Traffic Modelling with Stochastic Lindenmayer Systems

Detlef Saß,
University of Stuttgart, Institute of Communication Networks and Computer Engineering (IKR)
Pfaffenwaldring 47, 70569 Stuttgart
detlef.sass@ikr.uni-stuttgart.de

Abstract

Traffic characteristics in communication systems have changed massively in the past. However, these changes are ongoing as this is a continuous process caused by constantly changing technologies, services and usage patterns. Simultaneously, research enabled a refined understanding of traffic characteristics. An important and traffic invariant characteristic is self-similarity and long range dependence. This is also present in the Internet backbone, impacting particularly performance evaluation studies due to its significantly slower decrease of correlations/variability. Therefore, it essential that traffic models reflect the relevant characteristics of real traffic appearing in current networks and provide additional flexibility for further changes. Lindenmayer System (L-System) based traffic models provide such a flexibility. In this paper, we extend the L-System model to the bit-rate level. We show the good modelling properties of the extended L-System model on basis of artificial traces and real traces of a broad-band user access rate scenario. We describe the developed modifications of the fitting process to accommodate the model extension.

1 Introduction

Characteristics of traffic in communication systems have changed massively in the past decades and are continuously changing, caused by changing technologies, services and usage patterns. Increasing penetration of broad-band and mobile user access, enabled quick adoption of new services, like file sharing, geo services or video services (e. g. YouTube). Further, the services' shares shift with time or popularity and, hence, impact on traffic characteristics. Also, the penetration of communication systems and their usage in daily life increased immensely, from a dial-in towards an always and everywhere online pattern. All this contributes to continuously changing traffic characteristics. Simultaneously, research enabled a refined understanding of traffic characteristics and also revealed new characteristics (e. g. scaling, burstiness, peakedness, mono-/multi-fractality). An important and traffic invariant characteristic is the self-similarity property [10, 18] or the long range dependence (LRD) [12] which is also present in the Internet backbone on larger time scales [9]. This characteristic impacts particularly performance evaluation studies due to its significantly slower decrease of correlations/variability over scaling regarding time or space, impacting queue length dimensioning, burstiness/peak rate values and simulations times [6].

Traffic models play a fundamental role for performance evaluation studies of communication systems. In order to achieve realistic results, the traffic models have to reflect the relevant characteristics appearing in current networks. Specially, the current continuing trend to an increased date-centric and increased packet-oriented network paradigm impose new challenges for traffic modelling in transport networks contexts (e. g. with the upcoming carrier-grade Ethernet services and technology). It is essential using realistic traffic models of current traffic characteristics even for transport network evaluation studies. Additionally, traffic models have to provide flexibility to adapt to further changes wrt. their application and traffic characteristics.

Lindenmayer System (L-System) based traffic models provide such a flexibility. They were only recently introduced to the context of communication networks. These traffic models can be basically deployed at different abstraction levels (e.g. packet, rate, flow, connection-level) due to the adaptability of the underlying stochastic, parametric L-System formalism. The latter has already been successfully deployed for modelling multicellular growth of organisms and modelling natural-looking plants for computer graphics.

The L-System models can be parameterised such that they reflect the traffic characteristics of an almost arbitrary trace. Such generated traffic matches well the trace distribution and the correlation structure of the trace. It also reflects the degree of self-similarity contained in the trace, even the multifractality to a certain extend.

So far, L-System models have been applied only to packet-level yet. Performance evaluation in the context of transport networks are seldom performed at packet-level but rather on a rate or higher level.

In this paper, we extend the L-System model to the rate-level, making it applicable to above evaluations. We show the modelling properties of the extended L-System model on basis of artificial traces and real traces which are captured at the uplink of a dormitory network, providing already broad-band user access rates. We describe the developed modifications of the fitting process to accommodate the model extension. Also, we discuss selected aspects and properties for the determination of the model parameters. Results show that the extended traffic model matches well the given traffic characteristics.

This paper is organised as follows. In section 2, we describe the fundamental Lindenmayer formalism. This formalism is used in section 3 for traffic modelling where we describe the model and the parameter determination. In section 4, we discuss the results of the model applied to different traces. We present the conclusions in section 5.

2 Lindenmayer System

The Lindenmayer System (L-System) is a formalism for the generation of complex, fractal structures. There, a complex structure is represented by a string of elements. The term element denotes any discrete constructional unit that repeatedly appears in the desired complex structure.

The biologist A. Lindenmayer introduced originally L-Systems [11] for describing the development of multicellular organisms and are now also used for describing and modelling natural-looking plants [14, 13]. They are especially powerful for describing fractal (self-similar) structures.

The central concept behind an L-System is that it works as a parallel rewriting system, which replaces successively any individual element of the current string by a subsequent sequence of elements. This is called a production step. Such a production step is repeated iteratively until the required complexity of the string is reached. The number of required iterations is called maximal iteration depth S. The replacements are governed by so-called production rules. At least one production rule must exist for each element, but also multiple rules may exist.

The L-System builds a formal grammar [14], equivalent to context-free Chomsky grammars [3] or Backus-Naur form [2]. It is characterised by an quadruple $G := (V, P, \pi(P), \omega)$, where

- V is the *alphabet* containing all possible elements λ_i,
- P is the set of all *production rules*,
- $\pi(p)$ represents the *rule probability* by which a rule $p \in P$ is selected from all possible rules of an alphabet element and
- ω is the initial start element, called *axiom*.

The L-System is called *stochastic* if multiple production rules exist for an element and the applied rule is selected randomly. If exactly one production rule exists for each element, then it is called *deterministic*. Further, production rules for an element can depend on additional parameters (e. g. time, iteration depth,...). Such L-Systems are called *parametric*. In the following, we use L-System as a synonym for a parametric, stochastic L-System . An illustrative example is the following:

$$\begin{array}{ll} \text{Alphabet:} & V = \{A, B\} \\ \text{Production rules:} & P = \{A \xrightarrow{1} AB, \\ & \quad B \xrightarrow{0.5} AA, \quad B \xrightarrow{0.5} BA\} \\ \text{Axiom:} & \omega = A \end{array}$$

where $\xrightarrow{\pi(p)}$ denotes the probability $\pi(p)$ that a production rule p will be chosen. For this example, a possible sequence of strings at different iteration depths is

iteration depth	string
$i = 0$	A
$i = 1$	AB
$i = 2$	ABAA
$i = 3$	ABBAABAB

This general and abstract formalism enables the application of an L-System in further, rather different domains. The key is the appropriate transfer and interpretation of the alphabet and rules for the new purpose.

3 L-System based Traffic Model

Salvador et al. [15] first brought the the L-System to the traffic modelling context. They used the L-System formalism to model traffic and proposed different variants of approaches [16], however, only on the packet-level. They showed that the L-System based traffic model reflects traffic characteristics of different types of traffic (aggregated traffic and application traffic). For this, they fitted the L-System based traffic model to traces. The generated traffic by such an fitted model reflected the probability density function, degree of self-similarity and correlation structure [17] very well.

Here, we extend the L-System based traffic model to the bit-rate/volume abstraction-level. This increases the flexibility and makes it more applicable for studies of metro and backbone networks scenarios as the packet-level is usually not reasonable in such contexts. In the following we will use simply rates for bit-rates and L-System (traffic) model for an L-System based traffic model.

3.1 Traffic Model

The basic traffic modelling idea is that the sequence of traffic rates represents the complex, chaotic (fractal) structure, which is generated by an L-System as described in section 2. The L-System traffic model will reflect the traffic characteristics which are captured in a rate trace of an overall length T, where the rates are calculated for an interval of length Δ. The generation of traffic with a parameterised L-System model is then not limited to this time T.

Now, the L-System formalism has to be transferred to the traffic modelling context and its parameters $V, P, \pi(P)$ and ω have to be interpreted appropriately. Further, additional interpretations and parameters are required.

The L-System is characterised by $G:=(V, P, \pi(P), \omega)$. Here, the alphabet $V:=\{\lambda_1, \ldots, \lambda_L\}$ defines all possible, non-negative rates which can occur in the modelled traffic, where L is the number of alphabet elements.

The generation follows the L-System principle:

- Starting with the initial element, i.e. axiom ω.
- In each production step a new sequence $X^{(i)}$ of elements is generated, where $i=0, \ldots, S$ is the current iteration depth, representing the number of applied production steps for generation of $X^{(i)}$.
- The production step $X^{(i)} \rightarrow X^{(i+1)}$ is iteratively applied until the final iteration depth of S is reached. So, $X^{(0)}=\omega$ and $X^{(S)}$ is the final desired sequence of rates which will be used for traffic generation.

Further parameters and interpretations are required for the traffic modelling context apart from the direct L-System parameters. Each string $X^{(i)}$, at iteration depth i, is mapped to a time-scale i, where the elements of this string, representing rates, are based on intervals of length t_i. The production rules generate for each element $X_k^{(i)}$, $k=1, \ldots, 2^i$, of $X^{(i)}$ two successor elements $(X_{2k-1}^{(i+1)}, X_{2k}^{(i+1)})$ such that the arithmetic mean of the two successor elements remains the same as the original element. After a production step, the number of string elements in $X^{(i+1)}$ is doubled and the time interval length is halved ($t_{i+1}:=t_i/2$). The finest time-scale, with a time interval length of Δ, is reached after S repetitions of the production step. By this construction, it is achieved that the total mean (over total time length T) is conserved and defined by the initial string ω.

Traffic shows different traffic characteristics over all different time-scales. These time-scales can be grouped to ranges, where traffic shows similar behaviour. A range $\mathcal{R}_s$ is defined by $\{i_s, i_s+1, \ldots, i_{s+1}\}$, where $i_s \in \{0, \ldots, S\}$, $i_s < i_{s+1}$ and $s \in \{1, \ldots, S\}$. For this, the production rules are parametric such that they are conditioned to be applied only for a range of time-scales. Therefor, a production rule is specified by

$$p_{i,j}\Big|_{k \in \mathcal{R}_s} : \lambda_i \overset{\pi_s(p_{i,j})}{\longrightarrow} (\lambda_j, 2\lambda_i - \lambda_j)$$

and, thus, is only valid if the time-scale k lies between i_s and i_{s+1}. Further, rule $p_{i,j}$ for element λ_i will be selected with probability $\pi_s(p_{i,j})$.

This construction process can be represented as a tree where the root is the axiom and the leaves are the elements of the final sequence at the finest time-scale S.

Here, the traffic is modelled following the *multiplicative principle* where an initial value (axiom) is iteratively subdivided in a mean conserving manner.

This means, that each final element (leaf) is a weighted product of all node values from root to leaf. It is different to the additive approach which is predominant in traffic modelling. The L-System model belongs to the class of cascade based traffic models. Also, the models described in [4, 5] belong to this class and can be described by an L-System .

3.2 Parameter Inference

The fitting approach follows mainly the approach presented in [15], however, we extended it to the bit-rate level and we modified the time-scale range determination to be more general applicable.

The parameter inference is based on a bit-rate traffic trace calculated at intervals of length Δ. With the overall time span T of the trace also the number of time-scales S is defined as

$$S := \left\lfloor \log_2 \frac{T}{\Delta} \right\rfloor .$$

The required model parameters are the alphabet, production rules and the grouping of consecutive time-scales to ranges.

For the alphabet $V := \{\lambda_1, \ldots, \lambda_L\}$, λ_1 and λ_L is determined as the minimal and maximal rate values occurring in the trace, respectively. The remaining elements are calculated as

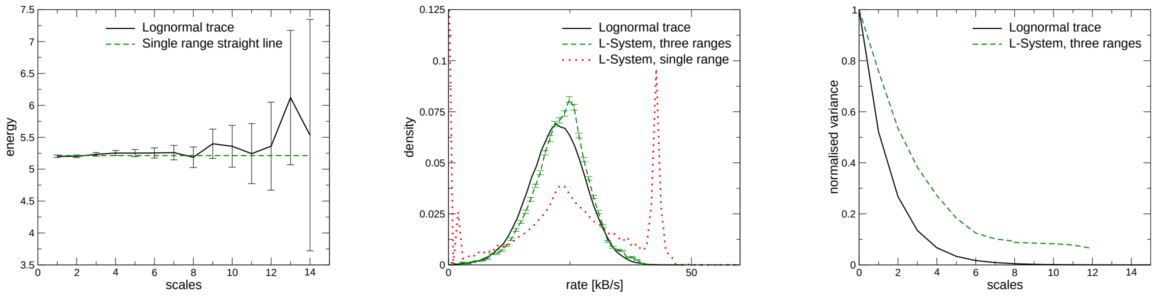
$$\lambda_i := \lambda_1 + (i-1)\frac{\lambda_L - \lambda_1}{L-1}, \quad i = 1, \ldots, L$$

which are $L-2$ equidistant spaced values between λ_1 and λ_L. The number of elements, L, defines the possible accuracy and thus how close the modelled values are at the real trace values. The decision of the number of element L is based on the form of the trace probability density function (pdf). A higher variable pdf requires a finer step size, thus more elements than a smoothly shaped pdf, where a coarser step size is sufficient.

The production rules define how elements are replaced in each production step. In principle, an element can be subdivided to any other mean-conserving pair. The subdivision is controlled by the probabilities of the production rules. However, the different variabilities of the rates in the real trace at different time-scales must be considered and, thus, the subdivision of rates. The variability is reduced at coarser time-scales (i. e. small i) and higher at finer time-scales (i. e. large i). This is achieved by grouping appropriate time-scales to ranges and inferring for each a set of production rules.

For determining this, the trace is analysed at the different time-scales of a ranges $\mathcal{R}_s$. Let denote $Y^{(i)}$ the trace at time-scale i and its $k-th$ element by $Y_k^{(i)}$.

Between two consecutive time-scales i and $i+1$, each interval k with its rate $Y_k^{(i)}$ at time-scale i is subdivided into two, halved successor time intervals with their corresponding rates $Y_{2k-1}^{(i+1)}, Y_{2k}^{(i+1)}$ at time-scale $i+1$. The probabilities π_s count these relative frequencies. The



(a) Log-scale wavelet diagram

(b) PDF for trace, L-System models (single range, three ranges)

(c) Variance-time plot

Figure 1 Lognormal trace and corresponding generated traffic of an L-System model

rates occurring in the trace are each rounded to the closest alphabet element.

In [15], Salvador et al. used for determination of the ranges only the wavelet scaling analysis [1] based on the log-scale diagram of second-order discrete wavelet coefficients over scales. There, a range corresponds to a set of consecutive time-scales which lie on a straight line (i.e. have a linear relation) wrt. to the confidence intervals in the log-scale diagram. However, we show in section 4 the need of further criteria.

The ranges are used to group consecutive time-scales which share a similar traffic behaviour regarding the scaling, variability and correlation. For each range, a single set of production rules is determined. The criteria for a group of time-scales building a range are

- linear sections in the wavelet log-scale diagram,
- same degree of variance based on the variance-time plot,
- balanced sample size of rate subdivisions across time-scales avoiding a predominance of subdivisions at a time-scale over the other for the probability estimation process,
- statistically large enough number of samples for the probability estimations in a range.

It has to be considered that the number of alphabet elements, production rules and ranges influence mutually the complexity of the parameter inference process. So, for an increased number of elements, an increased number of production rules and probabilities is necessary. Lower number of elements leads to a reduced accuracy of the generated wrt. to the original traffic. The same applies analogously for the number of ranges. However, the parameters can be tool-based inferred from a trace.

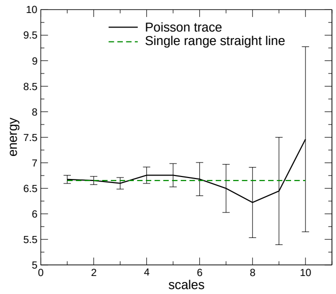
4 Results

In the following, we show numerical results of a fitted L-System traffic model on the bit-rate/volume level. In particular, we show the need of a further criteria for the range identification within the parameter inference process. Further, we compare the traffic characteristics of a trace with the characteristics of the traffic generated by an L-System traffic model which we fitted to traces. The matching self-similarity can be seen on the basis of the log-scale wavelet or the variance-time diagram.

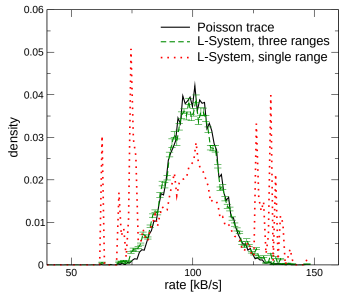
We use two types of traces to show different aspects. On the one hand, we use traces, which we artificially generate by probability distributions, to show limitations of the current range identification step. On the other hand we use a real trace collected in a productive network to show the good modelling properties of the fitted L-System traffic model.

We generate two artificial rate traces (rate interval length Δ=0.1s) which are based on a lognormal distributed (μ=1, σ=1) and negative exponential (λ=0.001) distributed packet inter-arrival times, respectively. In the following, we call the traces lognormal and Poisson, respectively.

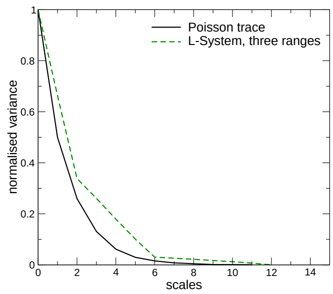
The measurements point for the real trace was the Internet uplink (upstream) of the dormitory network "Selfnet" of University of Stuttgart with approx. 1000 actively connected students. The students extensively use a wide variety of services (e. g. web, email, e-learning, gaming, e-commerce, voip, video conferencing) and new emerging Internet services are quickly adopted. Each student has a 100 Mbit/s access rate and, at the time of the collection, the uplink bandwidth was 100 Mbit/s. This environment of a dormitory network might be regarded as a possible future scenario for residential users with broad-band access. We cap-



(a) Log-scale wavelet diagram



(b) PDF for trace, L-System models (single range, three ranges)



(c) Variance-time plot

Figure 2 Poisson trace and corresponding generated traffic of an L-System model

tured the packet-level trace with our measurement platform I^{2}MP (IKR Internet Measurement Platform) [8], a hardware-supported (FPGA-based) passive measurement platform. From this, we build the rate trace (rate interval length of Δ=0.1s) which is used here. It contains S=17 scales and starts at late afternoon [8].

We implement the L-System parameter inferring procedures, including our extensions, in a set of routines. For the generation of L-System model based traffic, we use the IKR Simulation Library (IKR Simlib [7]) and implemented an L-System traffic generator. Also, we use IKR Simlib routines for the trace analysis.

We plot the log-scale wavelet diagram for the lognormal and the Poisson trace in Figures 1(a) and 2(a), respectively. A single straight line can be drawn for each trace which contains all scales within the confidence intervals. This was the reason to use on a first step an artificial trace which is governed by a clear known distribution and so excludes further biasing effects.

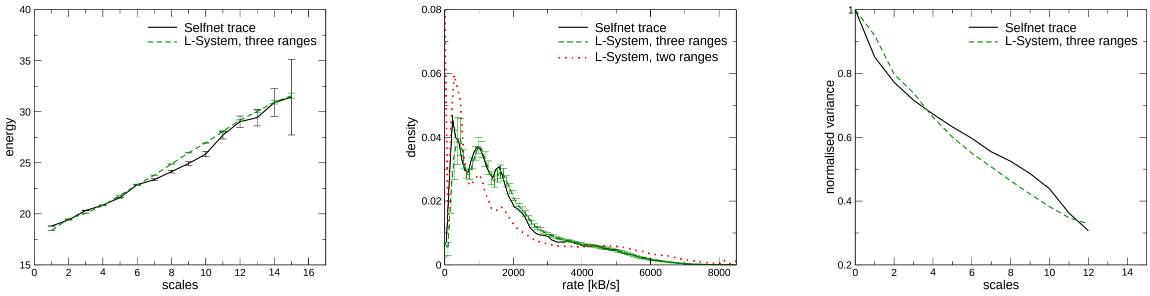
So, according to the procedure in [15], one range for the L-System model would be sufficient. We infer L-System model parameter (λ_1=0, λ_L=49, L=50, ω=22 for lognormal trace and λ_1=22, λ_L=147, L=126, ω=100 for Poisson trace). However, the fitted L-System with single range containing all time-scales (i=0, …, S) does not model well the traces' pdf, Figures 1(b), 2(b) (dotted, red curves). The curves show peaks at the borders of the distribution. These peaks are caused by indistinction of the different variabilities (and their associated different probabilities) present in the traces at higher and lower time-scales as the variance-time diagrams show (Figures 1(c), 2(c)). This indistinction causes the L-System model to tend massively to small rates already during the construction phase in the higher time-scales and leading to an overemphasis of the low and high rates (due to the mean conserving property) at the finest time-scale S.

As the number of subdivisions halves between two time-scales i and $i-1$, it is necessary that frequencies of subdivision at scales with a lower number of samples will not be covered and biased by larger frequencies of subdivisions at scales with a higher number of samples. We incorporate these two further aspects for the range determination. This leads to ranges [0,3], [3,7], [7,17] and [0,3], [3,7], [7,13] of the fitted L-System model for lognormal and Poisson trace, respectively. Now, this produces good matching results for the pdf (Figures 1(b), 2(b); dashed, green curves) and variance-time plot (Figures 1(c), 2(c); dahed, green curve) with same parameter set as before apart from the ranges.

We apply this also to the Selfnet trace, leading to λ_1=3, λ_L=9026, L=100, and ω=1809. The range determination based only on the log-scale wavelet diagram (Figure 3(a), solid curve) suggests range borders 6 and 10. This leads to a biased pdf (Figure 3(b); dotted, red curve). Considering also the other aspects, we determine the ranges [0,3], [3,10], [10,17] which leads to a good matching of the pdf (Figure 3(b)), log-scale wavelet diagram (Figure 3(a)) and a reasonable matching of the variance-time behaviour (Figure 3(c)).

5 Conclusions

In this paper, we present an extended Lindenmayer System (L-System) based traffic model. The original model was only recently introduced to the context of traffic modelling. An L-Systems is a general and powerful formalism (formal grammar). Especially, the stochastic, parametric L-Systems are excellent suitable to describe and to model fractal, random sequences with given constraints. This builds the underlying describing and con-



(a) Log-scale wavelet diagram

(b) PDF for trace, L-System models (two ranges, three ranges)

(c) Variance-time plot

Figure 3 Selfnet trace and corresponding generated traffic of an L-System model

struction method for the here described traffic model, where the traffic is generated according to the multiplicative (cascade) principle.

The model can be principally deployed at different abstraction levels due to the general L-System formalism, however, it was only deployed at packet-level, yet. Here, we extend the packet-level L-System traffic model to the bit-rate/volume abstraction level. We showed the good modelling properties of the extended traffic model on the basis of artificial traces (generated by a mathematical probability distribution) and real trace, captured at a large broad-band dormitory network at University of Stuttgart. The extended model can be fitted to a trace and is able to reflect well the trace's rate pdf, degree of self-similarity and variability over different time-scales. Further, we showed the need of additional criteria for the parameter determination within the inference process. For this, we proposed further criteria to improve the resulting parameter set.

All these extensions of the model leverage the L-System based traffic model to a broader application scope which is now also suitable for metro and backbone performance evaluation studies. Also, due to the flexible formalism of the underlying model, the L-System based traffic model is able to accommodate a wide range of different traffic characteristics and abstraction levels.

6 Acknowledgement

The author wishes to thank the dormitory network "Selfnet" of University of Stuttgart for enabling the author's measurements in their network. Further, he wishes to thank Joachim Scharf and Martin Köhn for their valuable discussion and helpful comments. This work is partly funded by the European Union through the IST Integrated Project NOBEL phase 2 (FP6-027305).

7 References

[1] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch. Wavelets for the analysis, estimation, and synthesis of scaling data. In K. Park and W. Willinger, editors, *Self-Similar Network Traffic and Performance Evaluation*, pages 39–88. Wiley, 2000.

[2] J. W. Backus. The syntax and semantics of the proposed international algebraic language of the Zurich ACM-GAMM conference. In *IFIP Congress*, pages 125–131, 1959.

[3] N. Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, IT-2(3):113–124, 1956.

[4] A. Feldmann, A. C. Gilbert, and W. Willinger. Data networks as cascades: investigating the multifractal nature of Internet WAN traffic. In *Proceedings of the ACM SIGCOMM '98*, pages 42–55, New York, NY, USA, 1998. ACM Press.

[5] J. Gao and I. Rubin. Multifractal analysis and modeling of long-range dependent traffic. In *Proceedings of ICC'99*, pages 382–386, June 1999.

[6] W. Gong, Y. Liu, V. Misra, and D. F. Towsley. Self-similarity and long range dependence on the Internet: a second look at the evidence, origins and implications. *Computer Networks*, 48(3):377–399, 2005.

[7] IKR. The IKR simulation library. Institute of Communication Networks and Computer Engineering, University of Stuttgart. `www.ikr.uni-stuttgart.de/IKRSimLib`.

[8] IKR. The IKR Internet Measurement Platform (I^2MP). Institute of Communication

Networks and Computer Engineering, University of Stuttgart, 2005. `http://www.ikr.uni-stuttgart.de/Content/I2MP/`.

[9] T. Karagiannis, M. Molle, and M. Faloutsos. Long-range dependence: Ten years of Internet traffic modeling. *IEEE Internet Computing*, 8(5):57–64, 2004.

[10] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 2(1):1–15, 1994.

[11] A. Lindenmayer. Mathematical models for cellular interactions in development, I & II. *Journal of Theoretical Biology*, 18:280–315, 1968.

[12] C. Park, F. Hernandez-Campos, J. S. Marron, and F. D. Smith. Long-range dependence in a changing Internet traffic mix. *Computer Networks*, 48(3):401–422, 2005.

[13] P. Prusinkiewicz, P. Federl, and R. Karwowski. L–System and beyond. Course at SIGGRAPH, July 2003.

[14] P. Prusinkiewicz and A. Lindenmayer. *The Algorithmic Beauty of Plants*. Springer-Verlag, 1990.

[15] P. Salvador, A. Nogueira, and R. Valadas. Modeling multifractal traffic with stochastic L-Systems. In *Proceedings of the IEEE Global Telecommunications Conference (Globecom)*, volume 3, pages 2518–2522, November 2002.

[16] P. Salvador, A. Nogueira, and R. Valadas. Joint characterization of the packet arrival and packet size processes of multifractal traffic based on stochastic L-Systems. In *In Proceedings of 18th International Teletraffic Congress (ITC-18)*, August 2003.

[17] P. Salvador, A. Nogueira, and R. Valadas. Framework based on stochastic L-Systems for modeling IP traffic with multifractal behavior. *Computer Communications Journal, Elsevier Science*, 27(18):1799–1811, 2004.

[18] D. Veitch, N. Hohn, and P. Abry. Multifractality in TCP/IP traffic: the case against. *Computer Network Journal, special issue "Long-Range Dependent Traffic"*, 48:293–313, June 2005.