Proceedings of the 7th International Conference on Applied Informatics
Eger, Hungary, January 28–31, 2007. Vol. 1. pp. 35–43.

Internet traffic v. technology change in 10 years*

György Terdik

UD, Dep. of IT

Abstract

More then 10 years now the modeling of High Speed Network Traffic has been a challenging problem for workers in statistics and engineering. One of the basic ideas in stochastic analysis which revolutionized our understanding of network traffic is the long-range dependence (LRD) and the self-similarity of the traces. In this talk we are analyzing traces measured by different Internet Providers during the last ten years or so. The question we address is the changing the parameters of self-similarity of the traces in accordance with the technology change. For this purpose the model of Smoothly Truncated Lévy Flights is applied.

1. Introduction

It is more then 10 years now that the first serious data based model of Internet traffic has been introduced, see [9]. After that it has been justified by several authors that the classical Erlang model for communication systems does not work in this case but the self-similar stochastic processes should be used for modeling instead. In the meantime there has been several changes in the technology as well as the usage of the Internet and the question arising whether these changes influenced the model of wide-area network traffic, [2, 6, 18, 3]. In this paper we introduce a particular model, namely the Smoothly Truncated Lévy Flights for understanding these possible changes.

1.1. Self-similar processes

Self-similar processes have particular importance in modeling network traffic traces. Self-similarity is a distributional property of the series in hand. The stochastic process $Y(t)$ is called self-similar if for all $a > 0$ real numbers,

$$Y(at) \stackrel{d}{=} a^H Y(t), \tag{1.1}$$

*This research is supported by the Hungarian NSF OTKA No. T047067.

with some positive H, where $\stackrel{d}{=}$ means the equality of finite-dimensional distributions. H is called Hurst exponent. The prototype of these processes is the standard Brownian Motion (BM). It is self-similar with Hurst exponent $1/2$ and the corresponding series, i.e. its increments series is independent identically distributed Gaussian series, hence it is a Lévy process. Now starting from BM there are two ways, at least, to follow keeping the self-similar property alive. One is Lévy process with $\alpha-$stable distributed increments. It is appropriate for modelling the inter-arrival time when $\alpha \in (0, 1)$. Several difficulties arising at statistical usage of these processes since there are no moment exists for $\alpha-$stable distribution if $\alpha \in (0, 1)$. We are taking over these problems applying truncated Lévy Flights which are can be close to the $\alpha-$stable distribution, as we shall show, and having moments at the same time.

The other way leads form BM to the Fractional Brownian Motion (FBM) which is a "linear" transformation of BM. FBM is self-similar it has stationary Gaussian increments and Hurst exponent from the range $(0, 1)$. Further nonlinear transformations of FBM provide stationary self-similar non Gaussian processes with higher order moments exist, for instance the Rosenblatt process and so on, see [11]. These processes are working well for modelling the high speed network traffic, i.e. the series of length. There are some delicate connections between long range dependence and self-similarity. Namely, self-similar processes have long range dependent increments and vice versa a long range dependent process is asymptotically self-similar. In this sense long range dependence can be considered a generalization of self-similarity. Some further more flexible generalizations are the multi-fractals, where the Hurst exponent might change with the aggregation level. We considered this problems in [15], where we analyzed traffic traces captured in two different networks: OC48 (2.5 Gbps) traffic collected by CAIDA (The Cooperative Association for Internet Data Analysis).

Through the investigations we prefer cumulants rather than absolute moments for describing the models because the scaling properties should not change with additive constants and with summing up of independent copies of a process. For non Gaussian processes cumulants contain some additional information about the distributional properties.

2. High speed network data

The following traces will be considered.

2.1. Traffic data repositories and traces therein

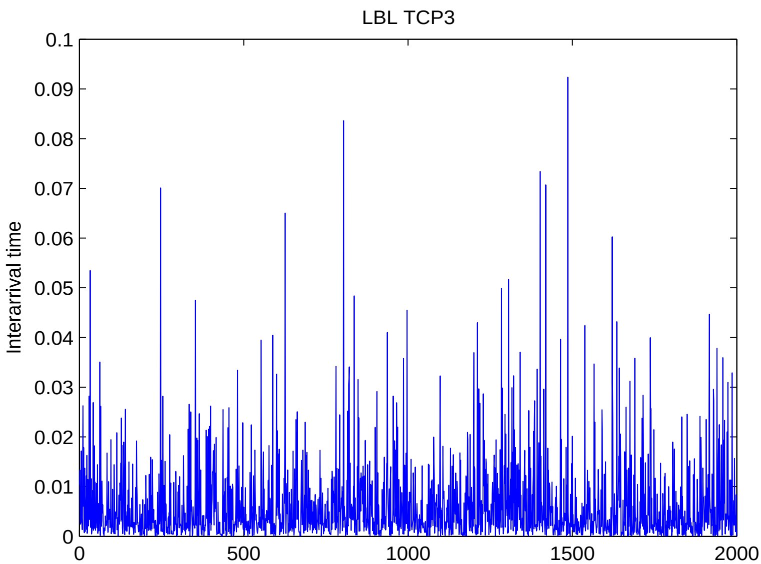
The **Internet Traffic Archive**[1] is a moderated repository to support widespread access to traces of Internet network traffic, sponsored by ACM SIGCOMM. There are the traces currently in the archive, last updated April 29, 2000.

[1] `http://ita.ee.lbl.gov/index.html`

2.1.1. Historical Traces

1. BC -x: 1 million-packet traces of LAN and WAN traffic seen on an Ethernet. The trace BC-pOct89 began at 11:00 on October 5, 1989, and ran for about 1759.62 seconds. These two traces captured all Ethernet packets. Ethernet packet arrival. Each line contains a floating-point time stamp (representing the time in seconds since the start of a trace) and an integer length (representing the Ethernet data length in bytes).The hardware clock had an actual resolution of 4 microseconds [10, 8].

2. LBL-TCP-3 - 2 hours of wide-area TCP packets. The trace ran from 14:10 to 16:10 on Thursday, January 20, 1994, capturing 1.8 million TCP packets (about 0.0002 of these were dropped). The tracing was done on the Ethernet DMZ network over which flows all traffic into or out of the Lawrence Berkeley Laboratory, located in Berkeley, California [13].



Inter-arrival times, LBL TCP3

3. LBL-PKT - 1 hour-long traces of all wide-area packets. The trace lbl-pkt-4 ran from 14:00 to 15:00 on Friday, January 21, 1994, and lbl-pkt-5 ran from 14:00 to 15:00 on Friday, January 28, 1994. Each captured 1.3 million TCP packets. The tracing was done on the Ethernet DMZ network over which flows all traffic into or out of the Lawrence Berkeley Laboratory, located in Berkeley, California [13].

4. DEC-PKT -n: The traces were gathered at Digital's primary Internet access point, which is an Ethernet DMZ network operated by Digital's Palo Alto

research groups. These traces each contain an hour's worth of all wide-area traffic between Digital Equipment Corporation and the rest of the world. Timestamps have millisecond precision. Timestamps a few microseconds apart were simultaneously timestamped by the kernel. Dec-pkt-4 started 14:00, Thu March 9th, 1995, 5.7 million dropped 1,488 packets [13].

2.1.2. MAWI Working Group Traffic Archive

Packet traces from WIDE backbone of Japan[2]. This is a traffic data repository maintained by the MAWI Working Group of the WIDE Project [1]. There are some 10Gbps connection to US as well. The traffic traces in hand are collected at the sampling point B. Samplepoint-B, daily trace of a trans-Pacific line (18Mbps CAR on 100Mbps link a link to CA), 2000, 2001, 2002, 2003, 2004, 2005, 2006. (This link was terminated on 2006/07/01). From each year we have chosen one day namely April 24th (Saint George Day).

2.2. Smoothly Truncated Lévy Flights

The Truncated Lévy Flights were introduced by Mantegna and Stanley [12] as models for random phenomena, which exhibit properties at small time-scales similar to those of self-similar Lévy processes. The Truncated Lévy Flights have distributions with cutoffs at large time-scales, i.e. they have finite moments of any order. Building on Mantegna and Stanley's ideas Koponen [7] defined the Smoothly Truncated Lévy Flights (STLFs), which had the advantage of a nice analytic form. Independently, the same family of distributions was described earlier by Hougaard [4] in the context of a biological application. The concept of the more general distribution, called tempered stable distribution, is due to Rosiński [14] (see e.g. [17] and [16] for a partial history of these works).

Since the interarrival times are positive, we consider STLF with a totally asymmetric distribution. It is given by the cumulant function (log of the characteristic function)

$$\psi_X(u) = a\Gamma(-\alpha)\left[(\lambda - iu)^{\alpha} - \lambda^{\alpha}\right], \tag{2.1}$$

where $\alpha \in (0,1)$ and $\lambda, a > 0$. A more general discussion of STLF is given in Appendix C. This distribution depends on three parameters: the *index* α, the *truncation* parameter λ, and the *scale* parameter a. These parameters provide some information about the position of the distribution in the following manner:

Property 1. If α and a are fixed and λ tends to zero, then the limit distribution is a totally asymmetric α- stable distribution and the corresponding Lévy process is self-similar.

Property 2. If λ and a are fixed and α tends to zero, then the limit distribution is Gamma with parameters (a, λ). In particular, if a is 1, then the limit is exponential, therefore the Lévy process is Poisson.

[2] `http://two.wide.ad.jp/`, `http://tracer.csl.sony.co.jp/mawi/`

Property 3. If λ and α are fixed, then for small a the distribution is close to the α- stable distribution and for large a the distribution is close to Gaussian. More precisely, moments of any positive order ϱ (including fractional) have the following asymptotics:

$$\log \mathbf{E}(|X|^{\varrho}) \sim \begin{cases} \min(\varrho/\alpha, 1)\log a + c_1, & \text{as} \quad a \to 0; \\ \varrho \log a + c_2, & \text{as} \quad a \to \infty. \end{cases}$$

For $m \geqslant 1$, the cumulants, derived from the cumulant function (2.1), are given in terms of the parameters α, λ, and a, namely,

$$\mathrm{cum}_m(X) = a\lambda^{\alpha-m}\Gamma(m-\alpha). \tag{2.2}$$

See [17, 16, 15] for details.

2.3. Estimating the parameters of $STLF_{\alpha}(a, 0, \lambda)$

Take the logarithm of

$$\mathrm{cum}_m(X) = a\lambda^{\alpha-m}\Gamma(m-\alpha).$$

and obtain

$$\log \mathrm{cum}_m(X) = \log a + (\alpha - m)\log\lambda + \log\Gamma(m-\alpha). \tag{2.3}$$

Plug the estimated cumulants $\widehat{\mathrm{cum}_m}$ into the left side of equation (2.3), then we have three unknowns a, λ, and α. In order to find the parameter values for the best fitting start with the system of equations when $m = 2, 3, 4$, i.e.

$$\begin{aligned} \log\widehat{\mathrm{cum}_2}(X) &= \log a + (\alpha-2)\log\lambda \\ &\quad + \log\Gamma(2-\alpha), \end{aligned} \tag{2.4}$$

$$\begin{aligned} \log\widehat{\mathrm{cum}_3}(X) &= \log a + (\alpha-3)\log\lambda \\ &\quad + \log\Gamma(3-\alpha) \\ &= \log a + (\alpha-3)\log\lambda \\ &\quad + \log(2-\alpha) + \log\Gamma(2-\alpha), \end{aligned} \tag{2.5}$$

$$\begin{aligned} \log\widehat{\mathrm{cum}_4}(X) &= \log a + (\alpha-4)\log\lambda \\ &\quad + \log\Gamma(4-\alpha) \\ &= \log a + (\alpha-4)\log\lambda \\ &\quad + \log(3-\alpha) + \log(2-\alpha) \\ &\quad + \log\Gamma(2-\alpha). \end{aligned} \tag{2.6}$$

The difference of the first two equations (2.4–2.5) gives

$$\log\widehat{\mathrm{cum}_3}(X) - \log\widehat{\mathrm{cum}_2}(X) = -\log\lambda + \log(2-\alpha)$$

$$= \log \frac{2-\alpha}{\lambda},$$

hence

$$\alpha = 2 - \lambda \frac{\widehat{\mathrm{cum}_3}\,(X)}{\widehat{\mathrm{cum}_2}\,(X)}.$$

Similarly from the last two equations (2.5–2.6)

$$\alpha = 3 - \lambda \frac{\widehat{\mathrm{cum}_4}\,(X)}{\widehat{\mathrm{cum}_3}\,(X)},$$

therefore we obtain

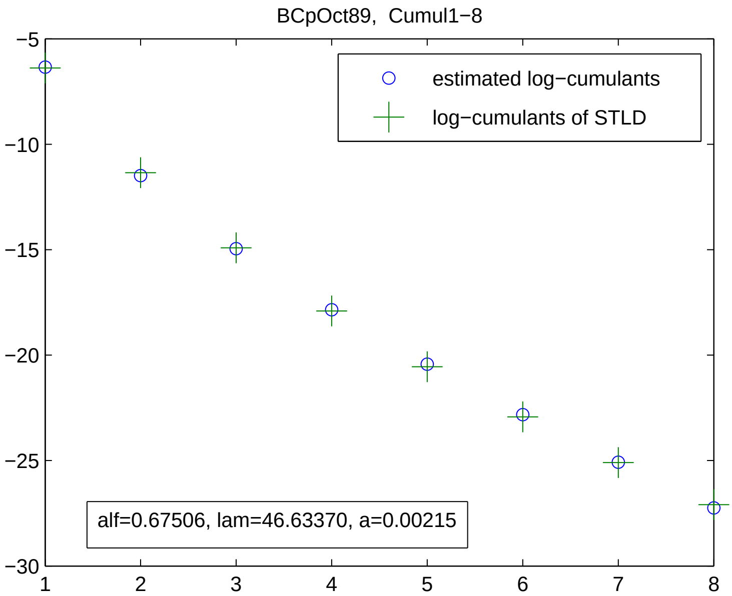
$$\begin{aligned}
\widehat{\lambda} &= \frac{\widehat{\mathrm{cum}_3}\,(X)\,\widehat{\mathrm{cum}_2}\,(X)}{\widehat{\mathrm{cum}_4}\,(X)\,\widehat{\mathrm{cum}_2}\,(X) - \left[\widehat{\mathrm{cum}_3}\,(X)\right]^2},\\
\widehat{\alpha} &= 2 - \frac{\left[\widehat{\mathrm{cum}_3}\,(X)\right]^2}{\widehat{\mathrm{cum}_4}\,(X)\,\widehat{\mathrm{cum}_2}\,(X) - \left[\widehat{\mathrm{cum}_3}\,(X)\right]^2},\\
\widehat{a} &= \frac{\widehat{\mathrm{cum}_2}\,(X)}{\widehat{\lambda}^{\widehat{\alpha}-2}\Gamma\,(2-\widehat{\alpha})}.
\end{aligned}$$

We obtain more precise estimations for the parameters, if we use these estimates as initial values and refine the estimates using nonlinear least squares, which minimizes

$$\sum_{m=1}^{8} \left[\mathrm{cum}_m\,(X) - a\lambda^{\alpha-m}\Gamma\,(m-\alpha)\right]^2.$$

3. Results and conclusion

We have fitted the STLF to all the traces mentioned above. For instance from the BCp89 trace estimated the parameters we plotted the theoretical and estimated cumulants up to order 8.



Estimated and Theoretical log-cumulants for the BCpOct89 data

The estimated parameters for the historical traces are the following.

Trace	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{a}$
BCpOct89	0.67506	46.6337	0.00215
DECpkt4tcp95	0.99249	7.64795	0.00001
LBLPkt5TCP94	0.56064	42.77707	0.01406
LBLTCP94	0.48692	70.00397	0.0209

Since the estimated scaling parameters relatively small we conclude that these traces are closer to the self-similar process then the following Wide Traces. Where α is smaller showing the tendency of approaching the Gamma distribution according to the Poisson process.

Trace	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{a}$
WideB2k1042414	0.14128	2286.746	0.31807
WideB2k2042414	0.11951	2979.278	0.38293
WideB2k3042414	0.06634	4608.418	0.4292
WideB2k4042414	0.1966	3996.299	0.26282
WideB2k5042414	0.0522	3060.139	0.53002
WideB2k6042414	0.11967	4436.569	0.31544

The conclusion is that changing of the technologies implies the some change in models, therefore we must monitor the traces and reevaluate our models continuously, [5, 2].

References

[1] Cho, K., Mitsuya, K., Kato, A., Traffic data repository at the WIDE project, In *Proceedings of FREENIX Track: 2000 USENIX Annual Technical Conference*, San Diego, CA, USA, June 18–23 (2000). USENIX Association.

[2] Figueiredo, D., Liu, B., Feldmann, A., Misra, V., D., Towsley, Willinger, W., On TCP and self-similar traffic, *Performance Evaluation*, (2-3):61, (2005), 129–141.

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

[4] Hougaard, P., Survival models for heterogeneous populations derived from stable distributions, *Biometrika*, 73(2), (1986), 387–396.

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

[6] Karagiannis, T., Molle, M., Faloutsos, M., Broido, A., A nonstationary poisson view of internet traffic, In *Proceedings of INFOCOM 2004*, vol. 3, (March 2004), 1558–1569.

[7] Koponen, I., Analytic approachto the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process, *Phys. Rev. E.*, 52, (1995), 1197–1199.

[8] Leland, W. E., Taqqu, M. S., Willinger, W., Wilson, D. V., On the self-similar nature of ethernet traffic (extended version), *IEEE/AC Transactions on networking*, 2(1), (1994), 1–15.

[9] Leland, W. E., Taqqu, M. S., Willinger, W., Wilson, D. V., Statistical analysis and stochastic modeling of self-similar data traffic, In J. Labetoulle and J. W. Roberts, editors, *The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks, Proceedings of the 14th International Teletraffic Congress (ITC '94)*, Elsevier Science B.V., Amsterdam, (1994), 319–328.

[10] Leland, W. E., Wilson, D. W., High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection, In *Proceedings of the IEEE NFOCOM'91*, Bal Harbour, FL, (1991), 1360–1366.

[11] Major, P., *Multiple Wiener–Itô integrals*, volume 849 of *Lecture Notes in Mathematics*, Springer–Verlag, New York, (1981).

[12] Mantegna and H. E. Stanley, R. N., Stochastic processes with ultraslow convergence to a Gaussian: The truncated Lévy flight, *Phys. Rev. Lett.*, 73, (1994), 2946–2949.

[13] Paxson, V., Floyd, S., Wide-area traffic: The failure of poisson modeling, *IEEE/ACM Transactions on Networking*, 3(3), (June 1995), 226–244.

[14] Rosinski, J., Tempering stable processes, Technical report, Department of Mathematics of the Univ. of Tennessee in Knoxville, Tennessee, (2004).

[15] Terdik, Gy., Gyires, T., Stochastic modelling of network traces, Technical Report, UD, (2006).

[16] Terdik, Gy., Woyczynski, W. A., Rosiñski measures for tempered stable and related ornstein-uhlenbeck processes, *Probability and Mathematical Statistics (PMS)*, Urbanik, (2006).

[17] Terdik, Gy., Woyczynski, W. A., Piryatinska, A., Fractional- and integer-order moments, and multiscaling for smoothly truncated lévy flights, *Physics Letters A*, 348, (2006), 94–109.

[18] Wisitpongphan, N., Peha, J., Effect of TCP on self-similarity of network traffic, In *Proceedings of 12th IEEE (ICCCN)*, Dallas, Texas, (2003), International Conference on Computer Communications and Networks, 370–373.