AirPress: High-accuracy spectrum summarization using compressed scans

Mariya Zheleva[1], Timothy Larock[1], Paul Schmitt[2], and Petko Bogdanov[1]

[1]Department of Computer Science, University at Albany SUNY, {*mzheleva,tlarock,pbogdanov*}*@albany.edu*
[2]Department of Computer Science, Princeton University, *pschmitt@cs.princeton.edu*

***Abstract*—Spectrum summarization is the analysis of a wide-band spectrum scan to determine the number of transmitters, their time-frequency characteristics, approximate modulation and legitimacy of operation. Spectrum summarization has emerged as a critical functionality to enable next-generation dynamic spectrum access technologies and legislation. Typically, spectrum summarization is performed in a cloud-based manner, requiring full-scan transmission from the spectrum sensors to the cloud. As spectrum scans generate large volumes of data, full-scan transmission quickly incurs prohibitively-high cost in terms of bandwidth and storage requirements. To address this problem we design AirPress, a spectrum scan compression method that leverages wavelet decomposition for lossy compression of spectrum data and allows up to 64:1 compression ratio of power spectral density traces without adversely impacting the spectrum summarization accuracy. We demonstrate the utility of AirPress on real-world spectrum measurements and show that it enables high-accuracy spectrum summarization of real-world transmitters while reducing the corresponding trace by 94%.**

I. Introduction

Dynamic Spectrum Access (DSA) has been a heavily-researched technology for next generation mobile wireless connectivity. Tangible progress towards DSA, however, hinges on deep understanding of spectrum use in support of DSA technology, spectrum policy and spectrum enforcement. To this end, industry, academia and the government have endeavored to establish an agenda for next-generation spectrum measurement infrastructures [1]. A recent survey on spectrum measurement objectives [2] identified a wide range of priorities. Spectrum measurements (i) should help incumbents and secondary users to make real-time decisions for spectrum use, (ii) should support validation of analytical methods and protocols, (iii) should assist in spectrum enforcement and (iv) should be able to serve multiple objectives. Thus, there is a need for a spectrum measurement infrastructure that can provide continuous spatial coverage of measurements, store scans longitudinally and summarize the spectrum occupancy including number of transmitters, their temporal and frequency patterns and the opportunity they grant for secondary access.

There exist several spectrum databases [3], [4], [5] but they are limited to TV bands and only provide information for occupied and idle channels. The FCC Spectrum Dashboard [6] covers a larger frequency range from 225MHz to 3.7GHz and provides information about spectrum allocation and assignment. However, it lacks real-time information about spectrum occupancy. The systems that best satisfy the spectrum inventory requirements are Microsoft's Spectrum Observatory [7] and CityScape [8], which monitor the spectrum from 30MHz to 6GHz and provide real-time spectrum occupancy information. Their spatial coverage, however, is limited to several locations and they do not provide detailed spectrum summarization. Practical challenges faced by a real-time spectrum inventory are related to storage and analysis of raw spectrum data. For example, a one second scan of a 600MHz spectrum band with a USRP sampling at 20Msps amounts to 23GB [9]. This large volume of data poses challenges in scalable storage and summarization of spectrum information.

To enable spectrum inventory at scale we propose AirPress, a method that compresses raw spectrum traces and thus enables large-scale spectrum scan collection, storage and processing. AirPress makes use of wavelet decomposition for lossy compression and, depending on the signal complexity, achieves up to 64:1 compression rate while maintaining small error rates. We demonstrate the utility of AirPress by analyzing controlled Wi-Fi and Bluetooth transmissions and a real-world wideband spectrum scan from 400MHz to 1GHz. Our analysis shows that different bands tolerate different compression levels, which creates an opportunity for adaptive compression towards a scalable spectrum inventory.

This paper makes several contributions: (i) we design AirPress, a spectrum compression technique that reduces the volume of spectrum scans by up to 94%, while preserving signal properties; (ii) we demonstrate that AirPress retains accurate spectrum summarization; and (iii) we harness AirPress to map spectrum compressibility across wideband spectrum measurements and show that compressability depends on the spectrum dynamics.

II. AirPress

In order to characterize the spectrum occupancy at a given location and enable advanced usage of available bands, we need high-resolution scans that preserve essential signal characteristics. While there is a variety of scanning sensors that allow different sampling rates and scan bandwidths, they all have the capability to produce large amounts of raw measurements. If stored in raw format, the data will quickly exceed the storage capabilities of the sensing node. Furthermore, this data generation rate is prohibitive for real-time offloading to a spectrum inventory database. Hence, a natural question arises: *Can the raw measurements be compressed in a manner that preserves the underlying spectrum characteristics, while reducing the overhead for storage and analysis?*

We propose to address the storage challenge using wavelet-based compression. Wavelets are a useful mathematical tool for hierarchical decomposition of signals that are efficient to compute and enable accurate signal reconstruction. Wavelets have been employed in diverse domains including image analysis [10], databases [11] and in wireless networks research for wideband spectrum sensing [12]. In AirPress, we adopt a one dimensional wavelet decomposition applied to a scan of signal power $p^t(f)$ over a range of frequencies at a given time instant t, where p is a function over n discrete frequency values. The shape of $p^t(f)$ has local regularities as transmissions correspond to contiguous frequency regions of constant power modulo noise and empty bands correspond to noise-level power. The *Haar* wavelet transform is a good choice for decomposing such impulse-like signals, and hence we focus on this basis.

The wavelet decomposition of a power scan $p^t(f)$ is a mapping from the n-dimensional original signal $p^t(f)$ to a set of coefficients w^t that correspond to summaries of the signal at different resolutions. The full decomposition w^t has n coefficients and can be used to reconstruct the original signal $p^t(f)$ exactly. Due to the local regularities of the signal many of the coefficients are close to zero and a lossy reconstruction can be obtained by maintaining only a synopsis of the decomposition $\tilde{w}^t$ containing a subset of k coefficients, i.e. $|\tilde{w}^t| = k$. A similar approach has been adopted for approximate query answering in databases [10]. One can also show that for a budget of k coefficients to compute a synopsis, keeping the coefficients of largest absolute value is optimal when minimizing the sum of squared error between the reconstructed and original signal [11].

Our compression approach applies a wavelet transform of the original scan $p^t(f)$ and computes a synopsis $\tilde{w}^t$ of pre-specified size k that can be used to reconstruct an approximation of the original scan $\tilde{p}^t$, and answer various queries regarding occupancy. The savings in storage in our scheme are k/n-fold, i.e. the compression rate is k/n. Of note is that while more aggressive compression leads to drastic reduction in spectrum scan size and smoother signals, it can also eliminate some of the original signal properties. Thus, a tradeoff exists between compression level and the truthfulness of the reconstructed signal. We explore this tradeoff in § III.

III. Experimental Evaluation

We now evaluate the proposed spectrum compression approach based on both controlled scans of a single transmitter and a wide-spectrum scan in an urban area. We evaluate the distortion of the reconstructed signal with respect to the original (in terms of sum of squared error) for increasing number of budget coefficients. We also quantify the effects of compression on the detection of transmitters from raw measurements.

Data and implementation. We use two USRP setups to collect traces for the purpose of this evaluation. The first consists of a USRP B210 with an MP 08-ANT-0860 antenna and a Lenovo Thinkpad X230 with an Intel core i7 CPU and 16GB of RAM to collect controlled Wi-Fi and Bluetooth transmissions. The second consists of a USRP N200 with a WBX daughterboard, an LP0410 antenna and a Lenovo ThinkPad X250 laptop with Intel i7 CPU and 8GB of RAM. This setup ran a custom Gnuradio script to collect a wideband trace from 400MHz to 1GHz. We collect three spectrum datasets. The first two are from individual Wi-Fi and Bluetooth transmissions. We select an interference-free environment and transfer a large file first over Wi-Fi (Channel 40) and then over the lower 32MHz of the Bluetooth band. We also use a wideband spectrum scan that was collected in a dense downtown area. The speed of wideband scans is limited by the bandwidth of the scanning radio. Thus, we scan the spectrum from 400MHz to 1GHz in 20MHz steps using a sampling rate of 20Msps. In order to avoid scalloping [13] loss we create a 75% overlap between consecutive steps. Of note is that even with such large overlap there were still some scalloping artifacts in the scan, which we further reduce by applying window-based smoothing. In order to ensure preservation of the signal properties the used window size was a fraction of 1/3000 of the scan size. We implement AirPress in Java and run our experiments on a Lenovo ThinkPad X250 with an Intel CORE i7 CPU and 8GB of RAM, installed with KUbuntu 15.04.

A. Error incurred by spectrum compression

For this experiment we use our Wi-Fi and Bluetooth scans, for which we calculate the Power Spectral Density (PSD) with an FFT size of 1024. In order to quantify their compressibility, we vary the number of coefficients used in the compression and report the resulting *Sum of Squared Errors (SSE)* of the signal reconstruction with respect to the original.

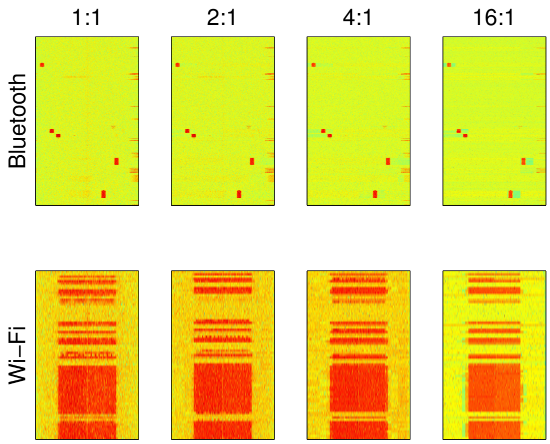
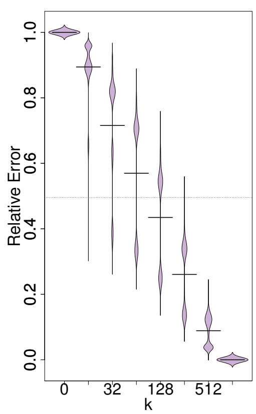
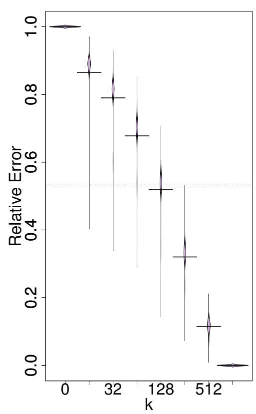


Fig. 1: Power over time (vertical axis) in a spectrum band (horizontal axis) corresponding to Bluetooth frequencies (top row) and Wi-Fi (bottom). Red color corresponds to high power while yellow to low power. The first column shows the temporal spectrum state before compression and the remaining columns after compression for increasing compression rates between 2:1 and 16:1.



(a) Wi-Fi



(b) Bluetooth

Fig. 2: Relative error reduction for increasing number of coefficients k in BT and Wi-Fi. The beanplots for every setting of k show the distribution of relative errors of compressing all snapshots in our scans.

Figure 1 shows the distortion of the original signal (first column) for increasing compression rates, i.e. decreasing number of coefficients used for the synopsis. Each figure is a heatmap of the power level (red:high, yellow:low) in time (vertical) and frequency (horizontal) space. The first row is from the Bluetooth band while the second from the Wi-Fi band. In both cases the actual transmitter shapes (rectangular red blocks) are well preserved; even for 16:1 compression ratio i.e. selecting the top $1/16$-th of all possible coefficients.

Figure 2 presents beanplots of the relative error of compression for increasing number of coefficients. We calculate the relative error as the fraction of SSE with a given number of coefficients and the SSE when representing the whole snapshot as its average, i.e. one coefficient. The horizontal lines in each bean show the

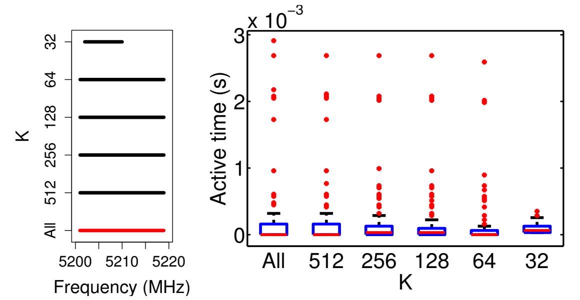


Fig. 3: (left) Bandwidth detection. The bandwidth of the Wi-Fi transmitter is successfully detected as we decrease the number of coefficients to 64. Past 64 coefficient the signal reconstruction deviates significantly from the original signal, which results in false bandwidth detection. (right) Active time detection. The average active time and the active time distribution is successfully detected as we decrease the coefficients to 64. False reconstruction of the signal with 32 coefficients result in inaccurate active time detection.

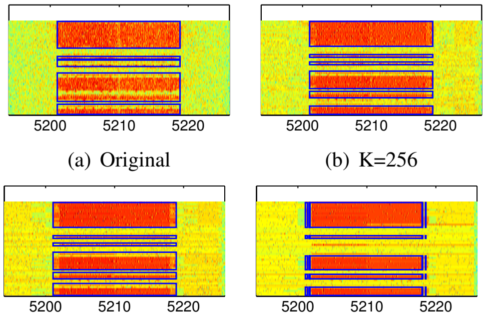
average relative error for that compression, whereas the beans show the distributions of relative error for each compression. The distribution of errors is bimodal in the case of Wi-Fi with the low-error mode corresponding to no-transmission snapshots and the high-error mode to transmission snapshots. In both the BT and Wi-Fi scans, the average relative error decreases as the number of coefficients grows. Also, reducing the data by an order of magnitude (from 1024 to 128) results in less than half the error compared to when all the data is represented as its average. The wide spread of error is promising for adaptive compression schemes that vary the number of coefficients depending on the state of the spectrum.

While these results bring insight into the SSE incurred by spectrum compression, they still do not answer the question of how does scan compression affect spectrum summarization. *What relative error can we allow before we begin loosing important details about the underlying signal characteristics?* We tackle this question more comprehensively in the following section, in which we quantify the effect of compression on transmission detection and behavior summarization.

B. *Effects of spectrum compression on summarization*

In this section we demonstrate AirPress's ability to preserve signal quality and enable detailed transmitter identification while performing a 16-fold spectrum scan compression. For this experiment we use the Wi-Fi spectrum scan. In order to summarize the spectrum use we leverage TxMiner [14], which is an unsupervised method for spectrum characterization.

We focus on detection of transmitter bandwidth and active time. The Wi-Fi transmission in question takes place in channel 40, which spans 20MHz between 5200MHz and 5220MHz. The transmission utilizes multiple consecutive time chunks of variable duration. As Figure 3(left) shows, we are able to successfully detect the bandwidth as the number of coefficients decreases



(a) Original (b) K=256 (c) K=128 (d) K=64

Fig. 4: Qualitative evaluation of transmitter detection with decreasing number of coefficients.

to 64. At 32 coefficients the signal reconstruction begins to differ drastically in comparison with the original, which leads to false bandwidth detection. Figure 3(right) presents results for active time detection with decreasing coefficients. Each boxplot presents the distribution of active times detected at the corresponding compression rate. As we can see, the mean of the detected active times remains unchanged as the coefficients decrease to 64. We see some of the outliers disappear and be replaced with several smaller outliers. This means that as we denoise the data with increasing compression some of the individual active intervals become more pronounced and are detected as separate intervals. As the number of coefficients reaches 32 the active time distribution changes dramatically, indicating a false detection of transmitter activity.

Lastly, we present an illustration of the Wi-Fi transmitter detection in Figure 4. The figure presents an annotated heatmap of the Wi-Fi transmission in question, where blue rectangles enclose the areas detected as active. As we can see, the same frequency band is detected as occupied across the different compression scenarios. The active time detection changes as we decrease the number of coefficients. We see that in the original data some active intevlas are detected together. As we apply compression with 256 and 128 coefficients, the pause between these intervals becomes less noisy and the detector is now able to identify them as separate active periods. As the number of coefficients decrease to 64 the edges of transmission become thinner, which results in failure to identify some short active periods and in fuzzy detection of the signal edges in frequency.

C. Wide-band spectrum compression

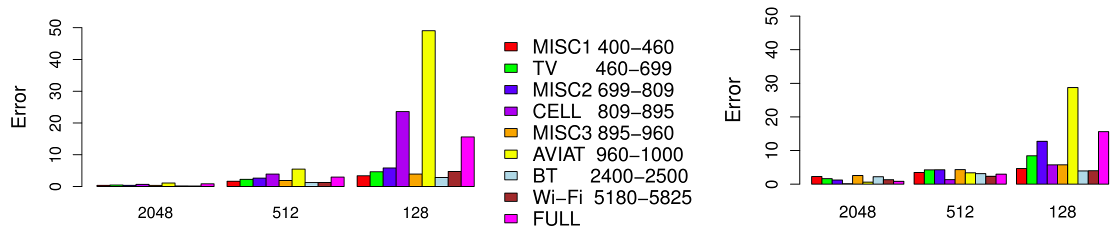
In order to demonstrate the utility of AirPress in creating a large-scale spectrum inventory we evaluate the compressibility of spectrum bands across a wide frequency range. We collect a spectrum scan from 400MHz to 1GHz with a step of 20MHz and a duration of 1 second. We add to this scan our traces from the controlled Wi-Fi and Bluetooth transmissions. We split the spectrum scan in several sub-bands according to the FCC's spectrum allocation charts as shown in Figure 5. We note that the miscellaneous (MISC) bands host a variety of technologies including air navigation, maritime, amateur radio, Earth exploration satellites, public safety, family radio and narrow-band PCS [6].

For this experiment we select a decreasing number of coefficients for each run. During each run we first compress the entire frequency band with the corresponding number of coefficients. We call this full-scan compression. We then redistribute the total number of coefficients to the individual sub-bands and compress these sub-bands separately. We call this split compression. We redistribute the coefficients based on two criteria. The first criteria is the sub-band size in frequency; that is larger bands will get a proportionally larger fraction of the coefficients. The second criteria is signal variance; that is, bands in which the signal varies more drastically will be budgeted with a larger number of coefficients. The goal of this coefficient budgeting is to demonstrate which factor influences spectrum scan compression: the size of the data or its richness.

Figure 5 presents our results for sub-band size (5(a)) and signal variance (5(b)). The y-axis presents average sum of squared error (Error) between the original and reconstructed signal for each sub-band, where reconstruction is obtained with 2048, 512 and 128 coefficients. As expected, for all sub-bands the error increases as we decrease the number of coefficients. Richer sub-bands such as Aviation and Cellular are less compressible when the objective is error minimization. Furthermore, the variance-based budgeting of coefficients leads to better compression than the size-based budgeting. Lastly, we note that the error from full-scan compression is smaller than the cumulative error from split compression. The reason for this is that there are idle fractions in each sub-band, which have to be assigned the same coefficients multiple times in the split compression. At the same time in the full-scan compression these idle fractions get assigned the same coefficients only once, which results in better compressibility and more efficient distribution of coefficients to dynamic bands. Our future analysis will study this hypothesis and design spectrum compression principles that regroup spectrum data based on expected activity as opposed to regulatory allocation.

IV. Related Work

AirPress provides a wavelet-based spectrum scan compression methodology that reduces the storage and upload bandwidth requirements, while retaining high-quality spectrum summarization. Related work to AirPress falls in three categories: (i) spectrum summarization, (ii) compressive spectrum sensing and (iii) application of wavelets to spectrum sensing.



(a) Sub-band size budgeting

(b) Signal variance budgeting

Fig. 5: Compression of the full spectrum (FULL) and individual functional bands using the same coefficient budget. We see that variance-based budgeting of coefficients leads to a better compression than size-based budgeting. The error of full-band compression is smaller than the cumulative error of split compression.

Spectrum summarization approaches develop unsupervised algorithms to determine spectrum utilization patterns [15] and to extract the number of active transmitters and their time-frequency activity [14] from wideband spectrum scans. These methods assume full spectrum scans and their feasibility has not been demonstrated in conjunction with spectrum compression.

Compressive sensing utilizes sub-Nyquist sampling to detect idle spectrum [16], [17], [18]. AirPress aims to reduce the upload and storage footprint of already collected scans and, thus, differs in goal and methodology from existing work. None of the existing work is targeted at detailed spectrum summarization; instead it identifies idle and occupied bands without providing further insight into number of transmitters and their time-frequency characteristics.

Applications of wavelets in spectrum sensing. Wavelet transforms were previously used by Tian et al. [12] for wideband sensing in order to identify available spectrum holes. Conversely, AirPress uses wavelets to compress spectrum information for use in a spectrum inventory. We store inherent time-frequency properties of individual transmitters, rather than identifying holes.

V. Discussion and Conclusion

We presented AirPress – the first method for wideband spectrum scan compression. AirPress enables spectrum inventory at scale, while preserving detailed signal characteristics for comprehensive analysis of spectrum occupancy. AirPress enables the design of new generation spectrum inventories that use *adaptive compression* to budget the amount of data they admit from various spectrum sensors. While the current prototype of AirPress makes use of 1D signal compression, we see possibility for improvement by the use of 2D wavelet decomposition as it may be better able to maintain characteristics in both the time and frequency domain. Furthermore, the current version of AirPress does not take into account historic information about spectrum compression, however, *delta-encoding* based on change over time would improve compresibility. Lastly, we believe that we can achieve improved transmitter characterization directly in the wavelet coefficient domain. Thus, while this paper presents several new and exiting results about spectrum compressibility, there are many more problems that remain unexplored and promise to bring forward the state-of-the-art in spectrum analysis at scale.

References

[1] "NSF Workshop on Spectrum Measurement Infrastructures." http://www.cs.albany.edu/~mariya/nsf_smsmw/, April, 2016.
[2] M. McHenry, "Spectrum measurements requirements survey." http://www.cs.albany.edu/~mariya/nsf_smsmw/docs/NSF_SMIW_survey_results.pdf, April 6-7 2016.
[3] "Spectrum Bridge ." https://spectrumbridge.com/.
[4] "TVFool." http://tvfool.com/.
[5] "Google Spectrum Database." https://www.google.com/get/spectrumdatabase/.
[6] "FCC Spectrum Dashboard ." http://reboot.fcc.gov/reform/systems/spectrum-dashboard.
[7] "Microsoft Spectrum Observatory ." https://observatory.microsoftspectrum.com/.
[8] S. Roy, K. Shin, A. Ashok, M. McHenry, G. Vigil, S. Kannam, and D. Aragon, "Cityscape: A metro-area spectrum observatory," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–9, July 2017.
[9] M. Zheleva, R. Chandra, A. Chowdhery, P. Garnett, A. Gupta, A. Kapoor, and M. Valerio, "Enabling a nationwide radio frequency inventory using the spectrum observatory," *IEEE Transactions on Mobile Computing*, 2017.
[10] K. Chakrabarti, M. Garofalakis, R. Rastogi, and K. Shim, "Approximate query processing using wavelets," *The VLDB JournalThe International Journal on Very Large Data Bases*, vol. 10, no. 2-3, pp. 199–223, 2001.
[11] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin, *Wavelets for computer graphics: theory and applications*. M. Kaufmann, '96.
[12] Z. Tian and G. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," in *CROWCOM*, (Mykonos Island, Greece), June 2006.
[13] R. Lyons, "Reducing fft scalloping loss errors without multiplication [dsp tips and tricks]," *Signal Processing Magazine, IEEE*, vol. 28, pp. 112–116, March 2011.
[14] M. Zheleva, A. Chowdhery, R. Chandra, A. Kapoor, and P. Garnett, "TxMiner: Identifying Transmitters in Real-World Spectrum Measurements," in *IEEE DySPAN'15*, (Stockholm, Sweden), 29 September - 2 October 2015.
[15] L. Shi, P. Bahl, and D. Katabi, "Beyond sensing: Multi-ghz realtime spectrum analytics," in *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15)*, (Oakland, CA), pp. 159–172, USENIX Association, May 2015.
[16] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *IEEE ICASSP'07*, April 2007.
[17] E. Candes and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, pp. 21–30, March 2008.
[18] J. Laska, W. Bradley, T. Rondeau, K. Nolan, and B. Vigoda, "Compressive sensing for dynamic spectrum access networks: Techniques and tradeoffs," in *DySPAN*, May 2011.