutions [56, 140, 148, 86, 29, 176, 59, 104]. While there has been work on protocols and techniques for faster and more reliable networks, research often does not extend far enough to keep such networks alive after the research is done [176].

Determining the topology of wireless networks was a significant topic [149, 197, 41] and there are numerous patents on planning cellular networks [28]. Within wireless, variations of topology planning considered hardware factors such as transmission power and directional antenna [149, 93]. The closest work to ours is an algorithm by Sen and Raman that attempts to minimize the overall network cost by considering tower height, antenna type, and transmit power [168].

When deploying their networks today, rural operators use a mix of incomplete planning tools. Some tools build upon terrain data to estimate line-of-sight between two locations, enabling an operator to perform rudimentary topography planning for relay sites [89], other tools [196, 177] provide numerous tools to manually plan, understand, and deploy wireless networks. Other systems, such as TowerDB and Celerate, attempt to simplify network management by juxtaposing geographic locations of devices with network information (e.g. IP-address, frequency, and SSID) [182, 84].

A few alternative planning models have also been proposed. IncrEase [36] is planning paradigm that incrementally introduces sets of additional transmission sites. In [94], the authors describe a mathematical model for automated network planning that considers economic and technical constraints. Several approaches have been proposed in cellular network planning for the placement of base stations (e.g. Andrews et al. [31]), but these generally focus on spectrum and interference rather than physical topography.

Chapter 9

System Design: Unsupervised Transmitter Detection for Spectrum Sensing

9.1 Introduction

Current spectrum management is based on exclusive allocation and assignment of radio spectrum to a given technology and operator. This paradigm results in artificial spectrum scarcity that leads to high cost for communication services, decreased network performance and deteriorated user experience. To put this in perspective, let us consider mobile cellular access, a service so critical to people's day-to-day life that over 110% of U.S. residents have active cellular subscription [13]. A critical resource to the scalability of mobile cellular networks is the amount of radio frequency spectrum they own, however, the offered load of mobile networks' large user population quickly saturates the available spectrum. This, in turn, leads to degraded network performance and high prices for limited monthly subscriptions. While mobile cellular bands are rapidly becoming over-

saturated, there exists an abundance of spectrum in other frequency bands that are largely underutilized [34].

These underutilized bands provide an opportunity for more efficient, shared spectrum access that has brought together policymakers, industry and academia to set an agenda for next-generation spectrum management [19, 95, 45, 119]. This new shared access paradigm mandates a rethink of current spectrum management practices and a corresponding re-design of mobile wireless technologies. ***A critical enabler of such advances is deep understanding of spectrum use***, both long-term as well as instantaneous. An actionable agenda for spectrum management including both measurement objectives and infrastructure was outlined at a recent NSF workshop on the topic [19]. It became clear that there is an urgent need for novel mechanisms for spectrum sensing and characterization that allow real-time and batch processing of spectrum scans to reveal an essential mix of characteristics such as number of occupants, their temporal and frequency usage patterns, mobility and their eligibility to operate. Without such measurement and characterization methodology, any effort to advance spectrum policy and technology will be uninformed and thus ineffective.

Current approaches to spectrum characterization are limited to detection of idle and occupied frequencies in a spectrum scan [199, 201, 180], however, they cannot provide a deeper functional spectrum characterization that teases out number of transmitters and their temporal and frequency usage patterns. More recent work has focused on cyclo-stationary analysis [165, 90] or signature matching [151] to detect transmitters. While such approaches provide deeper understanding of spectrum occupancy, they are limited within the predefined types of transmitters they can detect and are thus not well-suited for wideband spectrum characterization. Other recent work [205] addresses the problem of arbitrary transmitter characterization by the design of a custom machine learning technique that harnesses signal distribution properties to extract transmitter character-

istics. While this method is able to robustly identify transmitters, it is computationally expensive and cannot be employed for real-time characterization.

To address the limitations in existing work we design AirVIEW, an algorithm for unsupervised, wideband, high-sensitivity spectrum characterization. A key advantage of AirVIEW over existing approaches is its robustness to noise and its ability to detect transmitters even when their power is very close to the noise floor. AirVIEW operates on spectrum scans comprised of power spectrum density (PSD) measurements over frequency and time. *A sweep* in such scans is defined as a one-time vector of PSD values collected over a frequency band of interest. AirVIEW operates on individual sweeps as they are delivered by the sensor, and thus, streams through spectrum scans to determine active and idle frequencies within a spectrum sweep. Single-sweep frequency bands identified as active, are further processed through a two-step transmitter reconciliation procedure that identifies longitudinally-active spectrum occupants. For single-sweep spectrum analysis, AirVIEW employs wavelet decomposition [127] of PSD to separate transmitter edges from the noise floor. Once we compute the wavelet coefficients of a PSD vector, we construct the binary tree representation of these coefficients, which presents a powerful construct for further signal analysis [122]. We leverage multiscale products of wavelet coefficients at neighboring scales in the binary tree. These multiscale products were previously shown to amplify signal level transitions while suppressing the effect of noise [180, 157, 204].

Such an approach presents *ample opportunities* for high sensitivity transmitter detection in noisy spectrum data, that is not possible with classical methodologies that operate on raw PSD measurements. While promising, this wavelet-based approach *brings a plethora of new challenges* in order to be employed for characterization of arbitrary transmitters. First, in order to determine which transitions in the multiscale products represent an actual transmitter, we need an adaptive and data-driven approach to multiscale threshold estimation. Second, as detailed in Section 9.3, the accuracy of such a

method hinges on the careful selection of binary tree scales for multiscale product calculation. To this end, we observe that multiple single-sweep transmissions should be aligned in frequency and devise an unsupervised algorithm for adaptive optimal scale selection.

We evaluate AirVIEW’s performance on synthetically generated and real-world ground truth transmissions, such as TV and Wi-Fi and we show that it can robustly detect the temporal and frequency characteristics of transmitters even when their power is as low as 5dBm above the noise floor. We carry out a wideband measurement campaign (50MHz-4.4GHz) in an urban location in close proximity to an airport and a military airbase. We then employ AirVIEW to analyze the collected spectrum scans and demonstrate its ability to inform next generation Dynamic Spectrum Access (DSA) technology and spectrum policy through rapid characterization of (i) idle and occupied frequency bands, (ii) number of incumbents in each band, (iii) their temporal characteristics and (iv) the predictability of incumbents’ behavior. This chapter makes several key contributions:

- **Novelty:** We design AirVIEW, a high-sensitivity, unsupervised transmitter characterization algorithm that is robust to noise and is able to detect arbitrary transmitters.
- **Scalability:** Since it requires a single pass over the data, AirVIEW scales to wide spectrum scans at high temporal resolution, as demonstrated in our real-world evaluation.
- **Accuracy:** We provide empirical justification of our design and demonstrate the accuracy of AirVIEW on synthetic and real spectrum traces.
- **Applicability:** We leverage longitudinal, wideband spectrum scans from 50MHz to 4.4GHz to demonstrate AirVIEW’s applicability for detailed and unsupervised transmitter characterization of arbitrary, real-world transmitters.

In what follows we first present related work. Section 9.3 details unique challenges and opportunities in wavelet decomposition of spectrum scans and presents AirVIEW’s methodology. Section 9.4 presents a benchmark performance evaluation of AirVIEW on synthetic and real-world ground truth transmissions. Section 9.5 presents our AirVIEW-aided analysis of a wideband spectrum scan. Finally, we provide discussion and conclusions in Section 9.6.

9.2 Related Work

Prior work on spectrum analysis can be largely subdivided in such that deals with *activity detection* and other that performs *detailed transmitter characterization*. Activity detection aims to answer for a single sweep, which bands are active and which are idle. Such methods, however, do not provide further reconciliation of transmitter activity in multi-sweep spectrum measurements. The latter has been recently identified as an important advantage and has been tackled in several works for detailed transmitter characterization that utilize either supervised or unsupervised techniques.

Energy-based activity detection. Spectrum characterization has been largely explored in the past, however, the literature is limited in methods that provide detailed, robust and unsupervised transmitter characterization. Traditional approaches to spectrum characterization identify idle and occupied bands by the use of power thresholding [200, 125, 65], edge detection [180, 199] and compressive sensing [181, 50, 117]. While these approaches are computationally light-weight, they are only able to determine which parts of the spectrum are idle and which occupied, however, they are unable to attribute longitudinally-active bands to a single transmitter’s operation. Thus, such approaches cannot facilitate detailed transmitter characterization for next generation spectrum management. Furthermore, these existing techniques operate on the raw PSD measurements,

which limits their efficiency in noisy and low signal-to-noise regimes. Our work addresses these challenges by transforming the measured PSD signal into wavelet coefficient domain. Analyzing the signal in this domain reveals the underlying transmission structure and reduces the effect of noise caused by local variations. Furthermore, we develop a two-step transmission reconciliation technique that combines detected occupied bands into longitudinally-active transmitters to facilitate detailed spectrum characterization.

Wavelet-based activity detection. Prior work has theoretically justified wavelet analysis for secondary spectrum access [180], however, it has not been applied on real-world spectrum data. In contrast, our proposed method leverages wavelet coefficient analysis for robust detection of noisy, real-world signals (details to follow) and extends it to be able to detect occupied bands over multiple consecutive spectrum sweeps, which in turn allows for robust detection of transmissions.

Supervised (signature-based) characterization. Another related body of work is signature-based characterization which requires prior knowledge of transmitter parameters for detection [90, 151]. While such techniques enable detailed spectrum characterization, they suffer inherent limitations in the number and types of transmitters they can detect, which in turn makes them infeasible for wideband characterization of arbitrary transmitters. Our work departs from these early paradigms by developing robust and unsupervised techniques for rapid and high-sensitivity transmitter characterization.

Unsupervised characterization. Unlike energy detector methods, which require threshold setting, and signature based methods, which require prior knowledge of transmitter signatures, unsupervised methods do not require any user interaction or prior knowledge to function. Unsupervised methods are the most desirable for DSA since the goal is to allow devices and sensors to automatically control the spectrum hopping process. A recent work [205] is able to characterize the frequency spectrum using Rayleigh-Gaussian mixture models. While accurate, this method is computationally demanding

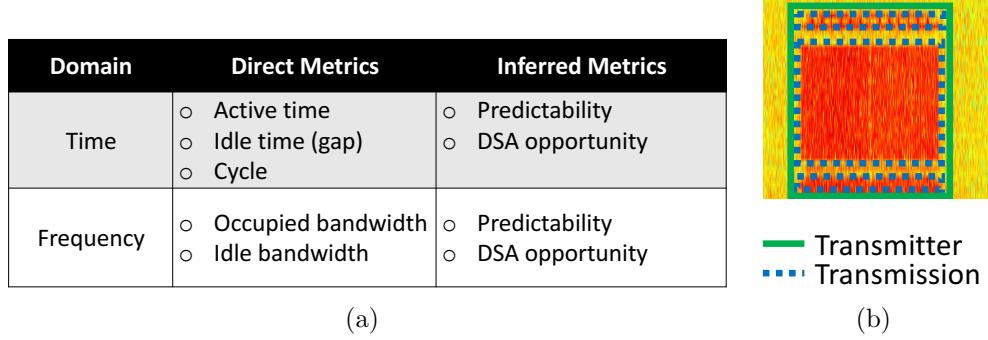


Figure 9.1: (a) Transmitter characterization metrics; (b) Illustration of a transmitter comprised of multiple transmissions.

and operates in batch mode as opposed to in a single pass, rendering it inapplicable for real-time transmitter detection at scale. Our goal is to devise an unsupervised method which is both robust to noise and scalable to real-world wideband sensing and detection.

9.3 Methodology

The primary goal of spectrum characterization is to map the time and frequency characteristics of transmitters with no or minimal prior information. To this end, we need a method for detection of arbitrary transmitters that scales to large volume of diverse spectrum data. Figure 9.1(a) summarizes key transmitter characteristics of interest. Depending on the application, the timescale requirements for transmitter characterization may vary. For example, if the goal of transmitter characterization is real-time dynamic spectrum access or spectrum enforcement then we need immediate characterization. Other applications, such as policy-making or DSA based on historic occupancy may allow offline batch processing of spectrum scans. Our goal is to address these varying requirements in a unified configurable framework.

Recent work proposed a method that is well-suited for offline batch processing of spectrum scans [205]. However, it requires a large time window of the scan in order

to detect transmitters and learn their temporal patterns, rendering it inapplicable for settings where detection and characterization needs to be performed on-the-fly as the sensed data arrives. To address this more challenging setting, we detect transmission edges at a given time instance, ensuring that they smoothly extend detected transmissions in the recent temporal horizon. To achieve this, our proposed method AirVIEW employs wavelet decomposition of the PSD measured over a given frequency band. This representation of the PSD allows us to *suppress the effect of noise* in detecting the frequency range occupied by an ongoing transmission. In addition, we maintain a summary of recent transmission ranges and impose *temporal smoothness* on newly detected ones such that their ranges agree across time. Note that AirVIEW is purely unsupervised and does not require prior knowledge of transmitter signatures and properties, making it applicable to detection in the wild.

In spectrum characterization we differentiate between a *transmission* and a *transmitter* (Figure 9.1(b)). We define a *transmission* as a single continuously-occupied time-frequency block of the measured spectrum. A *transmitter* is then a set of transmissions that are determined to be caused by the activity of the same radio-emitter. Under this definition, a broadcast transmitter will be characterized with a single transmission, whereas a TDMA, FDMA or frequency-hopping transmitter will be comprised of multiple transmissions. The task of spectrum characterization then can be split in two key sub-tasks: (i) robust identification of transmissions and (ii) efficient grouping of transmissions into a transmitter. An overview of AirVIEW’s workflow is presented in Figure 9.2. AirVIEW operates in three stages. As data arrives from the spectrum sensor, AirVIEW performs single-sweep transmission identification. A single transmission, however, may span multiple consecutive sweeps, thus the second step in AirVIEW’s flow is to combine aligned single-sweep transmissions into a multi-sweep transmission across time. Finally, AirVIEW combines similar-band transmissions into a transmitter to extract the unique

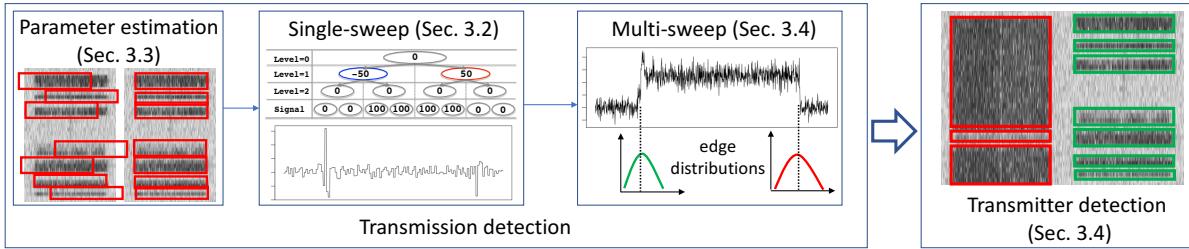


Figure 9.2: AirVIEW workflow.

characteristics of transmitters, e.g. cycle, active time, etc. In what follows, we first provide the necessary background on wavelet decomposition of spectrum scans. We then detail our proposed approaches to detect transmissions, and group them into transmitters. In addition, we discuss an unsupervised parameter calibration approach that takes advantage of the smoothness of transmissions in time.

9.3.1 Preliminaries: Wavelet decomposition of spectrum sweeps and multi-scale products

Wavelets are a useful mathematical tool for hierarchical decomposition of signals [44, 122]. The wavelet decomposition can be computed efficiently and it enables exact signal reconstruction. Wavelet analysis has been employed in diverse domains including image processing [51], databases [174] and in wireless networks for wideband spectrum sensing [180]. In what follows, we introduce the basic wavelet decomposition definitions that are necessary to discuss our method. For an exhaustive discussion of wavelets, the reader can refer to standard texts such as the one by Mallat et Al. [122].

Let $p_t(f)$ be the PSD signal over discrete frequencies $f \in F$ at time t , also referred to as a single time sweep or simply a sweep. Given a wavelet basis, the one-dimensional wavelet decomposition is a function $W : \mathcal{R}^{|F|} \rightarrow \mathcal{R}^{|F|}$ that maps the original PSD signal $p_t(f)$ to a set of real wavelet coefficients w_t of the same dimensionality. The coefficients

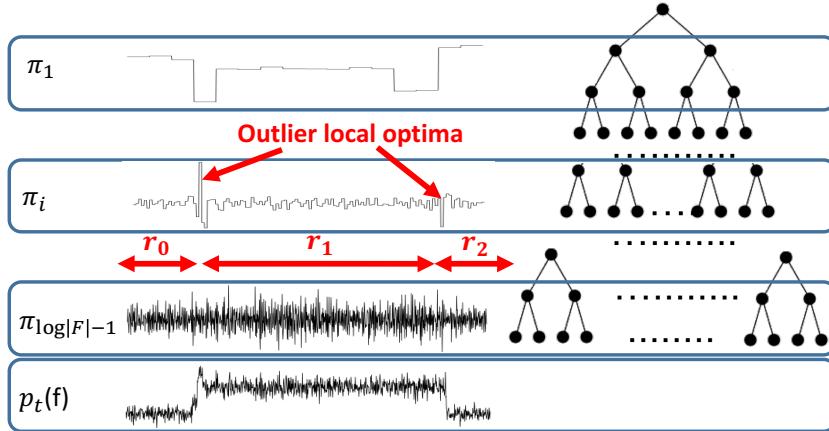


Figure 9.3: Example of wavelet decomposition with illustration of key features employed by AirVIEW.

can logically be thought of as a binary-tree hierarchy of increasing scales $s \in [0, \lceil \log |F| \rceil]$, where the 0-th scale corresponds to the root of the hierarchy and the maximum scale corresponds to the hierarchy's leaves. Thus, $w_t(s, j)$, are the coefficients at a given scale s of the tree, where j is the index of the coefficient at that scale. A sketch of this arrangement of coefficients is presented in Figure 9.3.

The wavelet basis is a family of functions with finite support that are translated and dilated according to the scale and are mutually orthonormal [122]. There is extensive research on different basis families, thus, AirVIEW allows for user-defined basis. For the analysis in this chapter we focus on the *Haar basis*, where the mother wavelet is a “square-shaped” function that is piece-wise constant [122]. The Haar basis was chosen since its shape matches well the sharp changes in power at the change points between transmission and no-transmission frequency regions. Other bases may be considered for different detection tasks within our framework.

To construct the Haar wavelet decomposition, one computes a sequence of averages at different scales according to the mother wavelet (details in [122]). The decomposition W is lossless as the original signal $p_t(f)$ can be exactly reconstructed by a reverse reconstruction function W^{-1} using the wavelet coefficients $w_t(s, j)$.

We note a few inherent properties of the wavelet binary tree (Fig. 9.3) that are essential to our method. First, low scale coefficients (close to the root) capture more drastic changes of the underlying signal, whereas high-scale coefficients (close to the leaves) capture finer changes (i.e. jitter due to noise). As a result, a lossy reconstruction based on the low-scale coefficients only (i.e. assuming high scales have zero coefficients) will produce a coarse representation of the signal. Such lossy schemes that take into account both the coefficient magnitude and scale have been adopted for approximate query answering in databases [51] and for image compression standards such as JPEG [185]. Our goal, however, is to detect frequency bands of active transmission as opposed to reconstruction or compression.

The multi-scale product π_s at level s is defined as:

$$\pi_s = \langle W^{-1}(w_t(s, \cdot)), W^{-1}(w_t(s + 1, \cdot)) \rangle, \quad (9.1)$$

where the right-hand side denotes the element-wise product of the signal reconstructions $W^{-1}(w_t(s, \cdot))$ using only coefficients at level s and $s + 1$ respectively. Such multi-scale products have been previously shown to “elucidate” edges of transmissions while suppressing the effect of noise [180, 157, 204].

A set of products and their corresponding scales in the wavelet tree are presented in Figure 9.3. Local optima in low-scale products (e.g. π_1 in the top of Fig. 9.3) clearly delineate frequency regions in which an abrupt power change occurs, however they lack “spatial” resolution in that they cannot pin-point the exact frequencies at which a transmission starts. Conversely, high-scale products are less coarse, albeit more susceptible to noise. The medium scale product π_i in the figure offers a good trade-off between robustness to noise (i.e. local optima still significantly noticeable compared to noise fluctuation), and resolution in the frequency range for precise detection of transmission

edges.

Intuitively, our detection algorithm identifies the most significant local optima in an appropriate level multiscale product to determine the start and end frequencies of transmitter activity captured in a single sweep (Sec. 9.3.2). The scale and threshold for optimum significance are adaptively determined in an unsupervised manner (Sec. 9.3.3). We also employ smoothed discovery by reconciling detected transmitters across time in Sec. 9.3.4

9.3.2 Single-sweep transmission detection

A *sweep* $p_t(f)$ in spectrum scans is defined as a single-time vector of PSD values collected over discrete frequencies $f \in F$ at time t . In order to perform transmission detection for a given sweep, AirVIEW leverages the multiscale products π_s derived from the wavelet decomposition of the original PSD $p_t(f)$, computed as described in Section 9.3.1. Candidate transmission edges are obtained as the top-ranked local maxima in the product signal and are then paired to get transmitters based on the average power of enclosed frequency bands.

Algorithm 1 lists the main steps of our single-sweep transmission detection. The method takes as an input the measured PSD signal $p_t(f)$, the product scale for detection s and a threshold scaling parameter β . Both s and β are learned as a pre-processing step as described in the following Section 9.3.3.

We first compute the wavelet decomposition $W(p_t(f))$ (Step 1) and the multi-scale product $\pi_s(f)$ at scale s (Step 2). As an example of product calculation, consider the product π_i in Figure 9.3, in which the position of outlier local maxima are annotated with arrows. We note that depending on the selected scale, the sign of the product value corresponding to a transmitter edge might be different. Since our analysis is concerned

Algorithm 1: Single-sweep transmission detection

Require: $p_t(f)$, s , β

Ensure: A set of transmission intervals $\mathcal{D} = \{[f_i, f'_i]\}$

- 1: Compute $W(p_t(f))$
- 2: Compute $\pi_s(f)$ and $\hat{\pi}_s(f) = |\pi_s(i) - \pi_s(i-1)|, i \in F$
- 3: Detect $X^\beta = \{x_i\}$: local maxima of $\hat{\pi}_s(f)$ exceeding $\tau_t = \mu_t + \beta\sigma_t$
- 4: Compute the average power $p_t(r_i)$ of frequency regions $R^\beta = \{r_i\}$ between local maxima X^β
- 5: $\mathcal{D} = \emptyset$
- 6: **for** $\forall r_i$ in descending power $p_t(r_i)$ **do**
- 7: **if** Neighbors $N(r_i)$ of r_i not in detected \mathcal{D} **then**
- 8: $\mathcal{D} = \mathcal{D} \cup r_i$
- 9: **end if**
- 10: **end for**
- 11: RETURN \mathcal{D}

only with the magnitude but not the sign of the product, we take the absolute pairwise difference between consecutive product values ($\hat{\pi}_s(f) = |\pi_s(i) - \pi_s(i-1)|, i \in F$) and perform our analysis on these pairwise differences. Note, that the product is a piece-wise constant function. The constant produce regions, induced by the coarse-scale reconstruction, are treated as single values in defining the consecutive difference. The differences are next considered to detect local optima: points of high absolute difference $\hat{\pi}_s(f)$ of the product multi-scale function.

Next (Step 3), we consider all local *maxima* x_i of the pairwise differences and keep only the β -outlier ones, i.e. those whose pairwise differences exceed the absolute threshold of the current sweep τ_t , calculated as the mean plus β standard deviations:

$$\tau_t = \mu_t + \beta\sigma_t \quad (9.2)$$

Here μ_t and σ_t are the mean and standard deviation of the pairwise difference of the multiscale products $\hat{\pi}_s(f)$ at scale s for the current sweep t . Intuitively, β is a threshold scaling parameter that determines how many standard deviations from the mean should

pairwise differences be in order to be ranked as a top local maxima, i.e. outliers.

The positions of outlier local maxima X^β partition the frequency domain into non-overlapping frequency regions $R^\beta = \{r_i\}$. To clarify this step, consider the example in Fig. 9.3. Assuming that only the annotated local maxima in the product π_i are deemed outliers w.r.t. β , we obtain three resulting partition regions r_0, r_1 and r_2 (also annotated in the figure). We compute the average power $p_t(r_i)$ from the original PSD signal for all regions r_i in R_β (Step 4) and consider the regions by decreasing average power (Steps 5-10). First, we initiate the set of detected regions \mathcal{D} (Step 5). Then, for each region r_i , we check if any of its immediate preceding and succeeding regions, termed neighbors $N(r_i)$, are already in the set of output transmissions \mathcal{D} . If the neighbors are not in \mathcal{D} they we include r_i in \mathcal{D} , else we proceed to the next region by decreasing average power. This sequence of steps is based on the premise that we have successfully detected transmission edges as outlier local maxima and that they will outline regions of transmission (high power) and noise (low power). Finally, we return the set of detected transmissions \mathcal{D} .

The complexity of Algorithm 1 is $O(F \log F)$, since the Haar wavelet decomposition can be computed by averaging with re-use in the hierarchy of size F and due to the need to sort the adjacent multi-scale product differences (Step 6).

Intuitively, our single-sweep detector segments the frequency space based on large differences in multi-scale product values and declares segments of high power transmissions. It has advantages over naive thresholding in the original power space, since multi-scale products are robust to noise. The two key parameters in our method are the threshold scaling parameter β and the scale of analysis s , as these parameters determine AirVIEW's accuracy. In the next section, we describe our unsupervised approach to estimate and optimal set of parameters s and β .

9.3.3 Alignment-driven parameter estimation

We make two important observations, that motivate our methodology for optimal selection of β and s . First, the magnitude of values in the multiscale product will depend on the inherent characteristics of the underlying PSD signal (including the degree of signal oscillations and the average transmitter power) and the selected scale s . Thus, in order to adequately select the top-ranked local maxima (Step 3 in Algorithm 1), we need an adaptive and data-driven approach for selection of the corresponding threshold scaling parameter β for AirVIEW to be accurate for arbitrary SNRs and signal variations. Second, low scale products amplify the edges, but lose the location specificity of the transmissions' edges (e.g. π_1 from Figure 9.3). High-scale products are more specific in the exact position of edges, though more susceptible to noise (e.g. $\pi_{\log|F|-1}$ from Figure 9.3). Hence, there is a trade-off between specificity of edge position and the accuracy of transmission detection that can be controlled by the careful selection of the product scale. We extensively investigate the trade-offs of β and s selection in both synthetic data with controlled signal-to-noise ratio and in real-world spectrum traces (Section 9.4). Our analysis shows that an approach that uses a fixed β and s leads to a sub-optimal performance. Thus, we design an unsupervised method for optimal threshold and scale selection.

Since, in general, we do not have access to supervision (i.e. annotated transmitters), we need to estimate the parameters of Algorithm 1 β and s in an unsupervised manner, i.e. simply using input data. To this end, we exploit a natural domain regularity, namely that transmissions occupy the same band in consecutive time sweeps. Hence, the intuition behind our parameter learning is as follows: *given a short interval of sweeps, find the parameter setting for β and s that results in maximally-aligned detected transmissions across time*. We show experimentally in Section 9.4 that maximizing de-

detection alignment also results in near-optimal parameters for detection of ground-truth transmissions. Hence, in the remainder of this subsection we focus on formalization and maximization of detection alignment.

Let $D_t^i(f_s^i, f_e^i), i \in [1, |D_t|]$ be a single detected active band in time t . Here, f_s^i and f_e^i are the start and end frequency of this detected band. Let also the set of all detected bands at t be denoted as \mathcal{D}_t . Intuitively, a well-aligned multi-sweep detection will result in f_s^i and f_e^i be the same (or as close as possible) across consecutive sweeps. Following this intuition, we formalize the alignment of transmissions detected in consecutive sweeps as the symmetric bi-directional average Jaccard similarity between maximally-aligned detected bands as follows:

$$J_t = \frac{1}{2} \left(\frac{1}{|\mathcal{D}_t|} \sum_{D_t^i \in \mathcal{D}_t} \max_{D_{t-1}^j \in \mathcal{D}_{t-1}} \frac{D_t^i \cap D_{t-1}^j}{D_t^i \cup D_{t-1}^j} \right) \quad (9.3)$$

$$+ \frac{1}{|\mathcal{D}_{t-1}|} \sum_{D_{t-1}^j \in \mathcal{D}_{t-1}} \max_{D_t^i \in \mathcal{D}_t} \frac{D_{t-1}^j \cap D_t^i}{D_{t-1}^j \cup D_t^i} \right), \quad (9.4)$$

where $|\mathcal{D}_t|$ denotes the number of detected bands at time t , $D_t^i \cap D_{t-1}^j$ denotes the number of frequencies in which the two detected bands overlap and $D_t^i \cup D_{t-1}^j$ denotes the number of frequencies in the union of the two bands. Intuitively, the definition of J_t averages the best possible Jaccard similarities for every t band with its predecessor $t - 1$ band and vice versa. Since, individual Jaccard similarities are constrained within the range $[0, 1]$, so is J_t .

Next, for a spectrum scan comprised of T sweeps, we calculate the overall detection alignment as the average J_t :

$$\mathcal{J} = \frac{1}{T-1} \sum_{t=2}^T J_t \quad (9.5)$$

\mathcal{J} also varies between 0 and 1, where a maximally-aligned detection corresponds to

$$\mathcal{J} = 1.$$

We leverage the so defined detection alignment \mathcal{J} for parameter estimation. Given a multi-sweep spectrum scan P_T^F over F frequency bins involving T spectrum sweeps, we seek to find the pair (β_o, s_o) that maximizes \mathcal{J} :

$$(\beta_o, s_o) = \arg \max_{\beta, s} \mathcal{J}(P_T^F, \beta, s). \quad (9.6)$$

Note, that in the definitions of \mathcal{J} , J_t and \mathcal{D}_t we deliberately omitted the parameters β and s in order to simplify the notation, however, all detected bands are acquired by executing Algorithm 1, and thus require the parameters as input.

Since we do not know anything about the function $\mathcal{J}(P_T^F, \beta, s)$, the simplest approach to its maximization is to discretize the space of values of β and perform a parameter sweep over combinations of parameter values. As we will demonstrate the accuracy of actual transmission detection and \mathcal{J} are highly correlated, thus rendering (β_o, s_o) a good choice for detection. We note that our parameter estimation approach is fully unsupervised, as it does not require an annotated ground-truth scan. In real-life spectrum characterization, we employ the approach in a training phase, in which AirVIEW learns the optimal (β, s) combination for a small subscan, and then applies this combination for the remainder of the spectrum characterization campaign.

The complexity of parameter learning is $O(BTF \log^2 F)$, where B is the number of discrete values β considered in the parameter estimation, T is the number of sweeps in the training scan, a factor of $O(F \log F)$ is added for each invocation of the sweep detector (Algorithm 1) and finally an extra factor of $\log F$ is added for the possible number of scales s in the scan, since the Wavelet tree height is logarithmic in F . As we demonstrate in the experiments, a small number of sweeps T is sufficient to train the parameters for further detection and relatively coarse granularity for β (i.e. small B) works reasonably

well. In addition, our empirical evaluation suggest that $\mathcal{J}(\beta, s)$ behaves similar to a concave function w.r.t. both of its parameters, so a simple hill-climbing approach can reduce the factor of $O(B \log F)$ to a constant without compromising quality if frequent and fast parameter estimation is warranted in non-stationary scenarios.

Several important questions arise with our approach to unsupervised parameter estimation.

- *Is the alignment of transmitter detection a good proxy metric for accuracy of transmitter detection?*
- *How long do we need to sense the spectrum before AirVIEW is able to robustly learn the optimal (β, s) combination?*

Our evaluation in Section 9.4.5 seeks empirical answers to these questions. Employing experimentation on ground-truth, synthetically-generated spectrum scans, we show that \mathcal{J} is indeed a good proxy for detection accuracy even in heavy-noise settings. Furthermore, we show that AirVIEW is able to robustly learn the optimal (β, s) combination with as few as 40 spectrum sweeps.

9.3.4 Reconciling transmissions across time

So far, we have described AirVIEW’s approach to detect a transmission in a single sweep t as well as a parameter training for a single sweep detector. As illustrated in Figure 9.1(b), a transmission typically spans multiple spectrum sweeps. In addition, the edge position of detected bands as described above might “shift” due to noise and also due to the uncertainty of the exact edge position introduced by multi-scale products. Hence, in AirVIEW, we reconcile detected instantaneous transmissions over time in a single pass, by enforcing temporal smoothness of transmission edges upon detection.

Algorithm 2: Multi-sweep transmission reconciliation

Require: $p_t(f)$, s , β , λ

Ensure: A set of temporal transmissions $\mathcal{T} = \{T = \{(D, t)\}\}$

- 1: Initialize temporal transmissions $\mathcal{T} = \emptyset$
- 2: **for** Consecutive t consider PSD signal $p_t(f)$ **do**
- 3: $\mathcal{D}^t \leftarrow SingleSweep(p_t(f), s, \beta)$ {Alg. 1}
- 4: **for** $\forall D \in \mathcal{D}$ **do**
- 5: **if** D matches a temporal transmission $T \in \mathcal{T}$ **then**
- 6: $T \leftarrow smooth(D, T, \lambda)$
- 7: **else**
- 8: Start a new transmission $T \leftarrow (D, t)$
- 9: $\mathcal{T} \leftarrow T$
- 10: **end if**
- 11: **end for**
- 12: Report transmissions in \mathcal{T} that were not extended in t
- 13: **end for**

Our multi-sweep transmitter reconciliation approach is detailed in Algorithm 2. It takes as an input a continuous stream of PSD sweeps, the scale s and threshold scaling parameter β (as in Algorithm 1) and an additional temporal smoothing parameter λ . The algorithm reports detected transmissions in contiguous sweeps $T = (D, t)$, where D is the frequency interval of an instantaneous transmission at time t .

We begin by initializing the list of active transmissions (Step 1). We then process the consecutive temporal PSD signals (Steps 2-11). For each time t we detect transmission \mathcal{D}^t (Step 3) and process each of those transmissions D one at a time (Step 4). We match the transmission interval D to the list of active transmissions \mathcal{T} and if the intersection of the D with the union of all D s in a given transmission T is at least half we declare it a match and add a temporally-smoothed version of D to T (Step 6). If no match is found, we initiate a new active transmission T and add it to the set of active transmissions \mathcal{T} (Step 8). After all instantaneous transmissions \mathcal{D}^t are processed, we report transmissions in \mathcal{T} that were not extended in time t (Step 12) and start processing the next scan $p_{t+1}(f)$.

In order to reconcile an instantaneous transmission (D, t) with its matched temporal

transmission T , we consider the edges of D and those of all preceding transmissions in T in function $\text{smooth}(D, T, \lambda)$ (Step 6). Let (f_i, τ_i) be the left edge positions of all transmissions in T including that of D at time t . We compute the time-decayed weighted average of those edge positions as $\bar{f} = [\sum_i w_i f_i / \sum_i w_i]$, where $w_i = e^{-\lambda|t-\tau|}$ is exponentially time-decaying weight giving preference to more recent instantaneous transmissions in T and $\lambda \geq 0$ is an exponential smoothing parameter. We perform similar smoothing average for the right edge \bar{f}' . A large smoothing parameter λ makes the contribution of past transmissions negligible and hence preserves the detected edges of D without smoothing. Alternatively, when $\lambda = 0$ all past edges in T are weighted equally and the edge averages are unweighted means of all past edges, while values in between result in exponentially-decaying importance of past transmissions. Note, that we use the independently detected edges of (D, t) for smoothing, but we report the smoothed versions of the instantaneous transmissions. As demonstrated in our experiments temporal smoothing using small non-zero values of λ helps minimize the “shifting” of consecutive transmissions caused by noise and low-scale s product detection.

Combining transmissions into transmitters. As illustrated in Figure 9.1, a single longitudinally-active transmitter may be comprised of multiple transmissions. We employ a simple frequency position-based combination of transmissions into transmitters. For all temporal transmissions T , we take the union of individual single-time transmission intervals and group T s whose union intervals overlap by at least half of their extent. Note, that this is a simple solution that cannot handle frequency-hopping transmitters and is oblivious to other transmitter properties such as power level and time interval between consecutive transmissions. Handling such cases is important for detection of arbitrary transmitters (i.e. frequency-hopping), however, it is beyond the scope of this chapter. We will consider such extensions in future work and focus on transmitters resident in a stable band for our experimental analysis.

9.4 Evaluation

In this section we evaluate the accuracy of AirVIEW in detecting transmissions in both real and synthetic data sets and in comparison with baseline methods.

9.4.1 AirVIEW implementation

Our current implementation of AirVIEW is a single core Java program and all experiments are executed on commodity desktop machines with the purpose of evaluating the utility of the method to characterize transmitters of various types and noise levels. Our eventual system implementation will, however, make use of recent advances in fast parallel wavelet decomposition for general purpose and specialized architectures such as FPGA and GPUs [73]. In addition, we plan to bring the decomposition and detection “closer to the sensor”, which will enable on-sensor and collaborative spectrum characterization [25].

9.4.2 Data and baselines approaches

We evaluate AirVIEW on both synthetically generated data, in which we control the signal-to-noise ratio, and in a real-world spectrum scans of TV channels, in which we have ground truth position of transmitters. We use two baselines for comparison. *Naive* is a simple thresholding scheme that detects outlier local maxima in the original PSD $p_t(f)$ signal as opposed to the multi-scale product π_s . Extraction of the transmissions is done in a similar way to AirVIEW, once edges are detected. A second baseline *Denoised* follows the general idea of lossy wavelet reconstruction by maintaining the most important coefficients based on magnitude and scale [51]. Such an approach is provably optimal with respect to minimizing the sum of squared errors of the reconstruction given a budget of maintained coefficients and is also shown to limit the effect of noise [51]. The lossy

“denoised” reconstruction is processed similar to Naive to recover transmissions.

9.4.3 Robustness to noise

One of the main challenges in wideband spectrum characterization is detection of transmission in a noisy environment, where noise is introduced both due to the environment and imperfections of the sensor. Hence, we first focus on evaluation of AirVIEW for varying signal-to-noise ratios (SNR). To control SNR, we synthesize realistic spectrum scans of similar characteristics to those we capture using sensors and vary the SNR by decreasing the mean signal power to levels very close to that of the noise. Similar to the noise in radio frequency signals [76] we add additive white Gaussian noise using a normal distribution with mean $-110dBm$ (our “noise floor”) and variance $4.0dBm$. We randomly select regions in time and frequency in which we inject transmissions of a desired power mixed with Gaussian noise of the same magnitude as non-transmission regions.

Since we know the *ground truth* position of instantaneous transmission in this synthetic data, our evaluation seeks to quantify how closely the real transmissions are recovered by the competing techniques. Let $\mathcal{D} = \{D_i\}$ be the set of detected instantaneous transmissions at a given time t (t is omitted in the notation for simplicity). Also, let $\mathcal{A} = \{A_i\}$ denote the set of actual ground truth transmissions at the same time. We define a true positive rate (TPR) measure based on how well we “cover” actual transmissions A_i by detected transmissions D_i , as follows:

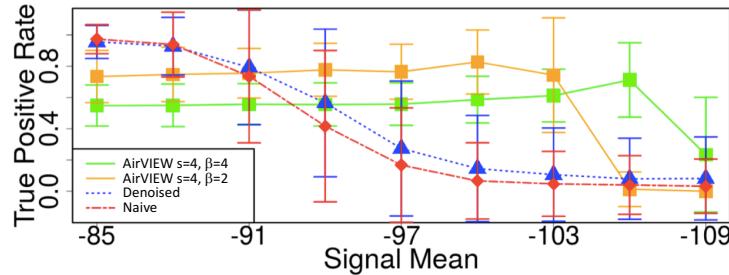
$$TPR = \frac{1}{|\mathcal{A}|} \sum_{A_i \in \mathcal{A}} \max_{D_j \in \mathcal{D}} \frac{|A_i \cap D_j|}{|A_i|} \quad (9.7)$$

where $|A_i \cap D_j|$ denotes the length of the intersection interval common to A_i and D_j . Intuitively, TPR will be 1 if all actual transmissions are “covered” by detected transmis-

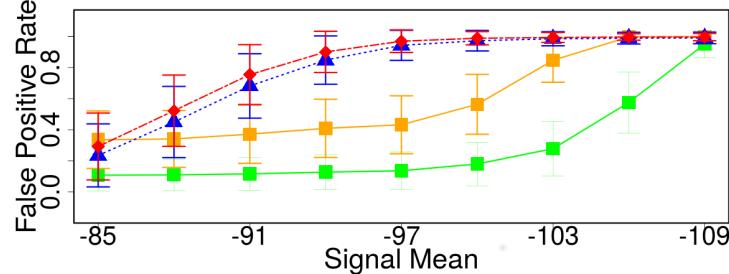
sions completely. On the flip side, we also want to measure the rate of false detections. The false positive rate is similarly defined as:

$$FPR = \frac{1}{|\mathcal{D}|} \sum_{D_i \in \mathcal{D}} \min_{A_j \in \mathcal{A}} \frac{|D_i| - |D_i \cap A_j|}{|D_i|}.$$

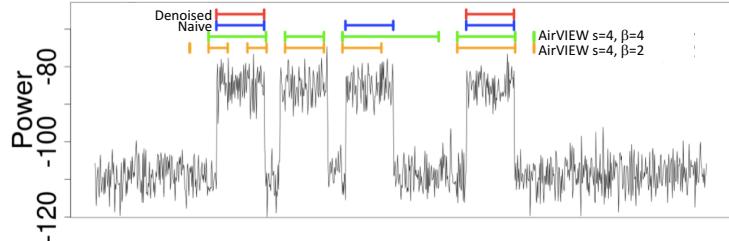
An ideal method will have a TPR of 1 and FPR of 0.



(a) TPR for varying SNR



(b) FPR for varying SNR



(c) Example detections at SNR -85:-110

Figure 9.4: (a) TPR and (b) FPR of two instantiations of AirVIEW ($s = 4/\beta = 2$ and $s = 4/\beta = 4$) and the competing techniques Naive and Denoised at varying signal levels with noise floor fixed at -110dBm. (c) an example single-time detection of a transmitter at -85dBm.

Fig. 9.4 presents the performance of AirVIEW on synthetic data with varying SNR.

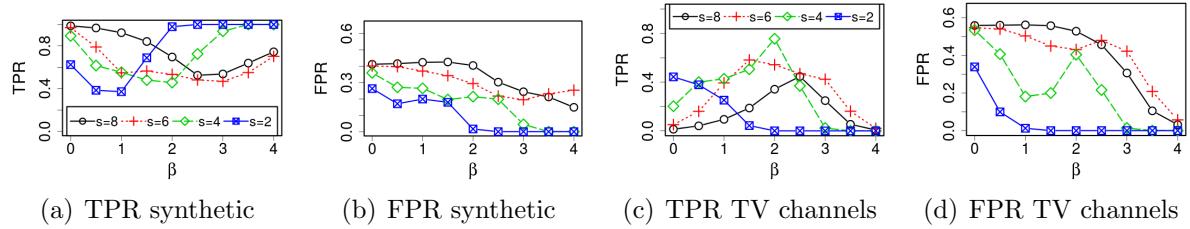


Figure 9.5: (a),(b): Average accuracy of detection in synthetic data using various parameter setting combinations of s and β . Over 1000 runs, 2 transmitters with mean power of -88 were randomly placed in the data and detection was applied using 4 different scales. (c),(d): Similar evaluation in real-world sensed TV band spectrum for constantly transmitting stations of known bands.

We compare the TPR and FPR of two variations of AirVIEW ($s = 4$, $\beta=2$ and 4), Naive and Denoised. At higher SNR levels (-85dBm, -88dBm and -91dBm), Naive and Denoised are able to detect transmitters with high accuracy. However, as the signal approaches the noise floor, their performance deteriorates quickly resulting in rapid increase of FPR and decrease in TPR. AirVIEW's TPR and FPR, on the other hand, remain robust to noise for very low signal to noise ratios (Figs. 9.4(a), 9.4(b)). It is important to note that the slightly lower TPR of AirVIEW in high SNR cases is due to its lack of specificity of edge position for low scales s . As a result AirVIEW is more conservative in placing the edges compares to Naive and Denoised. This behavior can be observed in Fig. 9.4(c), which presents a single-time detection of a -85dBm transmission. This drawback is well-compensated by AirVIEW's high robustness to noise in low SNR settings. Overall, AirVIEW exhibits satisfactory performance with high SNRs and significantly outperforms its counterparts Naive and Denoised when the signal to noise ratio is low. We note that AirVIEW's performance is dependent on the careful selection of s and β and provide a detailed discussion on their effect in the following section.

9.4.4 Effect of the scale s and threshold β

Beyond the robustness to noise we also study the TPR and FPR of our method for varying β and s . Fig. 9.5 summarizes the TPR and FPR achieved by AirVIEW when detecting transmissions in both synthetic and real-world PSD data. In synthetic data (Fig. 9.5(a), 9.5(b)), there is a clear trend of reaching perfect TPR and FPR when using relatively low scales ($s = 2, 4$) and $\beta > 2$ for $s = 2$ and $\beta > 3$ for $s = 4$. Higher scales are more affected by noise and thus do not exhibit as favorable TPR vs. FPR trends for any β . We collect our real-world groundtruth data over 25MHz in the UHF TV bands. This scan contains two active broadcasts, each of which we annotate as a single transmission with smooth edges in frequency. Our results in Fig. 9.5(c) and 9.5(d), indicate that, similar to the synthetic data, lower scales ($s = 2, 4$) and slightly lower thresholds $\beta < 2$ achieve the optimal TPR vs. FPR regime. Of note is that our real-world benchmark achieves slightly worse TPR and FPR as compared to synthetic data. This behavior can be attributed to the noisy nature of the sensed transmitters that causes some single-time scans to be detected as idle due to fading. At the same time, these scans were annotated as occupied.

9.4.5 Alignment-driven parameter estimation

Our evaluation thus far has focused on AirVIEW’s operation with fixed β and s . Such fixed-parameter operation, however, gains suboptimal performance. In line with this observation, we design a data-driven approach to parameter estimation, which we presented in Section 9.3.3. Some key questions arise with our data-driven approach. First, we use the detection alignment \mathcal{J} as a criteria to select the optimal (β, s) . In this section we justify this choice by demonstrating that detection alignment is a good proxy metric for detection accuracy. Second, we show that, indeed maximal detection accuracy

occurs at different β as s increases, which empirically demonstrates the need and benefits of adaptive parameter selection. Third, we quantify the amount of spectrum sweeps necessary for AirVIEW to robustly learn the optimal (β, s) . We present our evaluation methodology and results in turn.

Detection alignment is a good proxy for accuracy

We use two key metrics in the following evaluation: (i) alignment and (ii) accuracy. Our definition of alignment is as in Equation (9.5) of Section 9.3.3. We define accuracy as:

$$\text{Accuracy} = \frac{1}{|\mathcal{D}|} \sum_{D_i \in \mathcal{D}} \max_{A_j \in \mathcal{A}} \frac{|D_i \cap A_j|}{|A_j|} \quad (9.8)$$

Intuitively, our definition of accuracy encapsulates both TPR and FPR, making it an appropriate metric for overall evaluation of AirVIEW’s performance. Accuracy will be low if we fail to detect transmissions (i.e. our TPR is low) or if we falsely detect transmissions (i.e. our FPR is high).

Before we delve in our evaluation, we note that accuracy of transmitter detection can only be calculated if we have ground-truth data for transmitter activity. It is thus impossible to use accuracy for unsupervised transmitter detection (i.e. without prior knowledge of transmitter activity). Alignment, on another hand, does not require ground-truth and is thus ideal for unsupervised parameter estimation. Nevertheless, it is essential to evaluate the ability of alignment-based parameter estimation to gain maximal accuracy. Thus, our evaluation necessarily requires tight control over transmitter configuration and SNR regime.

To achieve such tight control, we run AirVIEW on synthetically-generated spectrum scans. Each scan is comprised of 30 sweeps, each of which contains 1024 PSD val-

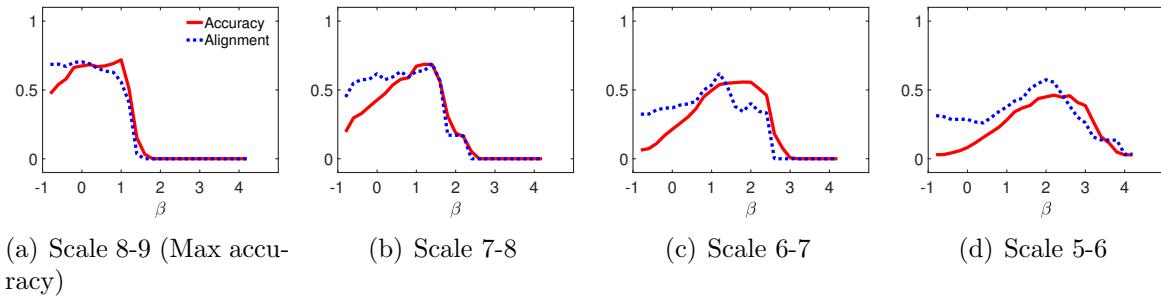


Figure 9.6: Accuracy and alignment at mean transmit power of -103dBm for various (β, s) combinations. We vary scale from 8-9 in (a) to 5-6 in (d). On each graph we vary β from -1 to 4 in increments of 0.2. Across all (β, s) combinations the accuracy and alignment follow a similar trend, which indicates that alignment is a good proxy for accuracy.

ues. Within each 30x1024 scan, we instantiate ten randomly-positioned, non-overlapping transmitters. Noise values in these synthetic scans were drawn from a normal distribution with a mean of -109dBm and a standard deviation of 2.0, whereas the transmitter values were drawn from normal distributions with decreasing means of $[-79 \ -85 \ -91 \ -97 \ -103]$ dBm and standard deviation of 2.0. We note that these cases present increasingly-challenging, yet realistic characterization scenarios. Specifically, all scenarios are challenging due to the high variation of the generated values. The last two scenarios are particularly challenging due to the low power of the generated transmitters. The scenarios are realistic, since the assigned standard deviations are informed by real-world spectrum measurements in the UHF band.

Figure 9.6 presents the accuracy and alignment for our most challenging scenario with mean transmitter power of -103dBm. Figures (a) through (d) present results for decreasing scales as we vary beta from -1 to 4 in regular increments of 0.2. Across all scales, we see that accuracy and alignment follow similar trends, which indicates that regardless of the selected (β, s) combination, *alignment is always a good proxy for accuracy*. We also note that for this particular scenario AirVIEW chooses scale 8-9 and $\beta = 0$ as optimal, since at this (β, s) combination the detection is maximally-aligned.

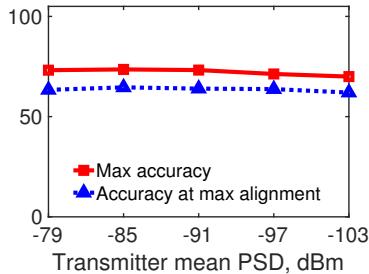


Figure 9.7: Maximum and achieved accuracy with decreasing transmitter PSD. Alignment is a good proxy for accuracy across all PSD regimes. Achieved accuracy is high, persistent and close to the max accuracy.

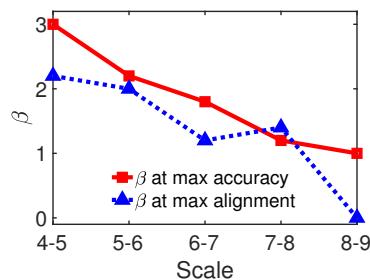


Figure 9.8: β at max accuracy and β at max alignment for increasing scales. As s grows, max accuracy and max alignment correspond to decreasing β s. This underlines the importance of adaptive parameter estimation.

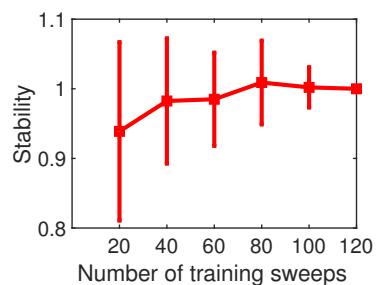


Figure 9.9: Stability of parameter estimation as a function of the number of spectrum sweeps used for training. AirVIEW learns the optimal parameters within 40 sweeps. Stability improves with more training.

We extend this analysis to multiple PSD values in Figure 9.7. The figure shows the maximum accuracy (red continuous line) and accuracy of the maximally-aligned detection (dotted blue line) for increasing mean PSD of the generated transmitters. This figure allows several important conclusions. First, since the accuracy at max alignment is well-coordinated with the max accuracy, we can conclude that alignment is a good proxy for accuracy across various PSD levels of detected transmitters. Second, we show that AirVIEW's accuracy is high, persistent and close to the maximal accuracy across various PSD regimes. Third, AirVIEW is highly-accurate even in the most challenging scenarios, where the transmitters' PSD are only 6dBm higher than the noise floor and the signal variance is as large as 2.0dBm.

Benefits of adaptive selection of (β, s)

Our next experiment demonstrates the need for adaptive selection of parameters (β, s) . Our results are presented in Figure 9.8, that plots the optimal β as the scale increases from 4-5 to 8-9. Red continuous line presents β at max accuracy, while blue

dotted line presents β at max alignment. We see that as the scale increases (i.e. as we get closer to the leafs in our binary tree of wavelet coefficients), the optimal β decreases for both max accuracy and max alignment. Thus, in order to maintain optimal characterization performance, AirVIEW needs to adaptively select (β, s) for a given spectrum sensing environment. A fixed selection of (β, s) will result in suboptimal characterization accuracy.

Training period

Finally, we evaluate the effects of training duration on AirVIEW's stability of spectrum characterization. Let P_T^F be a PSD spectrum scan comprised of T sweeps and F frequency bins in each sweep. Let $Accuracy_t$ be the transmitter detection accuracy over the entire scan duration T when parameter estimation was performed on a subset of all the sweeps of size t . Here, $Accuracy_t$ is calculated as in Equation (9.8). We define stability as:

$$Stability = \frac{Accuracy_t}{Accuracy_T} \quad (9.9)$$

Intuitively, this stability measure captures the relative difference in detection accuracy when training on partial vs. complete spectrum scan.

For this experiment, we execute 100 runs of AirVIEW on synthetic data. Each scan is comprised of 120 sweeps, each containing 1024 PSD values. Within each 120x1024 scan we injected 40 randomly-located, non-overlapping transmitters. Each of these transmitters spans 80 frequency bins and 5 sweeps. The noise values in this scan were drawn from a normal distribution with mean -109dBm and standard deviation of 2.0dBm. The transmitter values were drawn from a similar distribution with a mean of -85dBm.

Figure 9.9 presents average and standard deviation of stability across 100 runs for

an increasing number of training sweeps. We see that AirVIEW is able to learn the optimal parameter configuration (β, s) with as few as 40 training sweeps. Any additional training sweeps further stabilize the parameter estimation as indicated by the decrease of standard deviation.

Discussion. We note that for our training phase to be efficient, we need to observe as many sweeps with actual transmissions as possible. Our current approach does not take steps to ensure that the training sweeps contain actual transmissions. Instead, we blindly characterize every sweep until we observe a critical mass of transmitter occurrences that leads to stable parameter estimation. Such blind approach may incur unnecessary characterization overhead, especially in scenarios with temporally-sparse transmissions. To limit unnecessary computation, one may adopt a pre-processing step that provides a coarse estimation of transmitter availability. Such pre-processing may be based on average scanned power or basic clustering of the measured PSD (i.e. kmeans with K=2).

9.4.6 Detection in time

In Section 9.3.4 we introduced a method to “smooth” the detected edges of individual transmissions accounting for the “fuzzy” nature of transmissions over time caused by noise in measurements. We qualitatively present the effect of smoothing in a synthetic transmitter in Fig. 9.10. To generate this data, we first choose a starting location and bandwidth for the transmission and introduce “drift” patterns in the transmission location in the frequency domain. In each instantaneous transmission, we randomize the position of the edges from a fixed *model* position according to a normal distribution. To introduce drift, we bias the random offsets in a positive or negative direction for fixed windows of time.

Without any smoothing ($\lambda = \infty$) the transmissions shift frequently and disagree

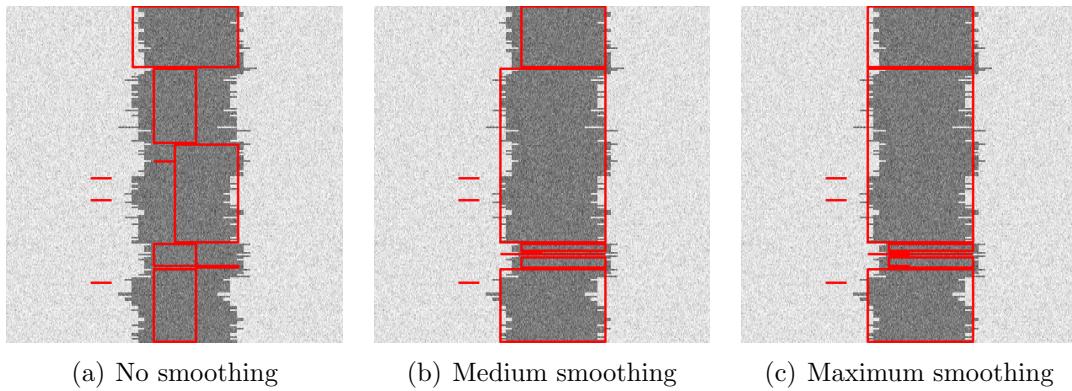


Figure 9.10: Examples of transmission detection with varying smoothing parameters: (a) $\lambda = \infty$, (b) $\lambda = 0.5$ and (c) $\lambda = 0.0$.

due to the drift (Fig. 9.10(a), where the detections are surrounded by red rectangles and the power level corresponds to the level of grey in the image). As we decrease λ (more pronounced smoothing) the instantaneous transmissions start to agree over time (Figs. 9.10(b), 9.10(c)). It is worth mentioning that this scenario presents a particularly challenging instance and real-world scans require moderate levels of smoothing.

9.5 Wideband characterization with AirVIEW

In this section we demonstrate several useful applications of AirVIEW to DSA technology and spectrum policy. We carry out a spectrum measurement campaign to capture a snapshot of a wide target frequency range from $50MHz$ to $4.4GHz$. We then detect transmissions with AirVIEW on the collected scan to evaluate the DSA opportunity over this wide measured band.

Measurement campaign. We collect a 6-second snapshot of the band from $50MHz$ to $4.4GHz$ in an urban area in close proximity to an airport and a military airbase. To be able to cover such a wide range, our measurement platform consists of two USRP N210s; one with WBX daughterboard ($50MHz - 2.2GHz$) and one with SBX daughterboard

($400MHz - 4.4GHz$). The two USRPs were connected through a splitter to a single multi-polarized, wideband antenna that operates in the range of $30MHz$ to $6GHz$. We used a quad-core Lenovo laptop with $16GB$ of RAM running Linux as a host for our software defined radio setup. We implemented a spectrum scanner on top of Gnuradio and configured it to sweep the target spectrum in $25MHz$ steps with dwell time of 6 seconds. At each step, the scanner ran with a sample rate of $25Msps$ and calculated the PSD with FFT size of 1024. This measurement campaign produced a total of $134GB$ of data, where each $25MHz$ chunk was swept 140,000 times. We note that a sampling rate of $25Msps$ and FFT size of 1024 result in very fine granularity of spectrum scans. However, depending on transmitter characteristics, such fine sensing granularity might not be necessary. We leave optimal adjustment of the sample rate and FFT size in each band as a future work.

AirVIEW execution. We run AirVIEW on the above spectrum trace with the following parameters. For the single-time scan detection we set the scale s to 6 and β to 3. For our detection across time we set the smoothing parameter λ to 0.5. When $\lambda > 0$, all past transmissions participate in the smoothing, however, this may affect the running time by summing insignificant, exponentially-decayed values. Thus, we perform the smoothing with the 50 most recent transmissions. In order to combine transmissions into transmitters we set the minimum frequency overlap to 90%.

Wideband characterization with AirVIEW. We begin by characterizing the idle and occupied bands in Figure 9.11, which shows that 74% of all measured bands did not incur transmitter activity for the duration of our measurements. We then measure the number of detected transmitters in Figure 9.12. We see that 26% of all the measured $25MHz$ bands have at least one transmitter. For these occupied bands, we found that AirVIEW robustly identifies TV transmitters, whereby with two exceptions, it finds between 1 and 4 transmitters in a $25MHz$ chunk. In two of the bands, $547.5MHz -$

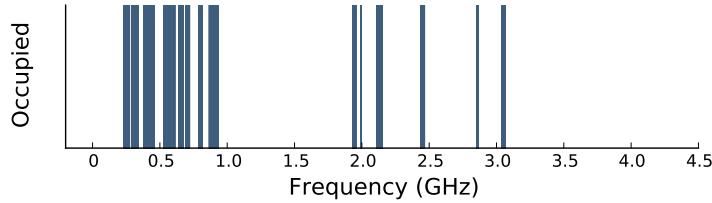


Figure 9.11: Occupied bands

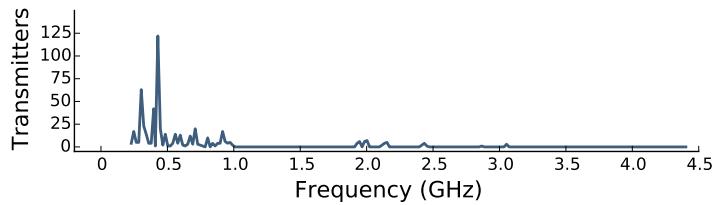


Figure 9.12: Number of Transmitters

$572.5MHz$ and $659.5MHz - 684.5MHz$ we observe 14 and 12 transmitters, respectively, which is due to the noisy nature of some of the TV channels sensed in these chunks that triggers the imperfections in our heuristic that maps transmissions to transmitters. AirVIEW also finds as many as 157 transmitters in the bands $225MHz - 328MHz$ and $335MHz - 400MHz$, which are allocated for federal/military use [15]. Finally, we find a large number of transmitters, 122, in the amateur radio bands ($420MHz - 450MHz$).

Beyond transmitter count, AirVIEW can detect transmitter dynamics in time and frequency. We demonstrate this by evaluating the temporal characteristics of the transmitters detected in our wideband scan. We base our analysis on three metrics: active time, cycle and gap, where the active time is measured as the duration in seconds between the beginning and the end of a transmission, the cycle is measured as the time between the beginnings of consecutive transmissions and the gap is measured as the time between the end of one transmission and the beginning of the next. These metrics allow us to quantify the fraction of time in which a transmitter is active in a given band, the fraction of time the band is available for secondary access, and whether the temporal

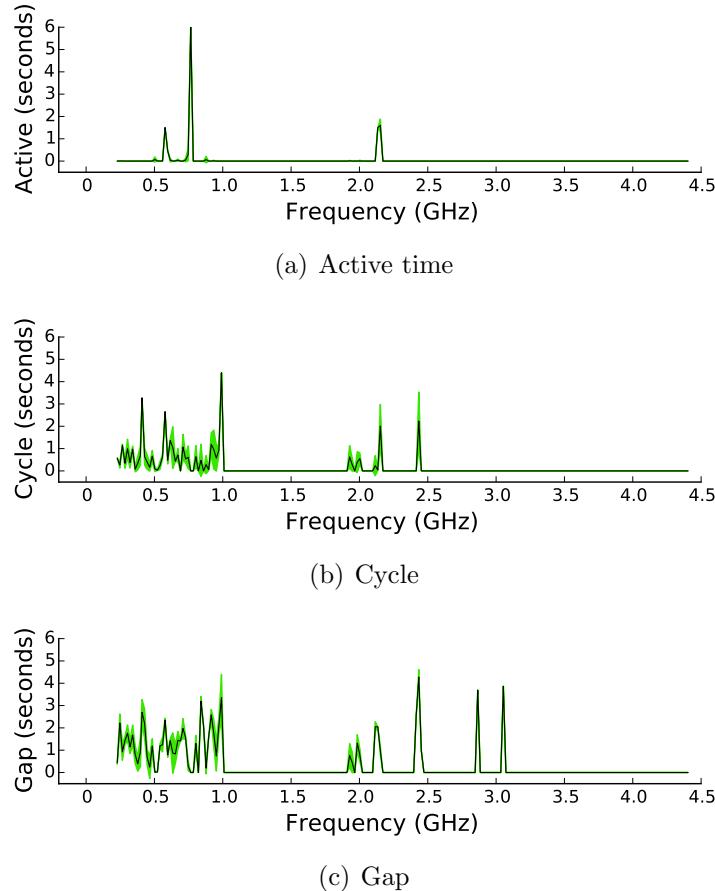


Figure 9.13: Temporal transmitter characteristics of a wideband spectrum scan. The average (black) and standard deviation (green envelope) of active time, cycle and gap illustrate that some intermittently-occupied bands have predictable incumbent patterns while in others the incumbent activity is very unpredictable.

behavior of primary users is predictable.

Figure 9.13 presents average (black) and standard deviation (green envelope) of active time, cycle and gap of transmitters detected in our wideband spectrum scan. Each value is averaged over all transmitters detected in a $25MHz$ chunk. In terms of active time (Fig. 9.13(a)), we see that TV transmitters are active either for the entire 6 seconds or appear as 3-second transmissions, due to intermittent deterioration in the scanned signal. Furthermore, as expected due to their broadcast nature, TV bands exhibit no variance in the active time, cycle or gap. Other bands such as some federal/military bands

($225MHz - 328MHz$ and $335MHz - 400MHz$) and the amateur radio band ($420MHz - 450MHz$) are characterized with smaller active times, however, their cycle size and gaps vary drastically and will thus be hard to predict. For example, the federal/military bands have a typical active time of less than a second and a large gap of $1 - 2$ seconds between consecutive transmissions. However the variance of the cycle and gap are as high as 0.31 and 0.25, respectively, which makes the DSA opportunity hard to predict.

We demonstrated that AirVIEW can perform transmitter detection and characterization and can, thus, play an essential role in DSA technology, spectrum policy and spectrum enforcement. In support of *DSA technology*, AirVIEW can answer what is the secondary access opportunity across various frequency bands. To this end, AirVIEW is able to identify unused spectrum bands for continuous secondary transmissions. Furthermore, where secondary users desire to co-exist with primary users, AirVIEW can evaluate the size and predictability of gaps in primary users' activity. In support of *spectrum policy*, AirVIEW can answer how efficiently is spectrum utilized by mapping the number of transmitters in a wide frequency range along with the size and predictability of their active time. *Spectrum enforcement* engines can leverage AirVIEW's single-pass detection to pinpoint transmitter patterns that deviate from the expected or the historically-observed patterns in a particular frequency range and alert spectrum enforcement authorities.

9.6 Discussion and conclusion

Real-world, rapid spectrum sensing and automated characterization have emerged as key challenges towards next generation mobile wireless networks. Spectrum scans, however, are plagued with high signal variation, which makes automated characterization extremely challenging. This reality is further aggravated when transmitters are sensed with low power level. To address these challenges and enable robust and rapid detection

of low-power transmitters in noisy spectrum scans, we design AirVIEW. AirVIEW is the first mechanism for unsupervised transmitter characterization that enables a single-pass transmitter detection and has very high sensitivity to low-power signals. Furthermore, AirVIEW leverages multiscale products of PSD wavelet decomposition in order to amplify the edges of ongoing transmissions and combat the adverse effects of signal fluctuations, typical for real-world spectrum measurements. We explored key trade-offs in parameter selection for our method and demonstrated AirVIEW’s ability to robustly detect synthetic and ground-truth transmitters. We also employed AirVIEW to create a map of spectrum use and characterize the DSA opportunity over a wide range of frequencies.

AirVIEW presents a significant departure from existing spectrum characterization and activity detection techniques across several criteria. First, by its operation on an alternative (wavelet-based) representation of raw spectrum scans, AirVIEW is able to combat the adversary effects of low-power and high-variability signals that has been a historic challenge in classical techniques. Second, classical spectrum characterization and activity detection methods operate in the raw PSD domain and are, thus, sensitive to a changing noise floor. AirVIEW overcomes this limitation by adopting a data-driven approach to parameter estimation that allows adaptive parameter tuning in the face of a changing noise floor to maintain high characterization accuracy. Third, AirVIEW employs post-processing techniques that robustly reconcile single-sweep transmissions in multi-sweep transmitters for holistic spectrum characterization. Last but not least, this capability is achieved in a fully-unsupervised fashion that does not require prior knowledge of spectrum properties or extensive training.

While AirVIEW achieves robust transmission detection, there are several open problems that shape a compelling future research agenda. Most notably, the current implementation of AirVIEW features a simple heuristic to group transmissions into transmitters that is only informed by the size and overlap of transmissions in frequency. This

approach fails to detect frequency-hopping incumbents and is oblivious to transmitters' power level and co-occurrence. A more robust technique is necessary for a single-pass detection of frequency-hopping transmitters.

In light of these important future directions, AirVIEW lays the foundations for next generation DSA technology and spectrum policy. We enable robust, single-pass transmitter detection in wideband spectrum measurements that is able to answer thus-far open questions related to number of incumbents and the opportunity they grant for secondary spectrum access.

Chapter 10

System Design: Internet Media Upload Caching for Poorly-Connected Regions

10.1 Introduction

The great majority of modern Internet content is built with the assumption of high-quality, always-on network connectivity. However, network infrastructure in developing regions typically lacks in the bandwidth available to users [43], and often exhibits poor latency and loss characteristics. While great strides have been made to improve connectivity through ICTD research, the gap between what media-rich sites demand and what rural network infrastructures are able to provide remains, and in many cases is growing.

Without full participation on the Internet, cultures unique to developing regions are at risk of being forgotten [63]. Web services that encourage participatory culture, often characterized as “Web 2.0,” are critical assets in the democratization of communication and provide a voice to people that previously went unheard. Traditionally, research

in the ICTD space has focused on improving connectivity or the user experience for content *consumption*. This is often accomplished using content prefetching and predictive caching techniques; or requiring users to navigate through software platforms that alter the experience from the “native” versions of Internet sites. While these solutions help narrow the digital divide, the next logical step is to enable users to provide insight into their lives as content *producers* through mainstream sites, not heavily modified sites designed to avoid bandwidth consumption.

In recent years, web content providers have actively sought to improve the user experience for those connecting in developing regions. Google, for instance, has extensively grown their content delivery infrastructure by placing caches on local and regional ISP networks across the world [49]. However, despite the physical proximity to cached content, researchers have found that lack of peering and Internet exchange point (IXP) presence for some Internet service providers causes traffic between African clients and nearby content servers to take circuitous routes, with many routes detouring through European routers before returning to Africa [79]. Facebook has also recently pushed to make their service more available to users in developing regions by engineering their mobile app to use less bandwidth as they found that the poor network conditions in developing regions ‘resulted in slow load times and constant crashes’ [46]. The ‘Facebook Lite’ [169] app only downloads high resolution copies of media if the user explicitly chooses to view a photo or video. The changes indeed make the Facebook app more efficient at consuming media in the face of poor network connectivity; however, uploading content remains an unsolved challenge. We strive to minimize bandwidth consumption for locally uploaded media by transforming and redistributing copies of local content at the network edge.

We perform in-depth, longitudinal analysis on network traces from two production networks in order to better understand network performance and user behavior in the context of media production and consumption. One network is in a village in Sub-

Saharan Africa and the other is in a tribal community in the United States. We find that web media content represents a large percentage of overall traffic in terms of bytes and performance of associated traffic is poor (i.e. large uploads from the local network often fail or are canceled entirely by users due to poor network performance). We further find high locality of interest in media as content uploaded by local community members is heavily consumed by users in the same network. Our findings lead to our proposed solution for allowing uploaded content to be captured, transformed, and redelivered from an Internet cache placed at the access network edge.

Our network trace analysis findings inform our design of VillageCache, a system designed for placement at the network edge that captures locally uploaded media, makes the necessary transformations, and delivers it to local users in order to avoid unnecessary bandwidth consumption on the Internet gateway link. We build a prototype with the assumption that users in such regions have a relatively fast local network connection (e.g. a WiFi LAN) such that they are able to locally transfer large files at reasonable speeds. Unlike a traditional web cache, our proposal avoids the requirement for a download of the object in order to be cached by transforming locally-uploaded media and directly injecting them into the cache for subsequent requests. This is critical as even a single download of high resolution media content is often difficult or impossible in constrained networks.

This chapter makes the following contributions:

- We perform an extensive packet trace analysis from two production networks, leveraging multiple techniques to shed light on usage and performance across multiple layers of the network protocol stack.
- Through our analysis, we find poor media performance for users in the observed networks. Additionally, we observe that moderate and large uploads are rarely successful over a constrained link. We find high locality of interest in locally produced

media. These findings provide key observations for our solution.

- We propose new HTTP headers to enable Internet caches to transform and store locally uploaded media for subsequent requests without requiring a single download from a web content server.
- We evaluate VillageCache in a lab environment that emulates realistic network properties. We show empirically that the system eases bandwidth usage and vastly improves the user quality of experience. Our solution allows users in poorly-connected networks to interact with modern Internet media content, which was previously impractical due to network conditions.

10.2 Traffic analysis

We examine network packet captures gathered from two production networks: one in the village of Macha, Zambia, the other in a tribal community within the United States in order to gain insight into Internet performance and user media consumption.

Macha, Zambia: The Macha network traces we analyze represent a full calendar year, from April 2011 through March 2012. As is somewhat common for networks in developing regions, we have a data hole, in this case during August through October as the network was not operational during that time. Internet access in Macha is available through a microwave terrestrial link shared over a local wireless network connecting approximately 300 community members. The Macha network is delivered using 802.11 devices acting as point-to-point links, hotspots, and layer-2 bridges. Public Internet access is gained through an Internet café. The bottleneck of the network is the shared Internet gateway link which has a committed speed of 2Mbps.

Tribal Digital Village (TDV): We also analyze six months of traffic collected in

2014 from the tribal digital village (TDV) network deployed in southern California¹. The network currently consists of roughly 350 homes subscribing to either 2 Mbps or 3 Mbps service. The Internet backhaul link is 500 Mbps with 200 Mbps provisioned at the gateway location and the remaining 300 Mbps provisioned for subsequent hops. Microwave links provide the backbone to reach locations up to roughly 80 km away. Last-mile access is provided using 2.4 GHz and 5 GHz WiFi equipment.

10.2.1 Traffic characterization

We capture traffic using tcpdump at the networks' gateway links and analyze the packets using Tstat [126] and Bro [141] to profile real-world usage. We first investigate TCP performance metrics for all traffic as well as traffic specifically related to YouTube and Facebook to profile usage of those services as they are the most popular services in terms of flows and bytes in the Macha network. Next, we explore the properties of Macha flows that are classified as ‘user-interrupted.’ Lastly, we characterize the user quality of experience (QoE) for observed YouTube video traffic. The focus of our analysis on the TDV network traffic is on Instagram media; Instagram is the most commonly accessed media type in this trace. Overall traffic performance analysis for this network is omitted for brevity. Our analysis motivates the design of VillageCache by revealing high locality of interest for locally-created media as well as poor network performance for consuming such media using the existing infrastructure.

TCP latency

We investigate latency in the Macha network, which dramatically affects TCP performance and packet loss. Our analysis considers average TCP round trip time (RTT), as calculated by Tstat, for each flow. As shown in Figure 10.1, both YouTube and Facebook

¹<http://sctdv.net/>

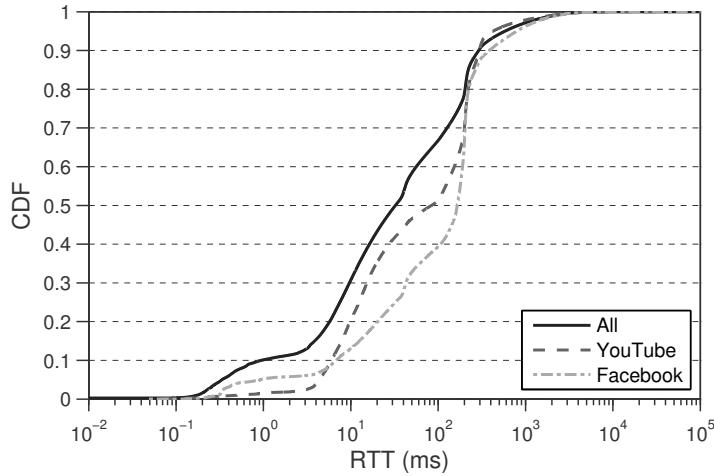


Figure 10.1: Observed RTT in Macha.

traffic typically experience longer RTT values than other traffic. The mean RTT for all traffic is 159.0ms, while the mean RTTs for YouTube and Facebook flows are 173.7ms and 232.1ms, increases of 9.2% and 46.0%, respectively. This is likely due to the physical location of YouTube and Facebook content and circuitous routes to and from the rural ISP. Note that the relatively large number of reported RTT values less than 10ms are due to HTTP errors such as request timeout and malformed requests that were rejected by a local, non-caching Internet proxy, leading to LAN-speed RTT values.

Upload interruptions

To better understand how upstream bandwidth limits impact users' ability to contribute content to the Web, we analyze user behavior in regards to canceling flows using the metric presented in [155], where a flow is determined to be 'interrupted' if all of the following conditions hold: (1) A TCP RST is sent from the client to the server; (2) no TCP RST or FIN is sent from the server to the client; and (3) the time gap between the last data segment from the server and the actual flow end is less than or equal to the average RTT of the flow.

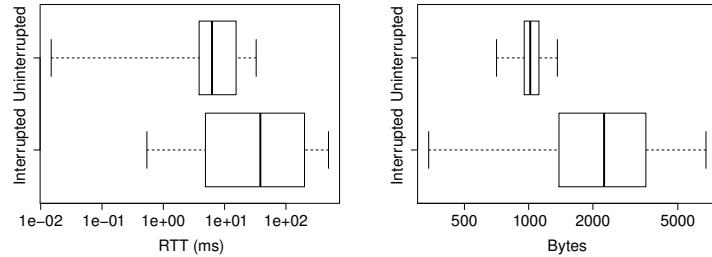


Figure 10.2: User-interrupted HTTP POSTs in Macha.

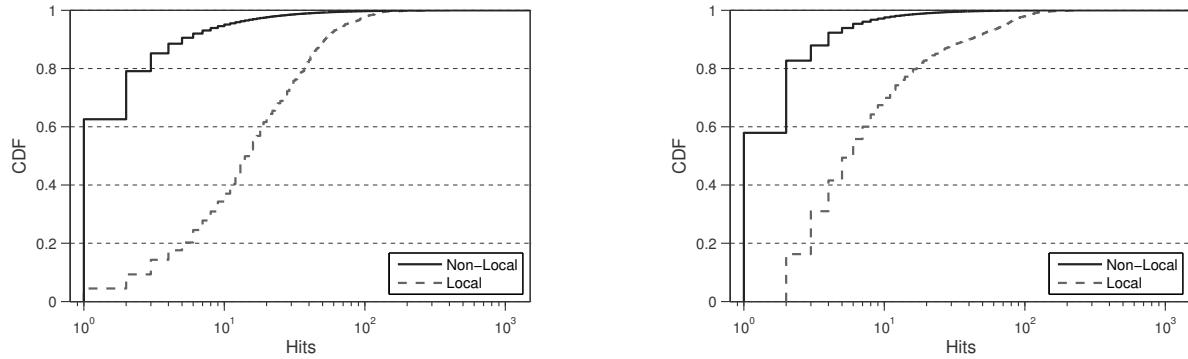


Figure 10.3: Local vs non-local Facebook image interactions in Macha.

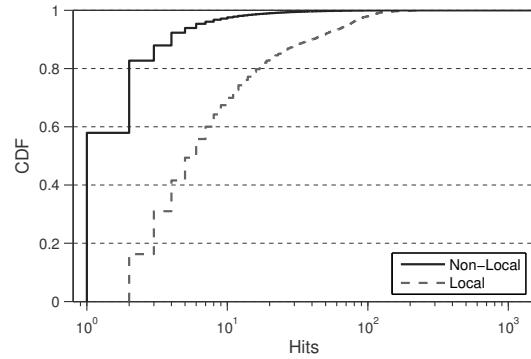


Figure 10.4: Local vs non-local Instagram media interactions in TDV network.

We focus on uploads by only investigating HTTP POST traffic; we compare RTT and bytes transferred between interrupted and uninterrupted HTTP upload flows. The Macha network traces include 9,324,212 POST flows, of which 61,232 were classified as interrupted. While this represents a small percentage of all POST flows, further analysis reveals these flows have significantly different characteristics, including bandwidth consumed, than uninterrupted flows. Figure 10.2(a) shows the measured TCP round trip time for interrupted flows is measurably different than flows that were allowed to complete. As shown, interrupted flows attain larger RTT values on the whole. The median for interrupted flows is 38.27ms while the median for uninterrupted flows is 6.23ms. Also of note is the difference in the upper quartile values between the two classifications. 25% of interrupted flows were longer than 200.99ms, while the longest 25% of uninterrupted flows were longer than 15.39ms. These results lead us to conclude that RTT plays a

significant role in whether the user will cancel an upload or allow it to finish.

Figure 10.2(b) shows a box plot for bytes transferred over interrupted and uninterrupted sessions. We find that interrupted flows tend to have included more bytes than uninterrupted flows, though with a wider variance. Generally, we see that poor network connectivity, such as what we have witnessed in the Macha network, results in user frustration (and ultimately the user canceling a session) during large transfers. These discoveries motivate a system in which large file transfers are avoided across the constrained gateway link.

Facebook media locality

We analyze the locality of interest for Facebook media to discover potential benefits of localizing uploaded content. We parse the traces to find the FBIDs for locally-uploaded Facebook images. To map FBIDs to Facebook uploads we employ a heuristic, where we search for image POSTs followed closely by an image GET, which includes the FBID, as the Facebook client downloads the image in order for the user to add metadata such as tags or friend identification. Using this information we are able to build a table of all Facebook image downloads for the year and determine whether those images were locally-created.

We examine the number of downloads per image for both local and non-local images. We define ‘Local’ as media that was uploaded by a local user and ‘Non-Local’ as media observed in the trace that did not originate in the local network. As shown in Figure 10.3, locally-generated images have a considerably higher hit-rate than those coming from external sources. We observe a mean hit-rate of 23.279 for local images, while non-local images are hit an average of 3.813 times. This high locality of content bolsters the argument for caching locally produced content at the network edge. The users most likely to download media are neighbors and friends of the user that uploaded the content.

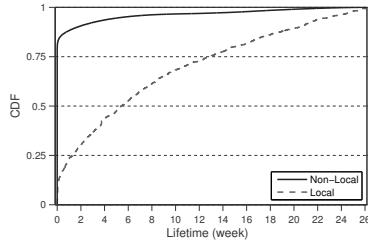


Figure 10.5: Lifetime for Instagram media in TDV network.

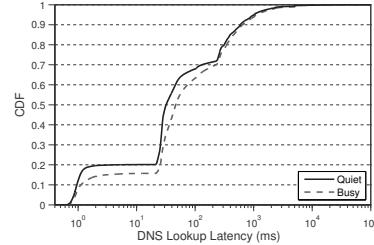


Figure 10.6: Macha DNS latency comparison.

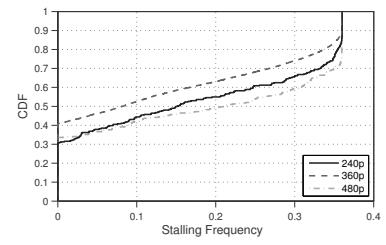


Figure 10.7: Macha YouTube QoE.

Tribal Instagram traffic

We also investigate Instagram traffic on the TDV network in order to illustrate locality of interest. Prior work has shown Instagram to be the most popular service in the TDV network [192]. Additionally, other services such as Facebook and Google have recently moved to SSL encrypted sessions, making detailed header analysis in this trace impossible.

Similar to our Facebook analysis, we find ID values for all Instagram media (classified as public) uploaded during the observation period and compare IDs for subsequent downloads of content. Figure 10.4 shows the number of hits for locally-created media versus non-local media. As with our previous results for Facebook, we observe much more interest in locally created content compared to non-local content.

Next, we look at ‘lifetime’ of local and non-local Instagram media. We define lifetime as the time delta between the first and last observations of a media object in the traces. Similar to the hits observed, Figure 10.5 shows locally-produced media experience much longer lifetimes compared to non-local media. This result further propels the argument for keeping uploaded media cached at the network edge, even in well-connected networks.

	Flows	GB	% Flows	% Bytes
YouTube	48,223	46.224	1.65%	36.00%
Facebook	571,530	9.156	19.61%	7.13%
Twitter	17,802	0.214	0.61%	0.17%
‘Other’ POST	202,857	1.375	6.96%	1.07%
‘Other’ GET	2,074,009	71.441	71.16%	55.64%

Table 10.1: Macha network traffic characteristics.

10.2.2 Potential cache benefits

We illustrate the potential benefits of localizing content by observing the bandwidth burden and latency caused by the web services we target.

Service-specific network load

We analyze all web traffic to explore the bandwidth usage in the Macha network and find that the top two services in terms of bytes consumed are YouTube and Facebook, with a long-tail distribution of other Internet sites. As shown in Table 10.1, roughly 43% of all HTTP bytes are associated to either YouTube or Facebook, with YouTube leading all web sites with around 36% of all bytes. The table also shows Twitter statistics as Twitter is a representative example of the sites falling in the long-tail. Of note is the relatively small number of flows associated with YouTube traffic, yet the large percentage of bytes associated with the service. This makes video traffic an obvious target for a localization since those flows represent very large per flow byte transfers. Facebook traffic exhibits the opposite behavior, with many flows and comparatively fewer bytes. Further analysis revealed this was mainly due to lightweight Facebook IM traffic and a chatty web API. We believe Facebook videos and images remain valuable targets for VillageCache as they represent heavy bandwidth burdens on the network given their size and the overall popularity of the service.

Effect of YouTube and Facebook traffic on overall network performance

Large TCP flows impact other flows' attainable throughput. We investigate the impact of YouTube and Facebook media TCP flows on the overall network performance. We focus on one month of traffic, April 2011, and categorize time intervals into two types: intervals during which YouTube or Facebook flows were present, and intervals with no such traffic. Using the discovered time intervals we categorize DNS flows. ‘Quiet’ flows are DNS queries that began during times when there were no YouTube or Facebook flows. Conversely, ‘Busy’ flows are DNS flows whose first frame was detected during YouTube or Facebook activity. Over the course of one month, we detect roughly 1.2M ‘Busy’ DNS flows and 800k ‘Quiet’ DNS flows.

We consider DNS lookup latency, which we define as the time between an original client request and a server answer, including valid as well as error responses. This approach is similar to previous work [107]. Since DNS traffic uses UDP as the transport protocol, latency measurements more accurately illustrate network conditions such as queuing delay as messages are encapsulated in a single packet and sessions are connectionless. DNS lookup latency is also a critical network metric as DNS performance strongly impacts client web traffic performance; each HTTP object must be resolved from a human-readable URL to an IP address. Figure 10.6 is a CDF plot of the DNS latencies observed. From the figure, we can estimate that the local network includes a DNS server corresponding to latency values of roughly 1ms. We can also observe the first-hop DNS server which corresponds to latencies from 20ms to up to around 200ms. Longer latency values are likely the result of DNS cache misses at the first-hop. In this case the performance of Busy and Quiet flows converge as congestion represents a relatively small portion of the overall lookup latency. We observe that the presence of YouTube or Facebook flows appears to have a detrimental effect on network performance when using

either the local or first-hop server. DNS lookups performed during ‘Busy’ times experience a median increase of roughly 20ms compared to DNS lookups made during ‘Quiet’ times. Given the UDP transport used by DNS traffic, we can posit that the differences are evidence of network congestion.

YouTube QoE

We observe the performance of YouTube in the Macha network to illustrate the poor quality of experience for users due to the constrained link. Prior work [92] has shown that the most important factor in quality of experience for YouTube videos is the phenomenon of video stalls.

We examine the videos detected in the trace files with regards to the frequency of stalling events. YouTube videos are encoded with variable rate codecs; however, we categorize the videos by resolution and use the corresponding average bit rate for comparison. Over the course of the year, Tstat identifies 51,232 requests for videos encoded at 360p (default) resolution, 1,857 240p requests, and 1,529 480p requests. Hoffeld et al. [91] finds that stalling frequency can be estimated for YouTube videos using the following equation, where x equals the V (video bit rate) divided by B (flow bandwidth):

$$F(x) = -1.09e^{-1.18x} + 0.36$$

The resulting stalling frequency value is defined as the ratio of the number of stalls and the duration of the video. Using this value we can estimate the maximum number of stalls that will occur given a video duration. Figure 10.7 shows a CDF of the stalling frequency values for videos observed in the traces. We see that videos encoded with all three resolutions perform rather poorly in the observed network. For example, the median stalling frequency for 240p videos encoded at 256kbps is roughly 0.2, given which

the model estimates a maximum of 12 stalls per 60 seconds of video. Overall, we observe that the QoE for YouTube video in the Macha network is quite poor and as such the network provides one example of an environment that would benefit from localizing the media.

Instagram object duration

An important consideration for Internet caches is the duration that an object must be cached. We examine user interaction with Instagram media in the tribal network traces to find the time deltas between interactions on each media object. We assume a cache will use a least recently used scheme (LRU). Through our analysis we find that the time between hits is less than two weeks roughly 90% of the time with a median of less than one day. These results suggest that a caching system could be effective while retaining only a few weeks of content.

Instagram media storage

We compute the storage that would be required to cache locally-produced Instagram media observed in the TDV network traces. Figure 10.8 shows the number of images and videos uploaded each month in the tribal network. We see that uploads are rather steady across the observation period. We use the Instagram API to download all locally uploaded media objects to discover the total bytes uploaded for each month. We observe a monthly mean of 369.06 MB over the six month observation period. Given these findings, we believe that a cache deployed in networks similar to the observed networks would have attainable storage requirements.

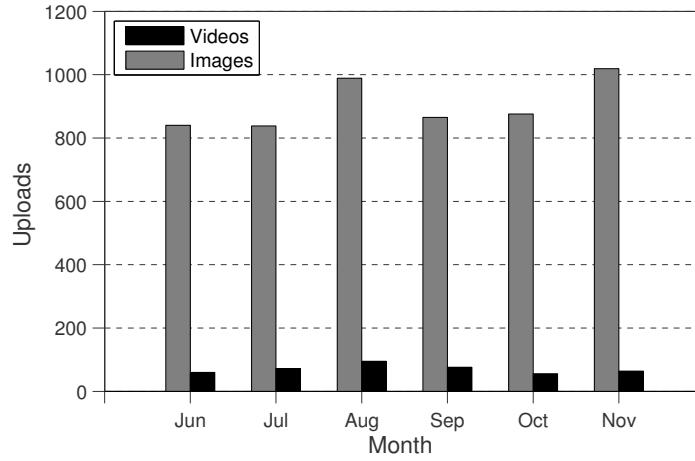


Figure 10.8: TDV network Instagram uploads.

10.3 VillageCache design

Our traffic analysis findings lead to the conclusion that a repository for locally uploaded media content at the network edge that can redeliver local content to mobile users without consuming Internet bandwidth would significantly ease congestion on the gateway link and offer improved performance. This particularly applies to Facebook and YouTube traffic as their associated traffic represents a large proportion of the bandwidth consumed in our Macha traffic analysis. Our design assumes that targeted networks are able to support fast local transfers and the Internet gateway is the bottleneck link. We envision the upload enabled cache will be placed at the Internet gateway, similar to our point of traffic collection in the observed networks, as such a location would typically allow for coverage of an entire rural community. Figure 10.9 gives an overview of the VillageCache system components.

10.3.1 Upload capture engine

User uploads can be captured using multiple methods, depending on the desired user experience and level of acceptable intrusion. We explore possible directions for upload

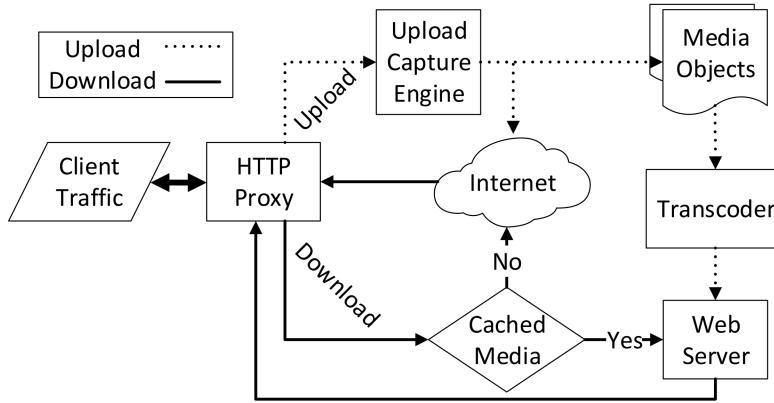


Figure 10.9: System architecture.

capture and discuss tradeoffs for each. All solutions result in a functionally equivalent system in which copies of uploaded media are kept locally and transformed in order to be transparently served back to users in subsequent requests.

Transparent upload scraper

The most desirable upload caching system is one that affords users the ‘native’ experience using web services. Ideally, the presence or absence of an upload cache is completely invisible to the user. Thus, we developed a transparent upload scraper mechanism. We chain a Squid proxy cache with a second proxy server running Tinyproxy on the same machine in order to transparently scrape file uploads. Only HTTP POST requests are forwarded to Tinyproxy while Squid services all other HTTP requests. We modify the Tinyproxy source to check the POST URL and content. If the POST is a Facebook or YouTube media upload, Tinyproxy saves the bytes contained in the POST request to a local file.

The transparent scraping method must include caveats including privacy and compatibility. First, as web services move toward SSL-encrypted traffic we must operate as

an SSL man-in-the-middle using the Squid’s ssl-bump mechanism explained further in section 10.3.3. This obviously raises privacy concerns as our proxy would have access to users’ encrypted upload traffic. In terms of performance, transparent scraping does nothing to improve network performance during the initial upload. User uploads remain constrained by the Internet gateway link. Lastly, we build the scraper assuming HTTP sessions with the web services include an object identifier at the end of the upload in order to properly name the object in the local cache. This relies on the targeted web services to include such information in HTTP session traffic.

Custom local media upload portal

Improved performance can be attained if we are willing to sacrifice the native web service experience for user uploads. Since large file uploads during times of bandwidth saturation result in long sessions and frustration for all, we allow users intending to upload media to YouTube or Facebook to explicitly *time-shift* their transfers by uploading to a custom local portal. In order to time-shift uploads, we require user authorization to work on their behalf using OAuth. We gain access to user OAuth tokens through a local upload portal page in which users log in to their respective YouTube or Facebook accounts and authorize the system. Users can upload video or images to the local page at LAN speeds as they avoid the gateway link. OAuth offers increased security versus locally saving user credentials, as the server only retains OAuth tokens and not clear-text credentials.

Upload time-shifting is accomplished by scripts that run during typically low bandwidth usage times such as late at night. The scripts use YouTube and Facebook API calls, along with the stored user OAuth tokens, to upload files without requiring any user interaction. We have targeted YouTube and Facebook for the initial design; however, similar techniques could be applied to any web service that uses OAuth authentication

and an upload API.

The custom portal solution offers desirable performance and avoids bandwidth saturation via local transfers and time-shifting. However, users are required to explicitly participate through a non-native platform. Additionally, time-shifting results in a delay before the uploaded media are available via the web service. While these tradeoffs are substantial, we believe the solution may be acceptable in highly constrained networks or those prone to disconnection.

Custom HTTP POST request/response

Given the inherent privacy and future compatibility drawbacks of our transparent scrape mechanism, we submit a simple, powerful solution to allow for locally-uploaded media to be captured and appropriately transformed at the network edge for subsequent requests. We propose the introduction of optional HTTP headers associated with POST request / response pairs. Modern Internet proxy caches (e.g. Squid) are readily capable of performing specialized operations for specific HTTP requests and headers, making our solution feasible.

The POST request must specify that some part of the request is potentially cacheable media to signify to the cache that the bytes contained should be kept in memory until the server response is received. The corresponding POST response includes flags that indicate whether the media should in fact be cached, and specifies transformations that must be performed on the object, such as name, file type (e.g. jpg, png, mp4, etc), resolution, along with traditional cache headers such as timeout. Using this information, the cache system can convert the object to the appropriate format in order to enable subsequent cache hits. Such a system essentially mimics the media transformations that are done by the web services. Our proposed HTTP headers include:

POST Request

- **media-upload-cache:** Boolean indicating object is cache-able and should be temporarily kept by edge device.
- **media-upload-range:** Byte range of the cache-able object.

POST Response

- **media-upload-no-cache:** If present, discard object without placing in cache.
- **media-upload-resolution:** Resolution(s) to target with transcoded media.
- **media-upload-codec:** Codec(s) to transcode media.
- **media-upload-mime-type:** IANA MIME type for cacheable objects.
- **media-upload-name:** String to name object when placed in cache.

This solution is powerful because web services can change their media API without notice. A general solution is not possible without cooperation from the service. Since we argue for *optional* HTTP headers, service providers not interested in enabling upload caching can simply not employ the headers in their POST mechanisms. Services that do have an interest in improving user experience in poorly-connected networks can simply add the appropriate POST request and response semantics to their service.

10.3.2 Transcoder

We transcode uploaded media to multiple formats on the cache server in order to serve a broad range of client requests. We use the open-source tools ffmpeg and imagemagick to convert the original videos and images to formats with the same resolutions and codecs that web services use which are readily available [195]. This means we do not have a bit-for-bit copy of media objects, rather we create equivalent local objects to what is served by the web services. In the case of YouTube sessions, the client player requests the specific video format using the ‘itag’ value present in the HTTP request. We examine this value and provide the corresponding version of the video back into the HTTP/HTTPS session

from the local server. As formats offered by YouTube evolve, we anticipate that ffmpeg will maintain the ability to produce files using the same codecs and container formats. Facebook images are converted to both WebP and JPEG formats of multiple sizes while videos are converted to MP4.

10.3.3 HTTP proxy

In order to redirect HTTP traffic to local resources when appropriate, we require a proxy that allows for analysis and modification of HTTP requests. For the prototype, we use Squid 2.7 running in transparent mode to provide both proxying and caching services to the local network. In order to function properly, the system must be placed at a location in the network hierarchy where all traffic destined for and returning from the targeted Internet services will pass through. This can be accomplished in a number of ways, including the use of firewall rules, or by placing the proxy inline with the gateway traffic. We chose Squid because it is widely deployed, well documented, and straightforward to alter in the case of necessary future changes. Lastly, Squid allows for HTTPS traffic proxying and redirection with the ssl-bump extension, a feature not widely available in other proxies.

Redirection to local media copies

Squid allows for HTTP redirection where a listener process examines HTTP requests and chooses whether or not to modify them. We leverage this functionality by building a custom redirector that determines whether or not a request is for cached media. In the case of Facebook images and videos, the ‘FBID’ contained within the query string is used. YouTube requests are parsed to find the ‘ID’ of the requested video. The redirector searches the local web server to check whether there is a local copy available in

the appropriate format for the requested file. If there is a local copy the URL request is altered to use the local web server; otherwise the URL request is passed on to the remote server unmodified.

It is important to note that the redirection engine only redirects image and video file requests and nothing else. A typical Facebook or YouTube HTTP session includes many requests for all of the objects on the page, including HTML, thumbnails, CSS, etc. These other objects are still retrieved from Internet servers rather than any local cache. While this consumes bandwidth, these objects are small in comparison to media content. Importantly, this behavior allows our solution to maintain content privacy settings for both services. For example, if a user does not have permission to access a YouTube video, the HTTP request does not include the video ID in the HTTP header. Instead, the request is of an image showing the user that they do not have the appropriate permissions to access that file. This functionality remains when our system is used, thus preserving YouTube's privacy settings.

HTTPS proxying

Many online services, including YouTube and Facebook, have recently moved toward SSL-encrypted HTTPS rather than unencrypted HTTP. This trend toward encrypted connections will continue and we anticipate that eventually most, if not all, Internet traffic will utilize HTTPS. To that end, we use the ssl-bump extension for Squid in which the proxy intercepts HTTPS sessions and examines them in a manner similar to unencrypted traffic. In order to make this work, clients are dealt certificates from our proxy machine as the server dynamically generates certificates for clients as connections are created. It is important to note that this technique is required for our prototype system only and not meant for a production environment. We feel this feature is valuable despite the potential increased visibility to users. If desired, ssl-bump can be disabled

altogether by changing a single line in the Squid configuration file.

10.3.4 Web server

In order for local redirection to function, we must store media files locally and make them available via a web server. As shown in Figure 10.9, the web server delivers local content back to the proxy during browsing sessions. We use the Apache web server as it allows for streaming modules for various video formats and is widely deployed. Also, as streaming formats continue to evolve it is likely that Apache will remain an early target for compatibility.

10.4 Evaluation

We construct VillageCache in a lab environment to assess its performance and the efficacy of localizing content distribution at the network edge. We use a server running Debian Linux 7.0 with Squid and connect to the Internet through the server with various mobile clients including Android tablets and phones via 802.11g/n WiFi. We place a machine upstream from the cache server in order to emulate the Macha network using *netem* and *tc* to throttle the link bandwidth as well as introduce latency similar to what we observed in Section 10.2.1.

10.4.1 File transfer time

We first explore file transfer times. We emulate a bottleneck 2Mbps link with an average RTT of 134ms that is uniformly distributed with a maximum variation of 20ms. We execute all file transfers 10 times with no competing background traffic and compare mean completion times.

Size (MB)	Local LAN (s)*	2Mbps Link (s)†
1	0.31	30.68
5	1.12	126.33
25	5.47	641.28
100	20.44	1,525.13
500	109.88	5,806.40
1,000	226.96	11,514.54
2,000	576.11	21,973.54

* Performance with cache

† Performance without cache

Figure 10.10: Average file download times.

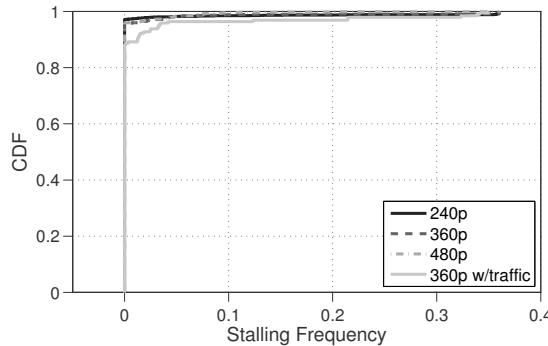


Figure 10.11: Cached YouTube QoE analysis.

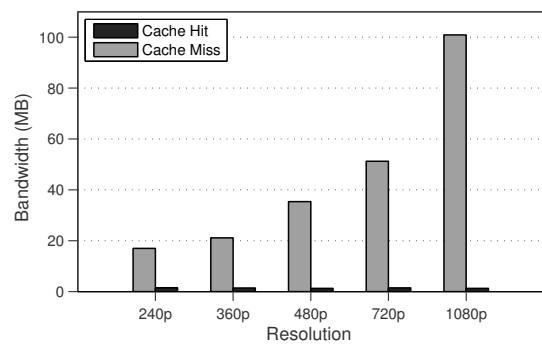


Figure 10.12: Cached video bandwidth savings.

The main goal of our system is to provide users in resource-poor networks the opportunity to fully participate in the modern Internet by improving media content performance. We compare HTTP GET times for various file sizes in two network scenarios to better illustrate the vast performance penalty when connecting via a constrained link. The first scenario consists of a client connected to cached resources via a local WiFi LAN. In the second, client traffic must go through the bottleneck link with parameters derived from prior packet analysis, as described above. We use varying file sizes up to 2GB. As expected, Table 10.10 shows that locally-stored web objects in poorly-connected networks are transferred in a fraction of the time it would take over a constrained link. Moreover, even moderate size file transfers across the constrained link take longer to complete than we can reasonably expect users to wait. Lastly, Internet caf  s in developing regions often charge for Internet access based on time online rather than bytes used. In such situations

it is clearly beneficial to localize media wherever possible.

10.4.2 YouTube

We evaluate the quality of experience for locally cached YouTube videos. These experiments were completed before a change in the HTTP object naming scheme utilized by YouTube. The current operation of YouTube makes object caching impossible, which led us to proposing the use of optional HTTP headers. Nevertheless, we present our experimental results to show the clear potential for caching such large media.

In our experiment automated browsers watch six 5-minute YouTube videos with different resolutions and video codecs passing through a constrained link. We vary the number of concurrent clients between 1 and 6. For each examined resolution (240p, 360p, and 480p) we run 10 sequences of 1 client, then two concurrent clients, and so forth up to six concurrent clients. The total is 210 videos viewed for each resolution with the performance analyzed in aggregate. We additionally run the experiment for 360p videos with 1Mbps UDP background traffic on the bottleneck link to examine the effectiveness of VillageCache during times of congestion.

As shown in Figure 10.11, the stalling frequency for each of the transfers is extremely low, particularly in comparison with Figure 10.7, resulting in stall-free playback for locally connected users. These results illustrate the dramatic improvement of the user experience when consuming files stored at the network edge rather than those stored in distant data centers.

Bandwidth savings

Bandwidth is a scarce and valuable commodity in poorly-connected networks. Our proposed system eases the heavy burden of media by avoiding the costliest transfer during

a session, the media file itself. We compare the bandwidth consumed on the Internet gateway link by a client connecting without the use of VillageCache versus that consumed by a cache-enabled client using proxy logs to show requested files and the corresponding response size in bytes. As shown in Figure 10.12, the bandwidth savings witnessed for cached videos is strikingly large.

These savings are what we would expect from a typical web object cache. However, a system such as ours includes additional benefits that are particularly critical to poorly-connected networks. First, cached locally uploaded videos *never* need to be downloaded from the web as they are immediately placed into the cache at the time of upload. Typical web caches require the object to be downloaded from the source *at least once* to be properly associated with the corresponding web resource. As we see with the file transfer time measurements (Table 10.10) and user interruption of flows (Figure 10.2) behind a constrained link, even one large file transfer may be prohibitive for some networks. Our proposal allows users in these networks to participate in media-heavy web services, which was previously rendered impossible due to poor network conditions. Second, we have shown that users heavily consume locally uploaded media. This creates a multiplying effect for bandwidth saved in our evaluation; the actual bandwidth saved is much higher since local media will be disproportionately downloaded more often than non-local media.

10.5 Related work

In recent years there has been much research in the improvement of Internet experience and information dissemination for users in networks with poor connectivity [54, 55, 58, 101]. Caching techniques are often explored when confronted with a constrained link. Common approaches such as prefetching and aggressive caching can be built into

browser plug-ins [54]. These types of systems help improve performance on individual computers. Our system proposal, on the other hand, is engineered with the cache placed in the network hierarchy where it can serve the entire community rather than individual machines. RuralCafe [57] is designed to allow for web searching over intermittent links in developing regions. TEK [179] also focuses on enabling web search from poorly connected networks using email. We have somewhat similar goals; however, we target a different context in media upload and consumption rather than web search.

Cloudlets [159] have been proposed to reduce delay, particularly for mobile clients. However, it is assumed that the cloudlet computer or cluster is well-connected to the Internet. We approach the problem assuming poor connectivity to the Internet from the local network.

Our work is informed by VillageShare [105], where content locality and time-shifted uploads in a rural African network are explored. We leverage the fact that content produced in poorly-connected networks is heavily consumed by users within them. However, where VillageShare was conceived as a custom local filesharing application, our vision is to make common, unmodified applications work in the absence of strong network connectivity. In our design, interactions are indistinguishable with sessions where content comes directly from the Internet service from the user perspective. This behavior is intentional and based on the notion put forth by Wyche [198], aiming for “ICT interventions grounded in users’ existing practices rather than introducing new and unfamiliar ones.”

Lastly, Google and Facebook have recently proposed high-profile projects whose aim is to provide Internet connectivity to people in the world’s most remote places using technologies such as balloons or drones. Our work is complementary and still beneficial in such environments as caches help most with constrained (e.g. high latency, low bandwidth, high loss, etc) links to the Internet, a likely result of such projects. Connecting the most remote places on Earth will require a multi-pronged approach; localizing content

where appropriate is one of many potential avenues for exploration.

10.6 Discussion and conclusion

In this chapter we have presented analysis of media consumption in two real-world networks to demonstrate the potential for localizing Internet media content in poorly-connected networks. Our study concludes that users access media-rich sites even with poor connectivity but are discouraged from interacting with media content by network conditions. Using our findings we propose a simple solution to cache locally uploaded media at the network edge so as to avoid the requirement for even a single download for the object to be available to other local network users. We evaluate a system prototype of VillageCache in a lab environment and show that such a cache can drastically improve the user quality of experience when accessing local content as well as ease the load on the Internet gateway link.

Aside from improving the end user experience, our proposal is mutually beneficial for the content providers as it offloads low-bandwidth, slow-moving flows from their data centers. As more service providers look to gain users in the most remote places, alternative content distribution techniques are likely to be a major consideration.

Network-level localization of Internet media lends itself to multiple avenues of future research. Given the nature of poorly-connected networks, a delay-tolerant implementation of the system should be explored. An eventual consistency model could be implemented where local upload requests made during times of congestion could be time-shifted and uploaded to the cloud at a later hour when there is less traffic. Web services typically allow for OAuth authentication for API calls, allowing for a middlebox to gather local uploads and move them into the cloud on behalf of the users in a secure manner. Such a solution would alleviate the issues surrounding the initial content upload that we

observed in Section 10.2.1.

Another interesting area for exploration is peering between locally-deployed caches. Network-level localization is convenient for system design, however different distribution models could be investigated leveraging social network analysis across multiple local installations. ‘Similar’ networks could then pre-share content amongst themselves in order to preemptively localize content that is likely to be locally requested. A final example, continuing with the idea of multiple deployed systems in a region, is a mobile app that could act as a data mule for copying local files to other local systems when the user migrates between them.

It is critical to enable full participation on the Internet by all users, particularly those whose cultures are at risk of marginalization. To that end, we hope network-level localization of locally-produced content provides a model that enhances user quality of experience while minimizing reliance on constrained Internet gateway links. In order to provide content to people in the most distant reaches of the world, services must rethink content placement and distribution strategies. Allowing decentralization can enable usage from networks that previously could not support full Internet participation.

Chapter 11

Conclusion and Summary

11.1 Conclusion

Internet and cellular connectivity has typically been achieved through enormous infrastructure investments by commercial providers. Unfortunately, a considerable portion of people reside in locations that, due to economics or other circumstance, do not justify the investment needed to bring high-quality connectivity. This dissertation focuses on improving connectivity in areas where infrastructure is either nonexistent or performing poorly. Our presented research studies both cellular and broadband connectivity and is based upon two cornerstones:

1. **Characterization and analysis** of real-world networks in order to fully understand the inadequacies users currently face [161, 163, 167, 160, 121, 209, 210, 207, 208].
2. **Network systems design** that are informed by our analysis to solve the unique connectivity challenges in resource-limited environments [145, 165, 162, 164, 166].

Our work spans multiple layers of the telecommunications hierarchy, illustrating the broad scope of opportunities for further systems research present in resource-limited environments. We have characterized last-mile connectivity in multiple locations to gain new insight into performance and built systems based on our observations. We have documented the challenges of middle-mile network planning in the context of rural WISP networks and created a system for computationally planning topologies. Lastly, we have investigated the cellular network core of MVNO providers to illuminate routing inefficiencies.

11.1.1 Characterization and analysis

In order to increase connectivity, we must first understand the performance of networks as they operate today. We study both cellular as well as broadband networks in resource-limited environments so that we may identify avenues for improvement.

Cellular. We study multiple facets of cellular communications networks collected from five different settings. We investigate the commercial cellular infrastructure from three providers that serves a large refugee camp in Jordan with the goal of understanding the network user experience in terms of performance and availability. We find that even within the confines of a relatively small area such as a refugee camp, service quality can differ drastically, creating community-level digital divides. We also find that the quality of coverage offered by carriers is unequal and radio-resource congestion is asynchronous, creating the opportunity for users to potentially achieve improved connectivity by switching between carriers in response to congestion on a single provider. We perform similar radio resource trace collection and analysis in rural Guatemala and again discover examples, yet less severe, of infrastructure failing to meet user demand.

Whereas traditional data network topologies and routing policies are comparatively

transparent, cellular core networks generally remain a ‘black box’ for networking researchers. To this end, we investigate the performance of MVNO networks in comparison with the MNO networks whose infrastructure the MVNOs run upon. Despite sharing underlying infrastructure, we found MVNO traffic performance is negatively impacted by routing and peering policies. We find inefficient, geographically circuitous routes for MVNO traffic while MNO traffic paths are relatively direct.

To understand cellular network communications patterns in a developing region, we analyzed large-scale highly-anonymized datasets by Orange Telecom from their network in Côte d’Ivoire. In our analysis we were particularly interested in answering two questions: (i) *how rural cellular network usage differs from that in urban areas*, and (ii), *how do cellular communities evolve in egocentric user networks*. We find that rural residents tend to primarily communicate with users within the same geographic location, whereas urban users exhibit lower locality of interest. Our community evolution analysis indicates there is a potential for higher information dissemination by leveraging information relays in egocentric networks.

Contributions and impact: Our trace collection and analysis of the infrastructure in Za’atari provides a glimpse into a previously understudied network environment. We identify shortcomings of the camp-serving infrastructure and provide recommendations for improving connectivity for camp residents. Our analysis work was published in a highly regarded interdisciplinary conference [163]. Further, our work was included in a technical report [121] generated for the United Nations to aid in planning future refugee camps. Our MVNO work motivates future research as MVNOs have quickly gained traction in the marketplace. In order to achieve throughput goals for the next generation of cellular (i.e. 5G), network inefficiencies such as those we found in MVNO networks must be identified and remedied. Our work with the Orange dataset combines anonymized cellular network traces and population density data to extract the unique characteristics

of cellular network usage in rural versus urban areas in a developing context.

Broadband. We collect and analyze a unique trace, capturing usage in a rural network in Macha, Zambia. This trace, among other unique features, includes traffic before and after an eight-fold upgrade of the Internet gateway capacity. We analyze the ramifications of the increase on user behaviors and overall network performance [209, 210]. We find that TCP flow performance unexpectedly *decreased*, due to users attempting to consume more ambitious Internet content shortly after the upgrade. While the upgrade was substantial, it proved to still be insufficient in the face of increased user demand, leading to a deterioration of performance and poor user experience. Ultimately, user behaviors adjusted to the new, higher bandwidth limit by avoiding heavy multimedia content consumption.

Contributions and impact: We identify surprising aspects of Internet adoption that illustrate novel research problems. Our findings are published in a highly-regarded interdisciplinary journal [209], as well as the top conference in the field of ICTD [210]. Our work also contributes a long-term network trace that allows analysis of Internet usage and performance in developing contexts.

11.1.2 Network systems design

Given our characterization and analyses, we strive to engineer network systems to alleviate the challenges we have uncovered. While our systems are tailored for the contexts we have studied, they are generalizable to other resource-limited settings. We design cellular-based as well as broadband-oriented systems.

Cellular. In our refugee camp and Guatemala studies, we discovered commercial infrastructure that was unable to meet user demand during some parts of the day. We create HybridCell [165], a system based on local cellular network technology that allows

for dynamic user offloading to the local cell during times of congestion on the nearby commercial network. We build mechanisms for peaceful spectrum coexistence, commercial congestion analysis using software-defined radios, and an Android application that programmatically changes networks without requiring user input.

In some cases, a local cellular network is not feasible. Additionally, one of our analysis findings was that radio resource congestion was not synchronous across all of the carriers in a given location. Given these factors, we design SmartCell [162], and Android application that informs users of radio resource congestion on the commercial infrastructure they use. Given the information available through SmartCell, users are empowered to make usage decisions such as physically relocating to areas with better performing infrastructure, or to switch SIM card to another provider in the same location that may have been experiencing less congestion.

During our visit to the refugee camp, we noticed that many residents used a small hill in the camp in an attempt to find a weak signal from their home country provider. Refugee camps, unplanned settlements, and disaster shelters are often located in previously rural areas just at the edge of or outside of existing cellular coverage. We create PhoneHome [164], a cellular system designed to extend existing commercial coverage as well as localize traffic where possible for such scenarios.

Contributions and impact: Our work has gained attention in the networking research community. HybridCell was presented at IEEE Infocom, a top networking conference. SmartCell was published in IEEE Communications Magazine, and PhoneHome was an invited paper in ICCCN.

Broadband. We also build systems based around broadband connectivity in challenging environments. The WISP network model has become popular for providing broadband connectivity to rural areas due to the decreasing cost of commodity wireless equipment. However, one characteristic of WISPs is that they are often built and main-

tained by small staffs with little networking expertise. We build a prototype system [145] aimed at easing the burden of WISP network planning by generating backbone tower locations and topologies that take into account the target coverage areas as well as the underlying topography.

Radio spectrum is a scarce, finite resource that is necessary for wireless communications. Given recent advances in agile radio technologies, researchers and regulators are now working toward dynamic spectrum access models where spectrum can be shared amongst multiple technologies in a given location. In order for DSA to be attainable, we must have precise knowledge of spectrum usage. To this end, we create AirVIEW, a one-pass spectrum analysis algorithm that allows for arbitrary transmitter detection in the presence of noise.

As we studied the network performance in Macha, Zambia, we found users were discouraged to consume heavy multimedia content such as videos or photos as the Internet link lacked the necessary bandwidth to accommodate such traffic. We also found high locality of interest for locally-generated content. In response, we create VillageCache [166], a web cache that scrapes social media uploads and transforms them at the network edge in order to serve them back to local network users without requiring a download from the cloud. VillageCache can drastically reduce the burden on a bandwidth-limited gateway, and allows for time-shifting large uploads to times of the day in which the network is lightly used.

Contributions and impact: Our work on WISP networks was presented at the Hotnets workshop and the coauthors intend to use it as the foundation of a pilot project intended to bring together academic researchers and WISP operators with the goal of creating successful WISP networks while minimizing capital expenditure. AirVIEW, enables wideband spectrum characterization, leading to a scalable solution for building a spectrum occupancy database which is critical for DSA to move forward. VillageCache

was presented at the top development-focused computer science conference [166].

11.2 Summary

Internet and cellular connectivity is a modern necessity; as such it is critical for us to work towards global population coverage in order to facilitate participation from all. Commercial network expansions have successfully brought connectivity to the majority of the world, but the remaining unconnected areas present significant challenges due to, among others, sparseness of users and lack of purchasing power. Bringing adequate connectivity to such environments requires a full understanding of the current challenges so that we may build network systems that provide effective service. This dissertation advances the field by examining network usage and performance in previously unexplored settings. We use those insights to design alternative solutions for improving connectivity for these areas. Moving forward, we must continue to explore a wide-range of networking technologies that are designed for challenging environments, as monolithic, homogeneous systems solutions will continue to be uneconomic and inappropriate for a large portion of the global population.

Bibliography

- [1] <http://www.openbts.org>. [Online; accessed: 2016-04-01].
- [2] <https://www.endaga.com/>.
- [3] <http://www.nuand.com/blog/product/bladerf-x40/>. Online; accessed: 2015-05-01.
- [4] <https://github.com/ptrkrysik/gr-gsm>.
- [5] <https://osmocom.org/projects/baseband/wiki/GSMTAP>.
- [6] African Undersea Cables, <http://manypossibilities.net/african-undersea-cables/>.
- [7] Closing the network ‘coverage gaps’ in Asia. <https://gsmaintelligence.com/research/2015/06/closing-the-network-coverage-gaps-in-asia/>. Accessed: 2015-05-01.
- [8] Dr Math – remote math tutoring using MXIT in South Africa. http://www.elearning-africa.com/eLA_Newsportal/mixing-it-with-dr-math-mobile-tutoring-on-demand/.
- [9] European Commission urban-rural typology. http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Urban-rural_typology. Accessed: 09/02/2013.
- [10] FAQ - Project Fi. <https://fi.google.com/about/faq/#network-and-coverage-1>. Accessed: 2015-05-01.
- [11] FCC Makes 150 Megahertz of contiguous spectrum available for mobile broadband and other users through innovative sharing policies. <https://www.fcc.gov/document/fcc-makes-150-megahertz-spectrum-available-mobile-broadband>. Accessed: 2015-04-18.
- [12] GeoIP Products, MaxMind, <http://dev.maxmind.com/geoip/>.

- [13] ITU statistics – Mobile cellular subscriptions. http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2015/Mobile_cellular_2000-2014.xls.
- [14] Maximizing Mobile. <http://www.worldbank.org/en/topic/ict/publication/ic4d-2012>. Accessed: 2017-02-02.
- [15] National Telecommunications and Information Administration, Federal Government Spectrum Use Reports 225MHz 5GHz. <https://www.ntia.doc.gov/page/federal-government-spectrum-use-reports-225mhz-5ghz>.
- [16] Rural Population in Cote D'Ivoire. <http://www.tradingeconomics.com/cote-d-ivoire/rural-population-wb-data.html>. Accessed: 03/02/2013.
- [17] So much going on! <http://rhizomatica.org/2015/01/14/so-much-going-on/>. [Online; accessed: 2016-04-8].
- [18] Syria Regional Refugee Response. <http://data.unhcr.org/syrianrefugees/regional.php>. Accessed: 2017-03-16.
- [19] NSF Workshop on Spectrum Measurement Infrastructures, Illinois Institute of Technology, Chicago, IL, USA. http://www.cs.albany.edu/~mariya/nsf_smsmw/, April 6-7 2016.
- [20] 3GPP. Mobile radio interface layer 3 specification; Radio Resource Control (RRC) protocol. TS 04.08, June 2000.
- [21] 3GPP. Physical layer on the radio path; General description. TS 45.001, 3GPP, Sept. 2008.
- [22] I. Abdeljaouad, H. Rachidi, S. Fernandes, and A. Karmouch. Performance analysis of modern TCP variants: A comparison of Cubic, Compound and New Reno. In *QBSC*, Kingston, ON, Canada, May, 2010.
- [23] A. Achtzehn, J. Riihijärvi, I. A. Barriá Castillo, M. Petrova, and P. Mähönen. Crowdrem: Harnessing the power of the mobile crowd for flexible wireless network monitoring. In *HotMobile '15*, Santa Fe, New Mexico, USA, 2015.
- [24] R. Ahmed and G. Karypis. Algorithms for mining the evolution of conserved relational states in dynamic networks. In *ICDM*, 2011.
- [25] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE communications magazine*, 40(8):102–114, 2002.
- [26] I. Allagui and J. Kuebler. The Arab Spring and the Role of ICTs. In *International Journal of Communication*, Vol. 5, pages 1435–1442, 2011.

- [27] H. A. Alzoubi, S. Lee, M. Rabinovich, O. Spatscheck, and J. Van der Merwe. Anycast CDNS revisited. In *WWW '08*, Beijing, China, April 2008.
- [28] E. Amaldi, A. Capone, and F. Malucelli. Planning umts base station location: Optimization models with power control and algorithms. *IEEE Transactions on wireless Communications*, 2(5):939–952, 2003.
- [29] A. Anand, V. Pejovic, E. M. Belding, and D. L. Johnson. Villagecell: cost effective cellular connectivity in rural areas. In *Proceedings of ACM ICTD*, 2012.
- [30] R. Anderson, E. Blantz, D. Lubinski, E. O'Rourke, M. Summer, and K. Yousoufian. Smart connect: last mile data connectivity for rural health facilities. In *NSDR*, San Francisco, CA, 2010.
- [31] J. G. Andrews, F. Baccelli, and R. K. Ganti. A tractable approach to coverage and rate in cellular networks. *IEEE Transactions on Communications*, 59(11):3122–3134, 2011.
- [32] Y. Anokwa, C. Dixon, G. Borriello, and T. Parikh. Optimizing high latency links in the developing world. In *WiNS-DR*, San Francisco, CA, September, 2008.
- [33] A. Ayanso, D. I. Cho, and K. Lertwachara. Information and communications technology development and the digital divide: A global and regional assessment. *Information Technology for Development*, 20(1):60–77, 2014.
- [34] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh. White Space Networking with Wi-fi Like Connectivity. *SIGCOMM '09*, Barcelona, Spain, 2009.
- [35] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani. Energy consumption in mobile phones: A measurement study and implications for network applications. In *IMC*, Chicago, Illinois, Nov 2009.
- [36] G. Bernardi, M. K. Marina, F. Talamona, and D. Rykovavov. Increase: A tool for incremental planning of rural fixed broadband wireless access networks. In *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, 2011.
- [37] V. Blondel, G. Krings, and I. Thomas. Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. In *The e-journal for academic research on Brussels, Issue 42*, October, 2010.
- [38] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
- [39] J. Blumenstock and N. Eagle. Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda. In *ICTD*, 2010.

- [40] P. Bogdanov, M. Mongiovi, and A. Singh. Mining heavy subgraphs in time-evolving networks. In *ICDM*, 2011.
- [41] S. Bosio, A. Capone, and M. Cesana. Radio planning of wireless local area networks. *IEEE/ACM Transactions on Networking*, 15(6):1414–1427, 2007.
- [42] E. Brewer, M. Demmer, B. Du, M. Ho, M. Kam, S. Nedevschi, J. Pal, R. Patra, S. Surana, and K. Fall. The case for technology in developing regions. *IEEE Computer*, 38(6), 2005.
- [43] E. Brewer, M. Demmer, M. Ho, R. Honicky, J. Pal, M. Plauche, and S. Surana. The challenges of technology research for developing regions. *IEEE Pervasive Computing*, 5(2):15–23, 2006.
- [44] A. Bruce, D. Donoho, and H.-Y. Gao. Wavelet analysis [for signal processing]. *IEEE spectrum*, 33(10):26–35, 1996.
- [45] M. Buddhikot, P. Mahonen, and J. Zander. NSF Future Directions in Spectrum Management Research (Collocated with DySPAN 2015), Stockholm, Sweden. <http://dyspan2015.ieee-dyspan.org/program/workshops>, September 29 2015.
- [46] K. Burnham. Facebook overhauls android app for emerging countries. <http://www.informationweek.com/software/social/facebook-overhauls-android-app-for-emerging-countries/d/d-id/1278740>, 2014. [Online; posted 20-June-2014].
- [47] F. Calabrese and C. Ratti. Real Time Rome. In *Networks and Communication Studies – Official journal of the IGU’s Geography of Information Society Commission*, 20:3, 247-258, 2006.
- [48] F. Calabrese, Z. Smoreda, V. Blondel, and C. Ratti. Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. In *PLoS ONE* 6(7), 2010.
- [49] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan. Mapping the expansion of Google’s serving infrastructure. In *IMC ’13*, Barcelona, Spain, 2013.
- [50] E. Candes and M. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, March 2008.
- [51] K. Chakrabarti, M. Garofalakis, R. Rastogi, and K. Shim. Approximate query processing using wavelets. *The VLDB Journal/The International Journal on Very Large Data Bases*, 10(2-3):199–223, 2001.
- [52] V. Chandrasekhar, J. Andrews, and A. Gatherer. Femtocell networks: a survey. *IEEE Communications Magazine*, 46(9), Sep 2008.

- [53] R. Chaudhri, G. Borriello, and W. Thies. FoneAstra: making mobile phones smarter. In *NSDR*, San Francisco, CA, 2010.
- [54] J. Chen, D. Hutchful, W. Thies, and L. Subramanian. Analyzing and accelerating web access in a school in peri-urban india. In *WWW '11*, Hyderabad, India, 2011.
- [55] J. Chen, R. Power, L. Subramanian, and J. Ledlie. Design and implementation of contextual information portals. In *WWW '11*, Hyderabad, India, 2011.
- [56] J. Chen, L. Subramanian, J. Iyengar, and B. Ford. Taq: enhancing fairness and performance predictability in small packet regimes. In *Proceedings of ACM EuroSys*, 2014.
- [57] J. Chen, L. Subramanian, and J. Li. Ruralcafe: web search in the rural developing world. In *WWW '09*, Madrid, Spain, Apr. 2009.
- [58] M. Chetty, D. Haslem, A. Baird, U. Ofoha, B. Sumner, and R. Grinter. Why is my Internet slow?: making network speeds visible. In *CHI 2011*, Vancouver, Canada, 2011.
- [59] M. Chetty, S. Sundaresan, S. Muckaden, N. Feamster, and E. Calandro. Measuring broadband performance in south africa. In *Proceedings of ACM DEV*, 2013.
- [60] C. Childs. Interpolating surfaces in ArcGIS spatial analyst. <http://webapps.fundp.ac.be/geotp/SIG/interpolating.pdf>. [Online; accessed: 2015-11-16].
- [61] O. Community. OpenCellID Database. <http://opencellid.org/>. [Online; accessed 1-November-2015].
- [62] L. Cricelli, M. Grimaldi, and N. L. Ghiron. The competition among mobile network operators in the telecommunication supply chain. *International Journal of Production Economics*, 131(1), 2011.
- [63] N. M. Dauenhauer and R. Dauenhauer. Technical, emotional, and ideological issues in reversing language shift: Examples from southeast alaska. *Endangered languages: Current issues and future prospects*, pages 57–98, 1998.
- [64] M. de Bruijn, F. B. Nyamnjoh, and I. Brinkman. Mobile phones: The New Talking Drums of Everyday Africa. 2009.
- [65] F. F. Digham, M. S. Alouini, and M. K. Simon. On the energy detection of unknown signals over fading channels. *IEEE Transactions on Communications*, 55(1):21–24, Jan 2007.
- [66] J. Donner. The rules of beeping: exchanging messages using missed calls on mobile phones in sub-Saharan Africa. *International Communications Association*, 2005.

- [67] B. Du, M. Demmer, and E. Brewer. Analysis of WWW traffic in Cambodia and Ghana. In *WWW*, Edinburgh, Scotland, May 2006.
- [68] N. Eagle, Y. de Montjoye, and L. Bettencourt. Community computing: Comparisons between rural and urban societies using mobile phone data. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, pages 144–150, 2009.
- [69] N. Eagle, M. Macy, and R. Claxton. Network diversity and economic development. In *Science 21 May 2010: Vol. 328 no. 5981 pp. 1029-1031*, May 2010.
- [70] N. Eagle, A. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. In *PNAS, Vol. 106, No. 36.*, 2009.
- [71] Earth Resources Observation and Science (EROS) Center. U.S. Releases Enhanced Shuttle Land Elevation Data. Available at <http://eros.usgs.gov/>, 2013. Accessed: 2016-06-21.
- [72] ESRI and GISCorps. Globe 2G Cell Network Restoration Over Time - Typhoon Yolanda. <http://giscorps.maps.arcgis.com/home/item.html?id=451e24ab8f3c4199861269811a2a9f2e>, 2013. [Online; accessed 5-May-2016].
- [73] J. Franco, G. Bernabé, J. Fernández, and M. E. Acacio. A parallel implementation of the 2D wavelet transform using CUDA. In *2009 17th Euromicro International Conference on Parallel, Distributed and Network-based Processing*, pages 111–118. IEEE, 2009.
- [74] H. S. F. Fraser and S. J. D. McGrath. Information technology and telemedicine in sub-Saharan Africa. *BMJ*, 321:465–466, August 2000.
- [75] A. Ghasemi and E. Sousa. Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs. *IEEE Communications Magazine*, 46(4), Apr 2008.
- [76] A. Goldsmith. *Wireless communications*. Cambridge Univ Pr, 2005.
- [77] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. In *Nature, Vol. 453, No. 7196., pp. 779-782*, 05 June 2008.
- [78] GRASS Development Team. *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.0*. Open Source Geospatial Foundation, 2016.
- [79] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett. Peering at the internet's frontier: A first look at ISP interconnectivity in Africa. In *Passive and Active Measurement*, Los Angeles, CA, 2014.

- [80] E. Gustafsson and A. Jonsson. Always best connected. *IEEE Wireless Communications*, 10(1):49–55, Feb 2003.
- [81] S. Ha, I. Rhee, and L. Xu. CUBIC: a new TCP-friendly high-speed TCP variant. *SIGOPS Oper. Syst. Rev.*, 42:64–74, July 2008.
- [82] B. Haider, M. Zafrullah, and M. Islam. Radio frequency optimization & QoS evaluation in operational GSM network. In *World Congress on Engineering and Computer Science*, San Francisco, CA, October 2009.
- [83] M. Haight, A. Quan-Haase, and B. A. Corbett. Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Information, Communication & Society*, 17(4):503–519, 2014.
- [84] S. Hasan, Y. Ben-David, M. Bittman, and B. Raghavan. The Challenges of Scaling WISPs. In *Proceedings of ACM DEV*, 2015.
- [85] S. Hasan, K. Heimerl, K. Harrison, K. Ali, S. Roberts, A. Sahai, and E. Brewer. GSM whitespaces: An opportunity for rural cellular service. In *DYSPAN*, McLean, VA, USA, 2014.
- [86] K. Heimerl, K. Ali, J. E. Blumenstock, B. Gawalt, and E. A. Brewer. Expanding rural cellular networks with virtual coverage. In *Proceedings of USENIX/ACM NSDI*, 2013.
- [87] K. Heimerl and E. Brewer. The village base station. In *NSDR*, San Francisco, CA, USA, June 2010.
- [88] K. Heimerl, S. Hasan, K. Ali, E. Brewer, and T. Parikh. Local, sustainable, small-scale cellular networks. In *ICTD '13*, Cape Town, South Africa, Dec. 2013.
- [89] Heywhatsthat. <http://heywhatsthat.com/>.
- [90] S. S. Hong and S. R. Katti. Dof: A local wireless information plane. *SIGCOMM Comput. Commun. Rev.*, 41(4):230–241, Aug. 2011.
- [91] T. Hoßfeld, R. Schatz, T. Zinner, M. Seufert, and P. Tran-Gia. Transport protocol influences on youtube videotreaming QoE. *University of Würzburg, Institute of Computer Science, Tech. Rep.*, 2011.
- [92] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz. Quantification of YouTube QoE via crowdsourcing. In *ISM 2011*, Dana Point, CA, 2011.

- [93] Z. Huang, C.-C. Shen, C. Srisathapornphat, and C. Jaikaeo. Topology control for ad hoc networks with directional antennas. In *Computer Communications and Networks, 2002. Proceedings. Eleventh International Conference on*, pages 16–21. IEEE, 2002.
- [94] S. Hurley, S. Allen, D. Ryan, and R. Taplin. Modelling and planning fixed wireless networks. *Wireless Networks*, 16(3):577–592, Apr. 2010.
- [95] iCON Inaugural Workshop. NSF SAVI: Institute for Cognitive Networking (iCON), Stevens Institute of Technology, Hoboken, NJ. <http://www.cognitive-networking.org/>, January 19, 2016 2016.
- [96] S. Ihm, K. Park, and V. S. Pai. Towards understanding developing world traffic. In *NSDR*, San Francisco, CA, June, 2010.
- [97] D. Iland and E. Belding. EmergeNet: Robust, Rapidly Deployable Cellular Networks. *IEEE Communications Magazine*, 52(12), 2014.
- [98] International Telecommunications Union. G.114 : One-way transmission time. <http://www.itu.int/rec/T-REC-G.114/en>, 2003. [Online; accessed 10-May-2016].
- [99] International Telecommunications Union. The World in 2013; Facts and Figures. <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf>, 2013.
- [100] International Telecommunications Union. G.107 : The E-model: a computational model for use in transmission planning. <http://www.itu.int/rec/T-REC-G.107>, 2014. [Online; accessed 10-May-2016].
- [101] S. Isaacman and M. Martonosi. The c-link system for collaborative web usage: A real-world deployment in rural Nicaragua. In *NSDR 2009*, Big Sky, MT, 2009.
- [102] S. Isaacman and M. Martonosi. Low-infrastructure methods to improve internet access for mobile users in emerging regions. *WWW '11*, Hyderabad, India, 2011.
- [103] V. Jacobson. Congestion Avoidance and Control. *SIGCOMM Computer Communication Review*, 18(4), August 1988.
- [104] D. L. Johnson, E. M. Belding, K. Almeroth, and G. van Stam. Internet usage and performance analysis of a rural wireless network in Macha, Zambia. *NSDR*, San Francisco, California, 2010.
- [105] D. L. Johnson, V. Pejovic, E. M. Belding, and G. van Stam. VillageShare: Facilitating content generation and sharing in rural networks. In *ACM DEV 2012*, Atlanta, GA, 2012.

- [106] D. L. Johnson, V. Pejovic, E. M. Belding, and G. van Stam. Traffic characterization and internet usage in rural Africa. WWW, Hyderabad, India, March, 2011.
- [107] J. Jung, E. Sit, H. Balakrishnan, and R. Morris. DNS performance and the effectiveness of caching. *IEEE/ACM Transactions on Networking*, 10(5):589–603, Oct. 2002.
- [108] S. Kechiche. M2M and MVNOs driving US connections growth. <http://gsmaintelligence.com/research/2013/08/m2m-and-mvnos-driving-us-connections-growth/397/>. [Online; posted 15-August-2013].
- [109] M.-S. Kim and J. Han. A particle-and-density based evolutionary clustering method for dynamic networks. *Proc. VLDB Endow.*, 2(1):622–633, Aug. 2009.
- [110] A. B. King. Web site optimization, <http://www.websiteoptimization.com/speed/tweak/average-web-page/>.
- [111] G. Krings, M. Karsai, S. Bernhardsson, V. Blondel, and J. Saramaki. Effects of time window size and placement on the structure of an aggregated communication network. In *EPJ Data Science*, May 2012.
- [112] A. Kumar, J. Chen, M. Paik, and L. Subramanian. ELMR: Efficient Lightweight Mobile Records. In *MobiHeld*, Barcelona, Spain, 2009.
- [113] D. F. Kune, J. Kölndorfer, N. Hopper, and Y. Kim. Location leaks over the GSM air interface. In *NDSS*, San Diego, CA, USA, 2012.
- [114] S. Kyriazakos, G. Karetsos, E. Gkroustiotis, C. Kechagias, and P. Fournogerakis. Congestion study and resource management in cellular networks of present and future generation. In *10th IST Mobile Summit*, Barcelona, Spain, 2001.
- [115] S. Kyriazakos, N. Papaoulakis, D. Nikitopoulos, E. Gkroustiotis, C. Kechagias, C. Karambalis, and G. Karetsos. A comprehensive study and performance evaluation of operational GSM and GPRS systems under varying traffic conditions. In *11th IST Mobile Summit*, Thessaloniki, Greece, June 2002.
- [116] F. La Rue. Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression. United Nations, May 2011.
- [117] J. Laska, W. Bradley, T. Rondeau, K. Nolan, and B. Vigoda. Compressive sensing for dynamic spectrum access networks: Techniques and tradeoffs. In *DySPAN’11*, Aachen, Germany, May 2011.
- [118] L. Leung. Taking refuge in technology: communication practices in refugee camps and immigration detention. <http://www.unhcr.org/4d5b992b9.pdf>, 2011. [Online; accessed 1-May-2016].

- [119] W. Lou. Panel A: Challenges and Future Directions of Wireless Spectrum Research (IEEE INFOCOM), San Francisco, CA, USA. <http://infocom2016.ieee-infocom.org/program/panels>, April 13, 2016 2016.
- [120] F. Lu, H. Du, A. Jain, G. M. Voelker, A. C. Snoeren, and A. Terzis. CQIC: Revisiting cross-layer congestion control for cellular networks. In *HotMobile '15*, Santa Fe, New Mexico, USA, 2015.
- [121] C. Maitland, B. Tomaszewski, E. Belding, K. Fisher, Y. Xu, D. Iland, P. Schmitt, and A. Majid. Youth Mobile Phone and Internet Use January 2015 Za'atari Camp, Mafraq, Jordan. Technical report, Penn State University, Oct. 2015.
- [122] S. Mallat. *A wavelet tour of signal processing*. Academic press, 1999.
- [123] K. W. Matthee, G. Mweemba, A. V. Pais, G. van Stam, and M. Rijken. Bringing Internet connectivity to rural Zambia using a collaborative approach. In *ICTD*, Bangalore, India, 2007.
- [124] I. Mbiti and D. N. Weil. Mobile Banking: The Impact of M-Pesa in Kenya. Working Paper 17129, National Bureau of Economic Research, June 2011.
- [125] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar. Xg dynamic spectrum access field test results [topics in radio communications]. *IEEE Communications Magazine*, 45(6):51–57, June 2007.
- [126] M. Mellia, A. Carpani, and R. L. Cigno. Measuring IP and TCP behavior on a edge node. In *Globecom '02*, Taipei, Taiwan, 2002.
- [127] Y. Meyer. *Wavelets - Algorithms and applications*. 1993.
- [128] Microsoft Corporation. Microsoft Windows Server 2003 TCP/IP Implementation Details. <http://www.microsoft.com/en-us/download/details.aspx?id=13902>, 2007.
- [129] M. Minges. Mobile cellular communications in the Southern African region. *Telecommunications Policy*, 23(7):585–593, 1999.
- [130] S. M. Mishra, J. Hwang, D. Filippini, T. Du, R. Moazzami, and L. Subramanian. Economic analysis of networking technologies for rural developing regions. In *1st Workshop on Internet and Network Economics, Dec 2005*, 2005.
- [131] Mobile for Development Impact. Network Coverage. https://mobileddevelopmentintelligence.com/network_coverage, 2014. [Online; accessed 1-May-2016].

- [132] V. Ndou. E-government for developing countries: opportunities and challenges. In *The Electronic Journal of Information Systems in Developing Countries*, Vol 18, pages 1–24, 2004.
- [133] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004.
- [134] A. Nikravesh, D. R. Choffnes, E. Katz-Bassett, Z. M. Mao, and M. Welsh. Mobile Network Performance From User Devices: A Longitudinal, Multidimensional Analysis. In *PAM 2014*, Los Angeles, CA, USA, March 2014.
- [135] J. Onnela, S. Arbesman, M. Gonzalez, A. Barabasi, and N. Christakis. Geographic constraints on social network groups. *PLoS ONE*, 6(4):e16939, 04 2011.
- [136] M. Paik, J. Chen, and L. Subramanian. Epothecary: cost-effective drug pedigree tracking and authentication using mobile phones. In *MobiHeld*, Barcelona, Spain, 2009.
- [137] T. Parikh, N. Patel, and Y. Schwartzman. A survey of information systems reaching small producers in global agricultural value chains. In *International Conference on Information and Communication Technologies and Development*, Bangalore, India, December 2007.
- [138] N. Patel, D. Chittamuru, A. Jain, P. Dave, and T. S. Parikh. Avaaj otalo: a field study of an interactive voice forum for small farmers in rural India. In *CHI*, Atlanta, GA, 2010.
- [139] N. Patel, S. Klemmer, and T. S. Parikh. An asymmetric communications platform for knowledge sharing using cheap mobile phones. In *ACM Symposium on User Interface Software and Technology (UIST)*, Santa Barbara, CA, October 2011.
- [140] R. K. Patra, S. Nedevschi, S. Surana, A. Sheth, L. Subramanian, and E. A. Brewer. Wildnet: Design and implementation of high performance wifi based long distance networks. In *Proceedings of USENIX/ACM NSDI*, 2007.
- [141] V. Paxson. Bro: A system for detecting network intruders in real-time. *Computer Networks*, 31(23-24):2435–2463, Dec. 1999.
- [142] C. Peng, H. Zheng, and B. Y. Zhao. Utilization and fairness in spectrum assignment for opportunistic spectrum access. *Mobile Networks and Applications*, 11(4):555–576, 2006.
- [143] M. Pizzi. Logging on in Za’atari: Part I. <http://www.smex.org/logging-on-in-zaatari-part-i/>. [Online; accessed: 2016-04-25].

- [144] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye. IP geolocation databases: Unreliable? *ACM SIGCOMM Computer Communication Review*, 41(2), April 2011.
- [145] T. Pötsch, P. Schmitt, J. Chen, and B. Raghavan. Helping the lone operator in the vast frontier. In *HotNets '16*, Atlanta, GA, USA, 2016.
- [146] F. Qian, Z. Wang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck. Characterizing radio resource allocation for 3g networks. In *IMC*, Melbourne, Australia, Nov 2010.
- [147] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs. Cutting the Electric Bill for Internet-Scale Systems. *ACM SIGCOMM Computer Communication Review*, 39(4), October 2009.
- [148] B. Raman and K. Chebrolu. Design and evaluation of a new mac protocol for long-distance 802.11 mesh networks. 2005.
- [149] R. Ramanathan and R. Rosales-Hain. Topology control of multihop wireless networks using transmit power adjustment. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 404–413. IEEE, 2000.
- [150] Range Networks. <http://www.rangenetworks.com/>. [Online; accessed 30-April-2016].
- [151] S. Rayanchu, A. Patro, and S. Banerjee. Airshark: Detecting non-WiFi RF Devices Using Commodity WiFi Hardware. *IMC '11*, pages 137–154. ACM, 2011.
- [152] A. Reda, S. Panjwani, and E. Cutrell. Hyke: a low-cost remote attendance tracking system for developing regions. In *NSDR*, Bethesda, MD, 2011.
- [153] C. C. Robusto. The cosine-haversine formula. *The American Mathematical Monthly*, 64(1):38–40, 1957.
- [154] S. Rosen, H. Luo, Q. A. Chen, Z. M. Mao, J. Hui, A. Drake, and K. Lau. Discovering fine-grained rrc state dynamics and performance impacts in cellular networks. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 177–188, New York, NY, USA, 2014. ACM.
- [155] D. Rossi, M. Mellia, and C. Casetti. User patience and the web: a hands-on investigation. In *Globecom '03*, San Francisco, CA, 2003.
- [156] J. P. Rula and F. E. Bustamante. Behind the Curtain: Cellular DNS and Content Replica Selection. In *IMC '14*, Vancouver, BC, Canada, November 2014.
- [157] B. M. Sadler and A. Swami. Analysis of multiscale products for step detection and estimation. *IEEE Transactions on Information Theory*, 45(3):1043–1051, 1999.

- [158] Sandvine. Global Internet Phenomena Report: 2H 2014. <https://www.sandvine.com/downloads/general/global-internet-phenomena/2014/2h-2014-global-internet-phenomena-report.pdf>. [Online; posted 21-Nov-2014].
- [159] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies. The case for vm-based cloudlets in mobile computing. *IEEE Pervasive Computing*, 8(4):14–23, 2009.
- [160] P. Schmitt and E. Belding. Navigating connectivity in reduced infrastructure environments. In *Proceedings of the Second Workshop on Computing Within Limits*, LIMITS ’16, Irvine, CA, 2016.
- [161] P. Schmitt and E. Belding. Low on air: Inherent wireless channel capacity limitations. In *Proceedings of the 2017 Workshop on Computing Within Limits*, LIMITS ’17, Santa Barbara, CA, 2017.
- [162] P. Schmitt, D. Iland, and E. Belding. SmartCell: Small-scale mobile congestion awareness. *IEEE Communications Magazine*, 54(7), 2016.
- [163] P. Schmitt, D. Iland, E. Belding, B. Tomaszewski, Y. Xu, and C. Maitland. 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, Ann Arbor, MI, USA, 2016.
- [164] P. Schmitt, D. Iland, E. Belding, and M. Zheleva. Phonehome: Robust extension of cellular coverage. In *International Conference on Computer Communication and Networks (ICCCN)*, Waikoloa, HI, Aug 2016.
- [165] P. Schmitt, D. Iland, M. Zheleva, and E. Belding. HybridCell: Cellular connectivity on the fringes with demand-driven local cells. In *IEEE INFOCOM ’16*, San Francisco, CA, USA, Apr. 2016.
- [166] P. Schmitt, R. Raghavendra, and E. Belding. Internet media upload caching for poorly-connected regions. In *ACM DEV*, London, UK, Dec. 2015.
- [167] P. Schmitt, M. Vigil, and E. M. Belding-Royer. A study of mvno data paths and performance. In *PAM 2016*, Heraklion, Crete, 2016.
- [168] S. Sen and B. Raman. Long distance wireless mesh network planning: problem formulation and solution. In *Proceedings of the 16th international conference on World Wide Web*, pages 893–902. ACM, 2007.
- [169] V. Shankar. Announcing Facebook lite. <http://newsroom.fb.com/news/2015/06/announcing-facebook-lite/>, 2015. [Online; posted 4-June-2015].
- [170] D. H. Shin and M. Bartolacci. A study of MVNO diffusion and market structure in the EU, US, Hong Kong, and Singapore. *Telematics and Informatics*, 24(2), 2007.

- [171] A. S. Sife, E. Lwoga, and C. Sanga. New technologies for teaching and learning: Challenges for higher learning institutions in developing countries. In *International Journal of Education and Development using ICT*, Vol. 3, Issue 2, pages 57–67, 2007.
- [172] R. Sinha, C. Papadopoulos, and J. Heidemann. Internet packet size distributions: Some observations. Technical Report ISI-TR-2007-643, USC/Information Sciences Institute, May 2007. Originally released October 2005 as web page <http://netweb.usc.edu/~rsinha/pkt-sizes/>.
- [173] J. Sommers and P. Barford. Cell vs. WiFi: On the Performance of Metro Area Mobile Connections. In *IMC '12*, Boston, Massachusetts, USA, November 2012.
- [174] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin. *Wavelets for computer graphics: theory and applications*. Morgan Kaufmann, 1996.
- [175] L. Subramanian, S. Surana, R. Patra, S. Nedevschi, M. Ho, E. Brewer, and A. Sheth. Rethinking wireless for the developing world. In *Proceedings of ACM SIGCOMM HotNets*, 2006.
- [176] S. Surana, R. K. Patra, S. Nedevschi, M. Ramos, L. Subramanian, Y. Ben-David, and E. A. Brewer. Beyond pilots: Keeping rural wireless networks alive. In *Proceedings of USENIX/ACM NSDI*, 2008.
- [177] Swiftfox. <http://www.swiftfox.net/>.
- [178] P.-N. Tan, M. Steinbach, and V. Kumar, editors. *Introduction to Data Mining*. Addison Wesley, 2005.
- [179] W. Thies, J. Prevost, T. Mahtab, G. Cuevas, S. Shakhshir, A. Artola, B. Vo, Y. Litvak, S. Chan, S. Henderson, et al. Searching the world wide web in low-connectivity communities. In *WWW '02*, Honolulu, Hawaii, 2002.
- [180] Z. Tian and G. Giannakis. A wavelet approach to wideband spectrum sensing for cognitive radios. In *CROWNCOM*, Mykonos Island, Greece, June 2006.
- [181] Z. Tian and G. B. Giannakis. Compressed sensing for wideband cognitive radios. In *IEEE ICASSP'07*, Honolulu, HI, April 2007.
- [182] Tower DB. <http://towerdb.invaneo.org/>.
- [183] Default TTL Values in TCP/IP, <http://www.map.meteoswiss.ch/map-doc/ftp-probleme.htm>.

- [184] University, Center for International Earth Science Information Network -. CIESIN -. Columbia and FAO, United Nations Food and Agriculture Programme -. and CIAT, Centro Internacional de Agricultura Tropical -. Gridded population of the world, version 3 (gpwv3): Population count grid, future estimates. <http://dx.doi.org/10.7927/H42B8VZZ>, Nov 2005.
- [185] B. E. Usevitch. A tutorial on modern lossy wavelet image compression: foundations of jpeg 2000. *IEEE signal processing magazine*, 18(5):22–35, 2001.
- [186] A. Vakali and G. Pallis. Content delivery networks: status and trends. *IEEE Internet Computing*, 7(6), November 2003.
- [187] N. Vallina-Rodriguez, A. Auçinas, M. Almeida, Y. Grunenberger, K. Papagiannaki, and J. Crowcroft. Rilanalyzer: A comprehensive 3g monitor on your phone. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, pages 257–264, New York, NY, USA, 2013. ACM.
- [188] N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, N. Weaver, and V. Paxson. Beyond the Radio: Illuminating the Higher Layers of Mobile Networks. In *Mobisys 15*, Florence, Italy, June 2015.
- [189] J. A. Van Dijk. Digital divide research, achievements and shortcomings. *Poetics*, 34(4):221–235, 2006.
- [190] P. van Hoorik and F. Mweetwa. Use of Internet in rural areas of Zambia. In *IST Africa*, Windhoek, Namibia, May 2008.
- [191] L. Vannini and H. le Crosnier. *NET.LANG: Towards the multilingual cyberspace*. Creative Commons, C&F edition, March 2012.
- [192] M. Vigil, M. Rantanen, and E. Belding. A first look at tribal web traffic. In *WWW '15*, Florence, Italy, 2015.
- [193] W. W. Vithanage and A. S. Atukorale. Bassa: a time shifted web caching system for developing regions. NSDR '11, Bethesda, Maryland, USA, 2011.
- [194] A. Wesolowski, N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee. Quantifying the impact of human mobility on malaria. In *Science*, Vol. 338, No. 6104. (12 October 2012), pp. 267-270, October 2012.
- [195] Wikipedia. Youtube — wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/YouTube#Quality_and_formats, 2015. [Online; accessed 19-September-2015].
- [196] WISPTools. <http://wisptools.net/>.

- [197] Y. Wu, P. A. Chou, Q. Zhang, K. Jain, W. Zhu, and S.-Y. Kung. Network planning in wireless ad hoc networks: a cross-layer approach. *IEEE Journal on Selected Areas in Communications*, 23(1):136–150, 2005.
- [198] S. P. Wyche, A. Forte, and S. Yardi Schoenebeck. Hustling online: understanding consolidated facebook use in an informal settlement in Nairobi. In *CHI 2013*, Paris, France, 2013.
- [199] L. Yang, W. Hou, L. Cao, B. Y. Zhao, and H. Zheng. Supporting Demanding Wireless Applications with Frequency-Agile Radios. *NSDI'10*, San Jose, California, 2010.
- [200] Y. Yuan, P. Bahl, R. Chandra, P. A. Chou, J. I. Ferrell, T. Moscibroda, S. Narlanka, and Y. Wu. Knows: Cognitive radio networks over white spaces. In *DySPAN 2007*, pages 416–427, April 2007.
- [201] T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Communications Surveys Tutorials*, 11(1):116–130, 2009.
- [202] K. Zarifis, T. Flach, S. Nori, D. Choffnes, R. Govindan, E. Katz-Bassett, Z. M. Mao, and M. Welsh. Diagnosing Path Inflation of Mobile Client Traffic. In *PAM 2014*, Los Angeles, CA, USA, March 2014.
- [203] F. Zarinni, A. Chakraborty, V. Sekar, S. R. Das, and P. Gill. A First Look at Performance in Mobile Virtual Network Operators. In *IMC '14*, Vancouver, BC, Canada, November 2014.
- [204] L. Zhang and P. Bao. Edge detection by scale multiplication in wavelet domain. *Pattern Recognition Letters*, 23(14):1771 – 1784, 2002.
- [205] M. Zheleva, A. Chowdhery, R. Chandra, A. Kapoor, and P. Garnett. TxMiner: Identifying Transmitters in Real-World Spectrum Measurements. In *IEEE DySPAN'15*, Stockholm, Sweden, September 2015.
- [206] M. Zheleva, A. Paul, D. L. Johnson, and E. Belding. Kwiizya: Local Cellular Network Services in Remote Areas. In *MobiSys '13*, Taipei, Taiwan, 2013.
- [207] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. Bringing visibility to rural users in Cote D'Ivoire. In *ICTD '13*, Cape Town, South Africa, Dec 2013.
- [208] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. Community detection in cellular network traces. In *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes - Volume 2*, ICTD '13, Cape Town, South Africa, 2013.

- [209] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. Internet bandwidth upgrade: implications on performance and usage in rural zambia. *Information Technologies & International Development*, 11(2), 2015.
- [210] M. Zheleva, P. Schmitt, M. Vigil, and E. Belding. The Increased Bandwidth Fallacy: Performance and Usage in Rural Zambia. In *ACM DEV'13*, Cape Town, South Africa, December, 2013.