Metrics for Characterizing Complexity of Network Traffic

Janne Riihijärvi, Petri Mähönen and Matthias Wellens

Department of Wireless Networks, RWTH Aachen University
Kackertstrasse 9, D-52072 Aachen, Germany
email: {jar, pma, mwe}@mobnets.rwth-aachen.de

***Abstract*— We propose the study of different entropy metrics as measures of complexity of network traffic. Based on examples with synthetic data as well as different measurement traces we show that these metrics provide a new way of characterizing and studying traffic behaviour. Especially multiscale entropy analysis appears to be a powerful tool, capable of uncovering rich structures on different time-scales from different traffic measurements. We also discuss potential uses of these tools in design and validation of traffic models. Examples of potential application areas for these metrics to be studied in the future include traffic characterization, classification and anomaly detection.**

I. Introduction

One of the key objectives for any study of network traffic is discovery and quantification of different types of structure. Examples of early seminal results in this direction were the discovery of scaling behaviour or self-similarity [1] and overall deviation from simple Poisson modelling [2]. Since then massive amount of work has been performed in network traffic characterisation using classical statistics and modern time series analysis (see e.g. [3]–[5]). In this paper we propose to extend these approaches with techniques developed in the dynamical systems community capable of detecting the presence of *arbitrary* structures and quantifying complexity in network traffic. While the term "complex" is often used in this field for informal arguments for presence of structure or predictability of a phenomenon general techniques for quantifying complexity in networks have not been developed. Measuring complexity of network traffic has obviously an inherent research value, but most crucially it can provide tools for model checking and insights on how the network works and what are the underlying dynamics like. In this paper we argue that different *entropy metrics* provide an interesting new approach for studying complexity of network traffic in *quantitative* and *general* manner. These techniques can further be used for studying complexity and presence of structure at different time scales by means of *multiscale entropy analysis*. They are also very practical to apply while also having a solid theoretical foundation, with many properties amenable to analytical derivations. However, one should note that these techniques are tools for *analysis* and not fall into the trap of interpreting entropy structures themselves as a model or exhibit some deeper complex systems meaning requiring no further interpretation. One still needs to have a strong domain expertise to interprete the results appropriately.

While our focus in this paper is on network traffic characterization the techniques discussed are general, and applicable for studying other types of time series as well. Although we are focusing in this paper specifically to demonstrate the method with the internet traffic date series, the multiscale entropy analysis is not limited to any particular network types. In fact, we are currently extending our analysis towards traffic modeling of wireless and cellular data sets. Interesting applications in this space include automated traffic analysis for wireless sensor networks (especially heterogeneous ones) as well as traffic profiling in future base stations with direct internet connectivity capabilities for data traffic. We expect the increased hierarchy and structure of future wireless communications infrastructures to yield very rich multiscale structures in traffic that should be observable with the analysis techniques presented here.

In the literature the information entropy (or Shannon entropy) of different empirical distributions related to network traffic has been mainly considered as a tool for anomaly detection (see, for example, [6], [7]). The scope of the alternative entropy metrics considered here is broader. While information entropy mainly characterizes the uniformity of a probability distribution (or departure thereof), metrics derived from the *Kolmogorov-Sinai entropy* [8] measure the existence of arbitrary structures or patterns in the data. Thus a time series with uniformly distributed values but with high regularity (canonical example being the bit string "101010101010") would have high Shannon entropy (as usually calculated for time series; we shall briefly discuss other ways to apply Shannon entropy for time series below), but zero KS-entropy. We do not claim that these alternative entropy metrics are somehow "better" than Shannon entropy, but that they characterize different aspects of the traffic data hitherto almost completely unexplored by the community. Only references using techniques close to the ones reported on here are [9], [10], both of which apply methods considered obsolete by the dynamical systems research community. We stress that although methods based on entropy, particularly Shannon entropy, have been used for traffic modeling this is not the focus in this paper. The approximate and multiscale entropy analysis represents a

new set of entropy definitions, with complementary measures, which have been developed especially in Physics community.

The rest of the paper is structured as follows. In Section II we give an overview of the different entropies considered in the paper, highlighting the relationships and differences to Shannon entropy. In Section III we then apply these definitions to selected artificial data sets, the behaviour of which is well-known. This is done to give the reader a basic understanding of the behaviour of the metrics. We then apply these techniques to selected datasets in Section IV. Finally, we discuss further applications especially focusing on validation of traffic models in section V before concluding the paper in Section VI.

II. Entropies for Characterizing Complexity of Time Series

We shall now briefly discuss the various entropy metrics that we expect to be most useful in network traffic characterization. We begin with Shannon's information theoretical entropy as usually applied to time series. We then move on to discuss two alternative entropies, *approximate entropy* and *sample entropy* targeting the characterization of complexity rather than information content. Finally, we discuss multiscale analysis applied to entropy metrics as a potential means for uncovering structures with different time scales.

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* of X is then defined as the mean information content, yielding $H(X) \equiv -\sum_{i=1}^{n} p_i \log p_i$. 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. Entropy of a time series is usually defined by calculating H for the empirical probabilities obtained from the time series (or in the case of continuous values being measured, on some normalized histogram). Such an approach is perfectly valid in the case of a Markov process [11]. However, if the data has more complicated structure with successive samples being dependent, more intricate analysis is needed. The modern entropy metrics we shall introduce in the following accomplish this by essentially applying H not on the data itself, but rather on *sequences* of samples of different lengths.

B. Approximate and Sample Entropies

We shall now introduce the *approximate entropy* statistics as defined by Pincus [12] motivated by earlier work by Eckman and Ruelle on approximating Kolmogorov-Sinai entropy of a time series [13]. Let $u(1), \ldots, u(N) \in \mathbb{R}$ be our time series. Given a positive constant $m \in N$, we form the vectors $\mathbf{x}(i) \equiv \big(u(i), \ldots, u(i+m-1)\big)$. Each of these vectors has m consecutive values from our original time series starting from $u(i)$. Between two vectors $\mathbf{x}(i)$ and $\mathbf{x}(j)$ we define the distance $d\big(\mathbf{x}(i), \mathbf{x}(j)\big)$ as the maximum of the absolute values of the component-wise differences. We then count "similar" vectors by defining

$$C_i^m(r) \equiv \frac{1}{N-m+1} \#\{j \mid d\big(\mathbf{x}(i), \mathbf{x}(j)\big) \leq r\}. \quad (1)$$

The parameter $r \geq 0$ specifies the tolerance for two sequences to be considered similar. Next, we introduce

$$\Phi^m(r) \equiv \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r), \quad (2)$$

and, finally, define the *approximate entropy* by

$$\mathrm{ApEn}(m, r) \equiv \lim_{N \to \infty} \big(\Phi^m(r) - \Phi^{m+1}(r)\big). \quad (3)$$

Notice that $\mathrm{ApEn}(m, r)$ can be interpreted as the mean information content of the conditional probability for two subsequences of $u(i)$ 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

$$\mathrm{ApEn}(m, r, N) \equiv \Phi^m(r) - \Phi^{m+1}(r). \quad (4)$$

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 [14] have proposed a modified version of ApEn, called the *sample entropy*[1] and denoted $\mathrm{SampEn}(m, r, N)$. The interpretation of all the parameters and values is similar to the ApEn case, but their definitions yield slightly more robust statistics than for ApEn. Since the definition itself is slightly involved and our space is limited we shall not go through the details here. The interested reader is invited to consult [14].

C. Multiscale Entropy

To study the possible long-term structures in the data the multiscale entropy (MSE) analysis introduced by Costa *et al.* in [15], [16] can be used. Their approach consists of forming a coarse-grained time series

$$y_\tau(j) \equiv \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u(i), \quad (5)$$

where $1 \leq j \leq N/\tau$ and τ is called the *scale factor*, and then applying an entropy metric of interest on $y_\tau(j)$. As a special case $\tau = 1$ the entropy value for the original time series is obtained, whereas higher values of τ correspond to study of behaviour of averages over time windows of length τ. In the examples that follow SampEn is always used as the underlying statistic.

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

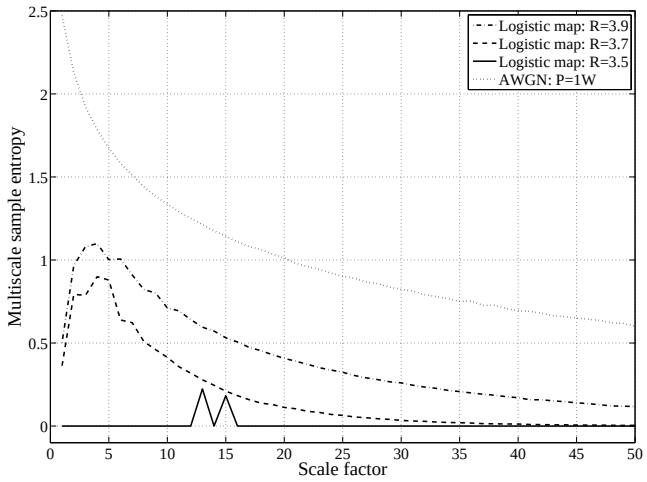

Fig. 1. Multiscale sample entropy computed for the artificial data sets ($m = 2$, $r = 0.15\sigma$).

III. Examples on Synthetic Models

We shall now show the values of the above entropy metrics for synthetic data. We consider white noise as an example of highly unstructured data, and successive iterations of the logistic map $x_{n+1} = Rx_n(1 - x_n)$ with different values for the parameter R. The latter is an interesting test case as it provides samples of adjustable complexity. 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.

Figure 1 shows the multiscale sample entropy for the introduced artificial time series and Table I lists the information theoretical and the approximate entropies for different parameter values. For the logistic map we see that for $R = 3.5$ very low approximate and sample entropies are obtained. This is fully consistent with the oscillating behavior discussed above. As R is increased more and more complicated behaviour is observed as is witnessed by the increase in both the approximate and sample entropies. Since the distribution of the values of the iterates tends to be close to uniform for large R, information entropy does not distinguish well between $R = 3.7$ and $R = 3.9$. However, the most striking difference is seen when white noise is studied. Due to its unpredictable nature both ApEn and SampEn indicate high level of complexity, whereas information entropy of white noise is actually lower than that of the logistic map for high R. All of these conclusions hold for longer timescales as well as can be seen from Fig. 1.

IV. Examples on Network Traffic

We shall now cover selected examples on how H, ApEn and SampEn behave when applied to various traffic data sets. We consider three types of time series. First consists of the IP source addresses (IP-src-addr) of the packets appearing in the trace and second of destination addresses (IP-dst-addr). In the third type we measure the number of packets per second occuring in the trace (IP-pps). For the first two types value of r was chosen to be less than one, so only perfect matches of addresses were considered. For IP-pps value of $r = 0.15\sigma$ was used, where σ denotes the standard deviation of the data. These values were confirmed to yield good results with a number of experiments, and are also consistent with the recommendations in the literature [12].

Four data sets of diverse types were considered. Two of these are traces from wireless networks deployed at two different conferences, namely SIGCOMM 2001 and OSDI 2006 [17], [18]. The second two correspond to fixed network traffic both observed during the course of A Day in the Life of the Internet project in 2007. First of these is backscatter traffic from DoS attack victims [19], with second trace containing traffic from WIDE 100 Mbps Ethernet transit link in Tokyo [20]. While not completely representative, already these time series and data sets prove our point that the entropy metrics presented are an interesting and useful extra measure for traffic characterization.

The results for Shannon entropy and Approximate entropy for different parameter values are collected in Table II. We see that the value of m, indicating the length of the patterns considered in the data for ApEn calculation has a major impact on the results. Especially the case of backscatter traffic is interesting. The IP source addresses behave qualitatively in similar manner as for other traces. The destination addresses on the other hand exhibit great deal of regularity, with ApEn falling close to zero from $m = 3$ onwards. From the nature of the traffic we can make the hypothesis that the DoS source or sources choosing the spoofed addresses are using relatively simple algorithm for selecting these addresses. Also in the Tokyo traces significant regularity can be seen for higher values of m, although not to the degree exhibited by the backscatter traffic.

We shall consider next the multiscale entropies for the different traffic traces. We focus our attention on time series indicating amount of traffic in packets per second, as the multiscale analysis has a clear interpretation in this case (for discrete data alternative techniques for aggregating data are required). Figure 2 shows the multiscale entropy for the SIGCOMM 2001 wireless trace. The behaviour of the entropy indicates relatively high degrees of complexity on time-scales of 5-25 seconds followed by rather significant reduction in entropy. Such timescales roughly correspond to a typical session behaviour especially for HTTP-driven applications, especially at the lower data rates used in 2001. In the substantially more busy wireless network at OSDI 2006 the behaviour of multiscale entropy is completely different, as can be observed from Figure 3. The lack of dependency on the timescale considered appears to be a new kind of indication on the self-similar nature of the traffic. The case of backscatter traffic depicted in Figure 4 features also little dependency on time scales considered (with the exception of slightly increasing values in large timescales), but overall the values for all m are substantially lower than in the above cases. Thus the

TABLE I

SHANNON AND APPROXIMATE ENTROPY ($r = 0.15\sigma$) RESULTS FOR ARTIFICIAL DATA SETS.

Data set	Information theoretical entropy		Approximate entropy				
	32 bins	512 bins	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
Logistic map, $R = 3.9$ (chaotic behaviour)	3.26	5.96	0.60	0.50	0.48	0.48	0.48
Logistic map, $R = 3.7$ (slightly chaotic behaviour)	3.21	5.91	0.48	0.38	0.36	0.36	0.35
Logistic map, $R = 3.5$ (oscillating behaviour)	1.39	1.39	0.00	0.00	0.00	0.00	0.00
White gaussian noise, P=1 W	2.80	5.56	2.61	2.48	1.77	0.53	0.06

TABLE II

SHANNON AND APPROXIMATE ENTROPY RESULTS FOR TIME SERIES CONSTRUCTED FROM TRAFFIC TRACES.

Data set	Information theoretical entropy		Approximate entropy				
	32 bins	512 bins	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
CAIDA backscatter, IP-src-addr	0.99	2.21	2.55	1.93	1.30	0.76	0.38
CAIDA backscatter, IP-dst-addr	3.42	6.14	0.74	0.03	0.01	0.01	0.01
CAIDA backscatter, IP-pps	1.03	3.34	1.02	0.82	0.67	0.58	0.52
OSDI 2006, IP-src-addr	2.33	3.23	1.97	1.46	1.09	0.82	0.61
OSDI 2006, IP-dst-addr	2.21	3.10	1.91	1.48	1.13	0.85	0.64
OSDI 2006, IP-pps	1.99	4.68	1.57	1.49	1.31	0.99	0.64
SIGCOMM 2001, IP-src-addr	1.85	2.55	1.68	1.32	1.04	0.78	0.58
SIGCOMM 2001, IP-dst-addr	2.37	2.99	1.82	1.42	1.07	0.74	0.50
SIGCOMM 2001, IP-pps	1.89	3.69	1.41	1.22	1.00	0.80	0.62
Tokyo DitLoI, IP-src-addr	1.93	4.09	3.26	1.39	0.64	0.27	0.11
Tokyo DitLoI, IP-dst-addr	2.16	4.57	3.35	1.40	0.49	0.16	0.06
Tokyo DitLoI, IP-pps	2.45	5.20	1.18	1.11	1.06	0.98	0.86

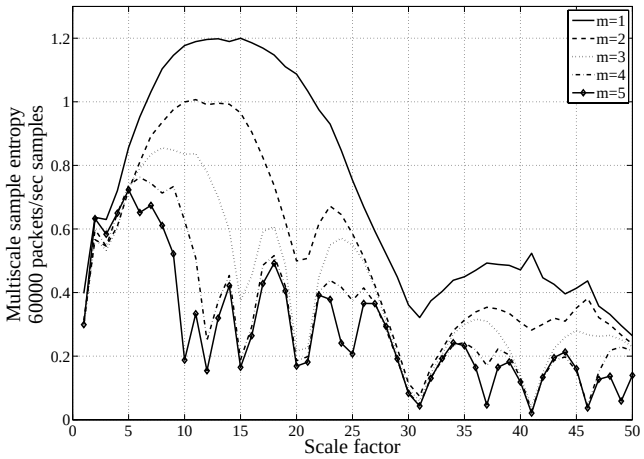

Fig. 2. Multiscale sample entropy of wireless traffic at SIGCOMM 2001.

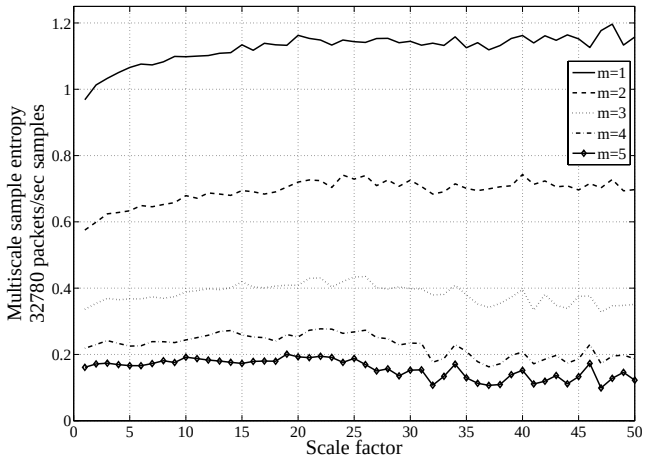

Fig. 3. Multiscale sample entropy of wireless traffic at OSDI-workshop 2006.

backscatter traffic would appear to have a higher degree of structure than found in the wireless traces. The results for the Tokyo trace (Figure 5) indicate much higher degree of complexity than in the backscatter traffic case.

V. COMPLEXITY OF TRAFFIC MODELS

Before concluding we briefly outline how the discussed entropy metrics can be used for validating models related to network traffic. The simplest models for a traffic-related time series would be processes consisting of i.i.d. samples of a random variable with density f. Pincus has shown in [12] that

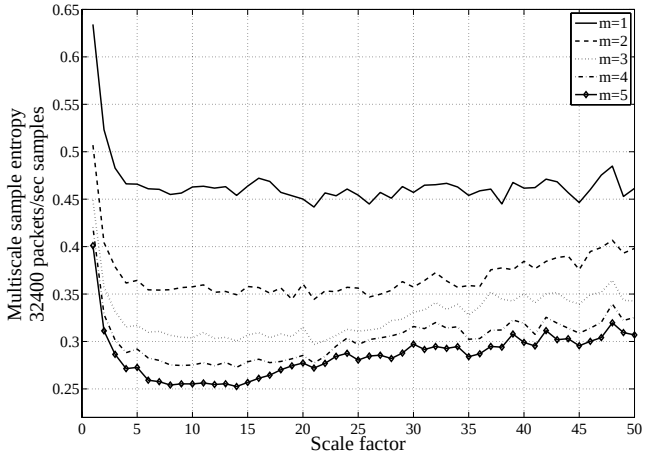

Fig. 4. Multiscale sample entropy of CAIDA backscatter traffic.

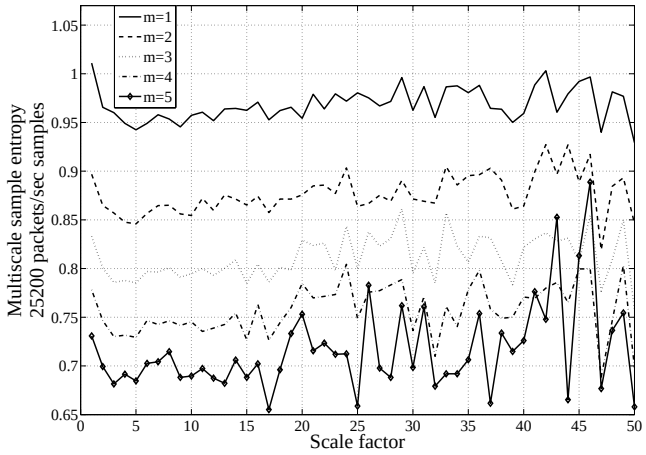

Fig. 5. Multiscale sample entropy of a wired border router in Tokyo.

for such a process

$$\mathrm{ApEn}(m,r) = -\int f(y)\ln\left(\int_{y-r}^{y+r} f(z)\,dz\right)dy. \quad (6)$$

As the right-hand side is manifestly independent of m, dependency of ApEn on m can be used to study correlations and dependencies in the data, and to help decide whether such a simple model suffices, or if a more intricate approach is required. In general comparison of ApEn with different values of m can be used to study the degree of dependency between successive samples. For example, $\mathrm{ApEn}(1,r,N) = \mathrm{ApEn}(2,r,N)$ indicates that the underlying process is (first-order) Markov, a fact that can well be exploited in modeling phase. The exact value of ApEn for stationary Markov processes is also known [21]. For example, in the first-order case with state space S we have for $r < \min_{x\neq y}|x-y|$

$$\mathrm{ApEn}(m,r) = -\sum_{x\in S}\sum_{y\in S} f(x)\pi_{xy}\ln\pi_{xy}, \quad (7)$$

where π_{xy} are the transition probabilities. Multiscale analysis allows to extend these considerations to different timescales, possibly indicating need for hierarchical or modulated models. This in one application of entropy metrics we intend to pursue in our future work.

VI. Conclusions

We have proposed the use of different entropy metrics for studying and quantifying complexity of network traffic and for discovering arbitrary structures in data. Based on examples with synthetic data as well as different measurement traces we have demonstrated that the methods indeed are a promising new way of characterizing and studying traffic behaviour. Especially multiscale entropy analysis has shown that rich structures on different time scales are present in the data sets studied. This also fits well to the current engineering and applied mathematics trend in networking community to consider multiscale and multiresolution analysis as an established technique for uncovering new structure in data. We also briefly outlined potential uses of these tools in model design and validation. We believe that these metrics will bring an interesting and powerful addition to the toolbox of traffic metrics in use today, with ample opportunities for future work. We expect the multiscale entropy analysis to provide significant new insights into the behaviour of network traffic and also fast method for data mining in large sets of measurement data for anomalous or unusual behaviours. Particularly interesting examples of potential application areas include traffic classification, anomaly detection and validation of traffic models. The computational complexity of these methods is relatively low, which means that these metrics can be used also for on-line applications in these areas. We also plan to study the implications of multiscale entropy analysis in network dimensioning and design problems, especially focussing on the implications of traffic complexity on, for example, active queue management and traffic shaping. More speculative line of work deals with traffic shaping based on entropy estimates in the vain of [22], in which a similar application in image processing domain is outlined.

Acknowledgment

The authors would like to thank RWTH Aachen University and the DFG for providing financial support through the UMIC excellence cluster. We would also like to thank European Union for providing partial funding of this work through the ARAGORN project. This work has been submitted by JR and MW in partial fulfillment of their Ph.D. degrees. PM acknowledges the kind invitation by ICT for submitting an invited paper.

References

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

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

[3] K. Claffy, G. Polyzos, and H. Braun, "Traffic characteristics of the T1 NSFNET backbone," *Proceedings of INFOCOM'93*, pp. 885–892, 1993.

[4] N. Brownlee and K. Claffy, "Understanding Internet traffic streams: dragonflies and tortoises," *IEEE Comm. Mag.*, vol. 40, no. 10, pp. 110–117, 2002.

[5] K. Claffy *at al.*, "A parameterizable methodology for Internet traffic flow profiling," *IEEE JSAC*, vol. 13, no. 8, pp. 1481–1494, 1995.

[6] A. Lakhina, M. Crovella, and C. Diot, "Mining anomalies using traffic feature distributions," *Proceedings of SIGCOMM 2005*, pp. 217–228, 2005.

[7] 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.

[8] J. Sinai, "On the notion of entropy of a dynamical system," *Dokl. Akad. Nauk. SSSR*, vol. 124, p. 768, 1959.

[9] A. Wagner and B. Plattner, "Entropy based worm and anomaly detection in fast IP networks," *Enabling Technologies: Infrastructure for Collaborative Enterprise, 2005. 14th IEEE International Workshops on*, pp. 172–177, 2005.

[10] M. Masugi, "KS-entropy-based analysis of IP-network traffic in terms of time variation of dynamical properties," *Nonlinear Analysis: Real World Applications*, vol. 7, no. 3, pp. 364–377, 2006.

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

[12] 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, March 1991.

[13] J. Eckman and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys*, vol. 57, no. 3, pp. 617–656, 1985.

[14] 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.

[15] 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, August 2002.

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

[17] R. Chandra, R. Mahajan, V. Padmanabhan, and M. Zhang, "CRAWDAD trace microsoft/osdi2006/pcap/s1 (v. 2007-05-23)."

[18] G. Voelker, "SIGCOMM 2001 Conference Wireless Trace (collection)."

[19] C. Shannon, D. Moore, and E. Aben, "CAIDA Backscatter-2007 (collection)."

[20] K. Cho, "WIDE-TRANSIT 100 Megabit Ethernet Trace 2007-01-09 (Anonymized) (collection)."

[21] S. Pincus, "Approximating Markov Chains," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 10, pp. 4432–4436, 1992.

[22] J. Starck and F. Murtagh, "Multiscale entropy filtering," *Signal Processing*, vol. 76, no. 2, pp. 147–165, 1999.