Applications of Rényi Entropy in Time-Frequency Representation Analysis and Signal Classification

Brandon W. Ardisson

December 3, 2023

1 Abstract

This work aims to introduce the generalized Rényi entropy and show how such an entropy measure can be used to characterize signal complexity through the consideration of time-frequency representations (TFR) as pseudo probability density functions. Properties of the specific parameterization of Rényi entropy that make it applicable to a wide range of TFRs will be discussed. To tie this concept in to practical application, an example will be presented, where researchers have demonstrated the ability to create signal signatures derived from such information measures and use these signatures to perform signal classification with high accuracy and precision. A summary of application performance along with a review of both trends and headwinds to this approach will be provided.

2 Introduction

Applying analysis concepts from information theory to complex non-stationary signal analysis is somewhat arduous procedure, but also quite intriguing. Entropy, often thought an abstract figure-of-merit, generally can be used to "rank" the level of complexity sourced from the system under inspection. Condensing the large and intricate analysis space traditional to non-stationary signal analysis, down to a single descriptive parameter provides means for relatively trivial implementation of machine learning classification systems. Understanding how engineers and scientists may better apply alternative, scalable, methods to the analysis of naturally occurring signals provides the motivation for this review. The application example, later discussed, particularly highlights how the methodology presented can be applied to fields that are ripe for improved processing efficiency and acceleration. At a high level, this topic should serve well due to its relevance to recent trends in machine learning/artificial intelligence, network traffic analysis, and cybersecurity.

3 TFR as PDFs

TFRs help to generalize the concept of time and frequency domains by facilitating the production of a combined time-frequency function, $C_s(t, f)$, that depicts how the frequency content of a signal, $s(t)$, changes over time. Conceptually, TFRs can be visualized as a musical score, a map of where "notes" of various frequency are spread in time. The analysis signal, $C_s(t, f)$, maps input signal energy to the time-frequency plane. $C_s(t, f)$ is obtained by through convolution of the input kernel function, $\Phi(t, f)$, with for example W_s, the Wigner distribution (WD), a specific TFR of the input signal $s(t)$ as depicted in Equation (1) [1].

$$C_s(t, f) \coloneqq \iint W_s(u, v)\Phi(t - u, f - v)\, du\, dv \coloneqq (W_s * \Phi)(t, f) \tag{1}$$

The kernel function, $\Phi(t, f)$, in an elementary manner, is a description of the time-frequency filter used to sample the signal $s(t)$. The kernel function is analogous to that of the window function commonly applied in discrete Fourier transform (DFT) analysis. The Wigner distribution, given by Equation (2), belongs to a class of TFR that is particularly useful in application, called the Cohen class of quadratic time-frequency representations (QTFRs). The WD is always real-valued, preserves time and frequency shifts, and satisfies the marginal properties corresponding to the analysis signal's instantaneous power, Equation (3), and spectral energy density, Equation (4) [2].

$$W_s(t, f) = \int_{-\infty}^{\infty} s\left(t + \frac{\tau}{2}\right)\ s^*\left(t - \frac{\tau}{2}\right)\ e^{-j2\pi\tau f}\, d\tau\,. \tag{2}$$

$$\int_{-\infty}^{\infty} C_s(t, f)\, df = p_x(t) = |s(t)|^2 \tag{3}$$

$$\int_{-\infty}^{\infty} C_s(t, f)\, dt = P_x(f) = |S(f)|^2 \tag{4}$$

Cohen class TFRs have properties that are very similar to the properties of probability density functions (PDF). Exploiting this similarity, TFRs applied to bivariate time-frequency signals can be scrutinized using methods reserved for information theory and more general probabilistic analysis. A particular fundamental property of generalized PDFs is shown by Equation (5), which indicates that the sum of all the individual probabilities (think components) of $p(x)$, accounts for the complete probability mass distribution over the domain $\forall\ x$ [3].

$$\int_{-\infty}^{\infty} p(x)\,dx = 1 \tag{5}$$

Cohen class TFRs aim to combine the marginal (1 dimensional) density concepts of instantaneous power, Equation (3), in the time domain, and spectral energy density, Equation (4), in the frequency domain. An outcome of this is that signal energy can be derived by integrating the subject TFR over the entire time-frequency plane as shown in Equation (6).

$$E_s = \int_{-\infty}^{\infty} |s(t)|^2\,dt = \int_{-\infty}^{\infty} |S(f)|^2\,df = \iint_V C_s(t,f)\,dt\,df \tag{6}$$

Inspection and comparison of Equation (5) and Equation (6) indicate a prominent parallel between TFRs and PDFs. All signal energy, and/or probable outcomes, are preserved across the integration domain.

4 Introduction to Rényi Entropy

William's et. al. in [4] continue to develop the argument for appropriation of information measures to the Cohen class TFRs by analyzing the well known Shannon entropy, Equation (7), and Rényi entropy, Equation (8), and their respective TFR marginal properties mentioned previously in Equations (3) and (4).

$$H(C_s) = -log_2 \iint C_s(t,f)\,dt\,df. \tag{7}$$

$$H_a(C_s) = \frac{1}{(1-\alpha)} log_2 \iint C_s^{\alpha}(t,f)\,dt\,df. \tag{8}$$

The well known Shannon entropy from information theory is a parameterization of the lesser known Rényi entropy. Note that Equation (8) to the second order, $\alpha \to 2$, reduces to that of Shannon entropy. Applying entropy analysis to TFRs introduces a unique way to characterize the underlying time-frequency signal. Certain properties of the Rényi entropy, later derived, add additional signal analysis capabilities to this methodology.

William's et. al. [4] provide an ideal signal stimulus for their analysis, a Gabor logon, or equivalently known as a Gaussian pulse, described by Equation (9). The Gaussian pulse is an ideal elementary signal element, it is not limited in the time or frequency domain, and both the product and convolution of 2 Gaussian pulse elements, results in a Gaussian element.

$$s(t) = \frac{1}{\sqrt[4]{\pi\sigma^2}} \exp(\frac{-t^2}{2\sigma^2}) \tag{9}$$

Evaluation of the time marginal, $|s(t)|^2$, and the frequency marginal, $|S(f)|^2$ are normalized and equal to 1. The same result is achieved when evaluating the Wigner distribution of the Gaussian test signal. Further, more rigorous evaluation of the Rényi entropy of the same components, results in a proof that shows the Wigner distribution of the stimulus provides no additional information beyond what is provided solely by the time and frequency marginals [4]. Due to this conservation, it is proposed that, as long as the underlying TFR is information invariant (information invariance as a result of time-frequency shifts and scaling), the entropy information measures are certified for use in TFR analysis. One of the objective observations of this certification is that the TFR information measure provides an assessment of the "number of equal energy resolvable elementary signals present" [4] in a TFR.

A demonstration of this result can be shown by evaluating the Rényi entropy, and for comparison a more traditional signal measure, the time-bandwidth product (TBP). For this demonstration 2 Gaussian pulse elements that are separated in time, but collocated in frequency, will be used for stimulus as depicted in Figure 1. In this initial experiment the time separation between the pulses is then stepped (increased) from 0 to several standard deviations in time spacing while the signal analysis measures are observed. The MATLAB code use to produce this plot is given in Appendix A.

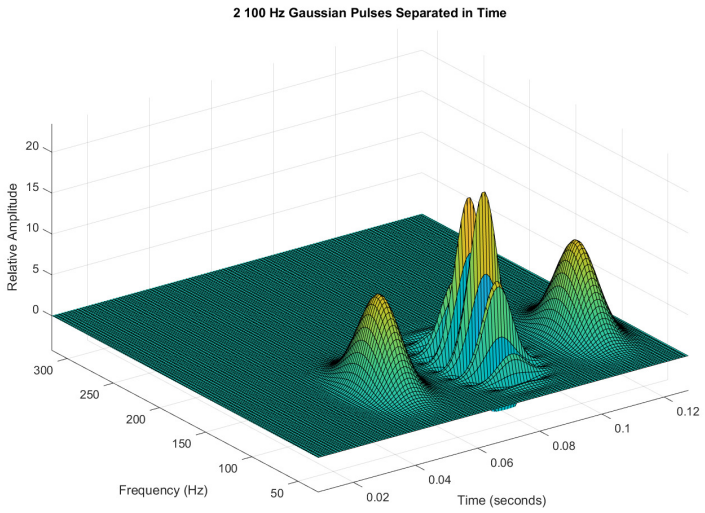

Figure 1: Gaussian pulse stimulus elements and cross-term interference component, both signals at $f = 100$ Hz and separated by a varying time delta as depicted in Figure 2.

Figure 2 shows that as the time separation is increased beyond 2 standard deviations (0.1 seconds), the information present levels off at approximately 1 bit (1.1 bits actual) above the 0 time separation level. The information gain is then, 1 bit, this is expected for $n = 1 \rightarrow 2$ signal elements since total information

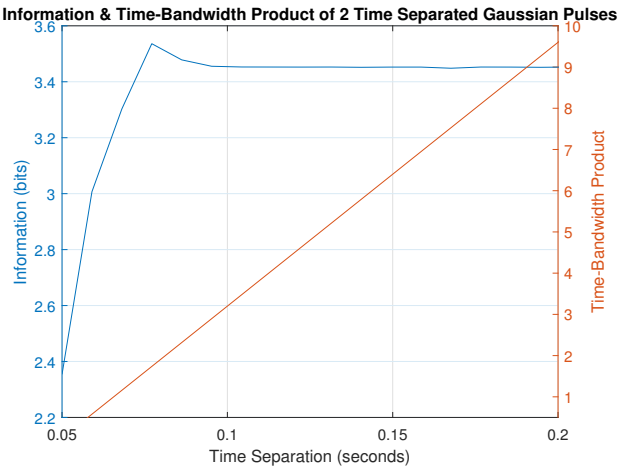

Figure 2: Rényi entropy and TBP, calculated from multiple Wigner distributions (WD) of 2 Gaussian pulses separated by various time deltas; simulated via MATLAB's Wavelet entropy function and with utilization of the TFR toolbox. See Appendix A for more detail.

is $log_2 2^n$. Meanwhile the TBP monotonically increases without bound, giving no information about the number of signals present. The same results are achieved, in the dual case, in which the signals are collocated in time, and the frequency variable is stepped in the same manner.

The approximation in experiment, and the information over estimation that occurs at 1 standard deviations in Figure 2 can be attributed to the cross-term interference common to QTFRs, also somewhat influenced by the choice of TFR. This is just one example of how the effects of cross-term interference influences overall TFR selection. We will see this consideration occur again in practical implementations discussed in the following sections.

5 Why Rényi Entropy?

Shannon entropy would be a natural candidate for this application - providing small entropy figures for signals of elementary components, and larger entropy values for diffuse TFRs with many complicated signals. Why then have we chosen to discuss Rényi entropy instead of the much more ubiquitous Shannon entropy? The choice of Rényi entropy for TFR signals analysis lies in the mathematical subtleties. In fact one of the more obvious violations is due to the presence of the logarithm operator and the fact that many TFRs, specifically Cohen class QTFRs, take on negative values [1].

Baranuik et. al. have shown that for each odd parameterization of Rényi entropy, $\alpha \geq 3$, there exist signals for which (8) is not defined due to violation of the analysis signal positivity requirement, Equation (10). Other parameterizations such as, non-integer values of α yield complex $C_s^\alpha(t, f)$ and are of limited utility. Integer orders of positive $C_s^\alpha(t, f)$ remain positive and do not present a hazard [1]. It is easy to show

that for sufficiently large, odd α, the positivity criterion given by Equation (10) fails. Uniquely, the 3rd order parameterization, $\alpha = 3$, has a substantial collection of numerical evidence indicating that 3rd order entropies are well defined for a large class of signals and TFRs.

$$\iint C_s^{\alpha}(t, f)\, dt\, df > 0 \tag{10}$$

Considerable effort has been put forth to find a signal for which Equation (10) fails when $\alpha = 3$. When a failing case presents, a small amount of Gaussian smoothing of the analysis TFR, W_s, is generally enough for (10) to stand. The preponderance of numerical data indicates that the examples for which the positivity requirement are violated when $\alpha = 3$ are quite exceptional [1]. A counterexample can be formed through a particular linear combination of Hermite functions [1], but this example is quite contrived. Users should be aware that corner cases do occur, but for the most part 3rd order Rényi entropy is suitable to quadratic, Cohen class, TFR signals analysis.

Apart from opening up the TFR input domain, Rényi entropy of the 3rd order (H_3), has other properties that make its use even more appropriate to our use case. One of these properties is the ability to perform *component counting*. If TFRs were "quasi-linear" – such that the resultant TFR would be a linear superposition of the elementary signal elements, then the analogy between TFRs and PDFs would predict an additive, or counting, behavior from Rényi entropy as seen in the prior demonstration.

Derivation and an analytic example that bolster the outcomes implied by Figure 2 begin by introducing, an ideal, a quasi-linear TFR, $I_s(t, f)$. This ideal TFR is compactly supported over some interval J_σ where $I_s(t, f) = 0$ for all $t \notin J_\sigma$ and for all f. For some 2 component signal $s + \tau s$, where $(\tau s)(t) \coloneqq s(t - \Delta t)$ represents a time shift by Δt, so long as $\Delta t > \sigma$ then we can express the TFR distribution as shown in Equation (11).

$$I_{s+\tau s}(t, f) = I_s(t, f) + I_s(t - \Delta t, f) \tag{11}$$

Rényi postulated that the Shannon entropy uniquely satisfies the axioms of symmetry, continuity, normalization, additivity, and the mean value condition of entropy, given by Equation (12) [5]. We can use this mean value condition (12) to determine the entropy content of the resultant TFR $I_{s+\tau s}(t, f)$ by considering the compactly supported TFR components $I_s(t, f)$ and $I_s(t-\Delta t, f)$ represented by $\{p, q\}$, as discrete probability distributions. The sum of all values of p, and q, over their support domain is indicated by $w(p)$ and $w(q)$. H^R is the normalized Rényi entropy of it's input argument, such that $H^R(I_{s+\tau s}(t, f)) \to H^R(p \cup q)$.

$$H^R(p \cup q) = \frac{w(p)H^R(p) + w(q)H^R(q)}{w(p) + w(q)} \tag{12}$$

Rényi questioned what would occur if the arithmetic mean implied by Equation (12) was replaced by some other mean value, leading to the introduction of a strictly *monotonic* and *continuous* function $y = m(x)$ modifying the mean behavior of Equation (12) to as is shown in Equation (13).

$$H^R(p \cup q) = m^{-1}(\frac{w(p)m[H^R(p)] + w(q)m[H^R(q)]}{w(p) + w(q)}) \tag{13}$$

Selection of an exponential function parameterized by the order α depicted by Equation (14) results in the following functional derivation of the normalized Rényi entropy of the order α, Equation (15). This outcome indicates that Rényi entropy behaves very similar to Shannon entropy, with only the mean value condition being relaxed from an arithmetic to exponential mean [1].

$$m(x) = 2^{(\alpha-1)x}\,; \qquad \alpha > 0, \alpha \neq 1 \tag{14}$$

$$H^R_\alpha(p) := \frac{1}{(1-\alpha)} log_2 \frac{\sum_i p_i^\alpha}{\sum_i p_i} \tag{15}$$

We can close out our 2 component ideal TFR example, by substituting $H^R_\alpha(p) = H_\alpha(p) - log_2||p||_2^2$ into Equation (13), and by recognizing that the ideal TFR signal energy is equal for the time shifted component, i.e.: $H_\alpha(I_{\tau s}) = H_\alpha(I_s)$ and $||s + \tau s||_2^2 = 2||s||_2^2$ we can algebraically resolve Equation (13) to Equation (16) [1].

$$H_\alpha(I_{s+\tau s}) = H_\alpha(I_s) + 1 \tag{16}$$

This result directly correlates with the previous example. Indicating that the 2 component signal yields exactly 1 additional bit of information over the single component signal.

Cross-components were ignored in the previous example. Cross components are particularalry troublesome, since they violate the linearity assumption made via our ideal TFR in (11) and undermine the accuracy of entropy data in general. However, this does not present an issue in application, since H_α is *cross-component invariant* under certain analysis conditions. Baranuik et. al. show that with the supporting time interval, σ, setting the component signal time spacing $\Delta t > \frac{1}{2}(\alpha + 1)\sigma$ results in cross-component cancellation [1].

In summary, the *component counting* property of H_3 holds when the input signals are sufficiently spaced in the time domain, and such that signal cross-components do not coincide with auto-components or other cross-components [1]. Additional properties of H_3 not detailed here include the amplitude and phase sensitivity to input components, and cooperation with common smoothing functions.

6 Application

One practical application of Rényi entropy for TFR analysis was presented by Marnerides et. al., titled *Internet Traffic Classification using Energy Time-Frequency Distributions* [6]. In this work, the authors exploit the concepts discussed this far, applying Rényi entropy analysis to wide-area network (WAN) traffic packet captures. In this application, TFRs are used to map aggregate transport layer volume information, such as byte and packet throughput, to the time-frequency (TF) plane. When entropy analysis is performed on this TF mapped volume data series, unique signatures can be formed to classify the type of data traffic, such as, HTTP, FTP, and DoS/DDoS/attack. We will discuss this example in more detail since it highlights both positives and negatives of the entropy based TFR analysis approach, as well as, recognizes current trends in cybersecurity, data analysis, and machine learning.

Firstly, the authors perform stationarity testing. This is an essential first step, as the outcome determines whether the data being analyzed calls for time-frequency representation analysis, or whether more traditional *wide-sense stationary* signal processing methods can be applied. Let $s(t)$, represent the packet or byte count time series. We can then derive the analytic signal $s_a(t)$ (17), through application of the Hilbert transform (18).

$$s_a(t) = s(t) + j s_h(t) \tag{17}$$

$$H\{s(t)\} = \frac{1}{\pi}\, p.v. \int_{-\infty}^{\infty} \frac{s(t)}{t-\tau}\, d\tau \tag{18}$$

Nonstationary, or time-variant, signals must have an instantaneous frequency (19), and/or amplitude, $|s_a(t)|$, that varies with time.

$$f(t) = \frac{1}{2\pi} \frac{d \arg\, s_a(t)}{dt} \tag{19}$$

Amplitude in this case represents that magnitude change in the byte or packet count at a specific time t. The amplitude of the frequency observed in one instance of a byte/packet arrival at time t is denoted given by $f(t)$. Time-series group delay, derived via application of the Fourier transform to $s_a(t)$, (20) is

also monitored, since it indicates the local time behavior of the frequency function.

$$t_s(\omega) = \frac{1}{2\pi} \frac{darg\, \mathcal{F}\{s_a(t)\}}{d\omega} \tag{20}$$

Stationarity test results taken from [6] are displayed in Figure 3. The instantaneous frequency and group delay for both a transport layer byte-series, TCP, and an application layer protocol, FTP, are shown.

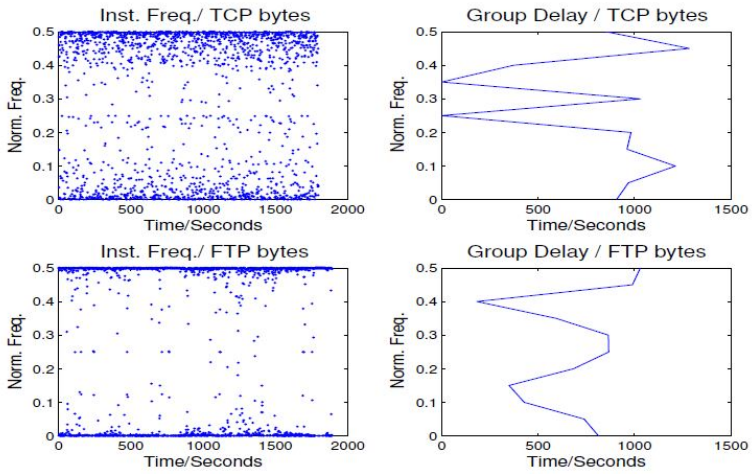

Figure 3: Stationarity analysis of TCP and FTP byte-series traffic, taken from [6].

Both the instantaneous frequency and group delay analyses indicate a non-linear behavior and indeed are time variant. This result indicates that the data being analyzed is non-stationary and suitable for TFR inspection. Additionally, the bi-modal behavior of the instantaneous frequency chart indicate that the signals being inspected have a multi-component nature [6].

Secondly, a suitable selection of Cohen class TFRs was selected and evaluated for performance on the test data. As has already been seen in Figure 2, where cross-component interference led to an overshoot in the entropy magnitude, TFR selection can have a deciding impact on the accuracy and representation of the data series in the TF plane. This is primarily due to the levels of auto- and cross- component noise leakage, as well as spectral amplitude accuracy of the desired signal. Chosen for analysis were the Wigner distribution (WD), Equation (2), the Smoothed Pseudo Wigner-Ville (SPWV), Equation (21), and the Choi-Williams (CW) distributions. The SPWV distribution is formed from the WD, by applying independent time and frequency windows, collectively represented by $\zeta(\tau)$, which is typically Gaussian in nature [2]. These added features work to attenuate auto- and cross- component interference and maintain a positive distribution [7].

$$SPWV_s(t,f) = \int_{-\infty}^{\infty} \zeta(\tau)\, s\left(t+\frac{\tau}{2}\right)\, s^*\left(t-\frac{\tau}{2}\right)\, e^{-j2\pi\tau f}\, d\tau \tag{21}$$

Contrasting the WD, and the SPWV, the CW TFR considers signal $s(t)$ to be comprised of multiple components, such that $s(t) = \sum_{a=1}^{N} s_a(t)$. The Choi-Williams TFR, given by Equation (22), is particularly well suited for mitigating spurious values contributed by cross-terms prevalent in multi-component signals. The CW distribution behaves remarkably well in satisfying intuitive notions of where signal energy should be concentrated, all while reducing spurious component interference and maintaining the common desirable properties of TFR distributions [7].

$$CW(t,f) = \frac{1}{4\pi^{\frac{3}{2}}} \iint \frac{1}{\sqrt{\frac{\tau^2}{\epsilon}}}\, e^{-[\frac{(u-\tau)^2}{(\frac{4\tau^2}{\epsilon})}]-j\tau\omega}\, s^*\left(u-\frac{\tau}{2}\right)\, s\left(u+\frac{\tau}{2}\right)\, e^{-j2\pi\tau f}\, du\, d\tau \tag{22}$$

Each of the 3 TFRs under consideration were evaluated using sample application layer FTP byte series captures, shown in Figure 4, is the WD time-frequency representation of this series. Due to the magnitude of the byte values, the WD TFR shows several components (due to auto- and cross- terms) that are not actually present in the sampled data set.

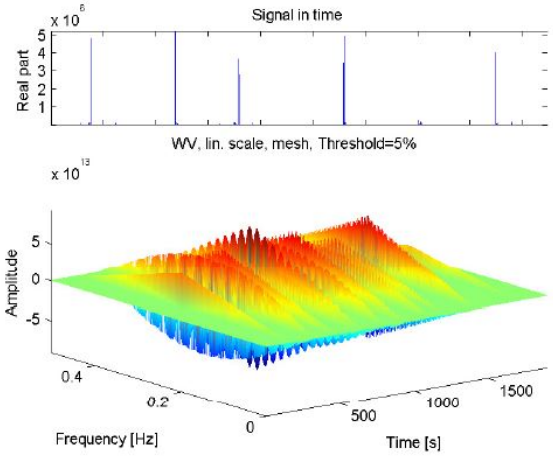

Figure 4: Time series and WD TFR of FTP byte sequence, taken from [6].

Figure 5 shows the Smoothed Pseudo Wigner-Ville TFR analysis of the same data set. The SPWV TFR improves on the WD by indicating the correct transmissions at the exact time. However, each of these TFRs have an amplitude-frequency relationship that does not correctly represent the true data. The triangular shaped components have centralized amplitude peaks that would imply that normalized middle range frequencies are associated with high amplitudes, whereas low and high frequencies are not.

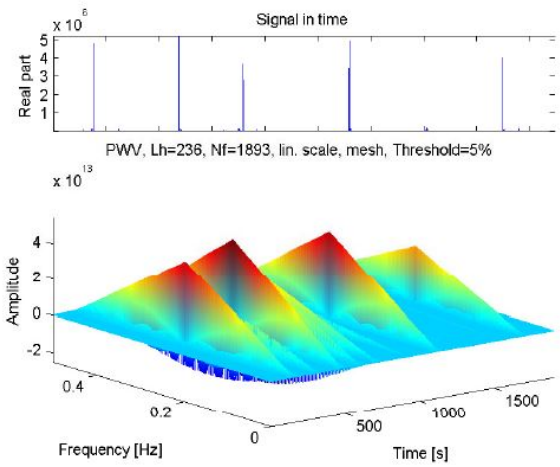

Figure 5: Time series and PSWV TFR of FTP byte sequence, taken from [6].

In Figure 6, the Choi-Williams TFR is used to analyze the same data set once more. Replacing the triangular component entries as seen in the previous figures, are now more constant amplitude spectral indicators. This TFR more accurately represents the amplitude and frequency relationship that is expected from high volume traffic flows that are commonly present on backbone wide area network (WAN) links [6].

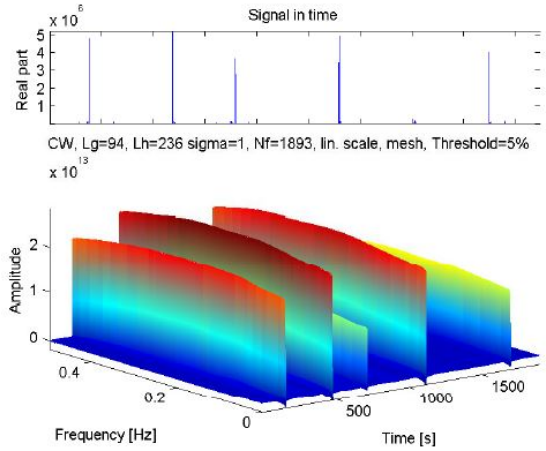

Figure 6: Time series and CW TFR of FTP byte sequence, taken from [6].

For the reasons discussed, the Choi-Williams TFR was selected as the primary TFR to be used for data analysis moving forward.

Thirdly, the authors trained a decision tree classification model based on traffic flow signatures derived through entropy analysis of real-world internet traffic. In this process both packet and byte segmented traffic flows were captured from the Widely Integrated Distributed Environment (WIDE) trace and the Keio trace. The WIDE trace is a US-Japan Trans-Pacific backbone link which carries internet traffic for WIDE

member organizations which work closely with the Internet Engineering Task Force (IETF) to develop and foster internet technology evolution. The Keio trace captures originate from the academic network at Keio University's Shonan-Fujisawa campus [6].

A supervised machine learning algorithm, the C4.5 decision tree classification scheme, was implemented [8]. This scheme initially requires a training data set as part of the learning process. For this training set, a segmented section of the larger Keio trace, and a temporal subset of the WIDE trace were selected. Additional samples of these captured traffic flows were classified through the Center for Applied Internet Data Analysis (CAIDA) maintained, CoralReef payload classifier. CoralReef is a port based traffic classification method that is still impressive in identifying legacy applications, further strengthened through the use of packet size and TCP flag attributes [8]. The CoralReef results were used as "ground-truth" data which is used as feedback during the initial training procedure [6].

Finally, the specific traffic flows (i.e. FTP, HTTP, SSH, SMTP, etc.) were extracted from the WIDE and Keio trace data to formulate the testing data set which will serve as input to the now trained C4.5 decision tree. The process, largely outlined in this paper, includes sampling the volumetric data from each test flow (packet/byte per unit time), mapping this to the time-frequency plane via the Choi-Williams TFR, followed by the computation of the Rényi information. Classification accuracy and precision were then compared between the newly computed Rényi information and the information values from the training set to evaluate system performance.

7 Results and Simulation Discussion

Figure 7 provides the accuracy results from the C4.5 decision tree classification model for both trace captures, differentiated by packet and byte segmented test sets.

Trace/Vol. feature	Accuracy
Keio-Bytes	92.7%
Keio-Packets	96.3%
WIDE-Bytes	88.6%
WIDE-Packets	91.8%

Figure 7: Rényi entropy based C4.5 decision tree classification accuracy results for various test data sets, taken from [6].

Interesting results are observed when comparing the packet based and byte based test set accuracy results. In both traffic captures the packet segmented data sets had a higher level of classification accuracy. Nonethe-less, the results shown in Figure 7 clearly indicate that Rényi information can be successfully used as a classification feature on real world non-stationary time series'. The higher accuracy result for packet segmented data flows is likely attributed by the fact that some application flows have standardize packet

sizes (number of bytes per packet) but not a fixed packet count. The packet size standardization would limit the entropy of the data set and reduce the classifier's efficacy [6].

Trace	Method	No. Features	Overall Accuracy
Keio	Energy_TF-D4.5(bytes)	1	92.7%
	Energy_TF-D4.5(packets)	1	96.3%
	BLINC	5	89.4%
	SVM	6-10	98+%
WIDE	Energy_TF-D4.5(bytes)	1	88.6%
	Energy_TF-D4.5(packets)	1	91.8%
	BLINC	5	58%
	SVM	6-10	98+%

Figure 8: A comparison of traffic classification methods from both trace data sets, differentiated by classification features and overall accuracy, taken from [6].

Figure 8 depicts the performance of the discussed example, when compared against other traffic classification methods available. Perhaps one of the prominent benefits to using the time-frequency based Rényi entropy classification method is due to the limited number of traffic features required to attain a similar degree of accuracy. This limits the amount of traffic flow pre-processing that must be completed to condition the data ahead of queuing the flow for classification. The reduction in count of necessary analysis features could make this approach well suited for edge applications that do not have the same computing power as network data centers for example. The TLS/SSL encrypted traffic results were not broken out in [6], though they would be interesting to analyze further due to the proliferation of end-to-end traffic encryption standards such as TLS 1.3, commonly used by the HTTPS appliation protocol.

A final comment is provided regarding the previously mentioned information overestimation from the first example, referenced by Figure 2. The results did not agree exactly with theory due to the cross-component influence on the measured entropy values. We can conclude that the time spacing between the input component Gaussian pulses was not sufficient for the cross-terms to diminish to zero, one of the properties attributed to Rényi entropy of the third order. This is primarly due to the fact that Gaussaian pulses have non-zero values across an infinite time and frequency domain, resulting in non-compact support. Non-compactly supported signal's auto- and cross- components will always have some degree of overlap (see Equation 23 in [1] for more detail on the asymptotic displacement vector).

8 Conclusion

The presentation and utilization of time-frequency representations, entropy analysis, and machine learning, to characterize real-world non-stationary datasets is an area that is ripe for exploration. Network acceleration and security can be improved through alternative classification methods and data-path segregation. We have

shown how non-stationary signals can be evaluated through the use of well understood information theory structures. Although not without pitfalls, Rényi entropy can be used successfully on real non-stationary data so long as the proper checks are confirmed a priori. The method and application presented had the highest accuracy per feature ratio of the alternative classification methods reviewed in [6]. Future work could include exploring edge implementations such as on an FPGA, evaluation of SSL/TLS traffic analysis, or regular expression pattern matching capabilities requested by network vendors looking to accelerate deep packet inspection (DPI).

9 Acknowledgments

I'd like to acknowledge the authors and maintainers of the open-source Time-Frequency Toolbox (https://tftb.nongnu.org/) which was used for data analysis and experiments shown in this work.

References

[1] R. Baraniuk, P. Flandrin, A. Janssen, and O. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1391–1409, 2001.

[2] F. Hlawatsch and G. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.

[3] T. M. Cover and J. A. Thomas, *Elements of information theory*, ser. Wiley Series in Telecommunications. New York: Wiley, 1991.

[4] W. J. Williams, M. L. Brown, and A. O. H. III, "Uncertainty, information, and time-frequency distributions," in *Advanced Signal Processing Algorithms, Architectures, and Implementations II*, vol. 1566, 1991, pp. 144–156. [Online]. Available: https://doi.org/10.1117/12.49818

[5] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, vol. 4. University of California Press, 1961, pp. 547–562.

[6] A. K. Marnerides, D. P. Pezaros, H.-c. Kim, and D. Hutchison, "Internet traffic classification using energy time-frequency distributions," in *IEEE International Conference on Communications (ICC)*, 2013, pp. 2513–2518.

[7] L. Cohen, "Time-frequency distribution - a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

[8] H. Kim, K. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, "Internet traffic classification demystified: myths, caveats, and the best practices," in *Proceedings of 2008 ACM CoNEXT Conference - 4th International Conference on Emerging Networking EXperiments and Technologies, CoNEXT '08*. ACM, 2008, pp. 1–12.

Appendices

A Entropy and TBP Graph Appendix

```
% clean up environment
clc; clear all; close all;

%% define sampling environment
time_scale_factor = 2;
time_samples = time_scale_factor * 128;
sample_frequency = 4 * time_samples; % prevent aliasing in TF domain
sample_period = 1/sample_frequency;

max_time = (time_samples - 1)*sample_period; % seconds
max_frequency = sample_frequency/2; % Hz, Nyquist rate

%% setup time steps to loop over
first_time_sample = 0.05;
time_steps = 23; % select this depending on resolution wanted
time_vector = linspace(first_time_sample, max_time, time_steps);
mean(time_vector);
std(time_vector);
% transform time samples to unit std deviation and mean = 0
% time_vector_t = (time_vector - mean(time_vector))/std(time_vector);
% mean(time_vector_t);
% std(time_vector_t);
% time_step = (time_vector_t(7)-time_vector_t(6))/time_steps;
```

```
% time_vector = first_time_sample:time_step:max_time; % time vector in
   unit std deviation steps

%% setup static frequency vector
center_frequency = 100; % Hz
k1 = center_frequency/sample_frequency;
k2 = k1;

%% gaussian pulse characteristics
% pulse 1
td1 = 0.05; % seconds, base time location of pulse 1
ntd1 = ceil(td1/sample_period); % index location in time vector

pulse_bw = time_samples/(8*time_scale_factor);
pulse_bw_hz = pulse_bw * sample_frequency/time_samples;
pulse_amplitude = 1;

%% create results array
swept_entropy = zeros(1, length(time_vector));
swept_tbp = zeros(1, length(time_vector));

%% loop
for i = 1:length(time_vector)

    ntd2 = ceil(time_vector(i)/sample_period); % index location in time
        vector
    par=[ntd1,k1,pulse_bw,pulse_amplitude;ntd2,k2,pulse_bw,
       pulse_amplitude]; % Parameter vector
    x=atoms(time_samples,par,0); % Generate sum x(t) of 2 gaussian
       pulses
    Wx= tfrwv(x); % Compute WD of x(t)
    t=linspace(0,max_time,time_samples);f=linspace(0,max_frequency,
```

```
        time_samples); % TF axis in s and Hz

    ent = wentropy(Wx, Entropy="Renyi", Scaled=0, Exponent=3); %
        calculate Renyi entropy

    bw = pulse_bw_hz;
    time = abs(td1 - time_vector(i));
    timeBandwidthProduct = bw*time;
    max_ent = max(ent); % grab this highest value in the array
    swept_entropy(i) = max_ent; % update entropy vector
    swept_tbp(i) = timeBandwidthProduct; % update tbp vector

end

%% generate plot
figure()
yyaxis left
plot(time_vector, swept_entropy)
hold on
xlabel('Time Separation (seconds)')
ylabel('Information (bits)')
hold on
yyaxis right
plot(time_vector, swept_tbp);
ylim([0.5, 10]);
xlim([0.05, 0.2]);
ylabel('Time-Bandwidth Product')
grid on
title('Information & Time-Bandwidth Product of 2 Time Separated
    Gaussian Pulses')
```