Measuring Complexity and Predictability in Networks with Multiscale Entropy Analysis

Janne Riihijärvi, Matthias Wellens and Petri Mähönen
Department of Wireless Networks, RWTH Aachen University
Kackertstrasse 9, D-52072 Aachen, Germany
email: {jar, mwe, pma}@mobnets.rwth-aachen.de

Abstract—We propose to use multiscale entropy analysis in characterisation of network traffic and spectrum usage. We show that with such analysis one can quantify complexity and predictability of measured traces in widely varying timescales. We also explicitly compare the results from entropy analysis to classical characterisations of scaling and self-similarity in time series by means of fractal dimension and the Hurst parameter. Our results show that the used entropy analysis indeed complements these measures, being able to uncover new information from traffic traces and time series models. We illustrate the application of these techniques both on time series models and on measured traffic traces of different types. As potential applications of entropy analysis in the networking area, we highlight and discuss anomaly detection and validation of traffic models. In particular, we show that anomalous network traffic can have significantly lower complexity than ordinary traffic, and that commonly used traffic and time series models have different entropy structures compared to the studied traffic traces. We also show that the entropy metrics can be applied to the analysis of wireless communication and networks. We point out that entropy metrics can improve the understanding of how spectrum usage changes over time and can be used to enhance the efficiency of dynamic spectrum access networks.

I. INTRODUCTION

Time series analysis has become one of the cornerstones of characterization and modelling of network traffic. Ever since the discovery of different variants of scaling or self-similar behaviour in traffic measurements [1] the research community has been very active in developing new analysis techniques and improved models. Automated detection of network anomalies has also become a highly active research field in its own right, and has started to adopt many state-of-the-art statistical techniques for automated reasoning about traffic conditions [2], [3]. However, one aspect that has received relatively little attention so far has been the *predictability* or quantifiable *complexity* of traffic and related models.

In this paper we apply a new class of *complexity metrics* for network traffic characterisation. The complexity of traffic is here understood as the difficulty in predicting its future behaviour, as formally measured by multi-symbol entropies. While entropies of various probability distributions associated to network traffic have been extensively studied and applied in the literature, almost all of this work has focused solely on the single-symbol, distributional properties of different traffic characteristics (for examples, see [3]–[6]). Such a treatment has already yielded very interesting results (see, for example, [7] for an application in anomaly detection),

but necessarily ignores large amount of information possibly present in the empirical time series (for a related discussion in an information-theoretic context, see [8]). We show through numerous examples that the introduced metrics can indeed uncover several types of structure from observed traffic, and offer complementary information to the classical metrics of time series analysis.

In order to demonstrate that these complexity metrics provide new information about time series, we explicitly compare the behaviour of multiscale entropies to usual metrics quantifying scaling and self-similarity. We carry out this comparison using both synthetic data traces generated from a number of well-known discrete time series models as well as measured traffic traces. The latter part is an extension of our earlier work [9]. We also discuss the applications of these complexity metrics, focusing in particular on validation of synthetic traffic models and anomaly detection. Based on experimental data we show that there is a significant difference in the complexity of anomalous network traffic compared to regular aggregate traffic observed at a router of an Internet service provider. In addition to the classical applications in traffic analysis, we also study measured traces of *wireless spectrum usage* for self-similarity and compare our findings to results from multiscale entropy analysis. We specifically show that the methods applied here provide insight in the difference between noise and interference and are capable of discovering reoccurring structures in the data.

The rest of the paper is structured as follows. In Section II we recall the necessary background on self-similarity and scaling behaviour, including a discussion on the fractal dimension and its relation to classical definitions of self-similarity. In Section III we introduce multiscale entropy analysis in detail, which is then applied on various time series models in Section IV. We then proceed by studying the complexity of different network traffic traces in Section V, also commenting on applications in anomaly detection and model validation. As another interesting application scenario we consider the analysis of spectrum usage in Section VI with an eye towards dynamic spectrum access and cognitive radios. Finally, we draw the conclusions in Section VII.

II. HURST PARAMETER AND FRACTAL DIMENSION

We shall begin with a background overview of self-similarity for time-series analysis. Since this material has

become standard, we shall be rather brief, and focus mainly on conventions and notation. The interested reader should consult many of the available overviews for further details [10]–[12].

Let now $u(t)$, $t \in \mathbb{N}$ be our time series, that is, a discrete-time real- or integer-valued stochastic process. We can usually distinguish between two broad classes of such processes arising in the applications. First class would consist of processes satisfying some kind of stationarity condition, and would correspond to, for example, traces of interarrival times or traffic volumes measured of windows of constant length. The second class would enumerate *cumulative* quantities, such as the total number of data received, the arrival times of individual packets, and so on. Following [11], we distinguish between these two cases also in notation, reserving $v(t)$, $t \in \mathbb{N}$ for processes of the cumulative type, and $u(t) = v(t) - v(t-1)$ for the corresponding *increment process*. Processes of the cumulative type clearly cannot be stationary, but the increment processes are often taken to be. The particular form of stationarity usually assumed is that of *weak stationarity*, meaning that $u(t)$ is taken to have constant mean, finite variance, and translation-invariant autocovariance. With these assumptions we can write the autocovariance of u as a function of the *lag* k only, namely as $\gamma_u(k)$. Finally, we define the *aggregated* or *coarse-grained* process $u^{(\tau)}$ by

$$u^{(\tau)}(t) \equiv \frac{1}{\tau} \sum_{i=\tau(t-1)+1}^{\tau t} u(i). \tag{1}$$

We use the same notation to distinguish between characteristics of the original time series $u(t)$ and the coarse-grained one. For example, the autocovariance function of $u^{(\tau)}(t)$ will be denoted by $\gamma_u^{(\tau)}(k)$.

There are several flavours of scaling behaviour or self-similarity that we shall now recall. Classically, a process $v(t)$, $t \in \mathbb{R}$ is called *self-similar with parameter H* if $v(at)$ and $a^H v(t)$ have the same finite-dimensional distributions. If in addition the increment process $u(t)$ of $v(t)$ is stationary, $v(t)$ is said to be *self-similar with stationary increments*. In applications weaker forms of self-similarity usually suffice, most common of these being different flavours of second-order self-similarity [11]. Given $H \in (1/2, 1)$, $u(t)$ is called *exactly second-order self similar* if for all $k \geq 1$

$$\gamma_u(k) = \frac{\sigma^2}{2}\big((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\big), \tag{2}$$

where σ^2 is the variance of $u(t)$. Similarly, $u(t)$ is called *asymptotically second-order self-similar* if the limit $\lim_{\tau \to \infty} \gamma_u^{(\tau)}(k)$ has the above form. For a demonstration how second-order self-similarity arises from the scaling behaviour of the increment process of a self-similar process with stationary increments, see [11]. Closely related to second-order self-similarity is *long-range dependence*, characterised by the non-summability $\sum_k r_u(k) = \infty$ of the autocorrelation function $r_u(k) \equiv \gamma_u(k)/\sigma^2$. Using (2) we see that the sum over $r_u(k)$ of a second-order self-similar process diverges precisely when $H \in (1/2, 1)$. There are, however, processes

that feature long-range dependence without being self-similar by the above definitions. Similarly, processes exist that are self-similar but not long-range dependent. However, since in applications second-order self-similarity has been the most prevalent, self-similarity and long-range dependence are often taken to be synonymous in the literature.

Another aspect of self-similar behaviour is the so-called *fractal dimension* [13], [14]. While fractal dimension can be defined for arbitrary sets, we shall focus here on the case relevant to time-series analysis, namely the fractal dimension of 1-dimensional curves, such as graphs of functions of a single variable. Let us denote the graph of a function f by $G_f \equiv \big\{ \big(x, f(x)\big) \,\big|\, x \in d(f) \big\}$, where $d(f)$ denotes the domain of f. The *Minkowski cover* of G_f is defined as

$$G_f(\varepsilon) \equiv \big\{ y \in \mathbb{R}^2 \,\big|\, y \in B_\varepsilon(x), x \in G_f \big\}, \tag{3}$$

where $B_\varepsilon(x)$ denotes a disc of radius ε centred on x. Then, denoting by $|A|$ the area of $A \subset \mathbb{R}^2$, the *Minkowski-Bouligand* dimension of G_f is defined as

$$D(G_f) \equiv \lim_{\varepsilon \to 0} \left(2 - \frac{\ln |G_f(\varepsilon)|}{\ln \varepsilon} \right). \tag{4}$$

Thus, the fractal dimension can be said to characterise self-similar behaviour in smaller scales.

Fractal dimension D and the Hurst parameter H are not, of course, entirely unrelated. For a large class of processes, called *self-affine*, they are related by a simple linear relationship $D + H = 2$. However, there also exist simple Gaussian processes that allow D and H to be specified separately. We shall use some of such models below, to demonstrate that D, H and the entropy metrics introduced in the next section are indeed generally independent characteristics of $u(t)$.

III. Modern entropy metrics

In this section we give an overview of the different entropy metrics applied in the rest of the paper. We begin by briefly recalling how Shannon's information entropy is commonly used for time series analysis. We then describe approximate and sample entropies in detail which explicitly take the temporal structure of the data into account. Finally, we describe how multiscale entropy analysis is carried out starting from the basic sample entropy.

A. Classical information entropy

For a discrete random variable X, taking values $\{x_1, \ldots, x_n\}$ with probabilities $\{p_1, \ldots, p_n\}$ we say that the *information content* of an outcome x_i is $-\log p_i$. The *information entropy* or *Shannon entropy* [15] of X is then defined as the mean information content, yielding $S(X) \equiv -\sum_{i=1}^n p_i \log p_i$ (we avoid the usual notation $H(X)$ to avoid confusion with the Hurst parameter). The choice of the base of the logarithm amounts to choice of units. In this paper we have used the natural units obtained by the use of natural logarithm. The entropy of a time series $u(t)$ is usually defined by calculating S for the empirical probabilities obtained from the time series (or in the case of continuous values being measured, on

some normalized histogram). Such an application characterises the shape of the distribution of the values $u(t)$ only, and does not use the temporal structure possibly present in the series.

B. Approximate and sample entropies

The above application of information entropy can be extended to characterise information content in the temporal structure as well by considering the distribution of *sequences* of values of $u(t)$. Particularly interesting example of such an approach is the *approximate entropy* statistic as defined by Pincus [16] extending earlier work by Eckman and Ruelle on approximating the Kolmogorov-Sinai entropy of a time series [17]. Given a positive constant $m \in N$, we form the vectors $\mathbf{x}(t) \equiv \big(u(t), \ldots, u(t + m - 1)\big)$. Each of these vectors has m consecutive values from our original time series starting from $u(t)$. Between two vectors $\mathbf{x}(s)$ and $\mathbf{x}(t)$ we define the distance $d\big(\mathbf{x}(s), \mathbf{x}(t)\big)$ as the maximum of the absolute values of the component-wise differences. We then count "similar" vectors by defining

$$C_s^m(r) \equiv \frac{1}{N - m + 1} \#\{t \mid d\big(\mathbf{x}(s), \mathbf{x}(t)\big) \leq r\}, \quad (5)$$

where N denotes the length of the sequence $u(t)$. The parameter $r \geq 0$ specifies the tolerance for two sequences to be considered similar, and is usually taken to be some multiple of the standard deviation of $u(t)$. Next, we introduce

$$\Phi^m(r) \equiv \frac{1}{N - m + 1} \sum_{s=1}^{N-m+1} \ln C_s^m(r), \quad (6)$$

and, finally, define the *approximate entropy* by

$$\mathrm{ApEn}(m, r) \equiv \lim_{N \to \infty} \big(\Phi^m(r) - \Phi^{m+1}(r)\big). \quad (7)$$

The approximate entropy $\mathrm{ApEn}(m, r)$ can be interpreted as the mean information content of the conditional probability for two subsequences of $u(t)$ that are similar at $m - 1$ points continue to be similar for all of m points. This is the connection to information entropy alluded to in the above.

The definition in terms of the infinite limit gives a parameter for describing the underlying process the time series is sampled from. For practical purposes this can be approximated by the statistics

$$\mathrm{ApEn}(m, r, N) \equiv \Phi^m(r) - \Phi^{m+1}(r). \quad (8)$$

Interpretation of the values of these statistics mirrors that of information entropy. Value of zero indicates complete regularity at the length scale of m observations, whereas higher values indicate higher complexity.

Richman and Moorman [18] have shown that ApEn suffers from certain statistical shortcomings, and have proposed an alternative, called the *sample entropy*[1], denoted as $\mathrm{SampEn}(m, r, N)$. They observed that ApEn-statistics are biased since in the calculation of the $C_s^m(r)$ self-matches of

the template vectors $\mathbf{x}(t)$ are included. In order to correct for this bias, they instead defined empirical probabilities $A^m(r)$ and $B^m(r)$ of $m + 1$ and m matches to the template vector, respectively, *excluding self-matches*, and defined the sample entropy statistic as $-\ln\big(A^m(r)/B^m(r)\big)$. The interpretation of all the parameters and values is similar to the ApEn case, but these definitions remove the bias from ApEn, and will yield more consistent estimators especially if only a short sample of $u(t)$ is available (the improvement is especially significant for $N < 10^4$ or so).

C. Multiscale entropy analysis

Varying m in the approximate and sample entropies allows the study of complexity at different timescales, measured as the number of successive samples considered. Unfortunately the amount of data needed to perform such an analysis scales exponentially as a function of m, making large-m analysis impractical. Entropy analysis on multiple time scales can, however, be carried out by computing the above statistics for the coarse-grained time series $u^{(\tau)}(t)$. This yields the multiscale entropy analysis of Costa *et al.* [19], [20]. We shall show below that such an analysis can indeed uncover structures occuring at a wide variety of length scales, making entropy analysis an interesting addition to the techniques available for studying and characterising network traffic.

IV. EXAMPLES FOR ARTIFICIAL DATA

We shall now apply the above metrics for a variety of processes, together with estimators for the Hurst parameter and the fractal dimension. Our objectives in this section are two-fold. First, we demonstrate that entropy metrics are indeed distinct from characterisations of self-similarity and fractal behaviour, and are sensitive to different types of structure in the process. Second, we try to give the reader a better intuitive understanding on the behaviour of the various metrics and parameters via suitably chosen examples. The processes studied are the logistic map, white Gaussian noise, fractional Gaussian Noise, the Cauchy class, and selected fractals.

The logistic map is defined by the iteration $x_{n+1} = Rx_n(1 - x_n)$, where R controls the behaviour of the samples. In particular it is known that for small enough R the samples either converge to a single fixed point ($R \leq 3$) or oscillate between few values ($3 < R < 3.54$). For larger R the behaviour of the samples becomes increasingly chaotic and more and more complex structures emerge.

Fractional Gaussian noise is simply the Gaussian process with autocorrelation function obtained from (2). Cauchy class is likewise a family of Gaussian processes, but this time with the correlation function

$$r(k) = (1 + |k|^\alpha)^{-\beta/\alpha}, \quad (9)$$

where $\alpha \in (0, 2]$ and $\beta > 0$. The realisations of the Cauchy class can be shown to have fractal dimension of $D = 2 - \alpha/2$ and Hurst parameter $H = 1 - \beta/2$, allowing the effects of D and H to be studied separately [21]. Generation of samples from these Gaussian processes was done by means of

---

[1]This term should not be confused with the calculation of the information entropy from the empirical distribution, which is also occasionally called sample entropy in the literature.

circulant embedding, an exact sampling algorithm introduced in [22], [23]. For comparison, we also cross-checked our results with fractional Gaussian noise generated by means of Paxson's algorithm [24]. Finally, we selected two classical fractals as examples. The first one is commonly referred to as the Kiesswetter curve and the second one is based on the Weierstrass-Mandelbrot function. Both were introduced and used as reference series for the estimation of the fractal dimension in [14].

The estimators for approximate and sample entropies were implemented in Matlab, together with the code required for multiscale entropy analysis. We used $r = 0.15\,\sigma\big(u(t)\big)$ throughout following the recommendations from the literature, although the results were confirmed not to be sensitive to small changes in r. For estimation of the Hurst parameter we used the wavelet-based estimator of Abry and Veitch [25], and for estimating the fractal dimension we used the variation method of Dubuc *et al.* [14]. In all cases the performance of the different estimators were cross-checked with the analytical results and numerical results available in the literature to ensure correct operation of the implementations. For the wavelet estimator of H we used the implementation of Veitch [26]. In contrast to several other Hurst parameter estimators, this method may also result in estimates $H \notin (1/2, 1)$. These estimates strongly indicate the lack of scaling and should not be considered during further analysis.

The results for Shannon entropy, SampEn, Hurst parameter and the fractal dimension for the considered models are given in Table I. First, for the logistic map we see that SampEn characterises correctly the reduction in complexity as R is reduced. Especially in the case $R = 3.5$ the complete predictability of the oscillatory behaviour reduced the SampEn to zero. As expected, the single-symbol Shannon entropy does not distinguish between $R = 3.9$ and $R = 3.7$, since the distribution of the values is close to uniform in both cases. The results for Hurst parameter and fractal dimension are also consistent with the lack of scaling or self-similarity.

For the white Gaussian noise Shannon entropy continues to give similar values as for the logistic map, but values of SampEn are dramatically higher. This is in accordance with the lack of structure in Gaussian noise. Fractional Gaussian noise behaves almost similarly in terms of different entropies (although for higher H SampEn analysis uncovers small amount of structure), but the emergence of scaling as H is increased is now clearly visible in the estimators for H and D. For the Cauchy class Shannon entropies continue to be similar for the different values of H and D, but SampEn values follow to a degree the behaviour of D. The estimators for H and D give results consistent with the changes in model parameters, although some bias appears to be present.

From the preceding examples it might appear that the SampEn is at least somewhat related to the fractal dimension D. This is quite intuitive since lower D indicates a smoother short-term behaviour. However, the example results on fractals depicted on the last rows of the table show that while in certain models SampEn and fractal dimension exhibit similar behaviour, they are fundamentally distinct. The fractals having a high amount of regularity lead to vanishing SampEn, while the estimates for H and D demonstrate the fractal yet short-range dependent nature of the process.

Results of the multiscale sample entropy analysis for the case $m = 2$ are shown in Figure 1. We see that the behaviour of SampEn at different length scales is very sensitive to the choice of the underlying process, and that no simple connection to H or D can be made in general. However, for specific families of processes relations between SampEn and H or D can indeed be identified. The complexity of fractional Gaussian noise on longer time-scales depends on H and the behaviour of the Cauchy class processes on shorter time scales is connected to the fractal dimension. For $D = 1.05$ low complexities are seen in shorter timescales, whereas for $D = 1.95$ almost white noise-like behaviour is seen for smaller scale factors. However, comparison to the results for the fractals again clearly demonstrates that no simple relationship between H, D and multiscale SampEn is present.

V. NETWORK TRAFFIC ANALYSIS

After introducing the entropy metrics we now continue with their applications in networking research. We start with the analysis of real-world network traffic traces and present two use cases of entropy metrics, namely the validation of network traffic models and the detection of network anomalies.

A. Analysis of real-world network traffic traces

We used publicly available data sets recorded in different scenarios as examples of measured network traffic traces. The first data set [27] was collected at the UCSD Network Telescope [28]. The gathered backscatter traffic sent by victims of Denial-of-Service (DoS) attacks is an example for abnormal network traffic. The second set of traces was taken throughout the *Day in the Life of the Internet* (DITL) [29] activity in Tokyo, Japan [30], and represents an example how aggregated network traffic typically looks like. Both data sets are available from the *Internet Measurement Data Catalog* (DatCat). The third example is the packet trace collected in the WLAN offered to conference attendees at SIGCOMM 2001 [31], which is available from CRAWDAD (Community Resource for Archiving Wireless Data At Dartmouth). The lower number of nodes turns the conference network into another interesting scenario. The analysis carried out here is a significant extension of our previous work presented in [9].

Out of the described data sets we extracted three traces each. The first consists of the interarrival times (iat) between consecutive packets. The second and third trace list the anonymized IP-source (src) and -destination (dst) addresses. We extracted traces of 10^6 samples and used the complete set for the estimation of the Hurst parameter and the fractal dimension. For the computation of the entropy metrics we shortened the length to 10^5 samples in order to keep the

TABLE I
VARIOUS ANALYSIS METRICS APPLIED TO ARTIFICIAL TIME SERIES.

Data set	Information theoretical entropy		Sample entropy			Hurst parameter	Fractal dimension
	32 bins	512 bins	$m=1$	$m=3$	$m=5$	Wavelet	Variation
Logistic map, $R=3.9$ (chaotic behaviour)	3.26	5.97	0.61	0.42	0.42	0.50	1.99
Logistic map, $R=3.7$ (slightly chaotic beh.)	3.27	5.97	0.52	0.37	0.32	0.48	1.98
Logistic map, $R=3.5$ (oscillating behaviour)	1.39	1.39	0.00	0.00	-0.00	-0.59	2.00
White Gaussian noise, P=1 W	2.60	5.37	2.47	2.47	2.47	0.50	1.81
Fractional Gaussian noise, $H=0.55$	2.64	5.41	2.47	2.47	2.44	0.55	1.80
Fractional Gaussian noise, $H=0.75$	2.65	5.42	2.38	2.37	2.36	0.75	1.76
Fractional Gaussian noise, $H=0.95$	2.66	5.43	1.94	1.91	1.91	0.95	1.69
Cauchy, $D+H=2$, $H=0.55$, $D=1.45$	2.66	5.43	1.44	1.44	1.44	0.65	1.58
Cauchy, $D+H=2$, $H=0.75$, $D=1.25$	2.69	5.46	0.70	0.70	0.69	0.81	1.49
Cauchy, $D+H=2$, $H=0.95$, $D=1.05$	2.73	5.50	0.18	0.22	0.21	0.99	1.35
Cauchy, $D+H\neq2$, $H=0.55$, $D=1.55$	2.65	5.41	1.71	1.70	1.69	0.67	1.64
Cauchy, $D+H\neq2$, $H=0.75$, $D=1.75$	2.65	5.42	2.02	1.96	1.95	0.86	1.72
Cauchy, $D+H\neq2$, $H=0.95$, $D=1.95$	2.53	5.30	2.40	2.36	2.35	0.96	1.78
Kiesswetter fractal, $D=1.5$	3.27	6.03	0.02	0.01	0.01	1.46	1.41
Weierstrass-Mandelbrot fractals, $D=1.45$	3.11	5.83	0.03	0.02	0.01	1.55	1.41
Weierstrass-Mandelbrot fractals, $D=1.25$	2.97	5.57	0.00	0.00	0.00	1.75	1.23
Weierstrass-Mandelbrot fractals, $D=1.05$	3.08	5.64	0.00	0.00	0.00	1.95	1.05
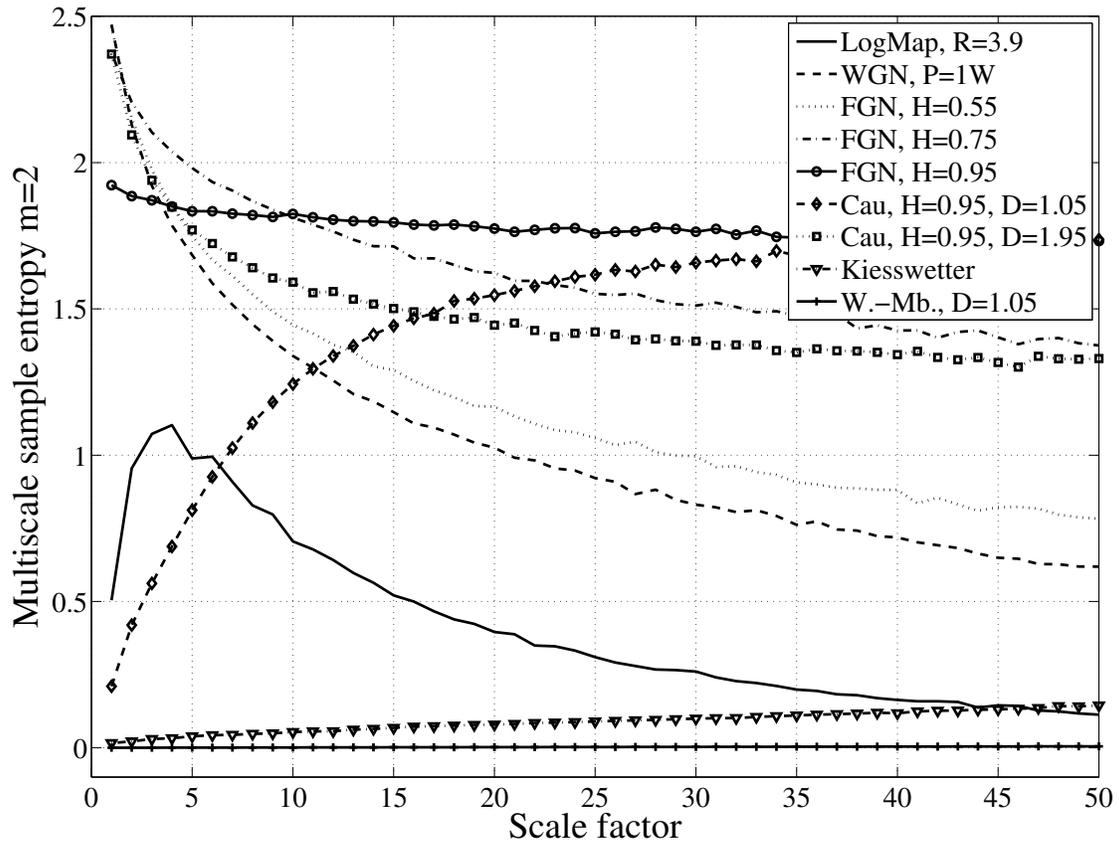
Fig. 1.   Multiscale SampEn computed for selected artificial data sets.

Data set	Information theoretical entropy		Sample entropy			Hurst parameter	Fractal dimension
	32 bins	512 bins	$m = 1$	$m = 3$	$m = 5$	Wavelet	Variation
CAIDA backscatter 2007, iat	0.00	0.80	1.47	0.96	0.34	0.95	1.65
CAIDA backscatter 2007, src	1.51	2.43	0.27	0.20	0.17	0.95	1.97
CAIDA backscatter 2007, dst	3.43	6.15	0.60	0.27	0.22	1.04	1.80
Tokyo DITL 2007, iat	0.11	1.51	0.93	0.92	0.90	0.66	1.77
Tokyo DITL 2007, src	1.42	3.17	0.74	0.66	0.60	0.86	1.96
Tokyo DITL 2007, dst	1.46	3.54	0.87	0.76	0.69	0.65	1.97
SIGCOMM 2001 WLAN, iat	0.74	3.05	0.49	0.16	0.11	0.70	1.69
SIGCOMM 2001 WLAN, src	2.10	3.38	0.51	0.31	0.27	0.54	1.87
SIGCOMM 2001 WLAN, dst	2.72	3.65	0.73	0.34	0.28	0.75	1.83
Self-similar pareto on/off traffic, $H = 0.55$	0.40	2.57	0.74	0.71	0.68	0.57	1.58
Self-similar pareto on/off traffic, $H = 0.75$	0.39	2.57	0.77	0.72	0.69	0.72	1.53
Self-similar pareto on/off traffic, $H = 0.95$	0.28	2.84	1.07	1.03	1.02	0.86	1.68

involved computational complexity at reasonable level[2].

Table II gives the same type of results for the network traces as previously discussed for the artificial time series. The SampEn gives rather low numbers compared to the artificial noise processes and shows that certain amount of structure is present in all traces. This basic result shows that there is potential in approaches trying to understand such structures and adapt the network and protocol behaviour appropriately.

The Tokyo iat trace shows self-similar behaviour as we expected based on the extensive literature on traffic analysis (see, e.g., [1], [10], [32]). The estimated Hurst parameter $H \in (1/2, 1)$ and a nearly constant SampEn are strong indicators for self-similarity. When looking at the CAIDA backscatter iat trace the SampEn decreases with m indicating the presence of structure especially when longer vectors are compared. This behaviour could be a useful indicator in the detection of network anomalies.

The fractal dimension is high for all examined traces indicating the lack of short-range dependence. This result confirms again that the SampEn does add further insight into the short-range network traffic characteristics compared to a sole analysis of the fractal dimension.

B. Network traffic model validation

We also generated synthetic traces from an often used model for network traffic based on the aggregation of multiple substreams generated by Pareto-distributed ON/OFF sources [33]. We created traces of same length as the measured ones using the open source implementation of the algorithm available at [34] and applied the discussed metrics to the traces.

The results are also shown in Table II and the estimated Hurst parameter and the SampEn confirm the self-similar behaviour. Also the Shannon entropy results indicate realistic

[2]For most source data sets the maximum length of reliably detectable patterns m_{max} is related to the length of the considered trace by $m_{max} \leq \log(N)$. Thus, the trace length of 10^5 is sufficient for $m_{max} = 5$ as considered here.
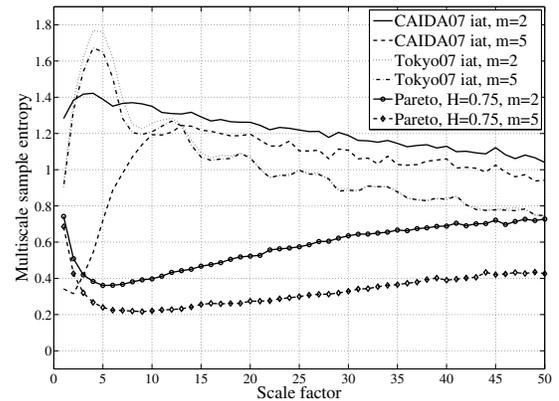

Fig. 2. Multiscale SampEn for different m-values computed for the iat time series taken by CAIDA and in Tokyo and for one synthetic trace.

modelling since they are lower compared to artificial time series as it was also the case for real-world traces. The main difference compared to fractional Gaussian noise is the increase in complexity with increasing H as measured by the SampEn. The model does also not exhibit any relationship between H and D. Both facts cannot be validated with real-world traffic because traces collected under comparable conditions but with different Hurst parameter estimates do not seem to be available.

After we have seen that the artificial model seems to reproduce the characteristics of the original time series well we continue with its multiscale behaviour. Figure 2 shows again that the entropy values for the original time series, i.e., scale factor $= 1$, are comparable to the Tokyo traces. They also tend to be close for large scale factors and thus show similar behaviour for longer time scales. This is most probably due to the fact that the self-similar behaviour was explicitly
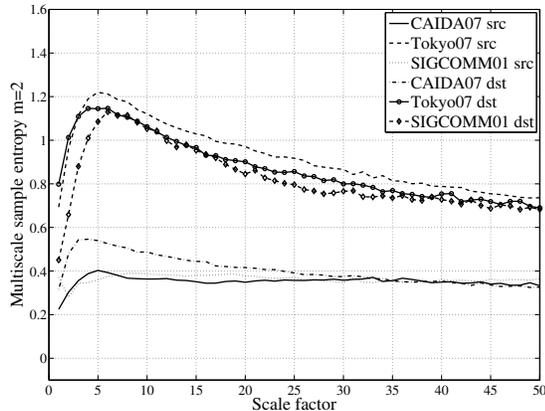
Fig. 3. Multiscale SampEn computed for all the src and dst time series.

considered during the model development. The complexity of the synthetic trace is lower for small and intermediate time scales. However, the consequences of this divergence between model and real-world traces are difficult to estimate and motivate further research.

C. Detection of network anomalies

The identification of abnormal behaviour in modern packet networks is an important research problem (see, e.g., [2], [3], [5], [7]) and the decreasing complexity of the CAIDA backscatter traces for longer patterns already indicated possible use of the entropy metrics in this application domain. The different impact of the pattern length m on the results for the CAIDA backscatter traffic and the Tokyo traces shown in Figure 2 could be another sign for abnormal situations.

In Figure 3 we compare the multiscale behaviour of all src and dst traces, which reveals the strongest indicator for abnormal behaviour in the backscatter case. The figure shows two types of behaviour. Both Tokyo traces and the SIGCOMM dst trace show higher complexity over all examined time scales and the SIGCOMM src and both CAIDA backscatter traces exhibit less complexity.

The Tokyo traces can be seen as a reference because they are taken in a fixed network that did not show any anomaly. Additionally, the aggregation level, at which the traces were taken, seems to be high enough since both the src and the dst traces show comparable complexity. For the SIGCOMM case the dst trace behaves similarly as the Tokyo trace but the special user group in the comparably small wireless network gives less complex results for the src trace[3]. The CAIDA backscatter src trace also shows less complex behaviour because the collected packets were sent by the limited group of DoS-victims.

---

[3]The mentioned selection of the parameter r for the multiscale sample entropy calculation based on the standard deviation of the examined trace normalizes the result to certain extent. However, the lower number of nodes and, especially, the integer-type of data still result in a lower estimated complexity.

When taking the dst traces into account we can also differentiate the SIGCOMM and the CAIDA trace. Although generated by a rather small user group the SIGCOMM dst traffic still shows similar complexity as in the Tokyo reference case. However, the CAIDA backscatter traffic, which is triggered by DoS-packets with spoofed source addresses, is directed back towards a limited set of spoofed addresses. The complexity of the process that the attacking nodes applied to select the spoofed address seems to be limited as is indicated by the multiscale entropy results. Such analysis may help to increase the confidence when detecting network anomalies, or to help in benchmarking other anomaly detection schemes.

VI. Spectrum usage analysis

The second use case class we have considered is to apply the entropy metrics to the area of dynamic spectrum access (DSA) [35]. This scenario shows the flexibility of the introduced methods and their applicability in different fields of communication research. We used spectrum usage data that we collected during an extensive measurement campaign. Each trace consists of power spectral density (PSD) samples measured every $\approx 1.8\,\mathrm{s}$ over $200\,\mathrm{kHz}$ channels for a frequency range of multiple GHz and lasts for multiple days up to two weeks. The measurement setup was based on a high-performance spectrum analyser; further details are given in [36].

The statistics over time of these traces are clearly different from packet traces because the medium access control and possibly applied fragmentation drastically change the observed behaviour. Additionally, our rather low sampling rate has significant impact since it is not guaranteed that each radio frame is detected.

As first step, we examined traces from multiple GSM900 channels with different communication loads. All Hurst parameter estimates are larger than one, $H > 1$, showing that self-similarity is not present. Even a small amount of usage in rarely used uplink channel dominates the statistics. For most traces we can conclude that the structure of the time series will be clearer if more traffic is transported by the cellular network. This result is valid also for longer time-scales.

Table III shows further results for selected spectrum traces. We discuss data collected at two measurement locations. One was located in a rather calm radio environment on a third-floor balcony in a residential area in Aachen (AB). The other location is on the roof of a university building with a good overview over downtown Aachen (AU) and represents a much busier radio environment.

For the case of the Universal Mobile Telecommunications System (UMTS) we analyse traces belonging to an unused channel and an adjacent one used by a nearby Vodafone UMTS base station. Since UMTS is based on signal spreading we expect characteristics similar to white Gaussian noise. Shannon entropy, SampEn, and fractal dimension partially confirm this expectation. However, the Hurst parameter estimates provide interesting insights. In the calm radio environment AB the unused channels show behaviour very similar to white Gaussian

TABLE III
VARIOUS ANALYSIS METRICS APPLIED TO MEASURED SPECTRUM TRACES.

| Data set	Information theoretical entropy		Sample entropy			Hurst parameter	Fractal dimension
	32 bins	512 bins	$m = 1$	$m = 3$	$m = 5$	Wavelet	Variation
UMTS AB, f=2108.0 MHz, unused	2.47	5.24	2.47	2.47	2.49	0.50	1.82
UMTS AB, f=2109.4 MHz, unused	2.26	5.03	2.47	2.48	2.53	0.50	1.81
UMTS AB, f=2112.0 MHz, used	2.31	5.08	2.34	2.25	2.22	1.27	1.84
UMTS AU, f=2108.0 MHz, unused	2.62	5.39	2.20	2.09	2.04	0.60	1.77
UMTS AU, f=2109.4 MHz, unused	2.69	5.45	2.19	2.07	2.01	1.37	1.81
UMTS AU, f=2112.0 MHz, used	2.26	5.03	2.39	2.33	2.32	1.13	1.79
Pager AB, f=465.2 MHz, unused	1.91	4.66	2.14	2.00	1.95	1.17	1.77
Pager AB, f=465.7 MHz, used	3.22	5.99	1.30	0.90	0.78	0.57	1.92
Pager AB, f=466.3 MHz, used	3.11	5.88	1.84	1.63	1.57	0.82	1.79
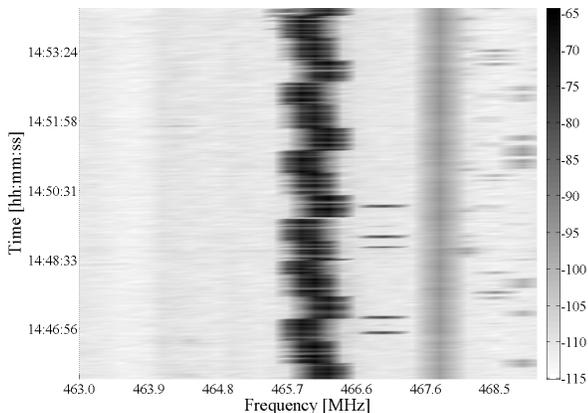
Fig. 4. Waterfall plot of the gray-scale coded power spectral density [dBm] for a representative time period.
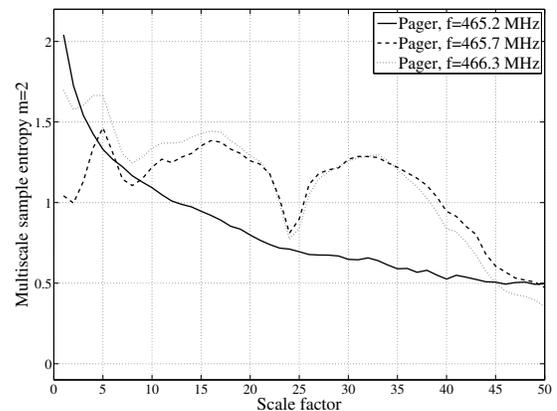

Fig. 5. Multiscale SampEn computed for spectrum traces taken in a Pager band.

noise. The used channel at AB gives a Hurst estimate $H > 1$ and shows that the spreaded signal is fundamentally different from noise. The SampEn is slightly lower for the used channel also indicating the presence of structure.

In the busy radio environment AU the characteristics for the used channel are similar to the calm radio environment. However, the traces received at the unused frequencies do not give $H = 0.5$ and also the complexity of the traces estimated by the SampEn is lower. The reason is that man-made interference dominates in these spectrum bands as is proven also by high measured PSD values [37].

Figure 4 shows the waterfall plot of the measured PSD for the frequency range used mostly by a narrowband pager service. Due to the clearly higher PSD the pager service is easily recognizable. It shifts the used frequency band regularly and shows periodic behaviour. The multiscale entropy analysis can unveil such behaviour as can be seen in Figure 5. The multiscale entropy drops clearly when the averaged time scale corresponds to half a period and again, around the full period. The third shown curve is based on data gathered in an unused

interference-dominated channel.

Deterministic signal properties such as the present periodicity could be exploited for more efficient secondary access. Possible enhancements include the intelligent selection of which channels should be sensed or how frequent a channel should be examined. However, we showed in earlier work that such deterministic behaviour is rather rare in current communication systems [38]. This may change with an increasing amount of machine-to-machine communication.

VII. CONCLUSION

In this paper we have introduced and applied multiscale entropy analysis to quantify the predictability or complexity of time series. We have specifically shown the value of the methods for network traffic and spectrum usage analysis. Our results both for artificial time series models and measured traces show that the multiscale entropy analysis can uncover traffic characteristics that techniques from classical time series analysis are not sensitive to. In particular we made a detailed comparison of multiscale entropy, self-similarity as measured by the Hurst parameter, and the fractal dimension of the time

series. We applied entropy analysis to measured traffic traces of different types, and discussed anomaly detection, validation of traffic models, and analysis of spectrum usage as highly interesting application areas. The variety of applications shows that entropy metrics are a valuable tool in the analysis of diverse components of communication systems.

While the results given here certainly appear promising, further work is required to study the limits and true potential of these techniques, especially in the context of on-line applications. Multiscale entropy analysis is more complex than Hurst parameter estimation but possible computational optimization is not fully explored, yet. Additionally, validation of online-capable approaches for anomaly detection is promising. Multiscale entropy analysis can also be potentially used off-line to help in ascertaining the ground truth when benchmarking alternative anomaly detection schemes. We are also extending our research in the area of dynamic spectrum access and cognitive wireless networks in order to exploit the information available from the entropy analysis for more efficient network operation.

References

[1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, Feb. 1994.

[2] M. Mahoney and P. Chan, "Learning rules for anomaly detection of hostile network traffic," in *Proc. of ICDM*, Melbourne, USA, Nov. 2003.

[3] A. Lakhina, M. Crovella, and C. Diot, "Mining anomalies using traffic feature distributions," in *Proc. of ACM SIGCOMM*, Philadelphia, PA, USA, Aug. 2005, pp. 217–228.

[4] A. Lall, V. Sekar, M. Ogihara, J. Xu, and H. Zhang, "Data streaming algorithms for estimating entropy of network traffic," *ACM SIGMETRICS Performance Evaluation Review*, vol. 34, no. 1, pp. 145–156, 2006.

[5] L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred, "Statistical approaches to DDoS attack detection and response," in *Proc. of DISCEX-III*, vol. 1, Washington, DC, USA, Apr. 2003.

[6] Y. Gu, A. McCallum, and D. Towsley, "Detecting Anomalies in Network Traffic Using Maximum Entropy Estimation," in *Proc. of IMC*, Berkeley, CA, USA, Oct. 2005, pp. 345–350.

[7] A. Wagner and B. Plattner, "Entropy based worm and anomaly detection in fast IP networks," in *Proc. of WETICE*, Linköping, Sweden, Jun. 2005.

[8] Y. Liu, D. Towsley, T. Ye, and J. Bolot, "An information-theoretic approach to network monitoring and measurement," in *Proc. of IMC*, Berkeley, CA, USA, Oct. 2005, pp. 159–172.

[9] J. Riihijärvi, P. Mähönen, and M. Wellens, "Metrics for Characterizing Complexity of Network Traffic," in *Proc. of ICT*, St. Petersburg, Russia, June 2008.

[10] A. Erramilli, M. Roughan, D. Veitch, and W. Willinger, "Self-similar traffic and network dynamics," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 800–819, May 2002.

[11] K. Park and W. Willinger, *Self-similar network traffic and performance evaluation*. Wiley New York, 2000.

[12] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch, *Theory and Applications of Long-Range Dependence*. Birkhäuser, 2002, ch. Self-similarity and long-range dependence through the wavelet lens, pp. 527–556.

[13] B. Mandelbrot, "Self-affine fractals and fractal dimension," *Physica Scripta*, vol. 32, no. 4, pp. 257–260, 1985.

[14] B. Dubuc, J. F. Quiniou, C. Roques-Carmes, C. Tricot, and S. W. Zucker, "Evaluating the fractal dimension of profiles," *Physical Review A*, vol. 39, no. 3, pp. 1500–1512, Feb. 1989.

[15] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 4, pp. 379–423, Jul. 1948.

[16] S. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991.

[17] J. Eckman and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, no. 3, pp. 617–656, 1985.

[18] J. Richman and J. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology- Heart and Circulatory Physiology*, vol. 278, no. 6, pp. 2039–2049, 2000.

[19] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale Entropy Analysis of Complex Physiologic Time Series," *Physical Review Letters*, vol. 89, no. 6, p. 068102, Aug. 2002.

[20] ——, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, p. 021906, Feb. 2005.

[21] T. Gneiting and M. Schlather, "Stochastic models that separate fractal dimension and the Hurst effect," *SIAM review*, vol. 46, no. 2, pp. 269–282, 2004.

[22] A. Wood and G. Chan, "Simulation of stationary Gaussian processes in [0, 1] d," *Journal of Computational and Graphical Statistics*, vol. 3, no. 4, pp. 409–432, 1994.

[23] C. R. Dietrich and G. N. Newsam, "Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix," *SIAM Journal on Scientific Computing*, vol. 18, no. 4, pp. 1088–1107, July 1997.

[24] V. Paxson, "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic," *SIGCOMM Computer Communication Review*, vol. 27, no. 5, pp. 5–18, 1997.

[25] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[26] D. Veitch, "Code for the estimation of Scaling Exponents," http://www.cubinlab.ee.unimelb.edu.au/~darryl/index.html (accessed on 2008-08-04).

[27] C. Shannon, D. Moore, and E. Aben, "CAIDA Backscatter-2007 (collection)," http://imdc.datcat.org/collection/1-0448-8=CAIDA+Backscatter-2007 (accessed on 2008-07-25).

[28] D. Moore, C. Shannon, G. M. Voelker, and S. Savage, "Network Telescopes: Technical Report," UCSD, April, Tech. Rep., 2004.

[29] CAIDA, "Day in the Life of the Internet," http://www.caida.org/projects/ditl/ (accessed on 2008-08-01).

[30] K. Cho, "WIDE-TRANSIT 100 Megabit Ethernet Trace 2007-01-09 (Anonymized) (collection)," http://imdc.datcat.org/collection/1-055M-0=WIDE-TRANSIT+100+Megabit+Ethernet+Trace+2007-01-09+\%28Anonymized\%29 (accessed on 2008-07-25).

[31] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "CRAWDAD trace ucsd/sigcomm2001/tcpdump/08292005 (v. 2002-04-23)," Downloaded from http://crawdad.cs.dartmouth.edu/ucsd/sigcomm2001/tcpdump/08292005, Apr. 2002.

[32] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, Jun. 1995.

[33] M. S. Taqqu, W. Willinger, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 2, pp. 5–23, Apr. 1997.

[34] G. Kramer, "Generator of Self-Similar Traffic (version 3)," http://wwwcsif.cs.ucdavis.edu/~kramer/code/trf_gen3.html (accessed on 2008-08-01).

[35] Q. Zhao and B. Sadler, "A Survey of Dynamic Spectrum Access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[36] M. Wellens, J. Wu, and P. Mähönen, "Evaluation of spectrum occupancy in indoor and outdoor scenario in the context of cognitive radio," in *Proc. of CROWNCOM*, Orlando, FL, USA, August 2007.

[37] M. Wellens, J. Riihijärvi, and P. Mähönen, "Evaluation of Spectrum Occupancy using Approximate and Multiscale Entropy Metrics," in *Proc. of SDR workshop*, San Francisco, CA, USA, June 2008.

[38] M. Wellens, A. de Baynast, and P. Mähönen, "Exploiting Historical Spectrum Occupancy Information for Adaptive Spectrum Sensing," in *Proc. of WCNC*, Las Vegas, NV, USA, Apr. 2008.