LTE Radio Analytics Made Easy and Accessible

by

Ezzeldin Omar Hussein Hamed

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

Signature redacted

Author . . . .
Department of Electrical Engineering and Computer Science
January 16, 2015

Signature redacted

Certified by . . . .
Dina Katabi
Professor
Thesis Supervisor

Signature redacted

Accepted by . . . .
Leslie A. Kolodziejski
Chairman, Department Committee on Graduate Students

LTE Radio Analytics Made Easy and Accessible

by

Ezzeldin Omar Hussein Hamed

Submitted to the Department of Electrical Engineering and Computer Science
on January 16, 2015, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science and Engineering

Abstract

Despite the rapid growth of next-generation cellular networks, researchers and end-users today have limited visibility into the performance and problems of these networks. As LTE deployments move towards femto and pico cells, even operators struggle to fully understand the propagation and interference patterns affecting their service, particularly indoors. This thesis introduces LTEye, the first open platform to monitor and analyze LTE radio performance at a fine temporal and spatial granularity. LTEye accesses the LTE PHY layer without requiring private user information or provider support. It provides deep insights into the PHY-layer protocols deployed in these networks. LTEye's analytics enable researchers and policy makers to uncover serious deficiencies in these networks due to inefficient spectrum utilization and inter-cell interference. In addition, LTEye extends synthetic aperture radar (SAR), widely used for radar and backscatter signals, to operate over cellular signals. This enables businesses and end-users to localize mobile users and capture the distribution of LTE performance across spatial locations in their facility. As a result, they can diagnose problems and better plan deployment of repeaters or femto cells. We implement LTEye on USRP software radios, and present empirical insights and analytics from multiple AT&T and Verizon base stations in our locality.

Thesis Supervisor: Dina Katabi
Title: Professor

Disclaimer

The work in this thesis was done in collaboration with another student Swarun Kumar. Most of the ideas in this thesis were developed during our discussions and brainstorming sessions. We both share credit for the design and system architecture.

Acknowledgments

This research was performed under the supervision of Professor Dina Katabi and in collaboration with my colleague Swarun Kumar. This work was published in ACM SIGCOMM 2014, and it was funded by the National Science Foundation.

I am grateful to my supervisor Dina for her great guidance support and continuous motivation. I am honored to be part of the NETMIT group and working under the supervision of Dina. It has been an amazing learning experience.

I have been fortunate to work with my colleague Swarun, I enjoyed all the days we spent understanding the LTE standard, writing the code, and building this interesting system. I have been also fortunate to work with Omid on building the Sparse FFT chip, and Rahul on building our wireless testbed. Working with you guys was so much fun.

I would like to thank Professor Li Erran Li for his great contributions to this work, John Romanishin for fabricating the rotating arm used in our experiments, and my friends in the NETMIT group for their insightful comments on our paper.

I am grateful to my friends Haitham, Fadel, and Amr for their valuable advice and great support. You have been always a great source of encouragement.

Finally, I can never forget the people who contributed to my success more than I did. My mother Samira, you are the cure to all my worries. My father Omar, I wish I have your patience. And my wife Aisha, you gave up on everything just to take care of me and our kids. Words are never enough to express my feelings to you, I love you all and I will do my best to never let you down.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cellular service has become an integral part of our life. Yet as users and researchers, we have little visibility into the performance and real problems of these networks. Even the little information we have is primarily from trace analysis sanctioned by mobile operators [14, 15]. The lack of open and transparent access into the operation and inefficiencies of the cellular physical layer limits our ability as researchers to contribute effectively to the development of these networks. It also limits the ability of policy makers to independently verify operators' claims of spectrum needs, and make informed decisions on licensed vs. unlicensed spectrum.

The need for increased visibility into the cellular PHY-layer is further emphasized by three recent trends. First, cellular deployment is moving towards small, femto, and pico cells [25], many of which will be deployed by a user to cover her home or small business. As a result, cellular operators no longer have full control over their LTE deployments, and struggle to understand the propagation and interference patterns affecting their service, particularly in indoor settings. Open, cheap, and ubiquitous radio monitoring can help deal with the challenging propagation patterns brought in by small cells. Second, the rise of LTE-based M2M applications motivates a more open access to LTE signal-based analytics. For example, Walmart, Home Depot, or Disneyland may leverage LTE signals and recent RF-based localization techniques to track how clients navigate their space and obtain business analytics. Also, oil and gas companies who are deploying LTE-supported seismic sensors [6] can leverage access to LTE radio propagation to better plan their sensor network and debug connectivity problems. Third, the FCC plans to repurpose certain bands (e.g., 3.5 GHz) for multi-tier spectrum sharing, including LTE small cell deployments [9]. Operating in a shared spectrum naturally fits with an open model for signal monitoring and analysis, where all the entities sharing the spectrum can better understand the problems and cooperate to find solutions.

All of these reasons motivate a more open access to the cellular PHY layer, particularly LTE. In this thesis, we ask the following question: Can we access the cellular PHY-layer of today's LTE deployments, without support from mobile operators? In particular, can we do this without requiring access to private user data or encrypted LTE channels? If such a service exists, it could contribute to better deployment of femto cells and repeaters, more businesses built on top of LTE networks, better informed spectrum policies, more efficient sharing of newly released bands, and an overall increase in transparency in an industry that is a major part of the world economy.

We introduce LTEye, an open platform for monitoring and analyzing the LTE PHY layer. LTEye is a passive sniffer that runs on off-the-shelf software radios (e.g. USRPs). It does not require provider support, and hence can serve end users, researchers, policy makers, or mobile broadband providers. LTEye extracts per-user analytics purely from the LTE control channels that contain meta-deta on uplink and downlink transmissions, without accessing private data or system parameters from encrypted data channels. Specifically, it tracks individual users based on their temporal PHY-layer IDs, without requiring or exposing their private information. It then intelligently links these IDs across control messages to generate transmission records for each user. Each record reports the transmission's resource utilization, modulation and coding rate, and frame loss rate. LTEye also records the wireless channel it perceives from base stations and mobile users within radio range. These channels are used to accurately monitor the 3-dimensional physical location of the users. LTEye maintains these records in a database called LTEyeDB. It processes the records to generate two dimensions of fine-grained analytics: temporal analytics, to track LTE performance over time, and spatial analytics, to characterize LTE service across spatial locations.

We implemented LTEye on USRP software radios [7]. We deployed LTEye in four locations in our campus to compare the temporal performance of two major LTE providers: AT&T and Verizon. Our results revealed several inefficiencies in these networks. First, both providers deploy a Frequency Division Duplexing scheme, which uses independent equally sized frequency bands for uplink and downlink traffic. However, LTEye reported that for both providers, the average utilization of downlink resources (25.2% - AT&T, 58.2% - Verizon) far exceeded that of uplink resources (0.6% - AT&T, 2.6% - Verizon). While it is expected that the downlink is higher in demand, our results reveal that the LTE uplink is a remarkable 20 to 40 times less utilized than the downlink. LTEye's analytics can therefore help policy makers encourage operators to adopt revised LTE standards that allow more prudent allocation of resources to the uplink and downlink[1], without relying on data from providers themselves to make the case.

Second, LTEye localized certain spots in our campus, where Verizon cellphones suffer poor link quality and often switch to 3G, despite reporting high signal power from the LTE base station. To investigate this, we moved our LTEye sniffers to these spots and found that they experienced high inter-cell interference (about 27 dB) from as many as five different base stations. To make matters worse, many of these base stations used overlapping channel estimation pilots that interfered, significantly impacting the decodability of these transmissions. These results help end-users better plan the deployment of femto cells to avoid such interference. Further, they benefit cellular providers themselves because they reveal interference problems that end-users face in indoors, hitherto inaccessible to providers. Interestingly, some of these PHY-layer inefficiencies may be unknown even to the operators as they are part of the PHY-layer implementations adopted by the base station vendors.

LTEye also benefits researchers by providing deep insights into the PHY-layer protocols deployed by cellular providers. While the LTE standard spells out much of the

[1] E.g. Asymmetric Carrier Aggregation [18] in LTE Advanced (3GPP Release 10) allows downlink resources to exceed the uplink.

PHY layer, the choice of rate adaptation algorithm is still left to individual operators. To gain insights into this algorithm, we analyzed LTEyeDB records of an AT&T base station in our locality. We found that even for static users with completely coherent channels and stable SNRs, the modulation and coding scheme changes significantly even between adjacent transmissions. More interestingly, the average modulation and coding of frames sent to a user changes, based not only on her wireless channels, but also on the network state as a whole. Specifically, if the network is scarcely utilized, the base station transmits to the user conservatively at low modulation on average, even if the wireless link is stable and supports much higher modulation. In contrast, as network utilization increases, under identical SNRs, the base station steadily increases its modulation to support more aggressive data rates to the user. Such analytics on the performance and design choices of today's cellular operators help researchers design better LTE protocols.

LTEye enables businesses and network administrators to continuously monitor the spatial locations of mobile users, and build a geographic heatmap of LTE coverage and performance within their facility. However, accurately localizing mobile users purely based on their LTE signals is a challenging task. This is because past work on accurate indoor localization proposes two classes of solutions that are ill-suited to LTE networks: First, localization using antenna arrays [29, 17] requires large bulky arrays, owing to the relatively low frequencies of LTE signals. Second, recent localization techniques using synthetic aperture radar (SAR) are less bulky, but are limited to signals transmitted and received by the same node (e.g. radar [11] or RFID backscatter [28] systems) and therefore do not apply to LTE signals. LTEye provides the best of both these solutions by extending SAR localization techniques to operate over communication signals as opposed to backscatter or radar signals. It also introduces a novel technique to handle errors due to multipath by identifying the shortest (or most direct) path.

Our evaluation of LTEye's spatial analytics in large indoor environments reveals a median accuracy in 3D localization of mobile users of 61 cm in line-of-sight and 85 cm in non-line-of-sight settings. Further, we visualize the LTE performance of the mobile users across locations, helping building managers find optimal locations for relays or femtocells.

Contributions: . This thesis contributes the following:

- The thesis presents LTEye, the first open platform to monitor and analyze per-user LTE PHY performance at fine temporal and spatial granularity.
- LTEye employs a new technique to identify and track individual users at the LTE PHY layer in a robust manner, without help from operators, and without requiring or exposing private user information.
- LTEye develops an innovative technique for accurate localization of users based on their LTE signals. This involves extending synthetic aperture radar (SAR) to operate over communication signals as opposed to backscatter and radar signals, and a novel technique for identifying the shortest and most direct path in the presence of multipath.
- LTEye's evaluation on AT&T and Verizon LTE deployments reveal deep insights on the inefficiencies, utilization patterns, and differences between these providers. LTEye also provides heat maps to characterize LTE performance across indoor locations, without GPS support.

Chapter 2

Related Work

(a) LTE Sniffing Equipment:. Devices such as Wavejudge, ThinkRF and IntelliJudge [24, 27] are wireless protocol sniffers to capture RF signals. They are mainly tools for wireless development and interoperability testing that provide visibility into the interaction between the PHY and protocol layers. Unlike LTEye, these devices need inputs from the cellular provider and do not perform localization or provide spatial analytics.

(b) Open LTE Implementations:. There have been efforts in developing open source implementations of LTE protocols, notably OpenAirInterface [8], and OSLD [12]. These initiatives enable running LTE base stations on software radios; they do not extract spatial or temporal analytics.

(c) LTE Measurement Studies:. Many recent LTE studies have been conducted using traces collected on participating smartphones or from inside LTE networks. Findings from these studies include: (1) The available bandwidth of LTE networks is highly variable and TCP is not able to fully utilize the bandwidth [15]; (2) LTE is significantly less power efficient than Wi-Fi [14]; (3) LTE latency is more consistent (less variable) than Wi-Fi [26]. Such studies focus on the higher layers of the stack, e.g., TCP throughput, transfer delay, and power usage. In contrast, LTEye focuses on the LTE radio layer; it provides fine-grained temporal and spatial information and does not require traces from the provider.

(d) Cellular Location-Specific RF Measurements:. Cellular operators need location-specific RF measurements to troubleshoot performance problems and plan future deployments. They typically obtain coarse location information by mapping a user to her serving cell. Operators then rely on drive tests to refine the spatial measurements. Drive tests are costly and constitute a big part of the network operating expenditure [16]. Further, they are increasingly inadequate as operators move toward femto cells, and need indoor coverage data. To reduce the cost and improve the spatial measurements, recent LTE releases propose mechanisms known as MDT [2]. MDT techniques localize a mobile phone either using in-network time measurement or by collecting location information using the phone's GPS. It is well-known however that in-network localization in cellular networks is not accurate (at hundreds of meters [21]) as time-delay measurements are only available for the serving cell of a mobile user. Even the E911 service using positioning reference signals can only guarantee 150m accuracy 95% of the time [20]. GPS measurements cannot be invoked often as they drain the user's battery. They also cannot capture indoor location.

(e) RF-based Localization Techniques:. Our work is related to past work on RF-based localization. This problem has received much recent interest resulting in highly accurate systems. ArrayTrack [29] and PinPoint [17] are an antenna-array based indoor location systems that tracks wireless clients at fine granularity. Chen et. al.[13] build antenna arrays using software radios synchronized with a reference signal. PinIt [28] is an RFID localization system that combines SAR with a deployment of reference RFIDs to achieve highly accurate localization.

Our design builds on this past work but differs in that it introduces two innovative localization techniques. First, we extend SAR to operate over communication signals exchanged between a transmitter and a receiver. This contrasts with the current approach for SAR, which is limited to backscatter and radar signals, where the transmitter and receiver are a single node with no Carrier Frequency Offset (CFO) or Sampling Frequency offset (SFO). Second, we introduce a new technique that when combined with SAR or standard antenna arrays, estimates the delay difference between the various paths traversed by the signal to identify the shortest path. In particular, we measure these delays in time based on phase offsets in the frequency domain. Hence, LTEye can resolve differences in delay below one time sample, unlike past work that estimates these delays via correlation in time [17].

Chapter 3

LTE Primer

In this chapter, we briefly introduce LTE concepts relevant to this thesis, at a high level. LTE networks are divided into multiple geographical regions called cells. Each cell contains a cellular base station that serves multiple mobile users. We focus on Frequency Division Duplexing, the mode of LTE most widely used by cellular operators. This LTE mode uses different dedicated carrier frequency bands for uplink and downlink transmissions. Hence, each base station uses a pair of frequency bands to communicate with users in its cell.

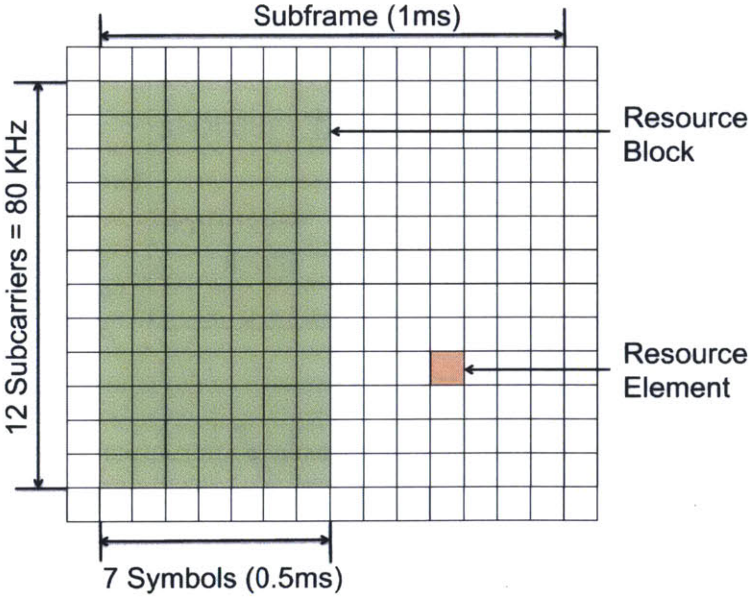


Figure 3-1: **Resource Block.** Resource Grid is divided into Resource Blocks, each 12 OFDM subcarriers and 7 symbols.

(a) Radio Resources. LTE's uplink and downlink transmissions are based on OFDM. While medium access and resource sharing is largely distributed in typical OFDM-based systems such as Wi-Fi, LTE centralizes much of resource allocation at the base station. In particular, base stations divide radio resources into multiple frames over time, each containing ten subframes, spanning 1 ms each. Resources in each sub-frame are divided both along time and frequency as a two dimensional time-frequency grid, as shown in Fig. 3-1. Each cell in the grid, called a resource element corresponds to one OFDM sub-carrier (15 KHz) over the duration of one OFDM symbol (66.7 μs).

The key task of an LTE base station is to apportion both uplink and downlink resources between different users along both time and frequency. It allocates resources to users at the granularity of *resource blocks*, each of which spans 0.5 ms (i.e., half a sub-frame) by 180kHz (i.e., 12 sub-carriers). To combat frequency-selective fading, the assignment of resource blocks to users both on the uplink and downlink is not fixed, but hops from transmission to transmission.

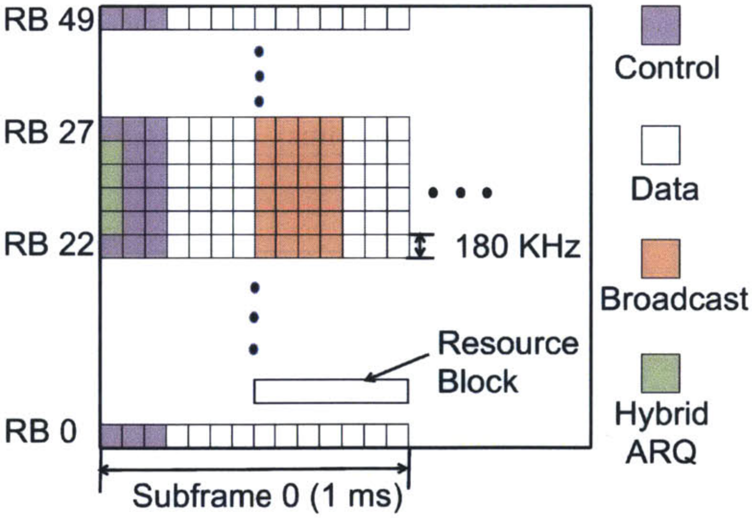


Figure 3-2: **Physical Channels.** Physical Channels are mapped to well defined Resource Elements in the Grid.

(b) Physical Layer Channels. As LTE centralizes PHY-layer control, base stations need to transmit both data and control information to the users. To this end, LTE partitions network resources into well defined channels, each responsible for different types of information. These channels are mapped to well-known resource elements of the grid, as shown in Fig. 3-2.

Broadly, LTE base stations use four main channels on the downlink: (1) A data channel to send users their downlink data. (2) A control channel to allocate network resources. (3) A broadcast channel for new users to learn system parameters. (4) A hybrid-ARQ channel to send ACKs to the users. Similarly, the mobile users on the uplink are allocated data and control channels to transmit their uplink data and control messages. Table 3.1 describes these channels in greater detail.

Downlink LTE Channels	
Name	*Description*
PBCH	Physical Broadcast Channel: Carries general information about the cell, like number of antennas on the base station and total bandwidth.
PDCCH	Physical Downlink Control Channel: Sends downlink control messages, e.g. for resource allocation on uplink/downlink.
PDSCH	Physical Downlink Shared Channel: Holds downlink data meant for users in resource blocks indicated by the PDCCH.
PHICH	Physical Hybrid-ARQ Channel: Contains positive or negative acknowledgments for uplink data.
Uplink LTE Channels	
PUCCH	Physical Uplink Control Channel: Holds control information from users to base stations, e.g. ACKs, channel reports and uplink scheduling requests.
PUSCH	Physical Uplink Shared Channel: Mainly carries uplink data in resource blocks indicated by the PDCCH.

Table 3.1: **LTE Channels**. Details the name and function of PHY-layer channels, as defined in LTE standards [3].

We point out here that the downlink control channel bears rich information on the LTE PHY-layer. Specifically, it contains multiple *downlink control information messages* in which the base station allocates resource blocks to specific users for every transmission on either the uplink or downlink. In addition, the control information also specifies the modulation and coding to be used in these blocks. Upon hearing a control message, a user accesses the relevant data channel to send (receive) her uplink (downlink) transmission.

(c) PHY User Identifier. The LTE PHY refers to each mobile user using a temporary unique ID called the Cell Radio Network Temporary Identifier (C-RNTI). The C-RNTI reveals no private information about the user. It is local to the users's serving cell, and is assigned when she enters the cell via a higher-layer connection establishment procedure[4].

A user continues to have the same C-RNTI as long as she is in the same cell and is not idle for more than the pre-configured tail timer value. The timer value is typically a few seconds to tens of seconds (12 seconds in the measurement result of [14]). Hence, the C-RNTI assigned to a user may change quite often if it transmits sporadically.

Chapter 4

LTEye

LTEye is an open platform to analyze LTE radio performance. Its design aims to satisfy the following key attributes:

- *Provider-Independent:* LTEye does not require any information or support from cellular providers. Thus, LTEye allows end-users, policy makers and third parties to make a fair assessment and comparison of the service quality of different providers, without relying on information furnished by the providers themselves.
- *Off-the-Shelf:* LTEye must be built on low-end off-the-shelf components, such as standard laptops and software radios. This makes LTEye more accessible to end-users and easy to be deployed in large numbers.
- *Secrecy Preserving:* LTEye cannot and does not decode private data or system parameters from encrypted uplink and downlink data channels, but gathers analytics purely from unencrypted control channels.

4.1 System Architecture

LTEye operates as a passive 2-antenna MIMO receiver. LTEye's architecture is a pipeline of two components: (1) the LTE Logger, and (2) the Data Analyzer. (see Fig. 4-1)

(a) LTE Logger: The LTE Logger sniffs on the LTE control channels to gener-

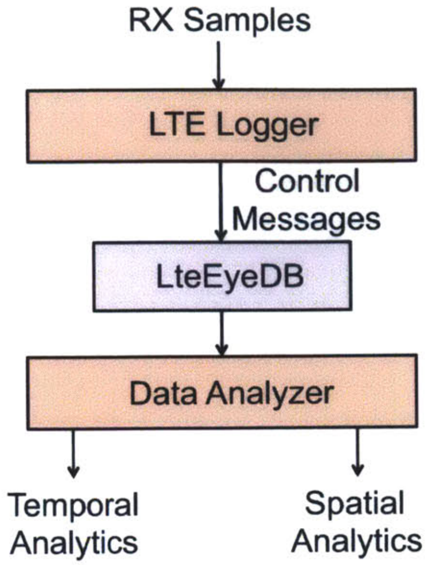


Figure 4-1: **LTEye's Architecture.**LTEye's Architecture is a pipeline of two components: The LTE Logger and Data Analyzer.

ate transmission records. The logger begins by listening to the broadcast channel to gather system parameters. It then sniffs the downlink control channel and performs LTE decoding, i.e., it demodulates the OFDM symbols, applies de-interleaving, descrambling and convolutional decoding to extract the actual bits of the downlink control messages.

The logger reads each of these control messages to populate LTEyeDB, a database of LTE information records tagged with their transmission time. Each transmission record consists of several fields retrieved from the control messages, as listed in Table 4.1. Many of these fields help characterize network performance and utilization, both for the user over time, and for the network as a whole, if viewed in aggregate. In addition, notice two important fields in the records: (1) Each record is indexed by the user's C-RNTI (i.e. her PHY-layer user ID), which are key to link multiple records belonging to the same user. (2) Records maintain the uplink channels seen by the LTEye sniffer over time. In §4.3, we show how these channels are essential to localize users.

Field	Description	Format
Time	Transmission Time	32-bit timestamp
C-RNTI	PHY-layer ID	16-bit Sequence
UL/DL	Uplink or Downlink	0/1
nrb	Number of Resource Blocks	0 to N_{RB}
alloc	Bit-Map of Resource Blocks	N_{RB} bits
MCS	Modulation and Coding Scheme	5 bits
isAcked	Acknowledgment	±1 (ACK/NACK)
UE-channel	Channel from user (if U/L)	$N_{sc}N_{rx}$ Complex Floats
BS-channel	Channel from Base Station	$N_{sc}N_{tx}N_{rx}$ Complex Floats
SNR, SINR	SNR and SINR of Base Station	Floating Point

Table 4.1: **Fields of LTEyeDB**. Where N_{RB}: Number of resource blocks, N_{tx}: Number of base-station antennas, N_{rx}: Number of sniffer antennas, N_{sc}: Number of OFDM subcarriers.

(b) Data Analyzer: The data analyzer processes the records in LTEyeDB to extract fine-grained analytics on both cellular base-stations and individual mobile users. It extracts two types of analytics: temporal analytics which describe LTE PHY metrics as functions of time (e.g., the per-user resource allocation over time), and spatial analytics which describe position-dependent RF metrics (e.g., the user location and the observed multipath effects). Sections §4.2 and §4.3 discuss both types of analytics in detail.

4.2 Enabling Temporal LTE Analytics

In this section, we describe some of the challenges in obtaining temporal analytics, without access to private user information or system parameters in encrypted data channels.

Uniquely Identifying Users: . To extract per-user temporal analytics, LTEye needs to uniquely identify each user served by the base station. One option is to require LTEye to sniff the LTE channel continuously for an extended duration of time to catch each user at the time she joins the cell and capture her C-RNTI (i.e. her PHY-layer user ID) assignment. Unfortunately this option has two limitations: First, the C-RNTI assignment is often transmitted from higher layers in the encrypted downlink data channel [4]. Second, low-end off-the-shelf equipment is unlikely to be able to continuously monitor and decode the LTE channel in real-time. Instead, LTEye sniffers should be able to periodically sniff the channel and obtain representative snapshots of the system. Hence, we need LTEye to uniquely identify all users, including those who joined the cell even when LTEye is not sniffing (i.e., LTEye did not hear their C-RNTI assignment).

To address this problem, we observe that a user's C-RNTI is used to scramble her control information on the downlink control channel. Specifically, recall from §3 that the control channel transmitted by the base station consists of multiple downlink control information messages, for different users. At the end of each message is a 16-bit sequence, which is the XOR of the checksum of the control message with the user's C-RNTI. Traditionally, a user de-scrambles this sequence by her C-RNTI to retrieve the checksum and validate correctness of the control information. In contrast, LTEye performs the *opposite* operation to retrieve the C-RNTI: It decodes each control message in the log including their corresponding scrambled checksums. For each of these packets, LTEye reconstructs the expected checksum and XORs them with the scrambled checksum to recover the C-RNTI. Of course, it is important to verify if the control messages and C-RNTIs that are decoded are actually correct. To do this, LTEye convolutionally re-encodes the retrieved control information message and compares it against the original coded control message to obtain the number of bit errors. It then discards control messages and C-RNTIs that report bit errors beyond a few bits.[1] Hence, our solution enables LTEye to map a C-RNTI to a user even if it was assigned when LTEye is not sniffing the channel.

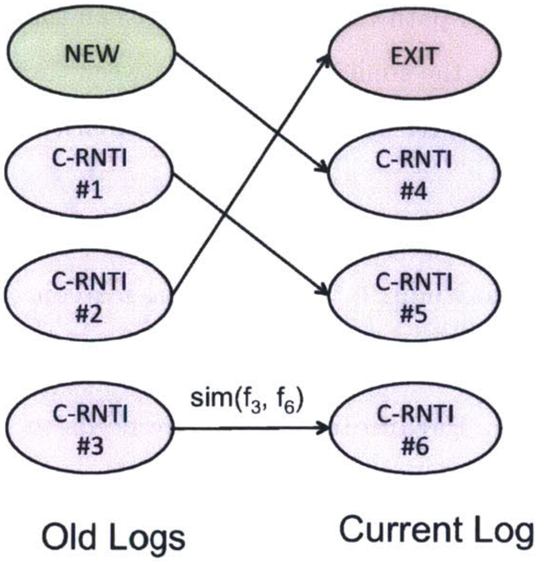


Figure 4-2: **Mapping C-RNTI.** Mapping C-RNTI from old logs to the current log. Three possibilities exist: (1) An old C-RNTI maps to a new C-RNTI; (2) A user with an old C-RNTI has left the cell; (3) A current C-RNTI is a new user.

Tracking User IDs: . LTEye's second challenge is to map C-RNTIs between logs. Recall from §3 that a user's C-RNTI may be re-assigned if she remains idle beyond a few seconds. Further, the same C-RNTI can be assigned to several users over time. Hence, while the C-RNTI is a natural PHY-layer user ID, LTEye must recognize when a user's C-RNTI changes and track the list of C-RNTIs assigned to it.

LTEye addresses this challenge by formulating it as a matching problem. The goal of this problem is to map C-RNTIs in the current (most recent) log produced by the LTE logger with the C-RNTIs in prior logs as shown in the bi-partite graph in Fig. 4-2. In addition, the graph has two additional nodes: EXIT and NEW, which account for C-RNTIs in the old log that have "exit" the system, and C-RNTIs in

[1] In our experiments, LTEye correctly retrieved 99.5% of C-RNTIs across locations.

the current log that are "new" to the cell, respectively. The weights in this graph must capture the similarity between the users associated with each pair of C-RNTIs. Specifically, we associate with each C-RNTI $\#i$ an RF fingerprint f_i that includes metrics such as the user's location (extracted by our localization method described in §4.3), its SNR and multi-path characteristics. We can then assign a weight to each edge (i, j) in the graph by the similarity metric between these fingerprints $sim(f_i, f_j)$. In §4.6, we design effective RF fingerprints and similarity metrics based on spatial analytics.

Given the graph and weights, we can now solve this matching problem using the standard Hungarian Algorithm[19].[2] The resulting matching either maps a C-RNTI to a user in a prior log, or identifies her as a new user.

Extracting System Parameters: . LTEye needs to *reverse-engineer* several PHY parameters, otherwise available to users via encrypted data channels. For e.g., the LTE standard allows several possible formats for downlink control information, each spanning multiple lengths [5]. Exhaustively searching for each of the possible format-length pairs in all downlink control messages is prohibitively expensive. Fortunately, operational LTE base stations employ only a small subset of these formats for all users (only three possible formats for AT&T and Verizon). As a result, LTEye learns the possible list of formats and lengths using the first few control messages to greatly reduce the search-space of formats for subsequent control information.

[2] We modify the algorithm to allow multiple C-RNTIs to map to NEW/EXIT simply by replicating these nodes. Edges at NEW/EXIT are weighted by a minimum threshold similarity.

4.3 Enabling Spatial LTE Analytics

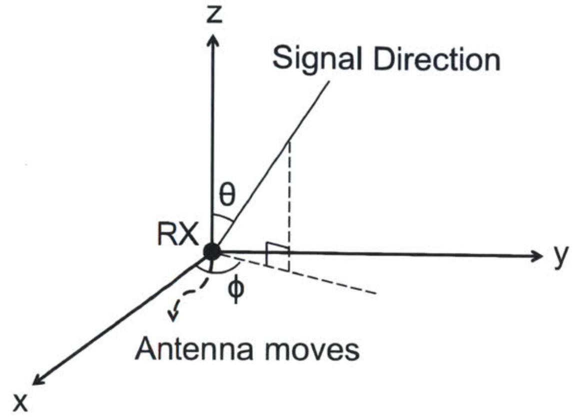


Figure 4-3: **Signal Direction Notation.** Definition of (ϕ, θ) in 3-D space.

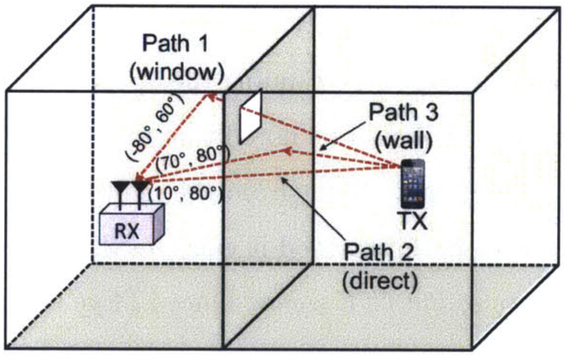


Figure 4-4: **Multipath (Layout).** Depicts an example layout of a transmitter and receiver in two rooms separated by a wall. The signal has three main paths: Path 1 is the strongest, and reflects from the ceiling through a window. Path 2 is the direct path penetrating a wall. Path 3 is the weakest reflecting at the farthest wall. The figure labels (ϕ, θ) for each path.

At the heart of LTEye's spatial analytics is the ability to localize the source of an LTE signal. To this end, LTEye performs the following functions:

- *Extracts the multipath profile of an LTE signal:* To begin with, LTEye extracts the multipath profile of a signal, which measures the power received along each spatial angle, i.e., along various signal paths. Fig. 4-5 shows an example multipath profile, where the signal traverses three paths corresponding to three local maxima in the graph.

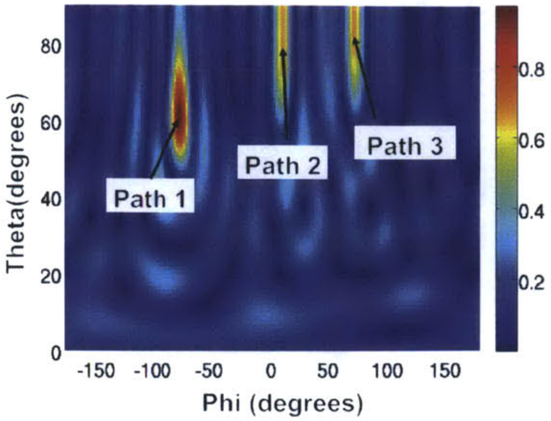


Figure 4-5: **Multipath Profilen.** Simulated multipath profile for (b) has peaks for each path at expected (ϕ, θ). The height of the peaks (in shades of red) corresponds to the relative power of the corresponding paths.

- *Identifies the direct path from the source:* Once LTEye has the multipath profile, it is natural to try to identify which path is in direct line of sight to the source (e.g., path 2 in Fig. 4-4). Finding the direct path is an important step in localizing the source. It should be noted that in some cases the direct path may be completely blocked and absent from the multipath profile. Our objective is to identify the direct path provided it exists in the multipath profile.
- *Localizes the source of the LTE signal:* Once LTEye finds the direct path to the source, it can localize the source to within a specific spatial direction. We can then deploy multiple LTEye sniffers to locate the source at the intersection of the direct paths as seen from these sniffers.

Past work on RF-based localization takes two approaches to build multipath profiles: The first approach, shown for Wi-Fi, uses an antenna array to steer its beam spatially and identify the power along each spatial direction [29]. However, LTE runs at much lower frequencies than Wi-Fi (700 MHz as opposed to 2.4 GHz), and hence an LTE antenna array will be 4 to 5 times more bulky than a comparable Wi-Fi array. The second approach uses synthetic aperture radar (SAR) [28], which uses a single movable antenna to emulate a virtual array of many antennas. As the antenna moves, it traces the locations of antenna elements in a virtual array.

SAR has traditionally been used in radar and RFID localization as it assumes backscatter signals where the transmitter and receiver are the same node and therefore have no carrier frequency offset (CFO) relative to each other. Hence, changes in the channel as the antenna moves are a function only of the antenna's location. In contrast, LTE signals are exchanged between an independent transmitter and receiver, with non-zero CFO. As the antenna moves, the channel changes both due to CFO and antenna movement. One option is to estimate and correct the CFO. Unfortunately this solution is fragile since any residual error in CFO estimation accumulates over the duration of movement and causes large localization errors. In the following section, we explain how we perform SAR over LTE signals without CFO estimation to realize the three functions in the beginning of this chapter.

4.4 SAR over LTE Signals Using Channel Ratios

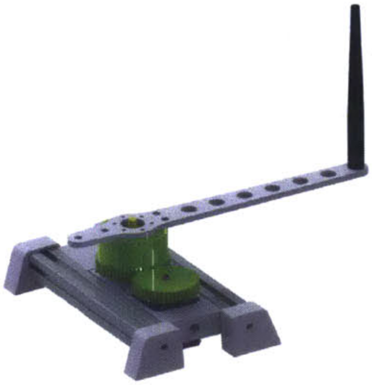


Figure 4-6: **LTEye Sniffer Prototype.** Prototype of the rotating antenna on a LTEye sniffer.

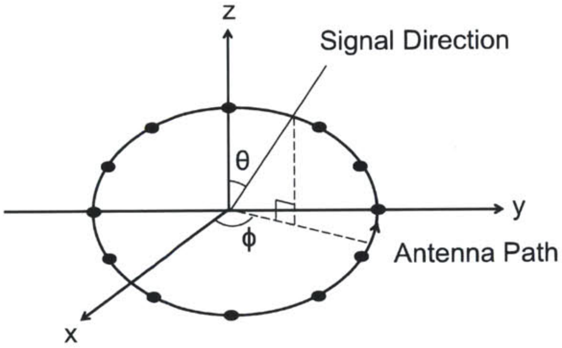


Figure 4-7: **Circular SAR.** Circular Synthetic Aperture Radar.

LTEye operates as a 2-antenna MIMO receiver, with one static antenna and another movable antenna. The movable antenna may be mounted on a rotating arm attached to the device, which is the approach taken in our implementation (see Fig. 4-6). The advantage of this approach is that it provides the 3-dimensional spatial direction (i.e. both the azimuthal angle ϕ and polar angle θ) of the various signal paths as shown in Fig. 4-3. Alternatively, the antenna may slide back and forth on an arm fixed to the body of the device.

LTEye uses its MIMO capability to perform SAR over communication signals, but without frequency offset estimation. Our key idea is that instead of applying the

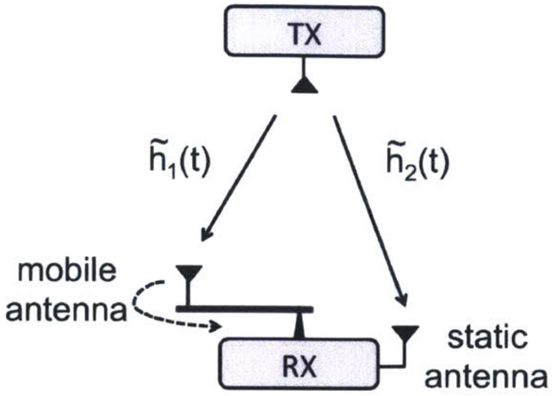


Figure 4-8: **Channel Notation.** Depicts $\tilde{h}_1(t)$ and $\tilde{h}_2(t)$, the measured wireless channels from a transmit antenna to the mobile and static antennas respectively.

SAR equations to the wireless channel of the moving antenna [11], we apply SAR equations to the *ratio* between the channel of the moving antenna to that of the static antenna. Taking the ratio of the two channels eliminates any effect of frequency offset since both MIMO antennas experience the same offset relative to the sender. However, since the ratio is between two antennas, one moving and the other static, it preserves how antenna displacement changes the channel of the moving antenna. This allows SAR to safely retrieve the multipath profile of the signal from the ratio, modulo frequency offsets.

Next, we mathematically show the validity of the above technique. Suppose that the receiver, placed at the origin, wants to measure the power of the signal $P(\theta, \phi)$ received from an independent transmitter along a spatial direction specified by the polar angle θ and azimuthal angle ϕ (see Fig. 4-3). According to the SAR formulation, this quantity can be measured as: [10, 22]:

$$P(\theta, \phi) = |h(\theta, \phi)|^2, \text{ where } h(\theta, \phi) = \sum_t a_f(t, \theta, \phi) h(t) \tag{4.1}$$

Here, $h(t)$ is the wireless channel to the moving antenna at time t, assuming zero frequency offset between the transmitter and receiver. The quantity $a_f(t, \theta, \phi)$ captures the relative motion of the transmitter and receiver, and is independent of the wireless channels. For e.g., if the antenna moves along a straight line (i.e., linear SAR)

$a_f(t,\theta,\phi)$ is defined as: $a_f(t,\theta,\phi) = e^{-j\frac{2\pi f}{c}x(t)cos(\phi)}$, where $x(t)$ is the antenna location at time t, and f is the frequency of the signal [28]. Similarly, if the antenna rotates with radius r (i.e., circular SAR) $a_f(t,\theta,\phi) = e^{-j\frac{2\pi f}{c}rcos(\phi-\phi_0(t))}$, where $\phi_0(t)$ is the angular position of the antenna at time t (see Fig. 4-7).

Past work on SAR require both transmitters and receivers to share a common reference clock. For e.g., SAR devices on airplanes or satellites both transmit signals and receive their reflections to image the topography of the ground[11]. Consequently, the measured channel $\tilde{h}(t)$ at the moving antenna is independent of frequency offset, i.e., $\tilde{h}(t) = h(t)$.

Unfortunately, when performing SAR between *independent* transmitters and receivers, the measured wireless channel $\tilde{h}(t)$ varies both due to position and due to carrier and sampling frequency offset between the transmitter and receiver, as well as any phase noise. In particular, we denote:

$$\tilde{h}(t) = h(t)e^{j\psi(t)} \tag{4.2}$$

Where $\psi(t)$ denotes the phase accumulated due to any carrier frequency offset, sampling frequency offset or phase noise between the reference clocks of the transmitter and receiver until time t. Thus, the key challenge to measure the power of the signal along any spatial direction, $P(\theta,\phi)$, as in the SAR Eqn. 4.1, is to eliminate this accumulated phase.

To resolve this challenge, LTEye is built on receivers that have at least two antennas: a static antenna, and a mobile antenna that moves along a known path. Let $\tilde{h}_1(t)$ and $\tilde{h}_2(t)$ denote the measured wireless channels from a given transmit antenna to a mobile and static antennas respectively (see Fig. 4-8). As both antennas are connected to the same reference clock, they both accumulate the same phase $\psi(t)$ until time t. Hence, from Eqn. 4.2, we have:

$$\tilde{h}_1(t) = h_1(t)e^{j\psi(t)}, \qquad \tilde{h}_2(t) = h_2(t)e^{j\psi(t)} \tag{4.3}$$

where $h_2(t) \approx h_2$ is relatively constant over a short duration since antenna-2 is static.

Hence, the ratio of the wireless channels is: $\tilde{h}_r(t) = \frac{\tilde{h}_1(t)}{\tilde{h}_2(t)} = \frac{1}{h_2}h_1(t)$; that is, the channel ratio is a constant multiple of the moving antenna channel without the phase accumulation from frequency offset or phase noise. Thus, we can perform SAR as in Eqn. 4.1 by substituting the channel ratio $\tilde{h}_r(t)$ for the value of $h(t)$. Hence, LTEye allows a wireless receiver to perform SAR over LTE signals without frequency offset or phase noise estimation.

Finally, we make a few important observations:

- The above approach can be readily extended to OFDM / OFDMA signals. Specifically. we apply Eqn. 4.1 with: $h(\theta, \phi) = \sum_f \sum_t a_f(t, \theta, \phi) h_{r,f}(t)/h_{r,f}(0)$, where the quantity $\tilde{h}_{r,f}(t)$ on subcarrier f of the OFDM signal is the ratio of the frequency (Fourier) Domain channels $\tilde{h}_1(t)$ and $\tilde{h}_2(t)$ measured on that subcarrier.[3]
- Our solution is resilient to movement of the transmitter that can be neglected relative to the movement of the rotating antenna. However, one can easily detect and exclude fast moving transmitters by checking if the channel of the static antenna $\tilde{h}_2(t)$ is coherent over a rotation of the moving antenna, using the coherence metric in §5.2.4.
- Our technique applies beyond LTE, and enables applying SAR to Wi-Fi and other communication technologies.

[3] Note that this is robust to frequency hopping by LTE users.

4.5 Identifying the Direct Signal Path

In this section, we describe how LTEye separates the direct path from the reflected paths in a multipath profile reported by SAR so as to localize the transmitter. Intuitively, the direct path is the shortest path among all paths traversed by the signal (even if the direct path is completely blocked, the shortest path is the path closest to the direct path). Thus, one may identify the direct path (or the path closest to the direct path) by measuring the delay difference between the various paths in the multipath profile of the signal.

Directly measuring time delays (e.g. by correlating with known pilot signals), however is not sufficiently accurate. Specifically, LTE receivers have a channel bandwidth of 10 MHz. However, electromagnetic waves travel at the speed of light. Hence an error of even one time sample for a 10 MHz sampling rate (i.e., each time sample spans 0.1 microseconds) translates to an error in path lengths of 30 m.

Below we explain how LTEye can measure *sub-sample delay differences* between the signal paths. The key idea is to exploit that delay in time translates into phase rotation in the frequency domain. Since LTE signals use OFDM, a time delay of the signal translates into phase rotation in the OFDM subcarriers. Yet, different OFDM subcarrier rotate at different speeds – i.e., higher OFDM frequencies rotate faster than lower frequencies. In fact, the phase rotation of a particular OFDM subcarrier f_i as a result of a delay τ is $\psi_i = 2\pi f_i \tau$. Thus, for each subcarrier, the difference in delay between two paths, p and q, for a particular subcarrier is:

$$\tau = \frac{\psi_{p,i} - \psi_{q,i}}{2\pi f_i} \tag{4.4}$$

One may also average across subcarriers to improve robustness to noise. Multiple sub-carriers can also resolve ambiguity if $|\psi_{p,i} - \psi_{q,i}| > 2\pi$ by correcting discontinuities in $\psi_{p,i} - \psi_{q,i}$ across frequencies. Thus, to identify the shortest path in a multipath profile, LTEye does the following:

- First, it computes the phase of the channel for each subcarrier for each path separately. This can be done by leveraging the fact that the SAR formulation defines $h(\theta, \phi)$, which provides not just the power, but also the phase of the channel component along different spatial directions (see Eqn. 4.1). Hence, we can simply measure the phase of this path as $\psi = arg[h(\theta, \phi)]$, for each OFDM subcarrier.

- Second, it computes the delay difference between each pair of paths using the Eqn. 4.4 above. It then identifies the shortest path as the one with least delay.

Once the direct path is found, the source can be localized along this path. Complete localization can be performed by using multiple LTEye receivers and intersecting their direct paths. If the direct paths do not intersect, the best estimate is the point that minimizes total distance (or equivalently, delay) to all LTEye receivers. The point can be found using simple geometric optimization omitted for brevity.

Next, we show how the multipath profile reported by SAR allows us to compute unique RF fingerprints for the users.

4.6 Measuring RF Fingerprints with Multipath Profiles

As described in §4.2, LTEye tracks the different C-RNTIs (PHY-layer user IDs) assigned to mobile users between logs. To this end, LTEye employs RF fingerprints to map C-RNTIs between logs. LTEye defines a user's fingerprint as the set of observed multipath profiles at each LTEye sniffer (See §4.4). The key advantage of this fingerprint is that it captures the user's location, multipath, and SNR, as perceived from LTEye sniffers. To measure similarity of two fingerprints, we employ dynamic time warping (DTW[23]), a technique that has recently been applied to capture similarity of two multipath profiles[28]. Given any two multipath profiles, DTW returns a cost function that varies inversely with their similarity. Hence, LTEye defines the similarity metric between two RF fingerprints as the inverse of the total DTW cost function[28] between each pair of profiles in the fingerprints.

One might wonder if LTEye's fingerprint matching algorithm scales, given that a cell may serve a large number of users. Fortunately, while LTE cells can serve hundreds of users, we observed that only a small fraction of these users (about 4%) are reassigned C-RNTIs between two logs.[4] Further, while capturing C-RNTI reassignments is essential to track individual users over time, it does not alter aggregate radio analytics in a statistically significant way, given that 96% of users in a cell retain their C-RNTI between logs.

[4] This is on average 1-2 users for AT&T and 3-4 users for Verizon.

Chapter 5

Results

5.1 Implementation

We implemented LTEye on USRP N210 software radios[7] and WBX daughter-boards. The USRPs receive in the 700MHz frequency range at a bandwidth of 10 MHz on uplink and downlink channels corresponding to either AT&T (734-744MHz) or Verizon (746-756MHz). We implement an OFDMA receiver for LTE signals that interfaces directly with the USRP Hardware driver (UHD).

To obtain temporal analytics, we decode the downlink control channel in a pipeline of two modules: The first module in C++ performs synchronization, channel estimation and QPSK demodulation. It logs the demodulated soft bits received over one second into a file, and repeats this process every three seconds. The second module in MEX (C++) and Matlab reads the file and performs descrambling, de-interleaving and convolutional decoding. It then processes the downlink control information messages to get per-user LTEyeDB records for uplink and downlink traffic, as in §4.1.

To obtain spatial analytics, we build prototype LTEye sniffers, each containing two USRPs connected to an external clock. We mount the antenna of one of the USRPs on a light-weight rotating arm fabricated by a 3D printer, with an adjustable radius of 15-30 inches, as shown in Fig. 4-6. The arm is driven by an off-the-shelf stepper motor rotate at 30-120 rotations per minute. We use an Arduino UNO board to rotate the stepper motor accurately at a constant speed and provide real-time feedback on the position of the rotating antenna. We implement a C++ module to use these positions and channel measurements, as in §4.3, to localize the users.

We evaluate LTEye's spatial analytics using five LTEye sniffers in multiple indoor testbeds, in both line-of-sight and non-line-of-sight settings. We employ ten Samsung Galaxy Note LTE smart phones as users. Unless specified otherwise, each user communicates over LTE with a mix of varying traffic patterns including browsing activity, file transfer, Skype calls, and video streaming.

5.2 Results on Temporal Analytics

In this section, we perform an extensive evaluation of LTEye's temporal and spatial analytics:

- We compare temporal analytics of two LTE providers and highlight their PHY-layer inefficiencies in §5.2.1 and §5.2.2.
- We provide insights on the LTE rate adaptation algorithm in §5.2.4.
- We apply LTEye's spatial analytics to two applications: detecting cheaters in an exam hall in §5.3.1 and visualizing a spatial heatmap of LTE performance in §5.3.2.
- We perform micro-benchmarks to evaluate the accuracy of LTEye's localization and RF fingerprints in §5.3.3 and §5.3.4.

5.2.1 Comparing Temporal Analytics of Providers

End-users can deploy LTEye sniffers to compare different providers in their locality in terms of their usage patterns, quality of service and congestion. In this experiment, we compare aggregate temporal analytics of two providers serving our campus: AT&T and Verizon.

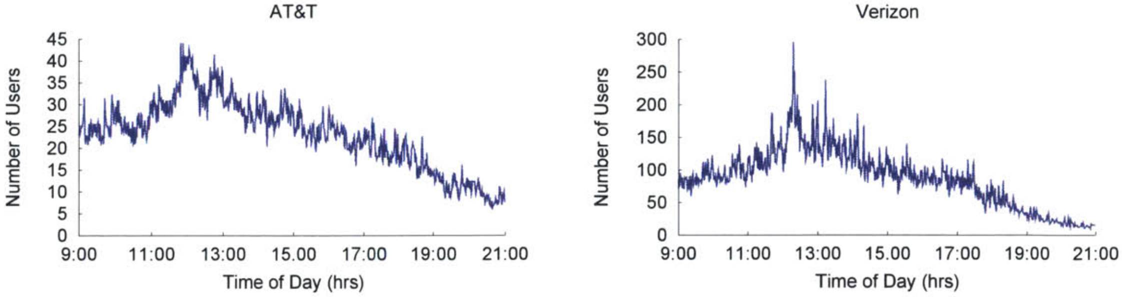


Figure 5-1: **Number of Users.** The plots show the Number of Active Users per second for AT&T (left) and Verizon (right), measured every minute over a typical working day from a representative base station.

Setup. We place LTEye sniffer in four locations in a large campus, each listening to the AT&T and Verizon base stations that serve that location. We populate

LTEyeDB over the duration of a representative weekday from 9:00am to 9:00pm. To reduce processing overhead, LTEye's logger collects traces for a duration of one second, every three seconds. It validates these traces by only accepting control information with low bit error rate as reported by the convolutional decoder. We then measure the following metrics for each one-second trace: (1) Number of Active Users (Fig. 5-1); (2) Mean Utilization of the Uplink and Downlink (Fig. 5-2); (3) Mean Link Quality measured as the number of bits per resource element (bits/RE) in the uplink and downlink (Fig. 5-3). We average each of these quantities over one minute intervals and plot them over time of day at a representative location. We also estimate the mean value of these metrics across locations over one week to infer aggregate trends (Table 5.1).

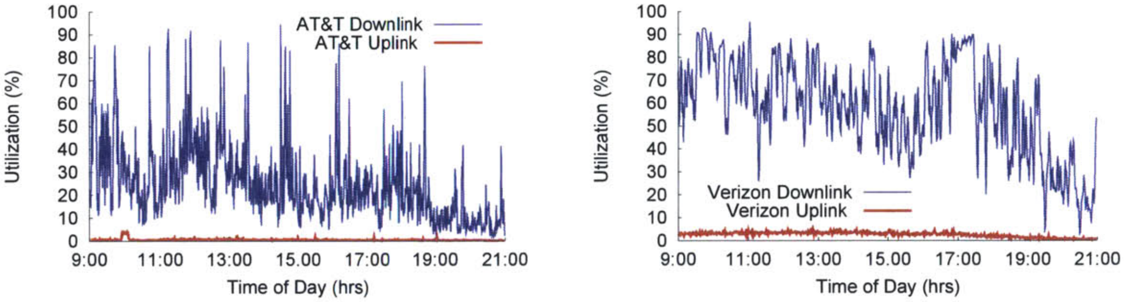


Figure 5-2: **Number of Users.** The plots show the Percentage of Utilized Resource Elements on the Uplink and Downlink for AT&T (left) and Verizon (right), measured every minute over a typical working day from a representative base station.

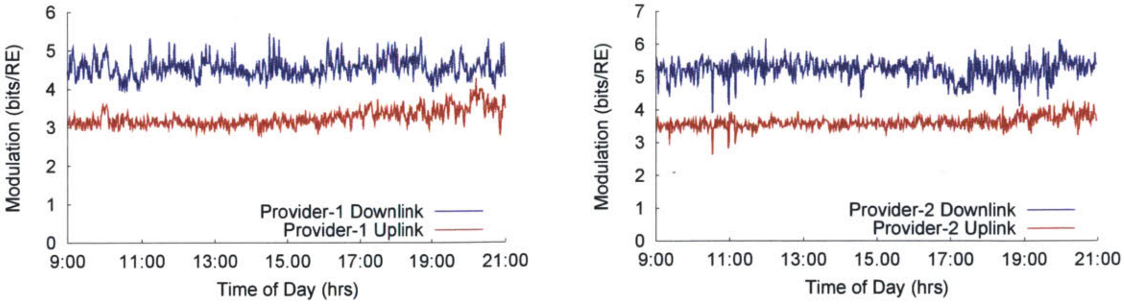


Figure 5-3: **Number of Users.** The plots show the Mean Number of Bits per Resource Element on the Uplink and Downlink for AT&T (left) and Verizon (right), measured every minute over a typical working day from a representative base station.

Metric	AT&T		Verizon	
	Mean	Std-dev	Mean	Std-dev
Number of Users	23.37	7.90	87.66	44.75
% Downlink Utilization	25.20	17.35	58.18	20.58
% Uplink Utilization	0.59	0.47	2.60	1.06
Downlink MCS (bits/RE)	4.56	0.26	5.23	0.30
Uplink MCS (bits/RE)	3.25	0.22	3.61	0.18

Table 5.1: **Aggregate Statistics**. Tabulates mean and standard deviation of statistics over four locations for AT&T and Verizon.

Number of Active Users. Fig. 5-1 measures the number of active mobile users in a representative AT&T and Verizon cell over different times of the day. We observe that for both providers, the number of users in the morning increases steadily, and peaks at around 12:00 pm, after which the number begins to decrease. The increase in activity at around noon may be attributed to a greater number of subscribers who access LTE outdoors as they leave for lunch. Across locations, we observe that Verizon has a greater number of active users on average at 87.7, while AT&T has 23.4 active users through the day (see Table 5.1).

Network Utilization. Fig. 5-2 plots the utilization of a representative AT&T and Verizon cell, over different times of the day. Specifically, We measure the percentage of resource elements used by uplink and downlink traffic. Two trends emerge: First, both providers often achieve high downlink utilization (over 80%) through the day. AT&T achieves such high utilization sporadically through the day (for 2% of the day), while Verizon is heavily utilized for a more significant fraction of the day (for 18% of the day). Second, the utilization of the uplink is significantly lower than the utilization of the downlink, both for AT&T and Verizon.

Specifically, the mean downlink utilization (25.2% - AT&T, 58.2% - Verizon), far exceeds uplink utilization (0.6% - AT&T, 2.6% - Verizon) even when averaged across locations. While it is expected that the downlink is higher in demand, our results reveal that the LTE uplink is an unprecedented 20 to 40 times less utilized than the downlink. This exposes the practical limits of Frequency Division Duplexing mode of the LTE standard used by both AT&T and Verizon, which provides independent equally sized uplink and downlink frequency bands. Hence, our results can help policy makers encourage operators to adopt revised LTE Advanced standards that permit unequal allocation of resources to the uplink and downlink (e.g., via asymmetric carrier aggregation[18]) without relying on data from providers themselves to make the case.

Link Quality. Fig. 5-3 measures the average quality of channels in the network for a representative AT&T And Verizon cell measured over different times of the day. In particular, we measure the mean number of bits transmitted per resource element (bits/RE), capturing the modulation and coding scheme (MCS) on the uplink and downlink. Our results show that the mean quality on the downlink (5.2 - Verizon, 4.6 - AT&T) exceeds that of the uplink (3.6 - Verizon, 3.3 - AT&T). As mentioned earlier, this is because users are limited in transmit power and number of MIMO antennas, when compared to the base station. Further, the mean link quality of Verizon is marginally higher when compared to AT&T.

5.2.2 Identifying PHY-Layer Problems and Inefficiency

Cellular Providers and independent researchers can use LTEye to diagnose problems and inefficiencies at the LTE-PHY layer. In this experiment, we identify such deficiencies by analyzing the LTEyeDB database populated for both AT&T and Verizon. In particular, we consider the traces gathered over four locations in our campus served by different base stations over one week, as explained in §5.2.1. Interestingly, many of these PHY-layer inefficiencies may be unknown even to the operators as they are part of the PHY-layer implementations adopted by the base station vendors.

Unnecessary Control Overhead. . As explained in §3, the LTE resource grid on the downlink is divided into three main PHY-layer channels: the broadcast channel, the control channel and the data channel. The control channel occupies the first 1-3 symbols of each LTE sub-frame resulting in a control overhead ranging from 7% to

21% of all downlink resources. Ideally, an LTE base station should adapt the number of control symbols used in each sub-frame depending on the amount of control traffic that is required to be sent.

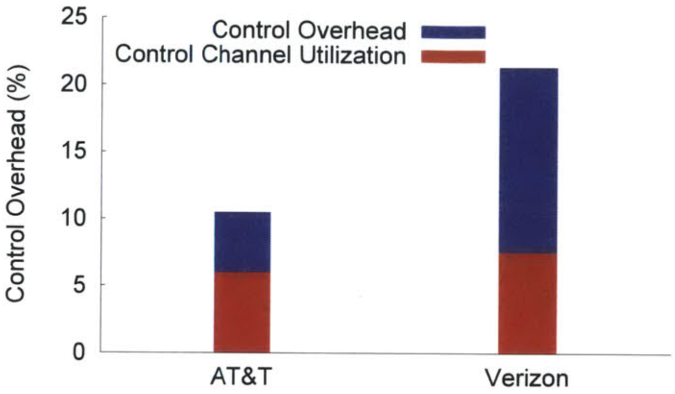


Figure 5-4: **Unnecessary Control Overhead.** Utilization of the control channel for AT&T and Verizon, across different number of symbols allotted to the control channel.

In practice, we discovered that the control overhead of AT&T base-stations was 10%, while that of Verizon base-stations was 21%. This is because, unlike AT&T, Verizon always uses three control symbols in each sub-frame, regardless of the amount of control traffic they contain. One might wonder if this is because Verizon's control channels are significantly more utilized, warranting the additional overhead. Fig. 5-4 plots the control overhead of both AT&T and Verizon, as well as how much of this overhead is utilized. Clearly, the overhead of Verizon is significantly larger, despite having only a marginally higher control traffic (7.5%) than AT&T (6.1%). As a result, we estimate that Verizon can gain as much as 10% of additional downlink resources for data, just by adapting its control overhead to control-traffic demand.

Inefficient Resource Allocation. . In this experiment, we analyze the downlink resource allocation mechanism of LTE base-stations during periods of high network utilization (over 80%). Fig. 5-5 plots the percentage of downlink data transmitted in each resource block across users, measured for both AT&T and Verizon. For AT&T, the graph remains flat across resource blocks, indicating that on average, a user is equally likely to get any of the downlink resource blocks. For Verizon however, we

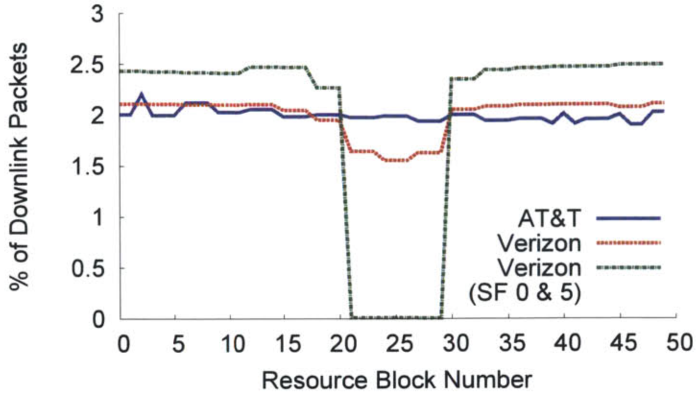


Figure 5-5: **Inefficient Resource Allocation.** Percentage of data allotted to different downlink resource blocks for AT&T and Verizon.

observe a peculiar dip around resource blocks 22-27. To investigate this, we noticed that Verizon avoids these resource blocks completely on subframes 0 and 5. This is because a few symbols in these resource blocks are dedicated for the broadcast channel. As a result, Verizon avoids allocating these resource blocks completely, leading the remaining symbols in these resource blocks to lie completely unused, even during peak hours of demand.

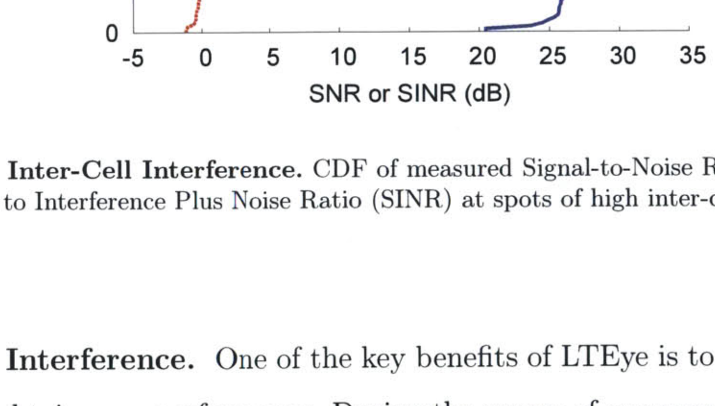


Figure 5-6: **Inter-Cell Interference.** CDF of measured Signal-to-Noise Ratio (SNR) and Signal to Interference Plus Noise Ratio (SINR) at spots of high inter-cell interference.

Inter-Cell Interference. One of the key benefits of LTEye is to provide insight into why users obtain poor performance. During the course of our experiments, LTEye localized users at certain spots that achieved poor link quality on the downlink (the lowest QPSK rate), but high quality on the uplink. To investigate this, we moved our LTEye sniffers and testbed mobile phones to these locations. Surprisingly, our phones at these spots reported very high RSSI[1] from the base station, yet often switched to 3G. We then used our LTEye sniffers at these spots to measure the signal-to-noise ratio (SNR) as well as the signal-to-interference plus noise ratio (SINR) on the downlink. Fig. 5-6 reports the CDF of these quantities across these locations. The figure demonstrates that these locations suffer from significant interference, with a mean SINR of 1.3 dB and a minimum of -1.2 dB, despite a high mean SNR of 29 dB. We realized the source of about 27 dB of interference is from neighboring Verizon base stations sharing same downlink spectrum. Specifically, our sniffer could sense as many as five distinct base station cell IDs at a single location. This is problematic as it affects pilot reference signals (known as cell-specific reference signals) that are critical for channel estimation. Base stations transmit these pilots in one of three different subsets of resource elements depending on their cell ID.[2] Given that five base stations are observed at a given location, some of these pilots will inevitably collide, significantly impacting the decodability of signals from those base-stations. Hence, these observations emphasize the need for effective placement and power control of base stations and small cells. They further highlight the importance of careful assignment of cell IDs to neighboring cells, to avoid interference between their pilots.

5.2.3 Insights into LTE Scheduling

In this experiment, we gather insights on the scheduling policy of an AT&T base station in our campus. We deploy a LTEye sniffer that listens to the downlink control channel of the base station. We then perform the following tests during periods of low network utilization, where the UEs in our testbed are the sole active users of the network.

[1] The phone reported a receive signal strength (RSSI) of around -85 to -95 dBm, where the noise floor is at -120 dBm.

[2] For 2-antenna base stations, reference signals on both antennas occupy one of three subsets of REs, based on cell ID modulo 3 [3].

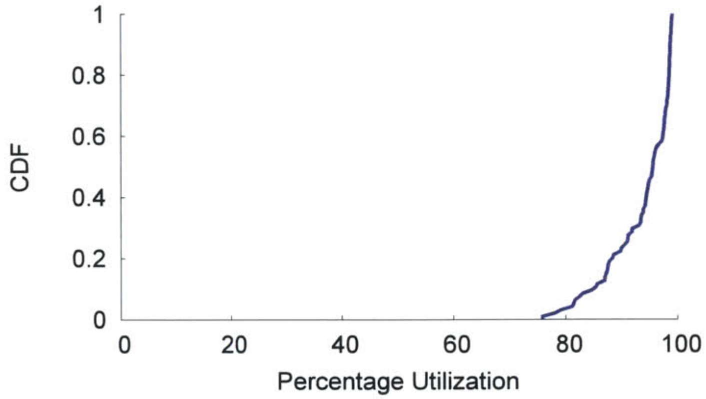


Figure 5-7: **Stress Test.** CDF of percentage downlink utilization in a user stress test.

Stress Test. We begin by performing a stress test on a single UE in the network. The UE downloads a large bitfile containing random strings via six simultaneous TCP connections to saturate the LTE downlink. We repeat the process multiple times across locations. Fig. 5-7 plots a CDF of the percentage of available downlink resources granted to the UE. The results show that UE is indeed allocated most (93.7% on average) of the downlink resources in the network.

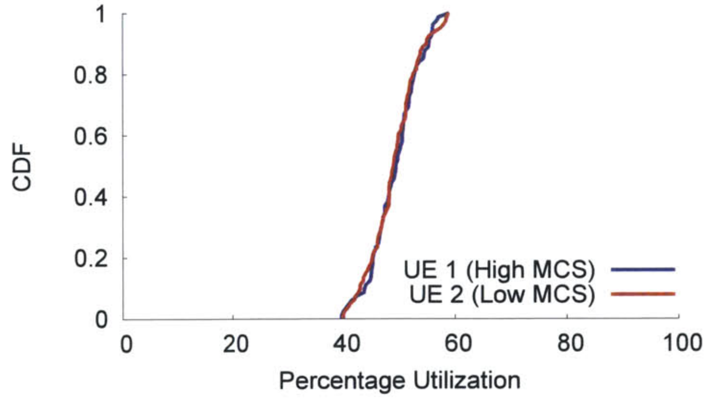


Figure 5-8: **Fairness Test.** CDF of percentage downlink utilization of two users sharing the network.

Fairness Test. Next, we evaluate the fairness of two UEs: UE-1 and UE-2 that perform the above stress test. We place UE-1 at a static location, while UE-2 is placed in a wide variety of locations with different channel quality. Fig. 5-8 plots a CDF of the percentage of available downlink resources granted to each UE. The results show that the UEs consistently obtain nearly equal shares of available resources (about 49% each), across experiments. This means that the static UE-1 achieves the same throughput across runs, regardless of UE-2's location or performance. In contrast, since UE-2 is placed in a wide range of locations, its throughput is dictated by its link quality. In other words, unlike Wi-Fi which is packet fair, LTE users with greater link quality achieve greater throughput compared to those with poor link quality.

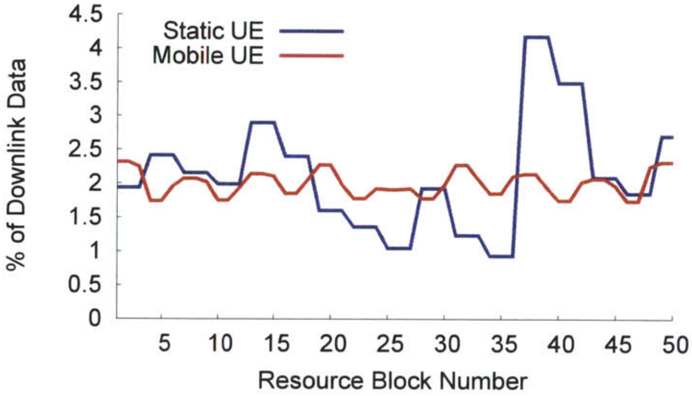


Figure 5-9: **Mobility Test.** Percentage of data transmitted in each resource block for a static and mobile user.

Mobility Test. Finally, we wish to understand the effect of the LTE scheduling algorithm for static and mobile users. We consider two UEs in an office building downloading a random bitfile over a 1 Mbps UDP connection. We move one UE in a random walk around a floor of the building, while the other is static. Fig. 5-9 plots the percentage of data transmitted on each resource block for the static and mobile UEs. Note that the distribution is relatively flat for the mobile user, but has distinct peaks for the static user. This is because, unlike the mobile user, the channel of a static user is relatively coherent, leading the base station to preferentially allot resource blocks with high channel quality to the UE. Hence, the LTE base station actively adapts its scheduling algorithm in response to the user's channels.

5.2.4 Insights into LTE Rate Adaptation

While the LTE standard describes much of the PHY-layer protocol and procedures, the rate adaptation algorithm is still left to the choice of individual operators. In this section, we show how LTEye can shed light on some interesting aspects of this algorithm for an AT&T base-station in our locality.

We consider a single user device that downloads a random bitfile over a 1 Mbps TCP link from a server (we control TCP rate by throttling the bandwidth at the server using tc[1]). We conduct our experiment during periods of low network utilization, where the mobile device in our testbed is the sole active user of the network. The user device is placed at a static high SNR location, which should support the highest modulation and coding on the downlink.

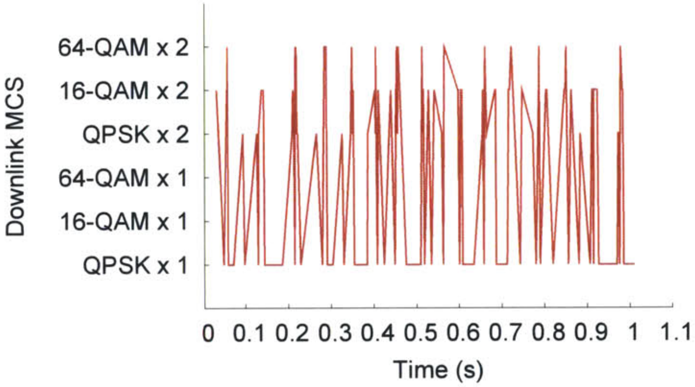


Figure 5-10: **Downlink MCS.** Trace of Downlink MCS for a user over time.

Fig. 5-10 plots the user's modulation and coding scheme for one second on the downlink. Surprisingly, the graphs indicate that the rate adaptation algorithm hops over a wide range of modulation, during the experiment. One possible explanation is that the base station is responding to loss of downlink packets. However, the control channel indicated no packet loss on the downlink over the entire experiment. A second explanation is that the wireless channel is not actually coherent over the duration of one second, even though the phone is static. To investigate this, we place a USRP at the user's location and estimate a channel coherence metric[3] capturing the base

[3] Given an initial channel $h(0)$ and current CFO-adjusted channel $h(t)$, channel coherence metric is $10\log_{10}|h(t)|^2/|h(t)-h(0)|^2$

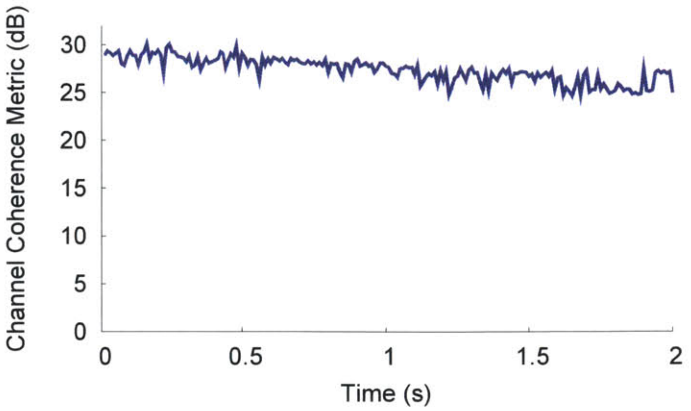


Figure 5-11: **Channel Coherence.** Channel Coherence metric of a USRP placed at the user's location over time.

station's SNR over two seconds, assuming the channel was estimated only once at time 0 (see Fig. 5-11). We observe that the channel is indeed coherent throughout the experiment. Hence, the rate adaptation algorithm is fairly complex, involving aggressive modulation exploration.

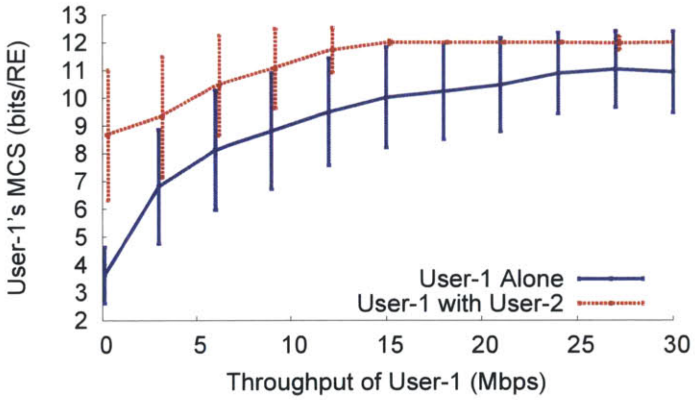


Figure 5-12: **MCS vs. Demand.** Mean downlink MCS (bits/resource element) across demand, with and without another high-demand user.

Next, we repeat the above experiment for different downlink demands (i.e. TCP throughput) under identical SNR, and plot the mean downlink modulation (in bits per resource element) as shown by the blue line in Fig. 5-12. Interestingly, as the downlink demand increases, the observed modulation also increases as well. In other words, the base station avoids transmitting packets to the user at high modulation (i.e. avoids risking higher loss probability), unless it is forced to, since the user demands high throughput.

Finally, we repeat the above experiment, this time adding a second user device (User-2) to the network, while first user (User-1) downloads a file at different TCP throughputs, as before. We allow User-2 to download a large bitfile containing random strings via six simultaneous TCP connections so as to saturate the LTE downlink demand. Interestingly, we now observe that packets to User-1 are sent at higher modulation, across demands (see the red line in Fig. 5-12). To understand why this is the case, note that by sending data to User-1 at higher modulation, the base station consumes less downlink resources per bit for User-1's data. This relieves more network resources that the base station can now assign to User-2 to serve its high demand. Therefore, base stations transmit packets at conservative modulation, only when this does not impact overall network throughput.

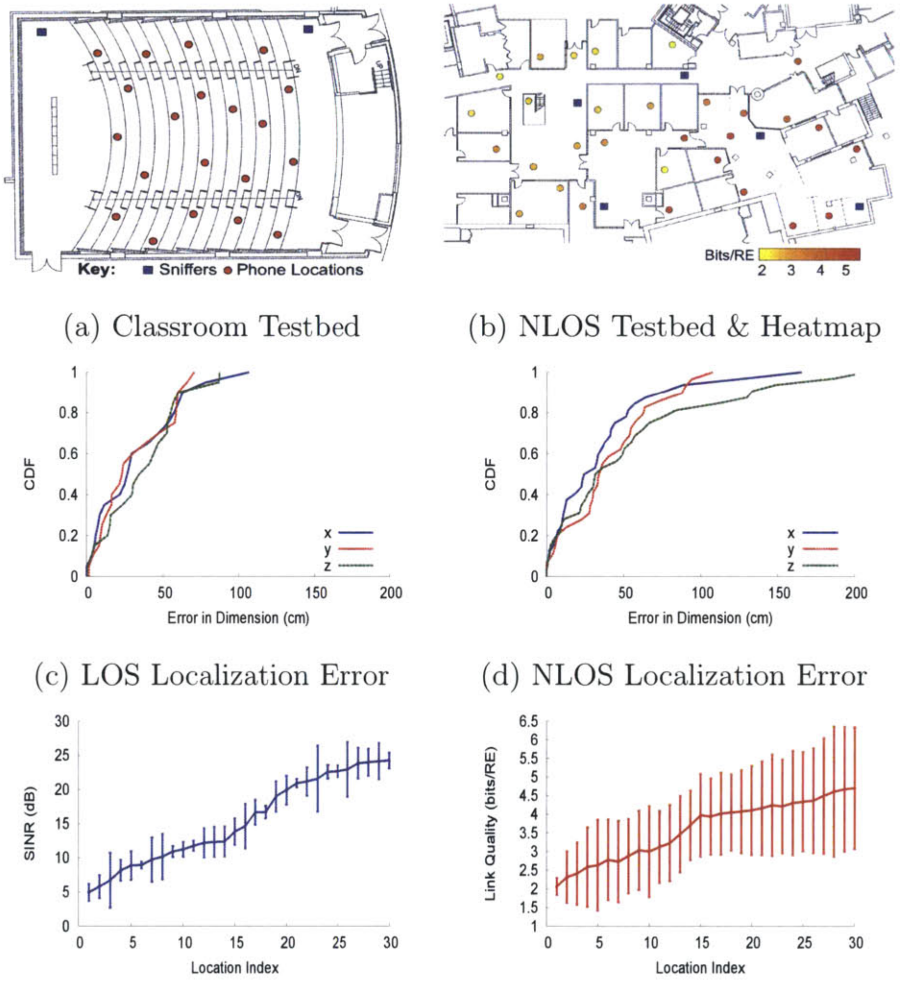
5.3 Results on Spatial Analytics

5.3.1 Detecting Cheaters in a Large Exam Hall

LTEye can enable new applications customized to the need of a particular community of users. For example, many modern exams follow an open book/material policy and allow students access to computers during the exams. However, students are asked to abstain from using the Internet to chat and collude online. Enforcing this policy over Wi-Fi is relatively easy by monitoring the Wi-Fi channels, turning the access point off, or even jamming the signal. However cheaters can still use their cellular service to chat with an accomplice. In this experiment, we demonstrate how LTEye's spatial analytics can help localize such cheaters in a large exam hall that accommodates up to 300 students.

Setup. We consider a large 24m × 17m exam hall as shown in Fig. 5-13(a) that seats up to 300 students. The exam hall has multiple chairs on a platform that slopes upwards from the podium. We place two LTEye sniffers on ledges close to the walls, as shown by the blue squares in the figure. The sniffers localize the 3-D location of ten active LTE cellphones accessing the Internet, placed in among twenty randomly chosen locations (the red circles in the figure).

Results. Fig. 5-13(c) plots a CDF of the error in each dimension of the estimated 3D-location of each cellphone. We observe a mean error in localization of 34 cm along each dimension and 61 cm in 3D displacement between the measured and actual location. Note that our errors are in 3-D space unlike past work [29, 28] and the experiments were performed in a large 24m × 17m area. LTEye identified the cheater's seat with 95% accuracy. Note that the cellphones are predominantly in line-of-sight to the LTEye sniffers, due to the nature of the exam hall. In §5.3.2, we estimate the error in localization for non-line-of-sight scenarios as well.



(a) Classroom Testbed

(b) NLOS Testbed & Heatmap

(c) LOS Localization Error

(d) NLOS Localization Error

(e) SINR across Locations

(f) Link Quality across Locations

Figure 5-13: (a),(b): Depicts our two testbeds a large exam hall and a floor of a large building. LTEye sniffers are denoted as blue squares and candidate phone locations as red circles. Further, the red circles in (b) are colored in shades of red, based on observed link quality from the base station; (c),(d): CDF of the error in 3D localization on each dimension in line-of-sight and non-line-of-sight scenarios respectively; (e),(f): Plots the measured SINR using USRPs and observed link quality from the downlink control channel across phone locations.

5.3.2 Visualizing LTE Performance over Space

As end-users, we have limited visibility into how LTE performance varies in different parts of our home or work place. In this experiment, we address this issue by synergizing LTEye's temporal and spatial analytics to visualize the performance of a cellular provider across spatial locations.

Setup. We deploy five LTEye sniffers on a 60m×34m floor of a large building, denoted by the blue squares in Fig. 5-13(b). We place ten phones in each of thirty randomly chosen locations shown as red circles in the figure. Note that several phones are in non-line-of-sight relative to all LTEye sniffers. We emphasize that for each client device, 3D-localization is performed using the spatial angles from at most three LTEye sniffers. The localization error can be further improved by incorporating spatial angles from additional LTEye sniffers.

Results. Fig. 5-13(d) plots the CDF of localization error along each of the three dimensions for phones that are in non-line-of-sight relative to all LTEye sniffers. Our results show a mean error in localization of 43.7 cm along each dimension and 84.6 cm in net 3-D displacement. Our algorithm to identify the direct line-of-sight path from §4.5 is crucial to localize phones in non-line-of-sight. Of course, the algorithm hinges on the fact that the line of sight path is, at the very least, observable in the multipath profile produced by SAR (See §4.4), even if it is not the most dominant path. Our experiments revealed that the line-of-sight path was always observed in the multipath profile of every phone in our large indoor testbed, including those furthest away from each LTEye sniffer in Fig. 5-13(b). Of course, while this may not generalize to every environment, our observations show the benefits of better penetration of signals in the 700 MHz frequency range through walls and obstacles, compared to Wi-Fi signals at 2.4 GHz or 5 GHz, and the higher transmit power of LTE devices in general.

Fig. 5-13(e) measures the mean and variance of Signal to Interference plus Noise Ratio (SINR) observed by a USRP placed in each of the thirty phone locations. The locations are sorted by mean SINR for ease of visualization. Fig. 5-13(f) plots the mean and variance of link quality for each phone (as bits per resource element) at the same locations, measured from LTEyeDB based on the downlink control channels. We observe that the link quality and SNR follow similar trends, showing that LTEye can effectively characterize the performance of the LTE network across spatial locations.

Fig. 5-13(b) visualizes the spatial distribution of link quality across phone locations, denoting positions of high quality as circles with darker shades of red, and low quality with lighter shades of red. The figure indicates that the link quality is strongest at locations to the bottom right, and weakest along locations to the top left. In fact, we found that placing an LTE relay at the top-left part of the floor significantly improves LTE service across the floor.

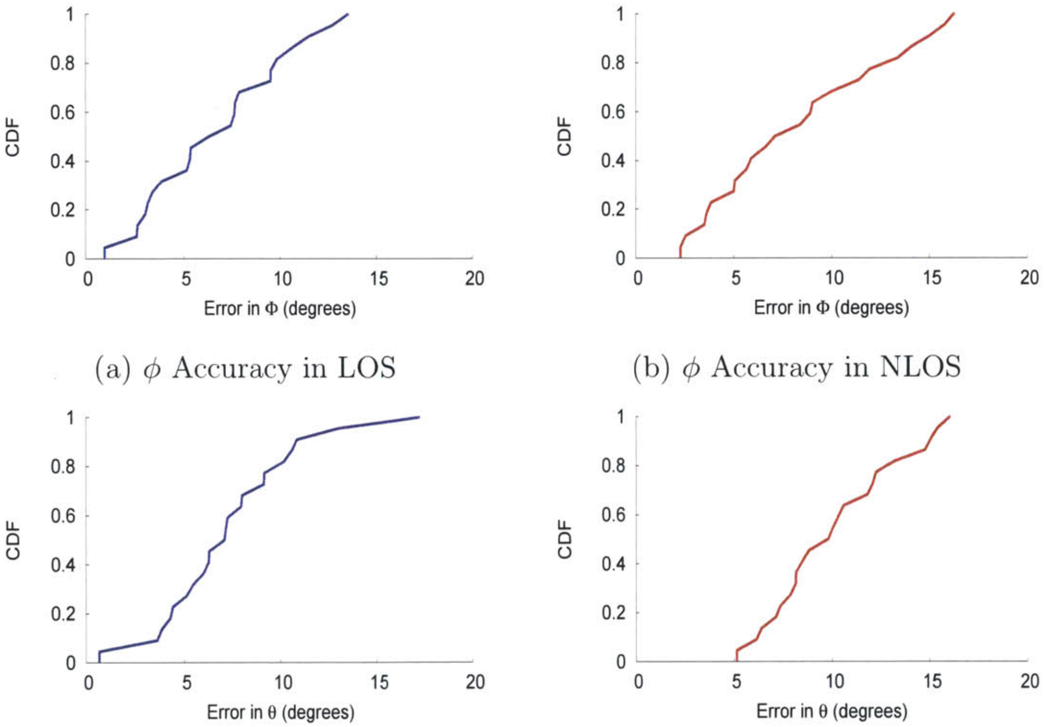
5.3.3 Measuring Accuracy of Observed Spatial Angles

In this experiment, we measure the accuracy of the polar angle θ and azimuthal angle ϕ, that are key primitives to LTEye's localization algorithm in §4.3, across spatial locations.

Setup. We consider an LTEye sniffer placed in one of five possible locations (denoted by blue squares) in a floor of a large building, as in Fig. 5-13(b). The LTEye sniffers are elevated on ledges close to the walls.[4] We place ten phones in several randomly chosen locations in both line-of-sight and non-line-of-sight, spanning the full range of spatial angles. We find the ground truth of the spatial angles by noting the actual 3D positions of the phones and the sniffer on a scaled high-resolution building floorplan.

Results. Fig.5-14(a)-(d) plot the CDF of error in θ and ϕ in line-of-sight (LOS) and non-line of sight (NLOS) locations. The figures show a low median error in both ϕ (LOS: 6.9°, NLOS: 7.8°) and θ (LOS: 7.2°, NLOS: 9.9°) across locations. Note that the accuracy can be further improved with multiple LTEye sniffers, particularly, in cases where the deviation in angles is large. Note that our algorithm to find the path of minimum delay was crucial for the accuracy of spatial angles in non-line-of-sight.

[4] LTEye cannot tell apart up from down as the rotating antenna path is symmetric. Placing sniffers on ledges removes this ambiguity.



(a) ϕ Accuracy in LOS

(b) ϕ Accuracy in NLOS

(c) θ Accuracy in LOS

(d) θ Accuracy in NLOS

Figure 5-14: **Accuracy of θ and ϕ**: Plots CDF of error in measured spatial angles in line-of-sight (LOS) and non-line-of-sight (NLOS).

5.3.4 Tracking C-RNTIs using RF Fingerprints

In this section, we evaluate LTEye's RF fingerprinting to map C-RNTIs assigned to the same phone. We consider the setup in §5.3.3 above and populate LTEyeDB across several experiments spanning ten minutes each. We track the correct C-RNTI mapping by constantly listening on the uplink from multiple USRPs placed close to each phone to recognize the high power signals that are sent during connection establishment. We also measure the inferred C-RNTI mapping from RF-fingerprints (See §4.2 and §4.6).

Results. We measure two quantities: (1) Precision: The percentage of new C-RNTIs which were correctly mapped to old C-RNTIs. (2) Recall: The percentage of correctly retrieved C-RNTIs-mappings among all actual C-RNTI mappings. Our algorithm achieves a high mean precision of 98.4±1.3% and mean recall of 96.7±1.4%, demonstrating the effectiveness of LTEye's C-RNTI matching algorithm. Note that we leveraged RF fingerprints to track C-RNTIs of users in §5.3.1 and §5.3.2 above.

Chapter 6

Conclusion

We presented LTEye, the first open platform to provide fine-grained temporal and spatial analytics on LTE radio performance, without private user information or provider support. LTEye employs a novel extension of synthetic aperture radar to communication signals to accurately localize mobile users, despite the presence of multipath. We empirically evaluate LTEye on software radios and provide deep insights on the LTE PHY and highlight shortcomings such as inter-cell interference and inefficient spectrum utilization.

Bibliography

[1] Linux Advanced Routing and Traffic Control. http://lartc.org.

[2] 3gpp. Radio measurement collection for Minimization of Drive Tests (MDT). http://www.etsi.org/deliver/etsi_ts/137300_137399/137320/11.03.00_60/ts_137320v110300p.pdf.

[3] 3rd Generation Partnership Project. E-UTRA, Physical Channels and Modulation, TS 36.211, v8.9.0. 2010.

[4] 3rd Generation Partnership Project. E-UTRA, Radio Resource Control, TS 36.331, v8.9.0. Apr 2010.

[5] 3rd Generation Partnership Project. E-UTRA, v8.8.0, Multiplexing and Channel Coding, TS 36.212. 2010.

[6] Bill Conley. M2M Implications of the 4G LTE Buildout. *Remote Magazine*, 2013.

[7] USRP N210. http://www.ettus.com. Ettus Inc.

[8] Eurecom. OpenAirInterface. http://www.openairinterface.org/.

[9] Federal Communications Commission. Enabling Innovative Small Cell Use In 3.5 GHZ Band NPRM & Order. FCC 12-148, 2012.

[10] R.G. Fenby. Limitations on Directional Patterns of Phase-Compensated Circular Arrays. *Radio and Electronic Engineer*, 1965.

[11] Patrick J. Fitch. *Synthetic Aperture Radar*. Springer-Verlag, 1988.

[12] FlexNets Group. Open-source long-term evolution (LTE) deployment (OSLD). https://sites.google.com/site/osldproject/.

[13] Chen Hsieh-Chung et al. Determining RF angle of arrival using COTS antenna arrays: A field evaluation. In *MILCOM*, 2012.

[14] Junxian Huang, Feng Qian, Alexandre Gerber, Z. Morley Mao, Subhabrata Sen, and Oliver Spatscheck. A Close Examination of Performance and Power Characteristics of 4G LTE Networks. MobiSys, 2012.

[15] Junxian Huang, Feng Qian, Yihua Guo, Yuanyuan Zhou, Qiang Xu, Z. Morley Mao, Subhabrata Sen, and Oliver Spatscheck. An In-depth Study of LTE: Effect of Network Protocol and Application Behavior on Performance. SIGCOMM, 2013.

[16] Huawei. SingleSON Whitepaper. http://www.huawei.com/en/static/SingleSON-75714-1-127574.pdf, 2012.

[17] Kiran Joshi, Steven Hong, and Sachin Katti. PinPoint: Localizing Interfering Radios. NSDI, 2013.

[18] Y. Kakishima, T. Kawamura, Y. Kishiyama, H. Taoka, and T. Nakamura. Experimental Evaluation on Throughput Performance of Asymmetric Carrier Aggregation in LTE-Advanced. In *VTC Spring*, 2011.

[19] Harold W. Kuhn. The Hungarian Method for the Assignment Problem. *NRL Quarterly*, 1955.

[20] Christine Laurendeau and Michel Barbeau. Centroid Localization of Uncooperative Nodes in Wireless Networks Using a Relative Span Weighting Method. *EURASIP Journal on Wireless Communications and Networking*, 2010.

[21] V Loscri et al. Carrier Independent Localization Techniques for GSM Terminals. In *PIMRC*, 2008.

[22] Sophocles J. Orfanidis. Electromagnetic waves and antennas. http://www.ece.rutgers.edu/ orfanidi/ewa/.

[23] Stan Salvador and Philip Chan. Toward Accurate Dynamic Time Warping in Linear Time and Space. *IDA*, 2007.

[24] Sanjole. The wavejudge wireless test system. `http://www.sanjole.com/our-products/wavejudge-test-system/`.

[25] Motorola Solutions. White paper: Femtocells - the gateway to the home. 2010.

[26] Joel Sommers and Paul Barford. Cell vs. WiFi: On the Performance of Metro Area Mobile Connections. IMC, 2012.

[27] ThinkRF. Product Concept Brief for Distributed and Remote LTE Analysis. `http://thinkrf.com/`.

[28] Jue Wang and Dina Katabi. Dude, Where's My Card?: RFID Positioning That Works with Multipath and Non-line of Sight. SIGCOMM, 2013.

[29] Jie Xiong and Kyle Jamieson. ArrayTrack: A Fine-grained Indoor Location System. NSDI, 2013.