Scaling Deep Learning Models for Spectrum Anomaly Detection

Zhijing Li[+], Zhujun Xiao, Bolun Wang[+], Ben Y. Zhao and Haitao Zheng
University of Chicago, [+]University of California, Santa Barbara

ABSTRACT

Spectrum management in cellular networks is a challenging task that will only increase in difficulty as complexity grows in hardware, configurations, and new access technology (*e.g.* LTE for IoT devices). Wireless providers need robust and flexible tools to monitor and detect faults and misbehavior in physical spectrum usage, and to deploy them at scale. In this paper, we explore the design of such a system by building deep neural network (DNN) models[1] to capture spectrum usage patterns and use them as baselines to detect spectrum usage anomalies resulting from faults and misuse. Using detailed LTE spectrum measurements, we show that the key challenge facing this design is model scalability, *i.e.* how to train and deploy DNN models at a large number of static and mobile observers located throughout the network. We address this challenge by building context-agnostic models for spectrum usage and applying transfer learning to minimize training time and dataset constraints. The end result is a practical DNN model that can be easily deployed on both mobile and static observers, enabling timely detection of spectrum anomalies across LTE networks.

CCS CONCEPTS

• **Networks** → Network monitoring.

ACM Reference Format:
Zhijing Li[+], Zhujun Xiao, Bolun Wang[+], Ben Y. Zhao and Haitao Zheng, University of Chicago, [+]University of California, Santa Barbara. 2019. Scaling Deep Learning Models for Spectrum Anomaly Detection. In *Mobihoc '19: The Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing, July 2–5, 2019, Catania, Italy*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3323679.3326527

1 INTRODUCTION

Cellular providers spend billions of dollars acquiring radio spectrum for network capacity and coverage. Yet spectrum management, specifically detection of faults from spectrum interference, remains a costly and ad hoc process, often involving manual diagnosis following customer complaints and operational failure logs. What makes detection hard is that interference can come from a variety of complex sources at any physical location, ranging from intentional spectrum misuse and misconfigured transmitters to RF leakage from cable plants and connectors. For example, interference from a misconfigured amplifier led to persistent quality-of-service issues for a tier 1 service provider [39].

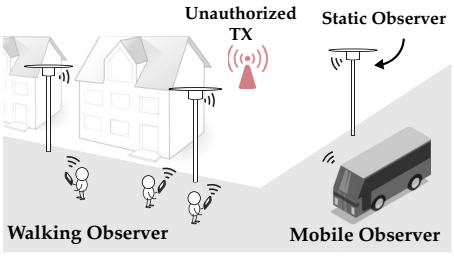


Figure 1: Spectrum anomaly detection by multiple observers.

These problems will grow in severity and scale in the near future. Advances in both reconfigurable hardware and spectrum usage policies make it easy to misuse spectrum without authorization. There is already evidence of these misuse attacks in China, where growth in unauthorized transmissions has prompted new initiatives to outlaw spectrum misuse [21]. Furthermore, cellular interfaces for IoT are coming, optimized for the network and energy needs of IoT devices. Adoption of these interfaces has the side effect of increasing security risks for LTE and nearby spectrum bands. A compromised device working on behalf of an attacker can perform jamming or denial of service attacks on cellular bands.

Clearly, cellular networks need robust and flexible tools to detect faults and misbehavior in spectrum usage, which we hereby refer to as *spectrum anomalies*. Despite open calls for automated management tools by the 3GPP standards body, current proposals are still limited to simplistic fail-stop fault models, and only on faults *within* the LTE infrastructure [1, 32]. Cellular carriers often evaluate physical spectrum usage by wardriving with specialized devices, and these activities are severely limited by high human and equipment costs [14].

Instead, we believe that cellular networks require *general* solutions capable of detecting a range of radio spectrum anomalies, from *transmissions at unexpected power levels*, to *interference from misconfigured devices* and *unauthorized transmitters*. Anomalies can appear anywhere in the physical network, and their detection requires a large-scale, distributed spectrum monitoring system.

In this paper, we explore the design of a general, scalable system for detecting spectrum anomalies in wide-area LTE networks. As shown in Figure 1, the system consists of two components: (1) a scalable, distributed spectrum monitoring system that measures physical spectrum usage using both static and mobile observers[2] distributed across the network, and (2) a general anomaly detection system that builds deep neural network (DNN) models using these measurements, and runs them at each observer as baselines

[1]Our proposed spectrum model, code, and test dataset are available at https://github.com/0x10cxR1/spectrum_anomaly_detection/.

[2]Spectrum monitoring requires both static and mobile observers to enforce coverage and scale. We assume that these observers are recruited by the carriers to perform spectrum monitoring and anomaly detection, and are well-behaved. This simplification allows us to focus on the problem of scaling DNN models for anomaly detection.

Mobihoc '19, July 2–5, 2019, Catania, Italy

ACM ISBN 978-1-4503-6764-6/19/07…$15.00
https://doi.org/10.1145/3323679.3326527

to detect spectrum usage anomalies. Our work builds on multiple prior efforts: some of which examined the feasibility of distributed spectrum monitoring using commodity devices [7, 30], while others validated the benefits of DNN-based spectrum anomaly detection using a single static observer [15, 33].

Why DNNs? Wide-area spectrum measurements collected by observers are highly complex, thanks to unpredictable signal propagation, frequent link adaptation, and traffic dynamics. Traditional models are unable to capture such complexity. DNN models, on the other hand, are known for automatically capturing complex patterns in the target data. Recent works have demonstrated the advantages of using DNNs to model spectrum usage [15, 33].

Despite significant progress, a large gap remains between current proposals and a feasible system for cellular networks. Since spectrum measurements generally depend on the context of the observer, *i.e.* time, location, and mobility status, each observer should ideally run a model tailored to the current context. But this renders our system impractical, given the amount of training overhead and run-time complexity it requires, *i.e.* it is impractical to assume that the system must build models for each physical location, and that each observer must change its model whenever it moves. A practical alternative is to explore a context-agnostic model for all observers, and whether such models can be trained, deployed and validated. *But will the elimination of "context-awareness" from DNN model designs degrade the accuracy of spectrum anomaly detection*?

We answer these questions through an empirical study on LTE networks, using detailed spectrum measurements across multiple LTE bands and cells. Our efforts lead to three key findings:

- Within each LTE cell, it is feasible to build a single, context-agnostic DNN model that accurately models normal spectrum usage pattern for the task of anomaly detection. Our DNN model does not use supervised learning to classify an event as normal or anomalous. Instead, we train a long-short term memory (LSTM) model on sequences of spectrum measurements. It recognizes events as anomalies when they deviate significantly from events expected or predicted by the model. Our model runs on both mobile and static observers to detect spectrum anomalies on the fly without any modification, putting a hard limit on the training overhead and run-time complexity. Deep autoencoder, another DNN model, can be designed to offer the same properties.
- Across LTE cells, the DNN model trained for a given LTE cell is not directly reusable at the other cells, but can be used to quickly train their models through *transfer learning*. Only a small amount of local spectrum measurements at the target cell is required. Our results show using transfer learning instead of training from scratch reduces required training data by a factor of 288.
- Since different LTE bands (frequency carrier, downlink or uplink) display different spectrum patterns, they require different DNN models customized for that band. The same transfer learning method can be applied to quickly train the model for a frequency band using existing models for other bands as a starting point.

Together, these findings demonstrate the feasibility of deploying a practical model for LTE spectrum anomaly detection on top of the distributed spectrum monitoring system. Specifically, the system first trains a general DNN model for normal spectrum usage, *i.e.* the teacher model, using past spectrum measurements from trusted observers. It then distributes this teacher model to each individual LTE cell's basestation, who uses a small amount of local spectrum measurements (contributed by trusted observers in the cell) to quickly calibrate the model, and distributes a unified, context-agnostic model to all the observers in the cell.

The above design has two key features. First, the spectrum DNN model is context-agnostic and can be easily deployed on a wide range of spectrum observers, static and mobile, and adapted using a minimal amount of local spectrum measurements. Each observer does not need to store a large number of models for each context, or switch to a new model whenever it moves. Instead, it runs the same DNN model regardless of its context, and only needs to switch to a new model when moving into a different cell. Second, the anomaly detection is general in that it avoids cellular-specific knowledge and can detect any events that affect spectrum usage.

2 PRELIMINARIES

To provide context for our later discussions, we present in this section the spectrum measurement dataset used in our empirical study, and our initial analysis on patterns in today's LTE spectrum usage. We also present existing models for spectrum anomaly detection, and evaluate their performance using our spectrum measurements in the presence of spectrum anomalies.

2.1 Analysis of LTE Spectrum Usage

Our Dataset. We performed signal measurements on three major LTE carriers in the US, including three downlink (DL) bands of AT&T (880MHz), T-Mobile (729MHz), and Verizon (749MHz), and one uplink (UL) band of AT&T (830 MHz). We used USRP N210 devices to capture 5 MHz spectrum within each LTE band.

While prior works on spectrum misuse detection [9, 20, 24, 33] only considered static observers, we performed measurements on LTE spectrum usage using both static and mobile observers (walking, driving). Our measurements were performed at two areas: a large university campus and an urban downtown area, separated by a distance of 8 miles. For each area, we verified that the observers were in the same LTE cell during the measurement period and the measurement range is within 1 mile.

Our measurements were performed between January and March 2018, and repeated in June 2018 to examine potential temporal variations. Specifically, we set up three static observers (well separated) in the university campus and collected measurements continuously for 7 days, and two static, well-separated observers in the downtown area for 3 days of continuous measurements. Walking and driving experiments were done for 45 min per day for 8 days. In total, the dataset contains more than 20 TB of signal data, where 32% of the data were collected at night.

Spectrum Usage ≠ RSS. Many prior works [9, 10, 20, 24, 42] have used measured received signal strength (RSS) as the base for spectrum anomaly detection, where an anomaly occurs if the current RSS deviates from a pre-defined range. Our measurement shows that RSS is not a viable base for mobile observers since it changes significantly and unpredictably over time. Figure 2 shows a random segment of RSS collected by a mobile observer over 10 minutes. Here the sudden rise of RSS values can be the result of multipath fading or interference from an unauthorized transmitter in proximity, which are indistinguishable using RSS data.

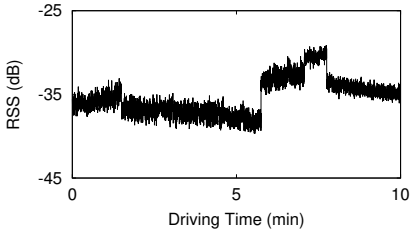
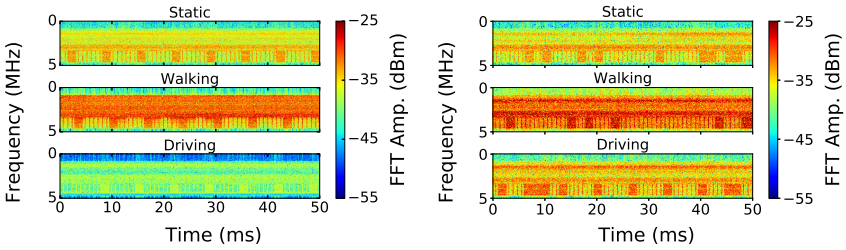


Figure 2: RSS varies largely over a 10-minute monitoring window, for a mobile observer.



(a) Initial spectrograms.

(b) Spectrograms after 10 minutes.

Figure 3: Spectrograms captured by spectrum observers under different contexts, based on the measurements on the 880MHz downlink band.

Time-Frequency Patterns of Spectrum Usage. Instead, we chose to analyze spectrum usage using the time-frequency spectrogram of the received signal. Spectrograms capture fine-grained signal amplitude over time at sub-frequencies, and are widely used for spectrum analysis. In absence of any anomaly, Figure 3 plots a spectrogram segment of 50ms at three observers (static, walking, driving) and another 50ms segment at each of the same observers about 10 minutes later. Despite the large difference in signal amplitude across users and time, we can observe visible temporal patterns from all six segments, in the form of bursts of high-power transmission along the time dimension.

We studied these patterns in detail and arrived at two key observation. *First*, the pattern is complex, especially in the temporal domain. Periodicity analysis shows that signal fluctuation peaks reside at 1200Hz, 160Hz and 60Hz, indicating that the key periodic pattern occurs every 0.21ms, 1.6ms, and 6ms. More frequencies of transmission bursts exist in addition to these main peaks, indicating more fine-grained temporal patterns beneath the obvious bursty patterns we observed. *Second*, the *short-term* patterns of the spectrum usage share some general shape. Carefully formed, they could serve as reliable "fingerprints" of normal spectrum usage.

Spectrum Patterns across Time, Cells and Bands. We also visually compared the spectrum usage patterns observed across time, LTE cells, and LTE bands. The short-term usage patterns are fairly consistent over time (by comparing observations in January-March, and June), differ slightly across cells (campus vs. downtown), but show more visible differences across LTE bands. We also performed periodicity analysis to confirm these observations.

2.2 Models for Spectrum Anomaly Detection

The above analysis suggests that it is feasible to build general spectrum anomaly detection by modeling the time-frequency patterns of normal LTE spectrum usage. The hypothesis is that the presence of a spectrum anomaly will produce visible changes to the patterns extracted from the measured signals, which trigger the detection of the anomaly. This type of anomaly detection prioritizes generality: training/building the model using normal spectrum usage *without* requiring any knowledge or labeling on anomaly instances.

There are multiple existing approaches of modeling spectrum usage patterns from the spectrogram, ranging from the classical methods of Kalman filter, one-class SVM [24, 28] to the recent proposal of neural network models (LSTM [33] and deep autoencoder [15]). Yet existing works only considered static observers.

We implemented and evaluated these approaches using our LTE measurements. A small portion of our measurements were conducted when anomalies were present. More details on these anomalies are described later in §6.1. For all the experiments, the observers were placed within 50m of the misuse transmitter.

Evaluation at Static Observers. For all the approaches (Kalman filter, one-class SVM, LSTM, deep autoencoder), we used as the model input the signal spectrogram over 256ms (we have tested other segment lengths between 32ms and 256ms and found that they do not change the conclusion). We trained the models using past spectrum measurements in absence of anomalies at each static observer. We also included a RSS-based method that uses a threshold to detect the presence of anomaly (Rule-based). The performance of the LSTM and deep autoencoder models are similar so we only included the LSTM result for brevity.

Figure 4 plots the results in terms of anomaly detection rate vs. false alarm rate. The results are similar across the four LTE bands so we only show the result in the 880MHz DL band for brevity. We see that the DNN model (LSTM in this case) largely outperforms the three non-DNN alternatives. This finding aligns with that of recent works [15, 33].

The reason behind the above result is that LTE spectrum patterns are complex due to the compound effect of traffic dynamics, RF propagation, and link adaptation. Additional complexity comes from possible correlations between feature dimensions. All of these make it difficult for traditional methods (*e.g.*, one-class SVM) to model the spectrum usage. Manual identification of good features that capture key spectrum patterns requires deep understanding of the data and much heavier efforts on feature engineering. And the complexity exacerbates when building the models for mobile observers.

3 SCALING SPECTRUM DNN MODELS

Our empirical analysis validates the observation of prior works [15, 33], where each static observer individually trains DNN models to detect spectrum anomalies. But *can such a context-specific model be deployed on a large-scale distributed monitoring system, where the spectrum observers are distributed across a wide area, and can be mobile or static*? Next, we answer this question empirically, testing whether models trained by a given static observer can be "reused" by another static observer at a different location and another mobile observer. Again we observe a consistent trend across all four LTE bands, and show the result for the 880MHz band for brevity.

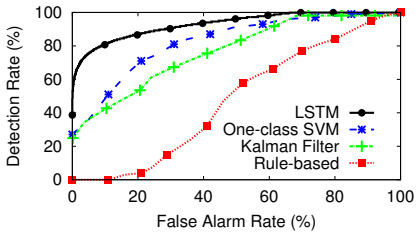
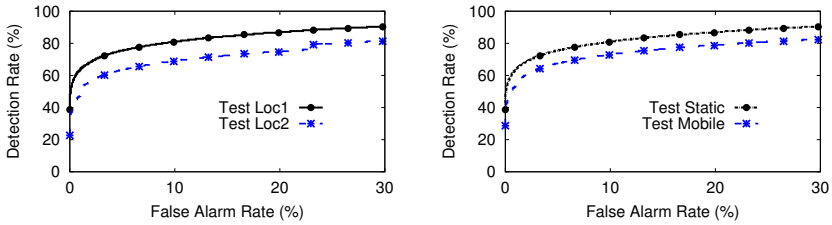


Figure 4: Anomaly detection performance of non-DNN (one-class SVM, Kalman filter, Rule-based) and DNN (LSTM) models, based on measurements at the DL 880MHz band.



(a) Test across locations. (b) Test cross context.

Figure 5: Anomaly detection performance when (re)using a LSTM model customized for a static observer at location 1. (a) running the model at location 1 and location 2. (b) running the model at location 1 and a mobile observer near location 1 (all based on measurements at the DL 880MHz band).

Test I: Reusing Models across Locations. Using our LTE measurements at three static observers (in the same LTE cell), we apply the same approach of [15, 33] to train, for each observer, the corresponding DNN models (LSTM and deep autoencoder). We then run the models customized for one observer at the other two observers both with and without the presence of spectrum anomalies. We considered a range of spectrum anomalies in the form of unauthorized transmissions used in §2.2.

Our results show that when reusing a spectrum LSTM model at a different location, the model is less accurate in capturing normal spectrum patterns. Thus the anomaly detection rate drops considerably (Figure 5(a)). The same applies to the autoencoder model.

Test II: Reusing Models at Mobile Observers. We also experimented with "reusing" models trained for a static observer at a mobile observer (walking at 3mph). Both observers were in close proximity (to reduce the impact of location change). Results in Figure 5(b) show a similar trend of performance degradation.

Our Focus: Scaling the DNN Models. Together, these experiments suggest that since wireless measurements depend on the context of the observer, *i.e.* time, location, and mobility status, ideally each observer should run a DNN model tailored to the current context. Unfortunately, this is impractical under our targeted scenario because the system must build models for each physical location in the network and user context, and each observer must change its DNN model whenever it moves. Such requirement leads to significant training overhead and run-time complexity.

This motivates us to explore a practical alternative: building a *unified* model for all the observers, with the goal of prioritizing scale and ease of deployment, minimizing training overhead, and maintaining reasonable accuracy. In the following sections, we tackle this new problem in two steps: first designing a single context-agnostic DNN model for anomaly detection in a single LTE cell (§4), then extending the single cell model to train models for many other LTE cells and bands using transfer learning (§5).

4 A SINGLE MODEL PER LTE CELL

In this section, we focus on designing a single DNN model for a single LTE cell, which will be deployed on all observers in the cell without any modification. Our hypothesis is that within a cell, the normal downlink spectrum usage seen by each observer comes from the same basestation, thus we could train the DNN model to capture a *unified* form of spectrum pattern that is context-agnostic, *i.e.* does not depend on mobility pattern and precise location within the cell. For uplink, each observer sees aggregated transmissions from many LTE users, and the normal spectrum usage could also display context-agnostic patterns. Thus our goal is to design models to automatically discover these context-agnostic patterns, and to validate whether they are sufficient for anomaly detection.

Our study considers two DNN models, LSTM and deep autoencoder. Both are known for capturing complex, temporal patterns in the target data that can be difficult to detect with simpler models [15, 18, 33]. In the following, we start with a brief introduction of the two models, and then describe the steps taken to build and train a context-agnostic version of these models using our spectrum data. We evaluate the models at both static and mobile observers, in terms of how they predict future spectrum usage. Later in §6 we evaluate the corresponding anomaly detection systems.

4.1 Background: LSTM and Deep Autoencoder

LSTM is a special type of Recurrent neural network (RNN), well-known for its capability of capturing comprehensive and intricate patterns embedded in sequential data. A LSTM model maintains an internal state in each RNN unit, and often consists of multiple stacked layers, forming an architecture similar to feed-forward neural network. This allows learning of complex relationships in sequential data. Normally another fully connected layer is attached at the end of the model for classification or prediction. Details on LSTM can be found in [18].

A stacked (or deep) autoencoder (details in [15]) is a DNN model designed to learn efficient data representation (or encoding) in an unsupervised way. It learns to compress the data from the input layer into representations, and then reconstructs the original data using the representations at the output layer. This process forces the autoencoder to extract the most useful features of the data.

Anomaly Detection. The above predictive models enable anomaly detection without prior knowledge of anomaly. The intuition is that since each model is trained using normal spectrum data, it *cannot* accurately predict data that contains anomaly, leading to large model prediction errors that trigger the anomaly detection. Figure 6 plots the anomaly detection process. We first train the model using spectrum observations in absence of anomaly, where given past values, the model predicts the next few values in the sequential data. Next, given a present spectrum observation, we use the model (and past observations) to predict the present spectrogram,

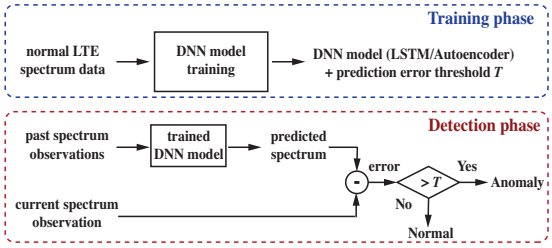


Figure 6: Anomaly detection using DNN models of spectrum usage.

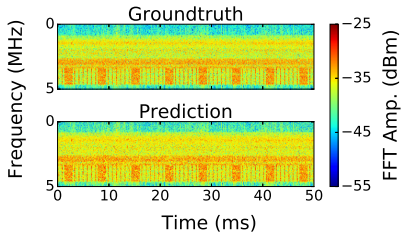


Figure 7: Spectrograms of actual and LSTM predicted LTE signal.

and compare it to the observed spectrogram. If the prediction error is larger than a threshold (details in §6), an anomaly is present.

Finding Clean Training Data. Ideally the model should be trained with measurements in absence of anomaly, which are hard to verify in practice. Fortunately, several works have confirmed that RNN, particularly LSTM can tolerate limited presence of anomalies in the training data without affecting anomaly detection performance [13, 16]. Under our target scenario, the system can choose training data from trusted observers who did not observe notable cellular service degradations at the time of data collection, thus the mass majority of the training data are in absence of anomaly.

4.2 Unified Models of Spectrum Usage

We now describe the unique steps we take to build and train a per-cell, unified predictive model on spectrum usage. While our description below is for LSTM, we apply the same process to build and train the deep autoencoder model.

Input to LSTM. We feed the raw spectrogram of the wireless signal into the LSTM model. Our intuition is that sufficiently powerful LSTMs operating on raw signals can extract meaningful patterns, while an alternative LSTM operating on aggregate statistics is vulnerable to poor choice of statistics, and could miss valuable dimensions of the data.

Given our LTE measurements, we configure the LSTM model to use $x = 25.6$ms of measured signal as input to predict the next $y = 6.4$ms. We chose these parameters because our spectrum analysis in §2 shows that the longest periodic pattern occurs at 6ms, thus a target frame of $y = 6.4$ms should be sufficiently large to include all the key patterns. We also experimented with other x values and found that 25.6ms offers the best performance under our target scenarios. We leave the optimization of x and y to future work.

Making the Model Context-agnostic. Our predictive model is context-agnostic, so that it can be deployed on all observers in the current cell, regardless of their physical location and mobility status. We take two steps to achieve that.

First, we apply *linear transformation* to expose the intrinsic spectrum usage patterns. As mentioned earlier, the input signal data displays a large variance across observing locations, which is an inherent property of radio propagation. Such high variance can cause LSTM (and autoencoder) to miss detailed temporal patterns and correlations among sub-frequencies, but focus solely on absolute power values. To expose these intrinsic spectrum patterns, we apply linear transformation, *i.e.*, mean-centering and scaling, to the input FFT amplitudes, and filter out input sequences that only contain noise (no signal at all). As a result, each input sequence to the LTSM model now has a zero mean and a variance of 1. This transformation is similar to the idea of *contrast stretching*, a common pre-processing technique in computer vision that exposes patterns by transforming pixel intensities to increase contrast [2].

Second, we use as training data a mixture of spectrum measurements collected by both static and mobile observers within the cell. Compared to context-specific models, this certainly minimizes the model training time and data requirements. One concern is that mixing training data from many sources can potentially increase the ambiguity between normal data and anomalies. For example, the normal spectrum usage seen at observer A could be similar to the anomalous spectrum usage seen by another observer B. When the training data includes those measurements from A, the trained model could misclassify B's observation of an anomaly as normal.

We took a detailed look at our measurement data (with anomalies), but did not identify any of such events. While our anomaly instances are limited in scale, this result suggests that the probability of such events is low in practice. An advanced attacker could form its misuse signals to imitate normal spectrum usage, but this is challenging since the observers will observe the aggregated signals from the attacker and the legitimate LTE transmitters.

Model Training. To train these models, we divide our per-cell LTE measurement data into two portions: one used for training, and one for testing. Both datasets contain measurements collected by static, walking and driving observers. The observers in the testing dataset do not appear in the training dataset. Overall, for each cell, the ratio of the training data volume and the test data volume is 2.5:1. Each model is trained to minimize the Root Mean Square Error (RMSE) between the predicted signal and the ground-truth signal (after transformation) in the training dataset.

4.3 Evaluation: Model Prediction Error

We evaluate how well the DNN models predict the spectrum usage of the immediate future, so that they can quickly detect anomalies that disturb the spectrum usage pattern. As an illustrative example, Figure 7 plots the actual spectrogram (on a randomly chosen 50 ms segment) and the output of the LSTM prediction model (after reversing the linear transformation). To reconstruct the 50ms segment from the prediction result, we cascade 8 segments of 6.4ms prediction results, each predicted from previously observed 25.6ms measurements. We see that the LSTM model is able to recover the key patterns in spectrum usage across sub-frequencies and time.

Evaluation Metric: Spectrogram Prediction Error (dB). We evaluate each model by the difference between the true signal spectrogram and the model prediction. Specifically, we calculate the RMSE between the true FFT amplitude (dBm) across sub-frequencies of the LTE band and the LSTM model prediction values after being inverse-transformed back to FFT amplitude (dBm). In a nutshell, the RMSE value approximates the amplitude spectrogram error in

(a) Across Location

Test \ Train	Loc 1	Loc 2	Loc 3	Unified
Loc 1	**2.71 (0.38)**	5.21 (12.74)	2.97 (0.56)	2.67 (0.32)
Loc 2	3.41 (0.64)	**2.46 (0.76)**	3.67 (1.09)	2.47 (0.24)
Loc 3	2.68 (0.46)	4.56 (7.03)	**2.63 (0.26)**	2.61 (0.29)
Mixed	3.70 (1.79)	4.86 (13.84)	4.07 (1.88)	**2.59 (0.23)**

(b) Across Time Periods

Test \ Train	Day	Night	Unified
Day time	**2.59 (0.23)**	2.74 (0.22)	2.53 (0.18)
Night	2.63 (0.23)	**2.71 (0.27)**	2.61 (0.22)
Mixed	2.59 (0.18)	2.74 (0.18)	**2.55 (0.22)**

(c) Across Mobility Context

Test \ Train	Static	Walking	Driving	Unified
Static	**2.52 (0.21)**	2.47 (0.23)	2.57 (0.33)	2.43 (0.23)
Walking	2.47 (0.29)	**2.62 (0.23)**	2.67 (0.27)	2.56 (0.18)
Driving	2.64 (0.29)	2.58 (0.26)	**2.59 (0.26)**	2.57 (0.23)
Mixed	2.66 (0.30)	2.54 (0.27)	2.61 (0.23)	**2.58 (0.19)**

Table 1: Prediction error (dB) of LSTM models under different training configuration. Numbers in parenthesis show the standard deviation of prediction error (dB). Here we show the result from the 880MHz band while the other bands lead to similar conclusions.

a single dB value. We refer to this metric as the prediction error (dB). Because our test data has many observers, we will present the mean and standard deviation of the prediction error across all the observations in the test dataset.

The prediction error directly links to the accuracy of anomaly detection. The smaller the prediction error is in absence of anomaly, the better the predictive model is and the higher accuracy the model has during anomaly detection. We evaluate the anomaly detection performance later in §6, which yields consistent results.

Unified vs. Customized Models. We evaluate our unified models by comparing them to models customized to individual observer's context. The results of LSTM and autoencoder are similar to each other: the prediction error of autoencoder is 3-8% higher than that of LSTM. We only show the LSTM results for brevity.

Table 1(a) shows the mean and standard deviation of the prediction error (dB) of our unified model and those of the models customized to three individual locations. The location-specific models, when running on a different location, produce large prediction errors (5.21 dB rather than 2.5dB). But the unified model is always as good as or even better than all the location-specific models.

We repeat the experiment in the time domain. Table 1(b) confirms that training data over day and night can also be mixed together when building the unified model. We also use data collected in June 2018 to further test our model (trained using measurements from January and March 2018), and the unified model consistently provides better prediction than those designed for specific time periods of the day. The difference between the models is less visible compared to that in Table 1(a), indicating that physical location has a much heavier impact on spectrum monitoring than time.

We also experiment with the mobility context. We group the measurements by their mobility context: static (mixed locations), walking, and driving ($\leq$ 25 mph). In addition to the unified model, we also trained mobility-specific models for each of the three contexts. Table 1(c) shows that the unified model and the mobility-specific models perform similarly. For both, the average prediction error is bounded by 2.62 dB with a very low variance (0.26).

Overall, the unified model achieves the best prediction performance, 2.58 dB (0.19), when tested at a diverse set of observers. This can be attributed to two factors. First, the model's timing configuration (using 25.6 ms data to predict next 6.4 ms data) allows LSTM to capture critical spectrum usage patterns, and yet remains small enough to make the model robust against context changes. Second, the linear transformation allows LSTM to focus on intrinsic patterns of signal spectrogram, which remains consistent across different mobility context, time periods, and locations.

It should be noted that a related challenge is whether and how such unified model per LTE cell can be used near cell boundaries, where an observer can potentially pick up signals from multiple basestations. When these basestations operate on the same frequency band, the observer could see signal patterns that are different from those at in-cell locations. This must be treated with care to minimize false alarms. As future work, we plan to address this issue using dedicated measurements at cell boundaries.

Model Complexity. We implement our LSTM and autoencoder models on a NVIDIA Titan X GPU, where it takes < 10ms for prediction on each data segment. As future work, we plan to implement our design on commodity mobile platforms such as smartphones and NVIDIA Jetson platforms. Existing works have successfully deployed efficient LSTM models on mobile devices [6, 27]. The LSTM model in [27] has 5 layers, each with 500 LSTM units, and runs efficiently on Nexus 5 Android smartphones. In comparison, our LSTM has fewer parameters (2 LSTM layers, 64 units each) and should also run efficiently on common mobile devices.

4.4 Models for DL and UL Bands

We take a closer look at the unified models built for each of the four LTE bands. Recall that our analysis in §2.1 shows that LTE bands display different spectrum usage patterns. The UL band (830 MHz) is particularly different from the DL bands.

Interestingly, the final model structure also differs between the DL and UL bands. For LSTM, the three DL bands share the same structure: 2 LSTM layers of 64 units plus 1 dense layer, while the UL band requires an extra LSTM layer. The same applies to autoencoder: the DL models have 4 dense layers while the UL model has 6 dense layers. This is somewhat intuitive since LTE DL signals originate from a single strong transmitter (basestation), while UL signals are aggregates of many weak transmitters. The UL spectrum patterns are more complex, requiring more neurons to learn.

Table 2 lists the prediction errors of the four bands using LSTM. The three DL bands perform similarly, while the UL band experiences larger prediction errors. We also verified the same model using spectrum measurement data collected a few months later (June 2018), and the results are consistent.

	880 MHz	750 MHz	730 MHz	830 MHz (UL)
Testing: Early 2018	2.58 (0.19)	2.70 (0.25)	2.54 (0.26)	3.14 (0.78)
Testing: June 2018	2.61 (0.21)	2.72 (0.24)	2.52 (0.25)	3.11 (0.73)

Table 2: Model prediction error (dB) of our unified LSTM model.

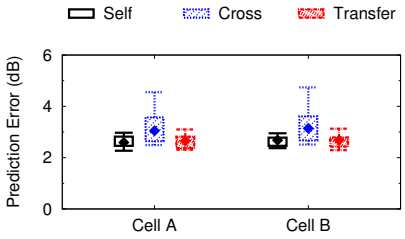


Figure 8: Prediction error of LSTM models transferred across different LTE cells in the 880 MHz DL band.

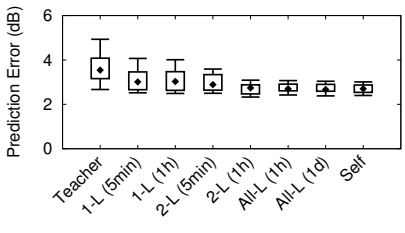


Figure 9: Prediction error of LSTM models with different transfer options (transfer model of DL 880 MHz to DL 730 MHz).

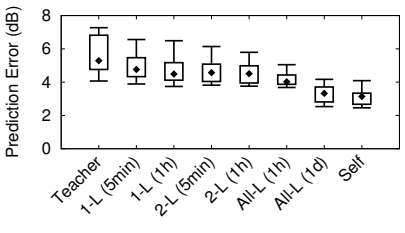


Figure 10: Prediction error of LSTM models with different transfer options (transfer model of DL 880 MHz to UL 830 MHz).

5 BEYOND THE PER-CELL MODEL

We now consider the problem of building spectrum models for many LTE cells and multiple bands. One can simply train a model for each LTE cell and band, but the training overhead and data collection requirements are practically prohibitive. Like other DNN models, LSTM and deep autoencoder require large amount of diverse training data and can take hours and even days to finish training. Currently, our models take 2 hours to finish with 1 day worth of training data and almost 24 hours with 8 days of training data. Practical deployment will likely need much larger and more diverse training data, and more frequent training to adapt the models. Thus it is critical to reduce the model training overhead across all the LTE cells. In the following we discuss and compare potential solutions to address the issue of training overhead.

5.1 Can Models be Reused?

Does reusing the model across cells and bands really work?

Test I: Reusing Models across LTE Cells. We apply the LSTM model trained for one cell to another (same network carrier, same frequency band, same technology, just a different basestation), and observe sizable performance degradation in both model prediction and anomaly detection. For the 880 MHz DL band, the average prediction error raises to 3.36 dB from the baseline of 2.58 dB and the standard deviation jumps from 0.19 to 0.56. This is likely because the two basestations are configured differently so their spectrum usage patterns differ.

Test II: Reusing Models across LTE Bands. We apply the LSTM model trained on the 880 MHz DL band to the 730 MHz DL band. The average prediction error and the standard deviation grow to 3.54 dB (0.71) compared to 2.70 dB (0.25). This is because the two bands show visible differences in spectrum usage patterns, which are captured by LSTM to produce a precise model for each. Autoencoder shows the same trend.

Together, these results show that models trained for a specific LTE cell and a specific LTE band are in general *not* reusable across cells and bands. This does not contradict with our conclusion in §4 where the per-cell model can be reused within the same cell. In a given cell, the spectrum usage pattern is fairly consistent.

5.2 Fast Training via Transfer Learning

To speed up model training at many cells, we consider an alternative solution, *transfer learning* [34], which adapts a pre-trained DNN model to a new scenario using limited training data. It leverages the underlying similarity between tasks associated with two models. By transferring model architecture and weights from a pre-trained model (*teacher*) to the new model (*student*), one can bootstrap and fine-tune the student model with limited training data.

Transfer learning is suitable for our problem because LTE cells share similar spectrum usage characteristics (§2), especially for DL bands since only LTE base stations are transmitting. Next we show that transfer learning can be used to quickly adapt a pre-trained LSTM model to a new LTE cell and even to a new LTE band, reducing training data volume by a factor of 288.

We note that a similar concept of "knowledge transfer" has been applied to wireless networking design, using knowledge collected by one basestation to help configure another basestation (*e.g.*, spectrum handoff [22], operating modes for energy saving [23] and content caching strategies [4]). Our solution is motivated by these existing works, but our contribution lies in the *novel application* of transfer learning to the problem of spectrum anomaly detection, and a detailed validation using real-world LTE measurements.

When applying transfer learning, we first build a student model by copying the teacher model, then use local spectrum measurement data to refine the model. Here the transfer process depends on k, the number of model layers "allowed" to be updated [34]. The simplest form of transfer learning updates on the last (dense) layer of the model. This is commonly used when the student model targets very similar task or domain characteristics as the teacher model. The more advanced ones allow more or all the layers to be updated. The student models can better adapt to new scenarios but require more local data to reach convergence.

Transfer across LTE Cells. To test the effectiveness of transfer learning, we transfer the LSTM model trained on cell A (using 1 day of training data) to cell B (same carrier, band, technology), and fine-tune B's model with 5 minutes of spectrum data collected in cell B. Here we chose 5-min because the tuning process converges. We consider the simplest transfer option of freezing all weights of the two LSTM layers and only updating weights in the last dense layer. For comparison, we train another LSTM model for B from scratch using the same amount of training data as of A (1 day).

Figure 8 shows the prediction error of the model trained from scratch (Self), the model of the other cell (Cross), and the transferred model (Transfer). The transferred model's error mean (2.65 dB) is extremely close to that of the model trained with the full data (2.59 dB). Yet this model only requires fine tuning the last dense layer using 5-min local spectrum data, comparing with 1-day worth of data for Self (a factor of 288 reduction). The same conclusion holds when we test the autoencoder model.

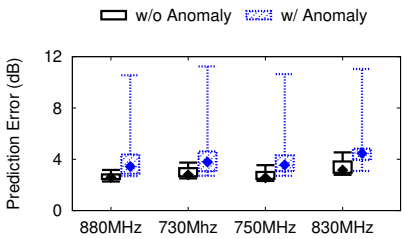


Figure 11: Quantiles of model prediction errors w/ and w/o anomalies.

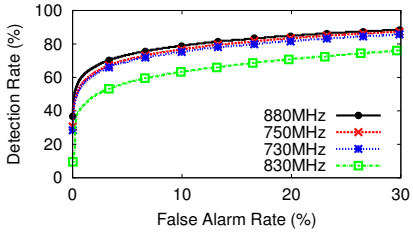


Figure 12: Detection rate vs. false alarm rate of unauthorized transmitters of different bands.

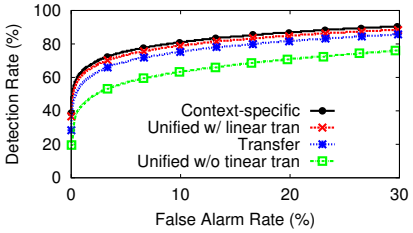


Figure 13: Our unified and transferred models are on par with the oracle design that uses location-specific models.

Transfer across LTE Bands. Since different LTE bands display different spectrum patterns, we expect more efforts to complete the transfer learning. Since our LSTM model has three layers, we experimented with three transfer approaches: 1L, 2L, All-L, respectively, to reflect the number of model layers it needs to fine tune. We also include the results of copying the teacher model (Teacher) and training from scratch (Self).

Figure 9 shows the quantile distribution of the LSTM model prediction error (in absence of anomaly) by transferring the model of the DL 880 MHz band to the DL 730 MHz band, with different transfer options. Since the underlying temporal patterns differ between these two bands, only fine-tuning the last dense layer is insufficient (the average prediction error is 3.13 dB compared to 2.70 dB of Self) even after adding more training data. In the end, fine tuning 2 layers with 1 hour of data achieves comparable performance of Self. Interestingly, fine-tuning all 3 layers with 1 day worth of training data slightly outperforms training from scratch (Self). Again the same trend applies to the autoencoder model.

We also seek to transfer the DL model (*e.g.*, 880 MHz) to the UL band (830 MHz). As mentioned in §4.4, when trained from scratch, the 830 MHz band requires more dense layers for both LSTM and autoencoder than the three DL bands. Thus direct transfer between the two types of bands might not be as effective as the above case. Figure 10 confirms that for LSTM, even after fine-tuning all the layers (of the transferred 880 MHz model), the prediction error is still not on par with that of Self (which needs an extra LSTM layer). Therefore, while transfer learning can potentially be applied to quickly customize LSTM (and autoencoder) models across bands, choosing the right teacher model can be a critical requirement. We leave this topic to future work.

Complexity. We implemented the transfer learning process on the NVIDIA Titan X GPU server. Across cells, the re-training took less than 5 minutes, and uses 5-minute of local data. Across bands, the re-training took less than 2 hours, using 1-day worth of local measurements. For both, training from scratch would take 24 hours, and use 8-day worth of spectrum measurements.

6 EVALUATION: ANOMALY DETECTION

In this section, we use real anomaly instances to evaluate our anomaly detection system built on the DNN models. Our evaluation seeks to answer the following questions: (1) whether our unified model performs as good as the (impractical) oracle system that builds context-specific model for each location, (2) how models trained via transfer learning perform in anomaly detection compared to those trained from scratch.

Choosing Anomaly Threshold. Our DNN models detect an anomaly if the difference between the measured data and the model predicted data exceeds a threshold. The threshold also determines the false alarm rate. To determine this threshold, we partition the model training data to two subsets: the training set and the validation set. After a model is trained, we use the validation data to calculate the statistics of the model prediction error. For our dataset, these errors can be modeled using a Gaussian distribution. From this distribution we can calculate, for each false alarm rate, the corresponding threshold on the prediction error.

Next, we discuss our experiments using two types of anomalies: unauthorized transmitters and misconfigured LTE basestations. For all the experiments, the detection rate of autoencoder is similar to that of LSTM (only 2-4% worse while keeping the same false alarm rate). Thus we only show the LSTM result for brevity.

6.1 Detecting Unauthorized Transmissions

We generated anomalies in the form of unauthorized spectrum usage, where an "unlicensed" transmitter (USRP N210) broadcasts various types of signals. Our anomaly instances include transmissions using the entire 5MHz band and OFDM signals (like LTE), using a portion of the 5MHz band (1, 2, 3MHz) with OFDM signals, and a narrowband misuse with QPSK and BPSK signals. When the anomaly is on, we set up a static observer and a walking observer in proximity (within 50 meters) who collect LTE measurements and perform anomaly detection. The walking observer follows a pre-defined route for all the anomaly instances. We did not have any driving observers since they quickly go out of range of our low power transmitters. In total, we performed 100 experiments, each of 5 minutes long.

Ethics. We are very aware of the potential impact of our experiments on cellular users, and took extensive precautions to ensure that these experiments had no impact on cellular users or basestations. First, we chose the second floor of an older campus building with heavy concrete walls and floors as our setting. We first measured signal propagation properties in the building by setting the transmitter frequency to 900MHz (closest unlicensed band), and using a spectrum analyzer to measure signal strength at numerous locations inside and outside of the building. We confirmed that the thick concrete walls and floors completely blocked signals beyond the immediate open hallway and adjoining offices and no signals were observed outside the building or on floors above or below.

Next, we scheduled experiments late at night and on weekends, when the campus building is generally unoccupied. Between experiments, one student walked the entire length of the hallway and checked to make sure no one else is on the floor. We did not encounter any other occupants of the building during our experiments. We also periodically used spectrum analyzers to (re)confirm that our transmissions are strictly constrained to the second floor.

Results: Anomaly Detection Accuracy. Figure 11 compares the quantile distribution (5%, 25%, 50%, 75%, 95%) of the model prediction error (dB) per spectrum observer, *i.e.* the RMSE between the measured and predicted spectrograms, across all the measurement instances, with and without anomalies. The presence of anomalies largely increases the prediction error. The two distributions are reasonably separated for the three DL bands, but overlap slightly for the UL band (830MHz). Next, Figure 12 plots the RoC result (anomaly detection rate vs. false alarm rate) for the four LTE bands. Here we average the detection result across all the measurement instances collected by the static and mobile observers, producing the average detection rate per spectrum observer. We see that for the three DL bands, the anomaly detection results are on par with each other, while the UL band is less effective. Yet these results are still significant better than non-DNN solutions (see Figure 4).

In our experiments, misdetection occurs when the anomaly's signal power is low and the observer is further away from the misuse. In practical deployment, one can improve the anomaly detection rate by deploying more observers for density and coverage, further pushing the need for a context-agnostic model.

Results: Unified vs. Customized Models. We compare our unified model (with linear transformation) and the basic version (without linear transformation), to an context-specific system that builds context-specific models for each location. We compute the anomaly detection result for the context-specific system by training a model for each static observer using its own past observations, and testing the model using the anomaly instances in range of the static observer. Note that our unified model is tested on all the anomaly instances and on both the static and mobile observers. As shown in Figure 13, our unified model (with linear transformation) performs as well as the context-specific model.

Results: Transferred vs. Self-trained Models. We compare the anomaly detection performance of models transferred from other cells with those of models trained from scratch. Figure 13 shows that the transferred model achieves almost identical performance while greatly reducing the training overhead.

6.2 Detecting TX Misconfigurations

We also study anomalies of basestation misconfiguration, implemented by modifying our LTE measurement traces. This will not produce any impact on cellular services. We consider two types of misconfiguration: (1) the misconfigured basestation stops transmitting signals at some or all sub-frequencies; (2) the misconfigured basestation suddenly changes its transmit power level. We produce both instances by modifying our LTE downlink measurement traces, replacing them with replays of measured noise signals or increasing/decreasing the amplitude of the received signals.

We note that these anomalies could also be detected by other methods, since (strong) static observers will likely detect changes in the spectrogram. Instead, we use these to show that our unified model can detect *general* types of anomalies beyond unauthorized transmissions. That is, the *same* model can detect both unauthorized transmissions and misconfiguration of basestations.

Detection Results. Table 3 lists the average detection rate per spectrum observer under 1% false alarm rate for different categories of misconfiguration. Here "x% F down" means transmissions on x% of frequency is replaced as noise. "ΔP =x dB" means transmit power is modified by x dB. Even at a very low 1% false alarm rate, the same unified model (as in §6.1) can effectively detect anomalies caused by misconfiguration, and the detection rate correlates with the severity of the anomaly. While linear transformation used to build our model "suppresses" the impact of transmit power level, our model can still detect sudden basestation power changes because it creates notable changes in the spectrum usage pattern.

	ΔP=3 dB	ΔP=5 dB	33% F down	66% F down	100% F down
880 MHz	58%	78%	52%	87%	100%
750 MHz	60%	81%	57%	90%	100%
730 MHz	53%	78%	54%	88%	100%

Table 3: Anomaly detection rate at 1% false alarm rate.

7 RELATED WORK

Anomaly Detection and Diagnosis in Wireless Networks. Existing works can be divided into three categories, depending on who runs anomaly detection and diagnosis. The first category involves system administrators [5, 17, 19, 31], who use system logs or Key Performance Indicators (KPIs) to detect network outages and performance degradations. The second leverages diagnosis by network clients [8]. The third category uses third-party devices to monitor and detect physical layer anomalies such as spectrum misuse, using metrics like signal strength variations [11, 37].

Our work falls into the third category. Our key contributions are the novel application of DNN (LSTM and deep autoencoder) and transfer learning to the problem of spectrum anomaly detection, the design of a context-agnostic DNN model, and the empirical study using measurements by both static and mobile sniffers.

Misuse Detection for Opportunistic Spectrum Access. Existing works have studied the issue of spectrum usage violation where a secondary user tries to transmit when a nearby primary user is inactive. They consider individual features of signal transmissions, including RSS spatial distribution [3, 24, 25, 29], RSS variation [10, 42], physical channel properties [12, 41], amplitude difference between direct and reflected paths [26] as well as airtime utilization [38]. Most designs were based on abstract propagation models, which do not capture real world settings.

Our work considers the issue of spectrum anomalies due to unauthorized transmitters and misconfiguration of LTE basestations. Our work differs from existing works by taking an empirical, data-driven approach. Instead of relying on a fixed set of features, we build DNNs to automatically extract features required for accurate anomaly detection, and develop context-agnostic DNN models.

Machine Learning for Signal Classification and Anomaly Detection. Early work focused on statistical hypothesis test [38] and threshold-based methods [20, 37]. Recently, ML models have been applied to the problem of signal classification (*e.g.* [35, 36]) and spectrum misuse detection [15, 24, 33]. [33] used a small scale study to show that LSTM outperforms Kalman sequence predictor. [15] developed an autoencoder model for spectrum anomaly

detection, based on a limited dataset (1000 samples) on the FM band. [24] applied one-class SVM to detect spectrum anomaly (via simulations) while [40] used supervised learning to train Hidden Markov Models. These existing works focused on anomaly detection by a single static observer without considering the impact of user mobility and location. They used either simulation or few measurement data for validation.

Our work has a much broader scope by developing robust, scalable anomaly detection capable of detecting previously unknown anomalies, for both static and mobile observers. We also collected detailed signal measurements on four LTE bands and under different user context to drive our empirical study.

8 CONCLUSION AND FUTURE WORK

We show that a scalable DNN model on LTE spectrum usage can be built and deployed on a wide range of observers (static nodes, walking users, buses), enabling real-time spectrum anomaly detection. Its performance matches the "oracle" design that trains customized models for each specific user context (location and mobility). The model remains constant for any observer in a single cell, and can be quickly trained and adapted using a small amount of local spectrum measurements. To the best of our knowledge, this is the first to show the feasibility of building practical and general spectrum anomaly detection systems for large-scale LTE networks.

Moving forward, we plan to explore several directions. *First*, we plan to explore other DNN models, and validate them using spectrum anomalies in the wild. Also, the frequency-temporal pattern of the prediction error could be used to distinguish different anomaly types. This needs to be further validated using anomaly instances in the wild. *Second*, spectrum measurements at individual observers can be noisy [30], or corrupted/modified by well-equipped adversaries. We plan to refine our system to be robust against such artifacts. *Finally*, the spectrum usage may change over time, *e.g.*, upon carrier upgrade. While our current measurements did not observe such changes, how to update the model to the new configuration will be an interesting direction to explore.

ACKNOWLEDGMENT

We thank the anonymous reviewers and our shepherd Joongheon Kim for their useful feedback on the paper. This project was supported in part by NSF grants CNS-1527939 and CNS-1705042.

REFERENCES

[1] Amirijoo, M., et al. Cell outage management in LTE networks. In *Proc. of ISWCS* (2009).
[2] Arici, T., Dikbas, S., and Altunbasak, Y. A histogram modification framework and its application for image contrast enhancement. *Trans. on Image Processing 18*, 9 (2009), 1921–1935.
[3] Bansal, T., Chen, B., and Sinha, P. Fastprobe: Malicious user detection in cognitive radio networks through active transmissions. In *Proc. of INFOCOM* (2014).
[4] Bastug, E., Bennis, M., and Debbah, M. Anticipatory caching in small cell networks: A transfer learning approach. In *Proc. of WAN* (2014).
[5] Bouillard, A., Junier, A., and Ronot, B. Hidden anomaly detection in telecommunication networks. In *Proc. of CNSM* (2012).
[6] Cao, Q., Balasubramanian, N., and Balasubramanian, A. MobiRNN: Efficient recurrent neural network execution on mobile GPU. In *Proc. of EMDL* (2017).
[7] Chakraborty, A., Gupta, U., and Das, S. R. Benchmarking resource usage for spectrum sensing on commodity mobile devices. In *Proc. of HotWireless* (2016).
[8] Chandra, R., Padmanabhan, V. N., and Zhang, M. Wifiprofiler: cooperative diagnosis in wireless lans. In *Proc. of MobiSys* (2006).
[9] Chen, R., Park, J.-M., and Reed, J. H. Defense against primary user emulation attacks in cognitive radio networks. *IEEE JSAC 26*, 1 (2008), 25–37.
[10] Chen, Z., Cooklev, T., Chen, C., and Pomalaza-Ráez, C. Modeling primary user emulation attacks and defenses in cognitive radio networks. In *Proc. of IPCCC* (2009).
[11] Cheng, Y.-C., et al. Automating cross-layer diagnosis of enterprise wireless networks. In *Proc. of SIGCOMM* (2007).
[12] Chin, W.-L., et al. Channel-based detection of primary user emulation attacks in cognitive radios. In *Proc. of VTC* (2012).
[13] Connor, J. T., Martin, R. D., and Atlas, L. E. Recurrent neural networks and robust time series prediction. *IEEE Trans. on Neural Net. 5*, 2 (1994), 240–254.
[14] Denisowski, P. Recognizing and resolving LTE/CATV interference issues. *White Paper, Rohde and Schwarz* (2011).
[15] Feng, Q., et al. Anomaly detection of spectrum in wireless communication via deep auto-encoders. *The Journal of Supercomputing* (2017).
[16] Guo, T., et al. Robust online time series prediction with recurrent neural networks. In *Proc. of DSAA* (2016).
[17] Gurbani, V. K., et al. Detecting and predicting outages in mobile networks with log data. In *Proc. of ICC* (2017).
[18] Hochreiter, S., and Schmidhuber, J. Long short-term memory. *Neural Computation 9*, 8 (1997), 1735–1780.
[19] Iyer, A. P., Li, L. E., and Stoica, I. Automating diagnosis of cellular radio access network problems. In *Proc. of MobiCom* (2017).
[20] Kaligineedi, P., Khabbazian, M., and Bhargava, V. K. Malicious user detection in a cognitive radio cooperative sensing system. *IEEE Trans. on Wireless Comm. 9*, 8 (2010), 2488–2497.
[21] Ke, J. Shanghai wants law on radio spectrum. Shine.cn, March 2018. https://www.shine.cn/news/metro/1803061282/.
[22] Koushik, A., Bentley, E., Hu, F., and Kumar, S. A hardware testbed for learning-based spectrum handoff in cognitive radio networks. *Journal of Network and Computer Applications 106* (2018).
[23] Li, R., et al. Tact: A transfer actor-critic learning framework for energy saving in cellular radio access networks. *Trans. on Wireless Communications 13* (2014).
[24] Liu, S., Chen, Y., Trappe, W., and Greenstein, L. J. Aldo: An anomaly detection framework for dynamic spectrum access networks. In *Proc. of INFOCOM* (2009).
[25] Liu, S., Greenstein, L. J., Trappe, W., and Chen, Y. Detecting anomalous spectrum usage in dynamic spectrum access networks. *Ad Hoc Networks 10*, 5 (2012), 831–844.
[26] Liu, Y., Ning, P., and Dai, H. Authenticating primary users' signals in cognitive radio networks via integrated cryptographic and wireless link signatures. In *Proc. of S&P* (2010).
[27] McGraw, I., et al. Personalized speech recognition on mobile devices. In *Proc. of ICASSP* (2016).
[28] Mehra, R. K., and Peschon, J. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica 7*, 5 (1971), 637–640.
[29] Min, A. W., Kim, K.-H., and Shin, K. G. Robust cooperative sensing via state estimation in cognitive radio networks. In *Proc. of DySPAN* (2011).
[30] Nika, A., et al. Empirical validation of commodity spectrum monitoring. In *Proc. of SenSys* (2016).
[31] Nováczki, S. An improved anomaly detection and diagnosis framework for mobile network operators. In *Proc. of DRCN* (2013).
[32] Onireti, O., et al. A cell outage management framework for dense heterogeneous networks. *IEEE Trans. on Vehicular Technology 65*, 4 (April 2016).
[33] O'Shea, T. J., Clancy, T. C., and McGwier, R. W. Recurrent neural radio anomaly detection. *arXiv preprint arXiv:1611.00301* (2016).
[34] Pan, S. J., and Yang, Q. A survey on transfer learning. *Trans. on knowledge and data engineering* (2010).
[35] Rajendran, S., et al. Distributed deep learning models for wireless signal classification with low-cost spectrum sensors. *arXiv preprint arXiv:1707.08908* (2017).
[36] Selim, A., et al. Spectrum monitoring for radar bands using deep convolutional neural networks. *arXiv preprint arXiv:1705.00462* (2017).
[37] Sheth, A., Doerr, C., Grunwald, D., Han, R., and Sicker, D. Mojo: A distributed physical layer anomaly detection system for 802.11 wlans. In *Proc. of MobiSys* (2006).
[38] Tan, K., Zeng, K., Wu, D., and Mohapatra, P. Detecting spectrum misuse in wireless networks. In *Proc. of MASS* (2012).
[39] Viavi. Interference hunting is fundamental to quality service. RCR Wireless News, Nov. 2016. https://www.rcrwireless.com/20161101/network-infrastructure/interference-hunting-fundamental-quality-service.
[40] Wei, H., Jia, Y., and Wang, L. Spectrum anomalies autonomous detection in cognitive radio using hidden markov models. In *Proc. of IAEAC* (2015).
[41] Xie, X., and Wang, W. Detecting primary user emulation attacks in cognitive radio networks via physical layer network coding. *Procedia Computer Science 21* (2013), 430–435.
[42] Zhang, L., Ding, G., Wu, Q., and Han, Z. Spectrum sensing under spectrum misuse behaviors: A multi-hypothesis test perspective. *IEEE Trans. on Information Forensics and Security 13*, 4 (2018).