On-line Error Detection through Observation of Traffic Self-Similarity

Wolfgang Schleifer
CERN European Laboratory for Particle Physics[*]
CH-1211 Geneva 23, Switzerland
Wolfgang.Schleifer@cern.ch

Manfred Männle
Institute for Computer Design and Fault Tolerance
D-76128 University of Karlsruhe, Germany
Manfred.Maennle@informatik.uni-karlsruhe.de

***Abstract* -- This paper presents a new and universally applicable approach of error detection in packet or cell communication networks. The error detection uses measured traffic load data. The advantage is that an error detection on a low layer decreases the probability of an undetected error on higher layers and makes a time-costly error detection on higher layers unnecessary. Most present systems use static traffic load thresholds for error detection. Compared with this technique, you can achieve a considerably higher sensitivity using the new approach presented in this paper.**

The basic idea of our detection algorithm is to exploit the property of self-similarity in network traffic. This analytical redundancy gives us a reference behavior of the network traffic load which we use to detect faulty behavior in the real network traffic load. For the error detection, we check the validity of the given self-similar property through a deviation indicator *Q* based on second order properties of the time series' distributions. This is, according to our observations, a sufficient condition for normal (error free) behavior.

1. Introduction

[*]Especially in the case of communication systems that use a common transmission medium, a single component error can lead to an unacceptably faulty system behavior. Such systems are for example Ethernet segments but also switched networks. Continuous error detection in network concentration points is one possibility for fast detection of errors and, in the case of a quick error handling after detection, prevents an unlimited error propagation.

The goal of this work is the development of a universally applicable approach for a fast and low cost error detection based on the analysis of the network traffic load in the concentration points of the network as for example the internal bus of a hub. We concentrate on developing a fast and highly sensitive approach to get a reliable error decision at

[*]. Parts of the work described in this article were done while W. Schleifer was with the Institute for Computer Design and Fault Tolerance, University of Karlsruhe.

acceptable network performance costs.

We exploit the self-similarity properties of packet traffic to detect errors. This analytical redundancy gives us a reference model of the normal network traffic load which we use to detect errors in the network under observation. When the deviation between the reference model and the real system traffic measurements is sufficiently small we may consider the system error free.

We assume the following error model: The effect of a fault is a significant increase or decrease of the network traffic load. This means that every possible low level failure such as corrupted packets, packet loss, server crash and so on leads to additional traffic or to a reduction in traffic. These effects are detectable by the change of the internal structure of the network traffic load. The validity of this assumption is confirmed by observations of errornous network traffic.

To validate our approach we perform a series of error injection experiments (modifications of traffic load of the internal bus of a hub in the campus network of the University of Karlsruhe) and also provide an example containing a real error.

2. Analytical redundancy

The pattern of normal (error free) network traffic load is affected by the interplay of deterministic and stochastic parameters of the network components. It can be described using the methods of fractal geometry. The basic principle of fractal geometry is the relation of objects in different scales. Mandelbrot showed in [2] that outlines of nature look similar in different scales. The references [4], [5], [6], [7], and [8] show that we can find a very similar phenomenon in computer network traffic loads. In analogy to fractal objects in nature, that look similar in different scales, the network traffic load time series seem to look the same in large and in small time scales [7]. Other references like [3] and [9] show that this property is independent from the topology, protocol and technology of the network used and also independent from the distribution of the traffic load.

2.1. Long-term correlation process

In contrast to the time-independent or time-limited dependent models, the network traffic load indicates a very high persistence [1]. This can be seen by the hyperbolically decreasing spectral function near the zero frequency (λ):

$$\lim_{\lambda \to 0} f(\lambda) = \infty . \qquad (1)$$

A stationary stochastic process has a short-range dependence if the autocorrelation function value ρ(τ) in the case of increasing time shift τ, decreases exponentially or faster [1]:

$$|\rho(\tau)| \le M\beta^{\tau} , \ \tau \in T , \ M > 0 , \ 0 < \beta < 1 . \qquad (2)$$

The condition for short-range dependence is that the autocorrelation function value ρ(τ) converges fast enough to zero so that the effect disappears at larger τ.

We have a long-range dependence if the autocorrelation function ρ(τ) converges hyperbolically for increasing time shift τ [1]:

$$\rho(\tau) \ge c_{\rho}\tau^{-\beta} , \ c_{\rho} > 0 , \ \tau \to \infty , \ 0 < \beta < 1 , \qquad (3)$$

$$\sum_{\tau = -\infty}^{\infty} |\rho(\tau)| = \infty . \qquad (4)$$

The absolute value of the autocorrelation functions ρ(τ) of long-range dependence processes (see equation (4)), in contrast to that of short-range dependence processes, do not converge .

The autocorrelation function shown in figure 2 does not converge to zero for larger time shift τ, i.e, the given measurements stem from a long term correlation process.

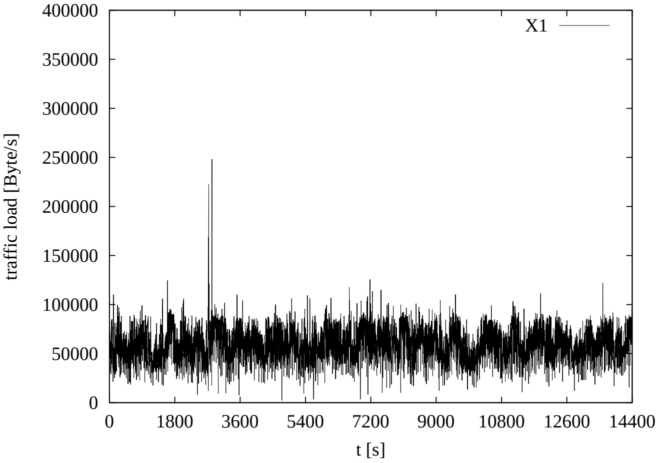

Figure 1: Example of a self-similar network traffic load time series (measured data).

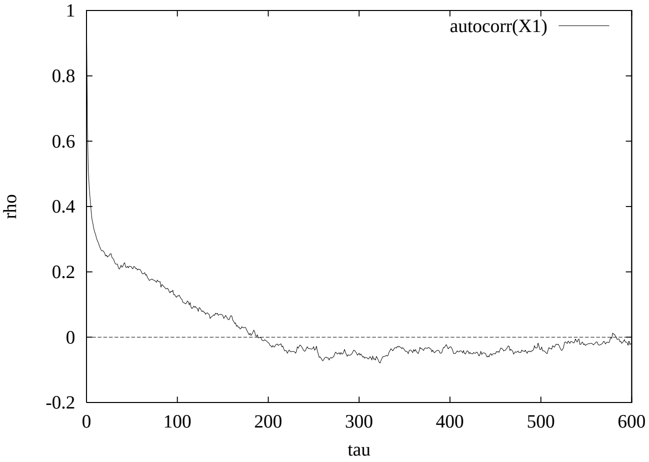

Figure 2: Autocorrelation function of a long-term correlation process.

2.2. Self-similar process

First order self-similarity means similarity of the time series in different scales. Second order similarity is restricted to the similarity of time series distributions. *Network traffic load* is assumed to show *second order self-similarity* [4], [6], [7].

A stochastic process X_t with time parameter *t* and zero mean shows a self-similar second order behavior with self-similar parameter *H* if, for a scale factor *a>0,* the rescaled process $X_r = a^{-H} \cdot X_{at}$ with the time scale *at* and the *a* sub-series of process X_t fulfill the equation

$$X_{at} = a^H X_t, \qquad (5)$$

where equality is understood in the sense of *equality of the second order properties of the finite-dimensional distributions* [1], [7].

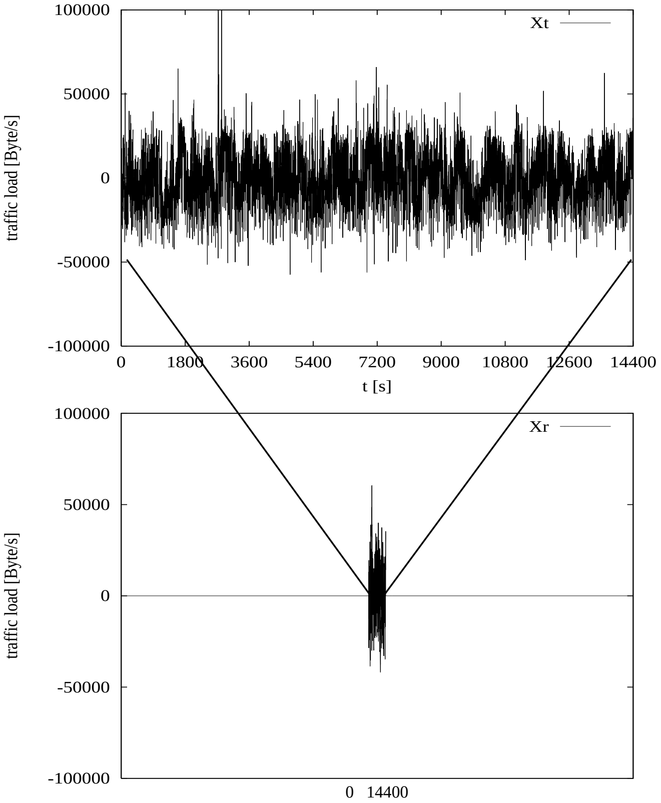

Figure 3: Example measurement X_t and its rescaled time series X_r.

The rescaled time series X_r for *a=30* of a measurement of real packet traffic is illustrated in figure 3.

The process $X^{(m)} = (X_k^{(m)}: k = 1,2,3,...)$ with the elements $X_k^{(m)} = 1/m\ (X_{km-m+1} +... + X_{km})$ with $k = 1,2,3,...$ has for all $m = 1,2,3,...$ the same variance

$$var(X^{(m)}) = \sigma^2 m^{-\beta}, \qquad (6)$$

where σ^2 is the variance of the process X_t and $\beta = 2(H+1)$.

2.3. Estimation of the self-similar parameter H

There are different ways to estimate the self-similar parameter *H*. The two most important and numerically most stable possibilities, the estimation of H by variance analysis and the estimation of H by R/S-statistic, are described in [10] and [7]. The resulting R/S-statistic for the experimental data series is shown in figure 4 (gradient of linear regression line), yielding $H = 0.85$.

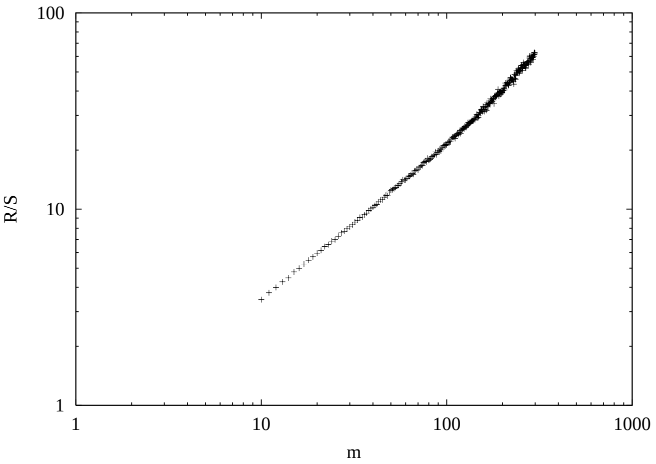

Figure 4: Estimation of the self-similar parameter H by R/S-statistic.

2.4. Off-Line Error detection

For error detection we check the validity of the self-similarity property of network traffic load given in equation (5). This is, according to our observations, a sufficient condition for normal (error free) behavior. The assumption is that in a given time series of network traffic load the self-similar parameter H can only change its value continuously and very slowly on account of the high number of traffic sources. We checked the validity of this assumption empirically with measurements in different hubs and found H reamaining (almost) the same even in the presence of errors.

For the validation of the self-similarity property we regard traffic load in two time scales, one higher scale X_t and one lower scale X_{at} with scale factor a. We use the transformation of the process with lower time scale

$$X_r = a^{-H} X_{at} \tag{7}$$

in order to calculate a reference distribution. We compare the distribution of a subperiods $X_1^s, \ldots, X_a^s$ of the time series (with the higher time scale of X_t) with the calculated reference distribution X_r.

If, with respect to their second order properties, the subseries X^S are indistinguishable from the reference series X_r, then X_t is self-similar [4]. To check this condition, we define the criterion Q which we derived from the χ^2 test [11]. First, we eliminate a possible first order deviation by normalizing the measured time series X_t to have zero mean. Next, we compute Q as

$$Q = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i}, \tag{8}$$

where k is the number of classes, n_i the number of values in the class i of the distribution of one of the a subseries X^s of X_t, and m_i the number of values in the class i of the distribution of the reference time series X_r. The number k of classes should be chosen that enough values will fall into each class (as a the thumb rule: at least five values per class). In our example with 480 values in X_r we chose k=16. We then choose an equidistant division of the interval $[-1.96\sigma, 1.96\sigma]$ for assignment of values to the k classes, where values outside the interval are assigned to the first or last class respectively and with σ as estimated standard deviation of X_r. The above partition heuristics is motivated by the fact that for normally distributed data 95% of it lie within that interval (X_r is usually not normally distributed, but the above estimation is in practice good enough to find a suitable class partition).

Note that the indicator Q may not have the same properties as Q^2 known from χ^2 test theory. This is not necessary, since in our approach Q is only used as an error indicator whose significant increase is interpreted to indicate an error in the network.

For off-line error detection, we compute the series of indicators $(Q_i)_i$ by comparing the reference distribution X_r with each subseries $X_i^s = (x_{(i-1)r+1}, \ldots, x_{ir})$, $i = 1, \ldots, a$.

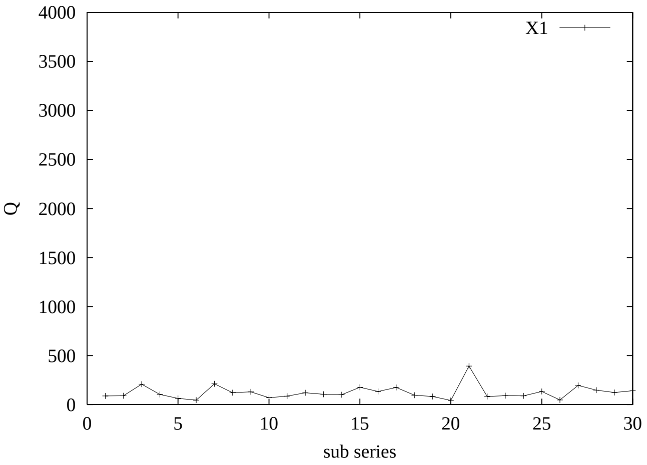

Figure 5: Example of the indicator Q for a self-similar second order behavior.

For the measured error-free data of figure 1, figure 5 shows the development of the indicator Q_i for all $a = 30$ sample subsets.

Error detection is finally performed by examination of *Q*. This can for example be done by a simple threshold. *Q* exceeding the threshold indicates the occurrence of an error in the network under observation. The threshold must be chosen appropriately in order to achieve a high error detection rate while keeping a low false alarm rate.

2.5. On-Line Error Detection

For off-line error detection the whole data set is split into a subseries whose distributions are then compared with the reference distribution. For *on-line* error detection, we consider at each discrete time *t* the subseries that consists of the latest *r* measured samples, with *r* the number of elements of the reference series: $X_t^s = (x_{t-r+1}, \ldots, x_t)$. By comparing the distributions of X_r and X_t^s we can compute a *moving* Q_t at each time *t* that indicates the occurrence of an error as soon as possible.

Since the moving *Q* describes a time series of *Q*-values we can also apply sophisticated techniques, for example the Page-Hinkley-test, a method to detect jumps in means of time series [12], [13].

3. Error Detection Experiments

3.1. Experimental Setup

To estimate the error detection capability of this approach, we implement a series of tests where we inject errors with a specified traffic load increasing or decreasing factor and a limited error duration. We then observe the deviation indicator *Q*. The following examples are based on one of the stationary parts of measurements on the hub (mbi.ira.uka.de) of the campus network at the University of Karlsruhe, shown in figure 1.

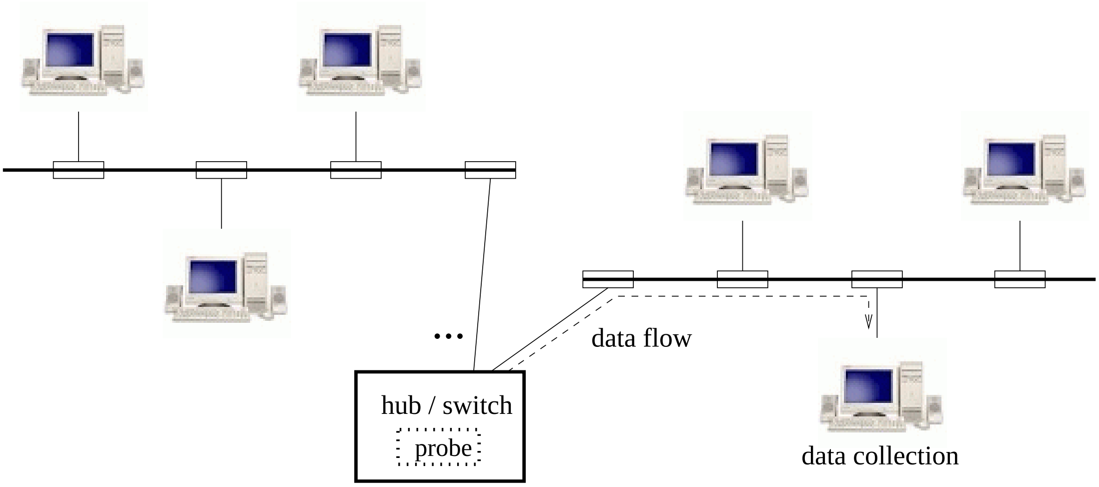

Figure 6: Experimental data collection setup.

Each hub contains several switching modules and a data acquisition probe from Newsbridge Networks which supports the full functionality of RMON1. The managing software used for data collection is the NetScout manager for SUN OS from Frontier, see figure 6.

As experimental parameters we use a scale factor of $a = 30$ and sample intervals of 1 second X_1 and 30 seconds X_{30} of the time series. In general, the sample interval and the scale factor *a* should be chosen in such a way that the resulting time series is stationary and that there are enough samples for the comparison of the distributions by equation (8).

3.2. Detection of Injected Errors

The injected errors increase or decrease the traffic load by a value $|F|$ during a limited time period T_e. The experiments in table 1 might simulate a few significant cases of real errors with different ratios R of the error to the traffic load average. The chosen error-values $|F|$ are typical for the abnormal behavior observed in our network during a long period. (In this example, R = 0.4 means about $|F|$=*25 KByte/s*.) We use values for the error duration T_e which are reasonable in relation to the length of the time series used. Independence from the time scale used is achieved by dividing the error duration through the sample period $T_e/\Delta t$. The errors are permanent during the given time period and lead to an increase or decrease in the traffic load, according to the sign in table 1.

TABLE 1 Injected errors.

error	sign	R	$T_e/\Delta t$	subseries	*comment*
F_1	+	0.4	480	15	one subseries affected
F_2	-	0.4	480	15	one subseries affected
F_3	+	0.8	200	15	short error
F_4	-	0.8	200	15	short error
F_5	+	0.4	1440	13, 14, 15	long error
F_6	-	0.4	1440	13, 14, 15	long error
F_7	+	0.4	3x480	7, 15, 26	several single errors
F_8	-	0.4	3x480	7, 15, 26	several single errors

In all measurements the location of an error in the time series is clearly detectable by an increased value of the deviation indicator Q (corresponding to equation (8)). In the test cases, the deviation indicator Q increases at least to triple the average value.

In figure 7 we show the result of the F_1 error injection experiment. The deviation indicator Q clearly increases for the subseries where the error was injected. The indicator Q of the error free subseries remains under a tolerance limit.

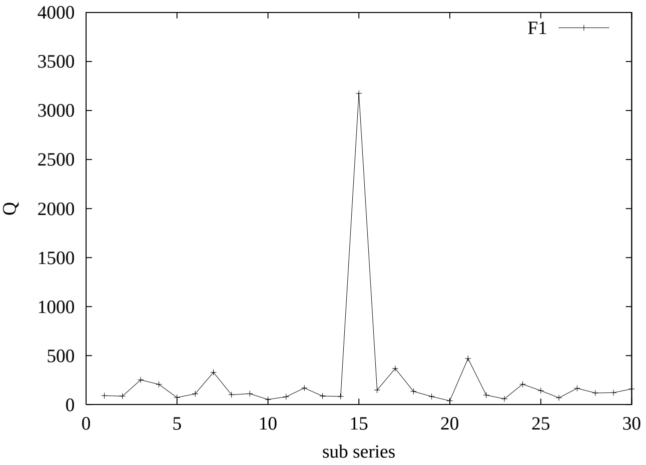

Figure 7: F_1 error injection experiment.

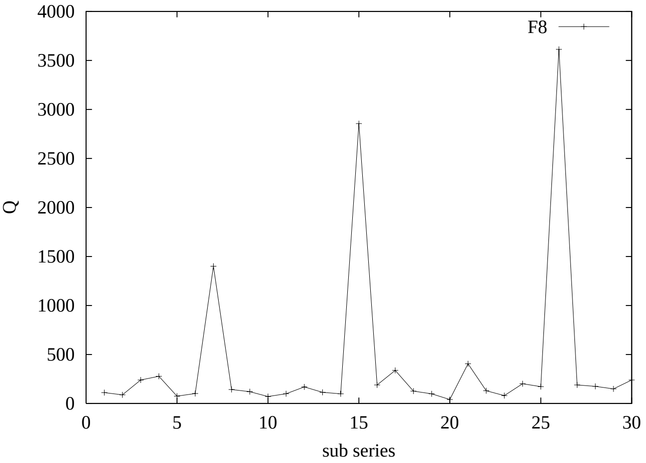

Figure 8: F_8 error injection experiment.

Several single errors, like error F_8, are indicated in each subseries where they occur (see figure 8). Depending on the original subseries, the injected error shows itself in different increases of Q. The deviation indicator Q seems to be very sensitive to the total error load $|F|$ and the error duration $T_e/\Delta t$, but we do not observe any difference for different signs of the error load, i.e., both, positive and negative errors are indicated in about the same increase of Q.

3.3. Detection of a Real Error

In figure 9 we provide another example of real traffic load of the same hub from another day, covering four hours at a sample rate of one pattern per second. At this day, a network error occurred at t=12075 s.

The results of off-line and on-line error detection are depicted in figure 10 and figure 11, respectively. Both, off-line and on-line detection show roughly the same increase in *Q*.

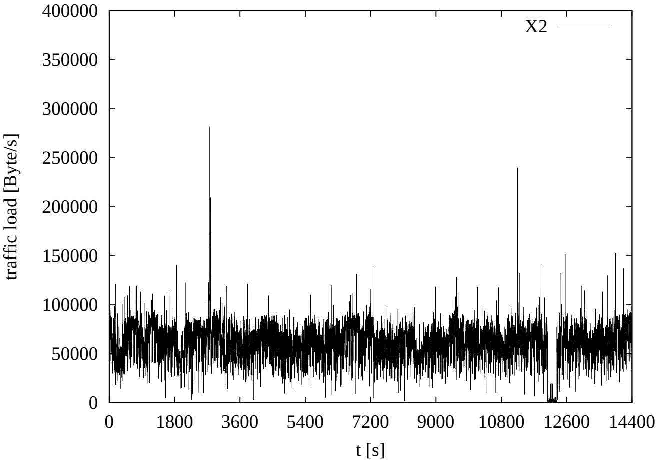

Figure 9: Measured data containing a real error.

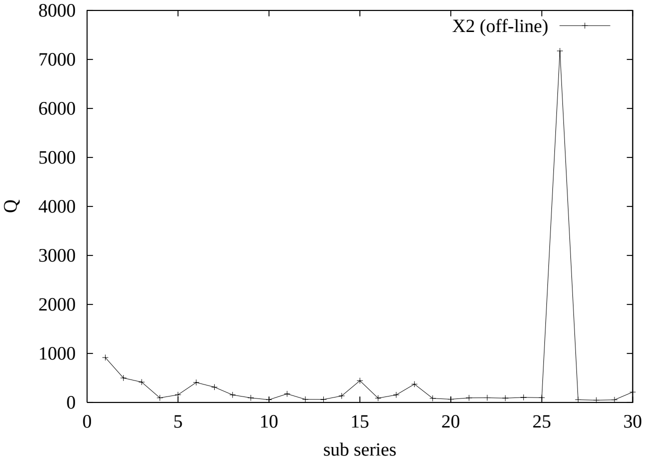

Figure 10: Indicator Q for the off-line error detection.

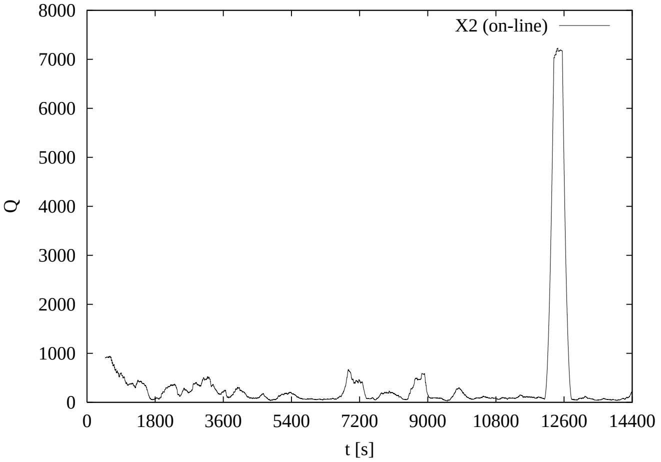

Figure 11: Indicator Q for the on-line error detection.

For the on-line detection of this example, the indicator crosses the level of $Q = 1000$ at t=12173 s and the level of 2000 at t=12213 s, leading to a detection latency of about 100 or 140 seconds, depending on the choice of the alarm threshold.

4. Conclusions

The idea was to empirically check whether the self-similarity properties of packet traffic are disturbed in presence of errors and whether such a disturbance can be detected. For this purpose we performed experiments on error free and errornous time series of Ethernet traffic measurements and designed error detection experiments.

We noticed that the real and injected *errors* did not or *did only marginally affect the Hurst parameter* H. Therefore a reference time series can be build even in the presence of errors. This was a (to us) surprising result that we did not expect.

Another main result was that the *self-similarity property is no more valid when an error occurs*. We can exploit this phenomenon for error detection by defining an indicator Q. We showed that we can locate an error in a traffic load time series by the significantly increased value of the Q that indicates the deviation of the traffic load distribution from the calculated reference distribution in means of the distributions' second order properties. According to our observations, the check of the validity of self-similarity property of network traffic in concentration points like hubs is a sufficient condition for normal (error free) behavior. This property seems to be independent from topology, protocol, and technology of the network and also independent from the distribution of traffic load. Therefore, the approach presented in this paper is universally applicable.

For off-line detection, the sample interval of the time series with lower time scale is in this approach equivalent to the error detection latency. The latency is considerably smaller the described on-line detection method.

The main advantage of this new approach is the considerably *higher sensitivity* compared with present error detection methods such as static traffic load thresholds. The implementation of the detection method is only based on traffic load data that are available in standard hubs. The computational cost is acceptable.

Future work should cover experiments to determine the most appropriate values for the time scale parameter *a* and the sampling rate of the traffic data measurement. Another issue will be to test the proposed detector under different network operating scenarios (like rush hour traffic, normal traffic, and week end traffic). These results shall yield a robust detector that is suited for industry application, for example to extend the RMON error detection functionality.

Acknowledgment

The authors thank Prof. Dr. W. Görke, chair of the Institute for Computer Design and Fault Tolerance at the University of Karlsruhe, Germany, where parts of this work have been performed. We also thank Prof. Dr. N. Henze for fruitful discussions on statistic detectors.

References

[1] BERAN, J.: ‘Statistics for long memory processes’ (Chapman & Hall, 1994)

[2] MANDELBROT, B.: ‘The fractal geometry of nature’ (Freeman, 1993)

[3] ERAMILLI, A., PRUTHI, P., WILLINGER, W.: ‘Recent Developments in Fractal Traffic Modelling.’ Proceedings of the St. Petersburg Regional ITC Seminar, June 1995, St. Petersburg, Russia

[4] WILLINGER, W., WILSON, D., LELAND, W.E., TAQQU, M.S.: ‘On the Self-Similar Nature of Ethernet Traffic (extended version).’ Proceedings of the IEEE/ACM Transactions in Networking, 1994, **2**, (1), pp. 1-15

[5] TAQQU, M.S., TEVEROSKY, V., WILLINGER, W.: ‘Is network traffic self-similar or multifractal?’ Technical report, Belcore and Boston University, 1996

[6] PRUTHI, P., ERAMILLI, A., WILLINGER, W.: ‘Self-Similarity in High-Speed Network traffic Measurements: Fact or Artifact?’ Proceedings of the 12th Nordic Teletraffic Seminar NTS12, 1995, Espoo, Finnland, pp.299-310

[7] WILLINGER, W., WILSON, D., LELAND, W.E., TAQQU, M.S.: ‘Self-Similarity in High-Speed Packet Traffic: Analysis and Modelling of Ethernet Traffic Measurements.’ Bellcore and Boston University, Statistical Science, February 1995, **10**, (1), pp. 67-85

[8] WILLINGER, W., WILSON, D., SHERMAN, R., W.E., TAQQU, M.S.: ‘Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level.’ IEEE/ACM Transactions on Networking, 1997, **5**, (1), pp. 71-96

[9] ALEXANDER, R., BROWNLEE, N., ZIEDINS, I.: ‘Modelling self-similar network traffic.’ Technical report, University of Auckland, 1995.

[10] BARTH, R.: ‘Fractale, Long Memory und Aktienkurse’ (Verlag Josef Eul, 1996)

[11] GIBBONS, J.D.: ‘Nonparametric Statistical Inference’ (McGraw-Hill, 1971)

[12] BASSEVILLE, M.: ‘On-Line Detection of Jumps in Means. In Detection of Changes in Signals and Dynamical Systems’, number 77 in Lecture Notes in Control and Information Science (Springer Verlag, 1986)

[13] DRABER, S., SCHULTZE, R.: ‘Detection of Jumps in Single-Channel Data Containing Subconductance Levels.’ Biophysical Journal, 1994, **67**, pp. 1404-1413