Gaussian traffic everywhere?

van de Meent, R.; Mandjes, M.R.H.; Pras, A.

Publication date
2006

Published in
Proceedings of the IEEE

Link to publication

Citation for published version (APA):
van de Meent, R., Mandjes, M. R. H., & Pras, A. (2006). Gaussian traffic everywhere? *Proceedings of the IEEE*. http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=5

Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Gaussian traffic everywhere?

R. van de Meent, M.R.H. Mandjes, A. Pras

REPORT PNA-E0602 JANUARY 2006

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Gaussian traffic everywhere?

ABSTRACT

It is often assumed that Internet traffic exhibits Gaussian characteristics, and this assumption has been validated in various studies of real Internet traffic. Less is known, however, about possible boundaries: at what timescales is traffic Gaussian and how much user aggregation is required for traffic to be Gaussian? The goal of this paper is to investigate these questions by analyzing hundreds of traffic traces, collected at four representative locations. To assess whether traffic is Gaussian, the paper starts with introducing an easy and fast procedure, based on earlier work of Kilpi and Norros. This procedure is used to investigate Gaussianity at timescales ranging from 5 msec to 5 sec. Our study shows that, if traffic is Gaussian at one timescale, it usually preserves this property at other timescales. The paper also investigates Gaussianity as function of the number of users. We conclude that, although it is impossible to give a hard number saying ‘above *N* users traffic is Gaussian’, it is fair to say that ‘only a few tens of users’ usually makes the aggregated traffic fairly Gaussian.

2000 Mathematics Subject Classification: 68M10, 90B18

Keywords and Phrases: Traffic modeling, traffic measurements, Gaussian models

Note: This work has been partly carried out in the framework of the network of excellence EURO-NGI, and was presented at the 2006 IEEE International Conference on Communications (ICC 2006), Istanbul. See also the strongly related paper "Evaluation of 'user-oriented' and 'black-box' traffic models for link provisioning", by the same authors.

Gaussian traffic everywhere?

Remco van de Meent
University of Twente
r.vandemeent@utwente.nl

Michel Mandjes
CWI
michel@cwi.nl

Aiko Pras
University of Twente
a.pras@utwente.nl

Abstract— **It is often assumed that Internet traffic exhibits Gaussian characteristics, and this assumption has been validated in various studies of real Internet traffic. Less is known, however, about possible boundaries: at what timescales is traffic Gaussian and how much user aggregation is required for traffic to be Gaussian? The goal of this paper is to investigate these questions by analyzing hundreds of traffic traces, collected at four representative locations. To assess whether traffic is Gaussian, the paper starts with introducing an easy and fast procedure, based on earlier work of Kilpi and Norros. This procedure is used to investigate Gaussianity at timescales ranging from 5 msec to 5 sec. Our study shows that, if traffic is Gaussian at one timescale, it usually preserves this property at other timescales. The paper also investigates Gaussianity as function of the number of users. We conclude that, although it is impossible to give a hard number saying 'above N users traffic is Gaussian', it is fair to say that 'only a few tens of users' usually makes the aggregated traffic fairly Gaussian.**

Keywords: Traffic modeling, traffic measurements, Gaussian models

I. Introduction

Traffic modeling in telecommunication networks has been and is still used for a variety of purposes. For instance, by having a thorough understanding of the offered traffic, it is possible to perform adequate bandwidth provisioning (i.e., to assign adequate bandwidth capacity figures in an economically viable way, at the same time satisfying a certain desired performance criterion, see, e.g. [1]).

Network traffic modeling has come a long way since the early days of telecommunications. In the IP world, various statistical models have been proposed to characterize traffic streams. The most simple model assumes Poisson arrivals of packets, but such a model has the undesirable feature that it fails to incorporate the (positive) correlations between packet arrivals observed in real traces. For this reason, the model with (a superposition of) ON/OFF sources is an attractive alternative: a broad variety of correlation structures can be modeled by choosing appropriate distributions for the ON- and OFF-times. A variant of the latter model is the so-called M/G/∞ input model, in which flows (groups of packets, with some general distribution) arrive according to a Poisson process, and generate traffic at a constant rate while being in the system. By choosing a heavy-tailed flow-size distribution, strong positive correlations can be obtained. There is vast body of literature on this topic; an overview of some approaches is given in, for instance, [2, Ch.3] and references therein.

Recently, the attention has somewhat shifted to Gaussian traffic models and multi-fractal analysis. The further development of this type of models was triggered by a number of measurement studies performed in the early 1990s, such as the famous Bellcore measurements [3]. These studies revealed extreme complexity and self-similarity in Ethernet traffic. Clearly, such phenomena on the link layer may relate to characteristics of traffic when regarding the higher layers in the protocol stack. For instance, Paxson and Floyd showed that [4] wide-area TCP traffic could also be modeled through a self-similar process.

A simple model with long-range dependency is a self-similar process characterized by a slowly (hyperbolically) decaying autocorrelation function. A stochastic model, advocated by Norros in [5], [6], that has many desirable properties (e.g. long-range dependency) is a self-similar Gaussian process: fractional Brownian motion (fBm). In recent years the fBm model (and other Gaussian models) found wide-spread use as a reference model for IP traffic.

Apart from the above (more or less) empirically derived motivations for Gaussian traffic models, there is also the argument of the Central Limit Theorem (CLT): by this theorem, the sum of a large number of 'small' independent (or weakly dependent), statistically more or less identical, random variables (users) has an approximately normal (i.e., Gaussian) distribution. Thus, one can expect that an aggregated traffic stream consisting of many individual communications may be modeled by a Gaussian stochastic process. However, the CLT argumentation does not apply to any timescale: on the timescale of transmission of (minimum size) packets, the traffic stream is always ON/OFF (either there is transmission at link speed, or silence) — which is obviously not Gaussian. Thus, apart from the number of users (referred to as 'vertical aggregation'), there should also be sufficient aggregation in time ('horizontal aggregation'). The necessity for some aggregation in both directions for traffic to be Gaussian, was pointed out by Kilpi and Norros in [7].

Contributions. Specifically, Kilpi and Norros examined the levels of aggregation that are required to justify Gaussian modeling [7]. In the present paper, we further explore this topic, in that we study the potential of Gaussian models, but also their limitations. These limitations relate to the 'minimal aggregation level' (both horizontally and vertically) needed to safely assume Gaussianity. In more detail, our contributions are:
(i) To assess whether or not traffic is Gaussian, we discuss the usability of an 'easy' goodness-of-fit test by comparing its results to a more standard test (Kolmogorov-Smirnov).

location	short description	time-span	traces	hosts	avg utiliz.
U	university residential network	5/2002 - 6/2002	15	1800	170 Mbit/s
R	research institute	5/2003 - 8/2003	185	250	13 Mbit/s
C	college network	9/2003 - 12/2003	302	1500	125 Mbit/s
S	server-hosting provider	12/2003 - 2/2004	201	100	23 Mbit/s

TABLE I
OVERVIEW OF THE MEASUREMENTS

One would hope that both tests lead to similar conclusions (as to whether or not reject Gaussianity) — it would clearly be an undesirable situation if different tests lead to different solutions.

(ii) We examine the Gaussianity of traffic on various timescales. As argued above, one expects that on small timescales Gaussianity cannot be assumed (and that is in fact what has been found in several studies, see, e.g. [8]). We systematically study the impact of the timescale, for different types of network environments. In addition, we investigate whether Gaussianity on one timescale does imply Gaussianity for other timescales (with the striking conclusion that, for a substantial range of timescales, it *does*).

(iii) We assess the impact of the number of users that are involved in the traffic stream, on the Gaussianity of the aggregate traffic. Here we give estimates as to the aggregation level from which on one can safely assume Gaussianity.

From a purely methodological perspective, the procedures followed are, to some extent, comparable to those described in [7]. We however felt that the dataset considered in [7] was somewhat limited, and that there was a need for a considerably more systematic study of these issues, using large numbers of traces, collected at several 'representative' locations. These locations differ in terms of link speeds, number of simultaneous users, applications being used, access speeds of users, the degree of user heterogeneity, etc. In other words, we believe that, due to the diversity of the selected traces, our conclusions have a stronger validity.

Organization. The paper is structured as follows: Section 2 gives an overview of the data sets that are used. Section 3 outlines the procedure to assess the Gaussianity of network traffic, cf. [7]. Part of this procedure is the assessment of the goodness-of-fit, for which we compare two approaches. Section 4 discusses the impact of the horizontal aggregation (timescale) on the Gaussianity, and Section 5 investigates the impact of the vertical aggregation (number of users). Section 6 provides the conclusions.

II. DESCRIPTION OF THE DATASETS

To assess the impact of horizontal and vertical aggregation on Gaussianity, we chose to rely on measurements of real network traffic. This choice is motivated by our objective to obtain insights that are representative for 'real environments', rather than just for simulation or lab environments.

To ensure that our data is representative for a variety of networks, we have performed an extensive number of measurements at 4 different locations. Each measurement dataset consists of 15 minutes worth of `tcpdump` records of all (Ethernet) frames passing a measurement point on the so-called Internet uplink (which carries both upstream and downstream traffic) of an organization's access network. Thus, all traffic flowing from and into the organizations' local networks is captured. A `tcpdump` record consists of a precise timestamp (of observing the packet at the measurement point) and (in our case) the headers of the packets up to the transport protocol layer. In total, some 700 traces were recorded; see Table I for an overview of the measurements[1].

Also listed in Table I is the estimated number of local hosts at each location, and the average aggregate bandwidth utilization over all traces. Noteworthy is that the hosts at all locations are connected via 100 Mbit/s FastEthernet links; the bandwidth capacities of the Internet uplinks are 1 Gbit/s, except for location S, which had an uplink capacity of about 50 Mbit/s.

The traffic at location *U* is a mixture of many applications, including WWW, e-mail and peer-to-peer traffic, both upstream and downstream, used by students. Location *R* may be characterized as an 'office': mostly WWW and e-mail traffic, from the rest of the Internet to the employees. Location *C*'s traffic is also mostly WWW and peer-to-peer traffic; the users are students and college employees. The traffic at location *S* is mainly WWW traffic, but the major part of the traffic flows from location *S* to the rest of the Internet — we will see that this is an important difference compared to location *R* later on in this paper.

III. ASSESSING THE GAUSSIAN CHARACTER

Network traffic modeling, resource provisioning, etc., often relies on the assumption that traffic is Gaussian (see, e.g. [9]): for any $t > 0$, the amount of traffic $A(t)$ offered in a arbitrary time window of length t is described by a normal distribution, parametrized by a mean Mt and variance $V(t) := \mathrm{Var}\, A(t)$. In other words:

$$A(t) \sim \mathrm{Norm}\big(Mt, V(t)\big) \ .$$

The goal of this paper is to assess the claim that network traffic is – at certain timescales – Gaussian.

Gaussianity assessment procedure. The procedure that we follow to assess the Gaussianity of network traffic is rather

[1] For reasons of confidentiality we cannot disclose the names of the organizations; we refer to them with the identifier listed in the first column of Table I.

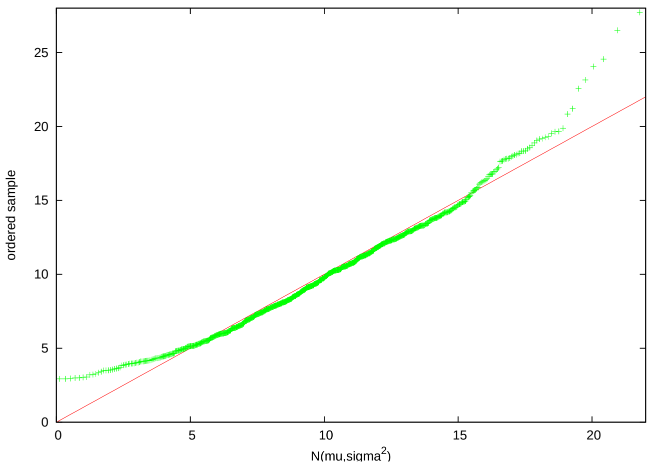

Fig. 1. Example quantile-quantile plot: $A(T)$ compared to Norm$\big(MT, V(T)\big)$ for $T = 1$ sec; location *R*, $\hat{M} = 9.8$ Mbit/sec and $\hat{V}(1) = 15.3$ Mbit^2/sec^2.

straightforward, similar to the procedure followed by Kilpi and Norros [7]: Quantile-quantile plots are made to compare the distribution of the observed traffic with a normal distribution, and this comparison is quantified using a 'goodness-of-fit' measure.

We determine whether $A(T) \sim$ Norm$\big(MT, V(T)\big)$ holds. We choose $T = 1$ sec to start with, motivated by our expectation that timescales of this order are relevant for performance as perceived by end-users of interactive applications like web-browsing. Later on in this paper we also consider other timescales.

Let A_i denote the amount of traffic offered in the ith interval of length T, and n the number of intervals. The (unbiased) estimates $\hat{M}$ and $\hat{V}(T)$ of the average and (sample) variance of the traffic rates in our traces can straightforwardly be determined:

$$\hat{M} = \frac{1}{nT}\sum_{i=1}^{n} A_i \quad \text{and} \quad \hat{V}(T) = \frac{1}{n-1}\sum_{i=1}^{n}\left(A_i - \hat{M}\right)^2 \ .$$

Note that the convergence of the estimator of the sample variance could be rather slow when traffic is long-range dependent, which can be expected for real network traffic [10, Ch.I].

To assess the Gaussianity of the network traffic, we use so-called quantile-quantile (Q-Q) plots. In these plots, the pairs

$$\left(\Phi^{-1}\Big(\frac{i}{n+1}\Big)\ ,\ \alpha_{(i)}\right)\ ,\ \ i = 1, 2, \ldots, n$$

are presented. Here Φ^{-1} is the inverse of the normal cumulative distribution function with mean $\hat{M}T$ and variance $\hat{V}(T)$, and $\alpha_{(i)}$ are the order statistics. When the traffic is 'perfectly Gaussian', all points in the Q-Q plot are on the diagonal.

In Fig. 1 an example Q-Q plot is given one of our traces. Clearly, most points are close to the diagonal. Hence, the traffic is 'fairly Gaussian'. Observe however that a fraction (up to, say, 20 out of 900) points is quite far above the diagonal at the 'high-end' of the spectrum, thus higher than the 'expected' value. This phenomenon is known as 'heavy (upper)tail', and is often seen in Internet traffic. For the context in which this study is performed, i.e., provisioning of network resources, the high-end of the spectrum is more important than the 'under-expectations' at the lower end. Thus, the heavy tail motivates conservative provisioning: if a Gaussian traffic model is assumed, one should be aware that the model may not be accurate when traffic rates are relatively high (compared to the average rate).

Figure 1 gives a first impression of the *goodness-of-fit* of some of the measured traffic compared to a Gaussian traffic model. To quantify the goodness-of-fit, we use the so-called *linear correlation coefficient* γ, which is generally defined as

$$\gamma = \frac{\sum_{i=1}^{n}\big(\alpha_{(i)} - \hat{M}\big)\big(q_i - \overline{q}\big)}{\sqrt{\sum_{i=1}^{n}\big(\alpha_{(i)} - \hat{M}\big)^2 \sum_{i=1}^{n}\big(q_i - \overline{q}\big)^2}}\ ,$$

where q_i are the quantiles of the model distribution, i.e. Norm$\big(\hat{M}, \hat{V}(1)\big)$, and $\overline{q}$ their average. Clearly, $|\gamma| \leq 1$, and γ equals 1 only if all pairs fall on the diagonal. For the examples in Fig. 1: $\gamma = 0.986$, corresponding to the 'fairly Gaussian' qualification above.

Remark on other Gaussianity tests. There is a vast body of literature on (alternative) tests for assessing normality (or another distribution) of measurement data, for instance the *Kolmogorov-Smirnov* test, see e.g. [11].

We have compared the values of γ that are computed from hundreds of our traces, with the outcome of the Kolmogorov-Smirnov test (modified for estimations of mean and variance, see [11, Sect.4.8]). The results are depicted in Fig. 2. It shows that, roughly speaking, if the correlation coefficient is high (say, $\gamma > 0.9$), then the Kolmogorov-Smirnov test (at significance level 0.05) supports the hypothesis that the underlying distribution of $A(T)$ is normal. In other words, the methods are in line with each other, and, as a consequence, it seems justifiable to use the 'easy' goodness-of-fit test based on γ, rather than the somewhat more involved Kolmogorov-Smirnov test.

Is traffic always Gaussian? One may wonder how representative the example is that was presented in Fig. 1. Therefore we look at all traces collected at all 4 our measurement locations. For each of these traces, we compute the goodness-of-fit γ, and we plot the results as to see how common certain values of γ are. The outcome is presented in Fig. 3.

From Fig. 3 we may conclude that, for all locations except *R*, in about 80% of the cases γ is above 0.9, suggesting fairly Gaussian traffic. For location *R*, we will later see that the somewhat reduced Gaussianity is likely caused by the fact that there are fewer users active at the same time.

IV. Timescale

In the previous section we investigated Gaussianity of traffic at a fixed timescale of $T = 1$ sec. In this section we will

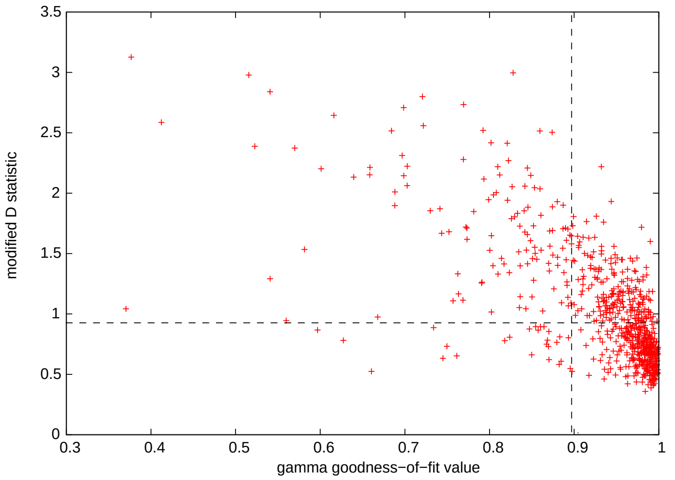

Fig. 2. Comparison between γ and D (Kolmogorov-Smirnov) statistic, using all traces.

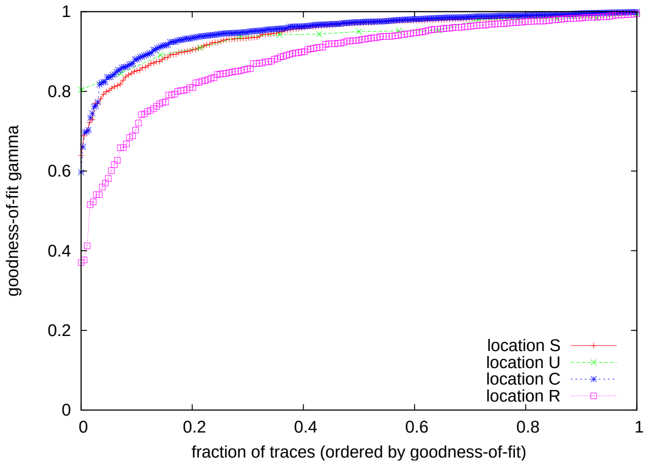

Fig. 3. Distribution of the determined linear correlation coefficient γ over all measurements

look into Gaussianity at other timescales, ranging from $T = 5\ \mathrm{msec}$ to $T = 5\ \mathrm{sec}$. The choice for this range of timescales T is motivated by our expectation that these dominate the 'user-perceived Quality-of-Service', and hence should be used for provisioning purposes (more precisely, the provisioning objective could be, for instance $\mathbb{P}(A(T) > CT) < \varepsilon$ for some predefined, small fraction ε, and T the timescale of interest).

An important question here is whether a computed value of γ at a given timescale gives a clear indication of γ at another timescale. Or in other words: if traffic is fairly Gaussian at a certain timescale, does that say anything about Gaussianity at other timescales? Suppose that a particular traffic stream exhibits strong Gaussianity at a timescale of, say, 5 seconds, and that such characteristic typically would be constant across timescales. If this is true, then, after having verified Gaussianity at timescales τ_1 and τ_2, one could also assume Gaussianity at intermediate timescales. Of course it would be tempting to also assume traffic to be Gaussian at

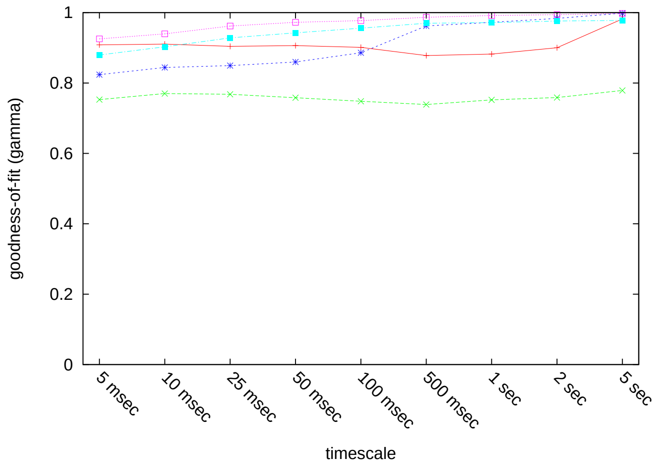

Fig. 4. Comparing Gaussianity at different timescales, for 5 example traces from location *R*.

smaller – possibly harder to measure – timescales as well, but, as remarked in the Introduction, this reasoning could be dangerous (as, at very small timescales, traffic is certainly *not* Gaussian).

First, we look at an example with only a few traces. We determine γ at various timescales; the results, with five traces from measurement locations *R*, are plotted in Fig. 4. The impression from the examples in Fig. 4 is that, because of the more or less horizontal lines, the Gaussianity is quite constant over different timescales.

Next, we investigate this for all traces. We introduce ν_γ as measure of the 'variation of γ'. More precisely, we define ν_γ as the square root of the sample variance of the γ_τ values at all assessed timescales $\tau_1, \ldots, \tau_n \in T$:

$$\nu_\gamma := \sqrt{\hat{\mathrm{Var}}\left(\gamma_{\tau_1},\ \gamma_{\tau_2},\ \ldots,\ \gamma_{\tau_n}\right)}\ ,$$

where we choose $T = \{5\,\mathrm{msec},\ 10\,\mathrm{msec},\ 25\,\mathrm{msec},\ 50\,\mathrm{msec},\ 100\,\mathrm{msec},\ 500\,\mathrm{msec},\ 1\,\mathrm{sec},\ 2\,\mathrm{sec},\ 5\,\mathrm{sec}\}$. The interpretation is that when ν_γ is low, the traffic is (more or less) equally Gaussian across multiple timescales.

We have computed ν_γ for all traces at all 4 measurement locations. After ordering them from low to high values of ν_γ, they are plotted in Fig. 5. Clearly, ν_γ is small in most cases: in over 95% of the traces, ν_γ is below 0.05. Thus we may conclude that γ is quite constant over different timescales; in other words: traffic that exhibits Gaussian characteristics at one timescale, is likely to be Gaussian at other timescales as well (for the timescales that we investigated, at least).

Finally we have computed the 'average Gaussianity' of all traces at various timescales, i.e., the average value of γ for all traces at a particular location, for various timescales. These are plotted in Fig. 6, together with error bars that represent the standard deviation of the computed γ values at a specific timescale.

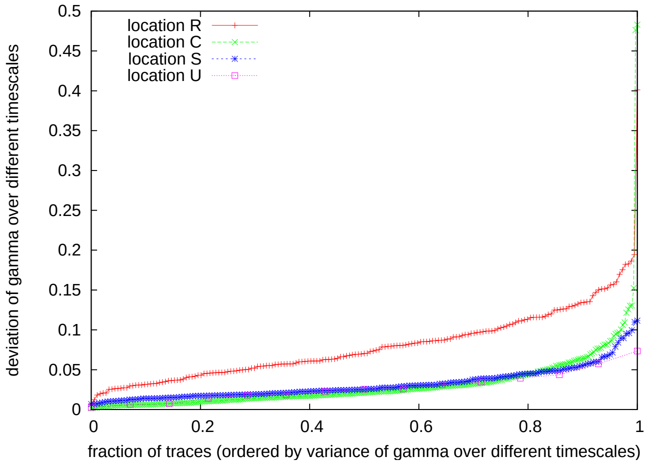

Fig. 5. ν_γ: goodness-of-fit γ over different timescales

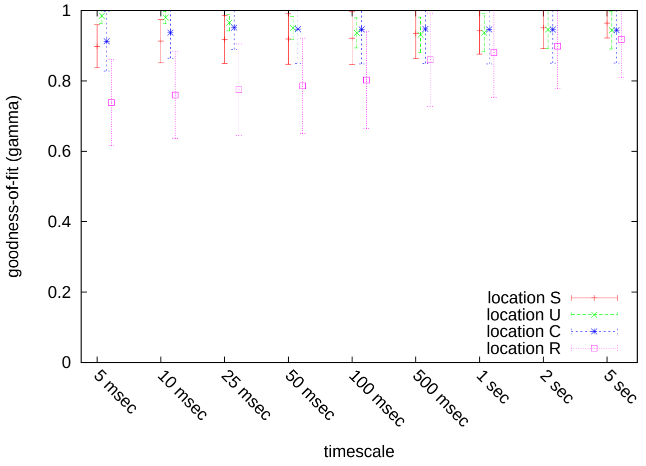

Fig. 6. 'Average Gaussianity' at different timescales for all locations

From the analysis in this section we may conclude that traffic is fairly Gaussian at the vast majority of the locations, for most traces, for most timescales from 5 msec up to 5 sec. Thus, traffic may be assumed to be fairly Gaussian in general. Also, generally speaking, if traffic is Gaussian at a particular timescale, it would still be Gaussian if the measurements were taken at other timescales.

V. Number of users

In the previous section we discussed the impact of the horizontal aggregation, i.e., the timescale, on the Gaussianity of network traffic. In this section we will look into the effect of the vertical aggregation, i.e., the number of users whose traffic is aggregated.

When traffic of a 'sufficiently large' number of users is aggregated, the resulting traffic mix exhibits strong Gaussianity [7]. We will now investigate in further detail what 'sufficiently large' means. We do this by comparing the Gaussianity of network traffic as function of the number of users involved.

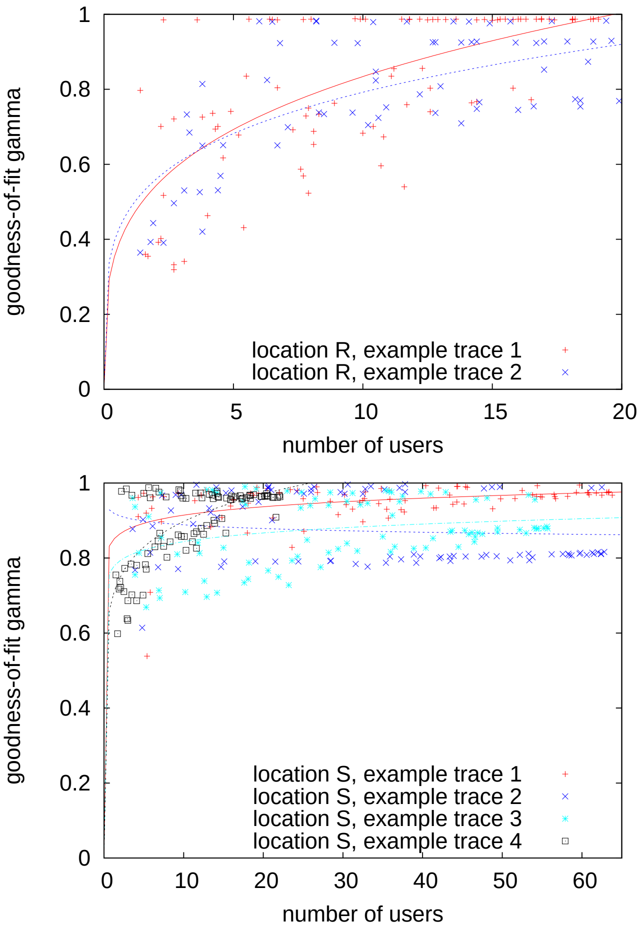

Fig. 7. Gaussianity compared to number of active users

In this section we will rely on the same traces as used in the previous sections. In these traces the traffic of a large number of users was aggregated, however. Therefore, we took from these traces only a subset of all packets, namely just the packets that relate to a random subset of users. In this way we can investigate the Gaussianity of a traffic in which just a fraction p of all users is aggregated (as a function of this p). Our procedure to reduce the number of users involved is as follows:

We process the trace per packet; when a new IP address within the local network address range is found, with a probability p all of this IP address' traffic in the trace will be taken into account, with a probability $1-p$ traffic of this user is not selected, thus reducing the number of users as desired. The experiment is repeated with the same p, evidently leading to different results due to the random nature of the selection process, as well as with different p. The experiments yield input to our 'Gaussianity quantification procedure' described earlier: for various numbers of 'active users', a Gaussianity figure is computed.

The number of 'active users' is defined as follows. Per T, e.g. 1 second, it is observed how many distinct IP addresses (within the local network address range) send or receive traffic

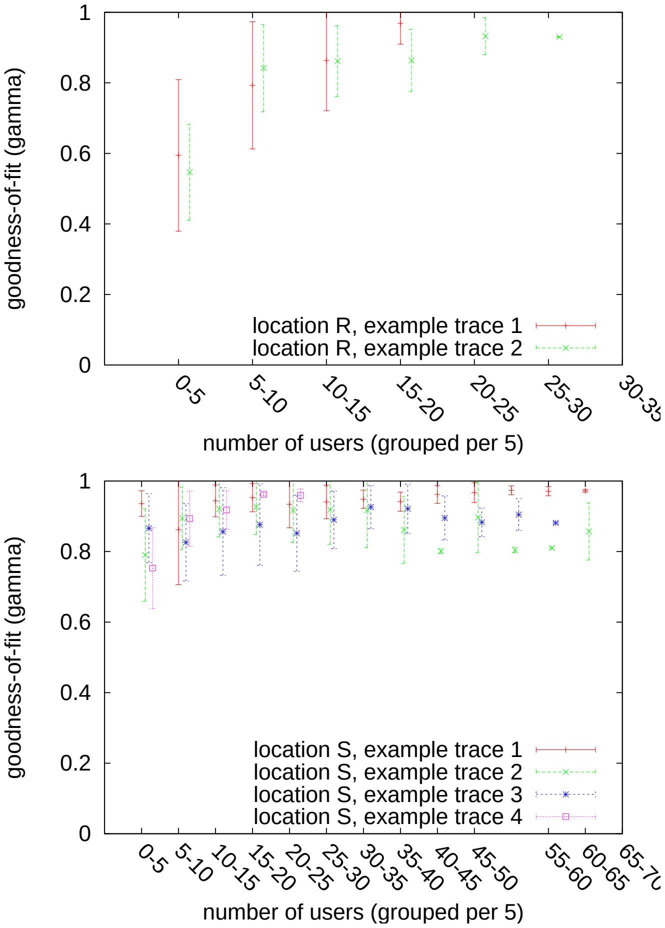

Fig. 8. Gaussianity compared to number of active users (grouped)

in that interval. The number of active users per experiment is then the average number of distinct IP addresses over all intervals (which is evidently not necessarily an integer number).

It is assumed that the traffic of the users that are not taken into account, does not influence the characteristics of the traffic of the users whose traffic *is* taken into account; this can be motivated by the relatively high degree of overprovisioning of the network links.

Figure 7 shows for locations *R* (top) and *S* how the number of active users relate to the Gaussianity of the network traffic, taking only a few example traces into account. As could be expected, Gaussianity increases with the number of active users. Also, for a given number of active users, there is typically quite some variation in the Gaussianity (between traces as well as within the same trace but for different experiments, i.e. with different subsets of selected users). Notably, there are cases when only a few users are active on average, but still the Gaussianity is almost 1. In these cases, a small number of users were dominating the trace and happened to be selected, and apparently their traffic plus the 'noise' of the others is Gaussian.

Figure 7 gives a first impression on the relation between the number of active users relate and the Gaussianity of the resulting aggregated traffic. For illustration purposes we have added (least squares) fits for the data plots; the fits stem from the formula $\alpha \cdot N^{\beta}$, where N denotes the number of users and α and β are scaling parameters. Next, we want to get a more thorough expression of this relation.

We compute the γ values for numerous experiments (as described above), and aggregate the results in two dimensions: (i) the number of users involved is grouped per 5, and (ii) the γ values are averaged and plotted together with an error bar indicating the standard deviation. The results are plotted in Fig. 8. The top picture shows the result for location *R*; below for location *S*. As the primary interest here is on the lower hand of the spectrum of the number of users (as, for large numbers of users, we already know traffic is quite Gaussian), we have limited ourselves here to the two locations with the least number of users.

From Fig. 8 it can be seen that, as expected, an increase in the number of users involved tends to increase the Gaussianity. It is not possible, however, to give a hard number saying 'above N users, traffic may be assumed Gaussian'. It seems justified to claim that 'only a few tens of users' makes the resulting traffic fairly Gaussian (at this timescale).

VI. Conclusions

This paper investigated Gaussianity of network traffic, using hundreds of real network traffic traces, collected at four different locations. These locations can be described as a university residential network, a research institute, a college network and a server-hosting provider.

To assess whether traffic is Gaussian, the paper builds on previous work of Kilpi and Norros. In particular it presents an 'easy' goodness-of-fit procedure, using the so-called *linear correlation coefficient* γ. Our study shows that the results obtained using this procedure are comparable to the more standard (and more difficult to use) *Kolmogorov-Smirnov* procedure. For all but one location, we found that, in about 80% of the cases γ is above 0.9, suggesting fairly Gaussian traffic.

An important conclusion of this paper is that traffic that is Gaussian at one timescale, is likely to be Gaussian at other timescales as well. More specifically, we investigated Gaussianity at timescales ranging from 5 msec to 5 sec, and introduced ν_{γ} to denote the 'variation of γ' over these timescales. In general we found small values of ν_{γ}: in over 95% of the traces it is below 0.05, implying that 'level of Gaussianity' remains stable over different timescales.

Finally the paper investigates Gaussianity as function of the number of users. Although it is impossible to give a hard number saying 'above N users traffic is Gaussian', it is safe to make the general claim that 'only a few tens of users' already make the aggregated traffic fairly Gaussian.

Acknowledgments

This work has been partially sponsored by the European Commission through the EuroNGI and E-NEXT Networks of Excellence.

References

[1] R. van de Meent and M. Mandjes, “Evaluation of ‘user-oriented’ and ‘black-box’ traffic models for link provisioning,” in *Proceedings of the 1st EuroNGI Conference on Next Generation Internet Networks Traffic Engineering*, (Rome, Italy), April 2005.

[2] P. Tran-Gia and N. Vicari, eds., *Impacts of New Services on the Architecture and Performance of Broadband Networks*. 2000. Final report of action COST 257.

[3] 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. 1, pp. 1–15, February 1994.

[4] V. Paxson and S. Floyd, “Wide area traffic: The failure of poisson modeling,” *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, June 1995.

[5] I. Norros, “A Storage Model with Self-Similar Input,” *Queuing Systems*, vol. 16, pp. 387–396, 1994.

[6] I. Norros, “On the Use of fractional Brownian motion in the Theory of Connectionless Networks,” *IEEE Journal of Selected Areas in Communications*, vol. 13, no. 6, pp. 953–962, 1995.

[7] J. Kilpi and I. Norros, “Testing the Gaussian approximation of aggregate traffic,” in *Proceedings of the 2nd ACM SIGCOMM Internet Measurement Workshop*, (Marseille, France), pp. 49–61, 2002.

[8] C. Fraleigh, F. Tobagi, and C. Diot, “Provisioning IP Backbone Networks to Support Latency Sensitive Traffic,” in *Proceedings of IEEE Infocom*, (San Francisco, U.S.A.), April 2003.

[9] M. Mandjes and R. van de Meent, “Inferring traffic characteristics by observing the buffer content distribution,” in *Proceedings of the 4th International IFIP-TC6 Networking Conference (NETWORKING 2005)* (R. B. et al., ed.), no. 3462 in Lecture Notes in Computer Science (LNCS), (Waterloo, Canada), pp. 303–315, May 2005.

[10] J. Beran, “Statistical methods for data with long-range dependence,” *Statistical Science*, vol. 7, pp. 404–416, November 1992.

[11] R. B. D’Agostino and M. A. Stephens, eds., *Goodness-of-fit techniques*. Marcel Dekker, Inc, 1986.