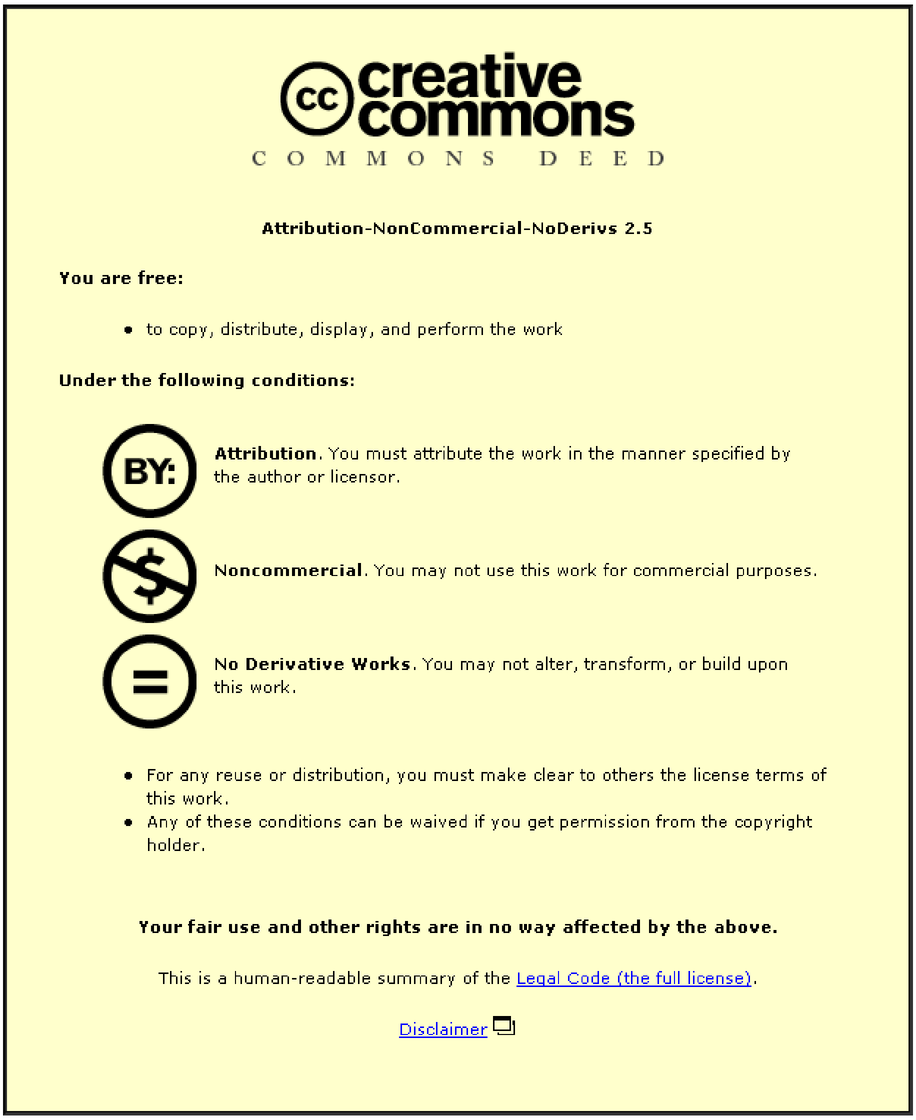

2009 Fourth International Conference on Systems and Networks Communications

Using Wavelets for Compression and Detecting Events in Anomalous Network Traffic

Konstantinos G. Kyriakopoulos, David J. Parish
Department of Electronic and Electrical Engineering
Loughborough University
Loughborough, LE11 3TU, U.K.
e-mail: {elkk, d.j.parish}@lboro.ac.uk.

***Abstract*—Monitoring and measuring various metrics of high-data rate networks produces a vast amount of information over a long period of time making the storage of the monitored data a serious issue. Furthermore, for the collected monitoring data to be useful to network analysts, these measurements need to be processed in order to detect interesting characteristics.**

In this paper wavelet analysis is used as a multi-resolution analysis tool for compression of data rate measurements. Two known thresholds are suggested for lossy compression and event detection purposes. Results show high compression ratios while preserving the quality (quantitative and visual aspects) and the energy of the signal and detection of sudden changes are achievable.

***Keywords*-Wavelets, compression, detection, computer networks, measurements**

I. Introduction

The performance monitoring and measurement of communication networks is an increasingly important area as demands on and threats to such networks increase. Monitored network data allows network managers and operators to gain valuable insight into the health and status of a network, and if interpreted correctly, can assist in planning upgrades and remedial action to keep the network operating in a near optimum manner.

Whilst such data is useful for real-time analysis, there is often a need to post-process historical network performance data. Thus, storage of the monitored data then becomes a serious issue as network monitoring generates significant quantities of data.

Furthermore, for the collected monitored data to be useful to administrators, these measurements need to be analysed and processed in order to detect interesting characteristics such as sudden changes. Identifying such characteristics in large amounts of data has been an interest of network researchers for many years and is not an easy task [1], [2], [3], [4], [5].

This paper considers the above issues and proposes the use of the Wavelet Transform as a multi-resolution analysis tool for compression of data rate measurements. The offered compression ratio is driven by the need to preserve several statistical characteristics of the original signal. These characteristics are: the general quality of the signal (PSNR), mean, standard deviation, scaling behavior and the long range dependency.

Furthermore, by applying a secondary threshold, an automated tool is presented that filters the wavelet domain coefficients and keeps only those that represent a significant change in the time domain. The tool generates graphs where the sudden changes are pinpointed [1]. This tool is meant to be a lightweight additional benefit from the wavelet coefficients as an initial step in detecting anomalous sudden changes.

Previous work by the authors on quality controlled compression has examined data rate signals monitored from a national ISP and delay signals generated in a testbed network simulating the time of day characteristics of a commercial network [6]. A comparison of two known wavelet threshold estimation techniques is presented in [7] and the performance of two common threshold application methods is discussed in [8]. The advantages of wavelets for compression and event detection are thoroughly discussed in [1].

This paper differentiates from previous work as it focuses on examination of anomalous data collected from the UCSD network telescope. The UCSD network telescope is a monitoring point that collects around 1/256th of all IPv4 address on the internet and is globally routed. A network telescope (a.k.a darknet) filters all legitimate traffic and allows only anomalous traffic including misconfigurations [9]. In this paper the examined traffic consists of data collected while the Witty worm was propagating [10].

In 2004 the Witty worm began attacking and affecting computer hosts by exploiting vulnerabilities in software designed to enhance, ironically, network security. In contrast to previous worms, Witty worm contained a destructive payload and infected in the first 30 minutes after its launch 12,000 computers. At the same interval, it generated 90 Gbps of UDP traffic. Witty worm initiates its attacks very rapidly as fast as possible allowed by the infected hosts internet connection. Witty worm not only propagates to other hosts having the same vulnerabilities but also overwrites sections of the computer's hard disk [11].

This paper shows the advantages and efficiency of the proposed algorithms (backed up by the results in section 3) applied on network traffic collected while anomalous activities were present (specifically the witty worm propagation). In contrast, previous work by the authors examined network traffic from experimental research networks, national ISP and emulated data in test beds.

The rest of the paper is structured as follows. In section 2 are

978-0-7695-3775-7/09 $26.00 © 2009 IEEE
DOI 10.1109/ICSNC.2009.70

IEEE computer society

presented the algorithms used while running the experiments and producing the results. This includes two thresholds for compression and event detection along with methodologies for estimating the long range dependence and the scaling behaviour of the signals before and after the compression. In section 3 the results are presented and finally, conclusions and ideas for future work are given in Section 4.

II. Methodology

In this section, the two thresholds for the compression and event detection tasks are presented. For the analysis part, the Haar wavelet was used as the mother wavelet for the analysis because it has the following advantages: It is conceptually simple, fast and is memory efficient [6].

The main methodology is presented in Fig. 1. After the wavelet analysis, the two thresholds are applied to the coefficients separately. Normalization and run length encoding are applied on the filtered coefficients in order to increase the compression ratio (C.R.) [6], [1].

The traces examined were collected from the CAIDA Dataset on the Witty Worm. The first two traces were used with the tcpstat tool in order to produce a data rate time series using an aggregation window of 1 sec.

A. Compression Threshold

[12] proposes an adaptive thresholding technique that is calculated from the absolute value of the non-zero wavelet coefficients. This scheme is not based on signal denoising but rather tries to statistically identify significant coefficients. The threshold is estimated from the following equation:

$$T = \begin{cases} 2 * \mu, & if \sigma > \mu \\ \mu - \sigma, & if \sigma \leq \mu \end{cases} \tag{1}$$

where σ is the standard deviation and μ is the mean of the absolute value of the non-zero detail coefficients.

B. Event Detection Threshold

In order to isolate the coefficients that reflect a sudden change in the original time series, a threshold based on the Donoho - Johnstone universal threshold (aka VisuShrink) is utilized [13]. For each level of decomposition the threshold is rescaled by a level-dependent estimation of the level's noise σ_{lev} [1].

Thus, the level dependent threshold is of the following form:

$$T_{lvl} = \sigma_{lvl} \times \sqrt{2 \log_e n} \tag{2}$$

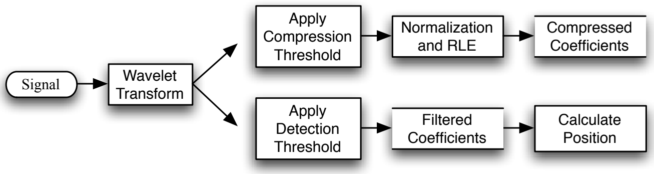

Fig. 1. Algorithm Flow Chart

Where n is the number of the total wavelet domain coefficients and σ_{lev} is the level-dependent noise standard deviation. As suggested by [13], the median absolute deviation ($\hat{\sigma}_{lvl}$) is used as a robust estimation for the noise standard deviation.

$$\hat{\sigma}_{lvl} = \frac{median(|cDetail_{lvl}|)}{0.6745} \tag{3}$$

where $cDetail_{lvl}$ are the detail coefficients for level *lvl*.

C. Estimating the Long Range Dependence

1) The Rescaled Adjusted Range Statistic: For the estimation of the Hurst parameter an algorithm was developed based on the rescaled range (R/S) statistic. The R/S statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation [14].

For a given set of n samples $(X_1, X_2, X_3, ..., X_n)$, with sample mean $X(n)$ and sample standard deviation $S(n)$, the classic rescaled adjusted range statistic for that particular set of samples is given by the following equation [15]

$$\frac{R(n)}{S(n)} = \frac{1}{S(n)}\Big[max(0, W_1, W_2, ..., W_n) - min(0, W_1, W_2, ..., W_n)\Big] \tag{4}$$

where

$$W_k = (X_1 + X_2 + ... + X_k) - k\bar{X}(n) \tag{5}$$

and $k = 1, 2, 3, 4, ..., n$

By combining 4 and 5 we can get the following equation [14]:

$$\frac{R(n)}{S(n)} = \frac{1}{S(n)}\Bigg[max_{1<k<n}\Bigg(\sum_{j=1}^{k}\big[X_j - \bar{X}(n)\big]\Bigg) - min_{1<k<n}\Bigg(\sum_{j=1}^{k}\big[X_j - \bar{X}(n)\big]\Bigg)\Bigg] \tag{6}$$

Hurst observed that the following equation well represents the relation between the R/S statistic expectation and the Hurst parameter [15]:

$$E\left[\frac{R(n)}{S(n)}\right] = an^{H} \qquad as \quad n \rightarrow \infty \tag{7}$$

where α is a constant. By taking the log transformation of equation (7) we have:

$$log\left(E\left[\frac{R(n)}{S(n)}\right]\right) = log(a) + Hlog(n) \tag{8}$$

The final step involves plotting the log of $R(n)/S(n)$ versus the log of the sample region $[1, n]$. This process produces the R/S plot also called a pox diagram.

By performing regression analysis, a least squares line is fitted to the points of the R/S plot. However, the edge values of the sample region are not considered. This is because for the smallest samples, the R/S values are biased due to short-range correlations; whereas for the largest samples the R/S

values are statistically insignificant [16], [14]. The slope of the regression line is an estimate of the Hurst parameter [15], [14].

2) Goodness of fit calculation: In order to estimate the error of the drawn regression line, the goodness of fit of the linear regression was calculated. The "goodness of fit", denoted by r^2, has no units and is given by the following expression [17]:

$$r^2 = 1 - \frac{SS_{reg}}{SS_{tot}} \tag{9}$$

where

$$SS_{reg} = \sum_{i=1}^{n} (y_i - y_i')^2 \tag{10}$$

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y_i})^2 \tag{11}$$

The variable i represents a sample and takes values between 1 and n, where n is the total number of samples.

- y_i is the actual log R/S value of sample i
- y_i' is the regression line estimate value for sample i
- $\bar{y}$ is the mean of log values of R/S statistic for the total range of samples

3) The Rescaled Range Statistic Block Algorithm: A variation of the classic R/S statistic was used as a Hurst estimation algorithm (Figure 2). This involved dividing the whole data set of n samples into $\lfloor n/N \rfloor$ non-overlapping sample blocks; where $1 \leq N \leq n$ is the sample block size. The calculations for the rescaled range value presented in section II-C1 apply recursively to each sample block and ultimately produce $\lfloor n/N \rfloor$ intermediate R/S values. These intermediate values are summed and divided by the number of sample blocks in order to find the average R/S value for the current sample block size. The whole process repeats for all values that the sample block size can take ($1 \leq N \leq n$) [15].

The above R/S statistic variation uses non-overlapping blocks of samples. However, there are other variations that use overlapping blocks or are limited to data sizes that are a power of two [18].

As was mentioned in section II-C1, for the regression analysis the edge values should not be considered. Thus, in order to exclude the edge values, a window that includes 90% of the data trace values was utilized. The window slides through the data trace and for each position of the window a regression line is calculated.

The Hurst parameter is estimated based on the gradient of the regression line with the best goodness of fit value. The R/S algorithm flowchart is presented in Fig. 2:

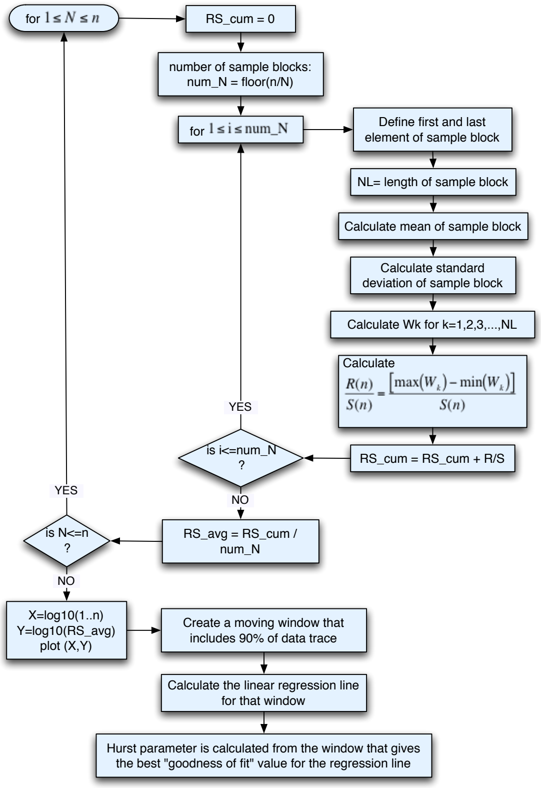
for $1 \leq N \leq n$
RS_cum = 0
number of sample blocks: num_N = floor(n/N)
for $1 \leq i \leq$ num_N
Define first and last element of sample block
NL= length of sample block
Calculate mean of sample block
Calculate standard deviation of sample block
Calculate Wk for k=1,2,3,...,NL
Calculate $\frac{R(n)}{S(n)} = \frac{[\max(W_k) - \min(W_k)]}{S(n)}$
RS_cum = RS_cum + R/S
is i<=num_N ?
YES
NO
RS_avg = RS_cum / num_N
is N<=n ?
YES
NO
X=log10(1..n) Y=log10(RS_avg) plot (X,Y)
Create a moving window that includes 90% of data trace
Calculate the linear regression line for that window
Hurst parameter is calculated from the window that gives the best "goodness of fit" value for the regression line

Fig. 2. Rescaled Range Statistic flow chart

D. Energy and Scaling Behavior

The *energy function plot* [19] was used in order to examine the preservation of the scaling behavior after the compression. A statistic known as the *energy function* E_j indicates the average energy of the arrival process contained at scale j and is defined by [19]:

$$E_j = \frac{1}{N_j} \sum_k |d_{j,k}|^2, \qquad j = 1, 2, ..., n \tag{12}$$

where N_j is the number of wavelet ("detail") coefficients at scale j.

The energy plot can be generated by plotting $log(E_j)$ against the scale j from finer to coarser scales. Intuitively, this plot illustrates the scaling behaviour of the underlying time series (such as a traffic arrival process) at different time-scales.

This approach has been used by many researchers [20], [21], [22], [23] for evaluating their traffic models with respect to the correct reproduction of the scaling structure of the modeled traffic. By comparing the busrtiness of the synthesized and original traffic at a variety of scales, researchers can evaluate how closely their model matches the correlation structure of the modeled network.

III. Results

The quality of the reconstruction signal was compared with the original by using the PSNR value calculated from

$$PSNR = 10 * log\left(\frac{MAX^2}{MSE}\right) \tag{13}$$

where MAX is the maximum value of the original signal and MSE is the mean square error calculated from

$$MSE = \frac{1}{N}\sum_{i=0}^{N-1}\left|x_i - \bar{x}_i\right|^2 \qquad (14)$$

where x_i is the i^{th} sample from the original signal, $\bar{x}_i$ is the i^{th} sample of the reconstructed signal and N is the total number of samples. Empirically, PSNR values more than 44 dB give very good visual results, PSNR values less than 35 dB loose some of the important signal characteristics while PSNR values less than 30 dB are not acceptable for such signals.

On average, for the four signals, the compression ratio (C.R.) starts at around 10:1 at level 1 of the decomposition and stabilizes around 17:1 at level 10. The PSNR quality performance begins at 48 dB and stabilizes around 40 dB at level 10. In Fig. 3 the reconstruction for the first signal is shown in comparison to the original. The error between them is also given for easier judgement. The example in Fig. 3 gave the smallest PSNR value but the highest C.R.

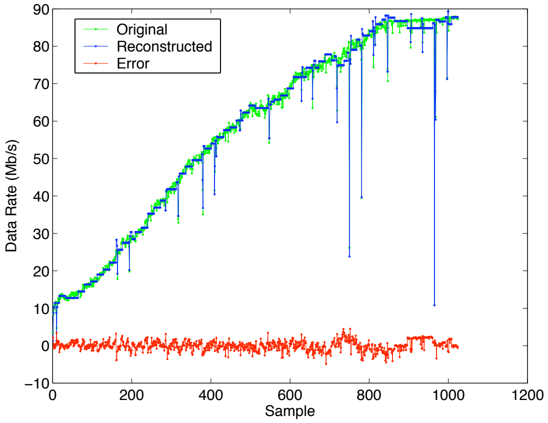

Fig. 3. Original and reconstructed signal 1 (at level 10) and the error (lower line) between them. PSNR = 36.5 and C.R. = 19.7

Table I shows the C.R. with respect to the original file size (C.R. orig.), C.R. with respect to the bzip2 compressed file size, PSNR values and absolute percentage relative error for the mean, standard deviation and Hurst values of each reconstructed signal at level 10. The proposed compression technique achieves very good results in compression and correctly captures the mean, standard deviation and visual quality (see Fig. 3) without compromising time properties

Table II shows the error of the energy for each scale of decomposition and for all examined signals in decibels. The reconstruction preserves very closely the scaling characteristics of both types of signals and even in the worst case scenarios, it offers better results than some advanced network traffic generation models. As a comparison, reference [20] provides a model with the aim of reproducing the scaling nature of real traffic. Even though the authors do not provide their error quantitatively, it can be seen from their figures that the error in some cases is several dB, much greater than that in table II where the error is less than 0.3 dB.

The proposed procedure detects the sudden changes in the examined data rate signals. In Fig. 4 and Fig. 5, the original examined signal is presented on top and the detected changes on the bottom of the figures. The significant coefficients produced after the thresholding are normalized and then plotted in the time instance that they represent.

In the first signal (Fig. 4), the data rate increases until it reaches around 90 Mbps. There are various sudden drops of data rate throughout the signal. The higher the detection value the bigger the change. In the second signal (Fig. 5), data rate has stabilized but sudden drops continue to exist. For both signals, the algorithm managed to detect these events.

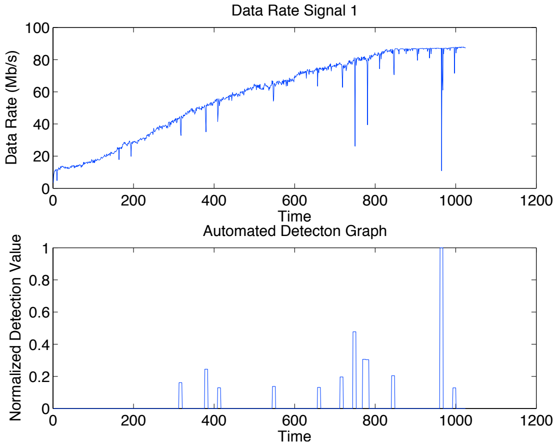

Fig. 4. Detecting Changes in Data Rate Signal 1

IV. Conclusions

In this paper the wavelet transform is used as a multiresolution analysis tool and two known thresholds are applied for compressing traffic measurements and detecting events in

TABLE I
Absolute Percentage Relative Error of Statistics for Reconstructed Signals at Level 10.

	Signal 1	Signal 2	Signal 3	Signal 4	Avg.
C.R. orig.	19.74	17.30	17.45	15.79	17.57
C.R. bzip2.	6.45	3.72	3.20	2.87	4.06
PSNR (dB)	36.54	39.93	42.07	42.40	40.24
Mean (%)	0.02	0.02	0.02	0.02	0.02
Std (%)	0.11	1.69	1.16	1.42	1.10
Hurst (%)	1.68	13.59	13.36	2.29	7.73

TABLE II
Difference of Energy in Decibels per Scale for Reconstructed Signals at Level 10.

	L1	L2	L3	L4	L5	L6
Signal 1	0.04	0.12	0.07	0.27	0.00	0.01
Signal 2	0.04	0.02	0.01	0.12	0.24	0.26
Signal 3	0.01	0.01	0.03	0.05	0.21	0.19
Signal 4	0.01	0.03	0.03	0.05	0.29	-0.03

anomalous communication network data traffic. The compression, even though lossy, preserves the significant and interesting characteristics of the original signal (mean, standard deviation, PSNR) without significantly affecting the temporal characteristics (long range dependency and scaling behavior).

On average, on the deepest analysis scale (level 10), the compression scheme offers a C.R. of 17.57 relative to the original file size and 4 times more compression than bzip2. The average PSNR on the same scale is 40.24 dB. The time adaptive characteristic of wavelet analysis makes it a suitable tool for examining an environment that is time varying such as communication network traffic. Additionally, wavelet analysis can perform a local analysis and provides both frequency and time resolutions, which are necessary for the anomaly detection procedure [1].

The compression algorithm is already implemented in a real-time computer network monitoring tool for on-line compression of time series measurements [6]. For this task, the open source CoMo tool was used. CoMo is a passive monitoring platform developed for the purpose of measuring performance metrics of high speed links and replying to real time queries [24].

The calculation time of the algorithm is not an issue for the on-line implementation of the compression and anomaly detection algorithms. This is because a few milliseconds of processing time are adequate in both cases. In order to produce a data rate signal from the captured data packets, aggregation of packet length needs to take place every second, i.e. each sample is per second. Capturing such a signal of, say, 1024 measurement points would require 1024 seconds. Thus, there is a window of around 17 minutes for the analysis and detection phases to complete.

As for future work, the secondary threshold for event detection could easily be imbedded in a CoMo module. CoMo will be responsible for capturing data packets and producing measurements of the network, while the module of the proposed algorithm will detect anomalies in the analyzed captured signal [1].

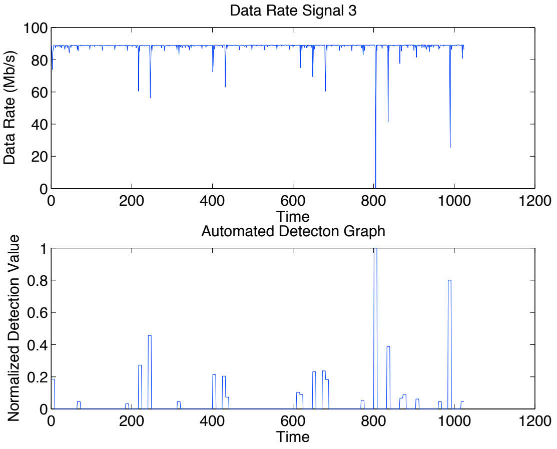

Fig. 5. Detecting Changes in Data Rate Signal 2

Acknowledgements

The authors would like to acknowledge CAIDA for providing the CAIDA Dataset on the Witty Worm - March 19-24, 2004, Colleen Shannon and David Moore, http://www.caida.org/passive/witty/.

References

[1] Konstantinos G. Kyriakopoulos and David J. Parish, "Automated detection of changes in computer network measurements using wavelets," in *Proceedings in ICCCN Workshop on Performance Modeling and Evaluation in Computer and Telecommunication Networks (PMECT 2007)*, Honolulu, Hawaii, USA, 13-16 August 2007, IEEE.

[2] Seong Soo Kim, A. L. Narasimha Reddy, and Marina Vannucci, "Detecting traffic anomalies through aggregate analysis of packet header data.," in *NETWORKING*, Nikolas Mitrou, Kimon P. Kontovasilis, George N. Rouskas, Ilias Iliadis, and Lazaros F. Merakos, Eds. 2004, vol. 3042 of *Lecture Notes in Computer Science*, pp. 1047–1059, Springer.

[3] A. Dainotti, A. Pescape, and G. Viorgio, "Wavelet-based Detection of DoS Attacks," in *Proceedings of IEEE Global Telecommunications Conference*, 2006.

[4] J.G.G.H.X. Yao and RKC Chang, "Anomaly detection of network traffic based on wavelet packet," in *Communications, 2006. APCC'06. Asia-Pacific Conference on*, 2006, pp. 1–5.

[5] G. Carl, R.R. Brooks, and S. Rai, "Wavelet based Denial-of-Service detection," *Computers & Security*, vol. 25, no. 8, pp. 600–615, 2006.

[6] Konstantinos G. Kyriakopoulos and David J. Parish, "A system for online compression of high speed network measurements," *International Journal of Internet Protocol Technology*, vol. 3, no. 2, pp. 95 –106, 2008.

[7] Konstantinos G. Kyriakopoulos and David J. Parish, "Wavelet compression techniques for computer network measurements," in *Proceedings of the Fourth IASTED International Conference on Signal Processing, Pattern Recognition, and Applications 2007 (SPPRA 2007)*, Innsbruck, Austria, 14-16 February 2007, IASTED, pp. 109–115, Acta Press.

[8] Konstantinos G. Kyriakopoulos and David J. Parish, "Wavelet compression of network delay measurements," in *Proccedings of the 3rd IADAT International Conference on Telecommunications and Computer Networks (IADAT-tcn 2006)*, Portsmouth, U.K., 27-29 September 2006, pp. 115–119, IADAT.

[9] Colleen Shannon, "Passive data collection: UCSD network telescope," http://www.caida.org/data/passive/network_telescope.xml Page last visited on 29/2/2008.

[10] Colleen Shannon and David Moore, "The caida dataset on the witty worm," March 19-24 2004, Support for the Witty Worm Dataset and the UCSD Network Telescope are provided by Cisco Systems, Limelight Networks, the US Department of Homeland Security, the National Science Foundation, DARPA, Digital Envoy, and CAIDA Members. http://www.caida.org/data/passive/witty_worm_dataset.xml.

[11] CAIDA, "The spread of the witty worm," Website, November 2008, http://www.caida.org/research/security/witty/.

[12] Savita Gupta and Lakhwinder Kaur, "Wavelet based image compression using daubechies filters," in *8th National conference on communications, I.I.T.*, 2002.

[13] David L. Donoho and Iain M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[14] David Nawrocki, "R/s analysis and long term dependence in stock market indices," *Managerial Finance*, vol. 21, no. 7, pp. 78–91, 1995.

[15] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1–15, 1994.

[16] O. Rose, "Estimation of the hurst parameter of long-range dependent time series," Tech. Rep., University of Würzburg, Institute of Computer Science, February 1996, Research Report Series 137.

[17] M. S. De Silva, *Emergence in Active Networks*, Ph.D. thesis, Loughborough University, 2004.

[18] Ian Kaplan, "Estimating the hurst exponent," Website, http://www.bearcave.com/misl/misl_tech/wavelets/hurst/index.html. Page last visited on 10/05/2007.

[19] Patrice Abry and Darryl Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Transactions on Information Theory*, vol. 44, no. 1, January 1998.

[20] Kashi Venkatesh Vishwanath and Amin Vahdat, "Realistic and responsive network traffic generation," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 111–122, 2006.

[21] Joel Sommers and Paul Barford, "Self-configuring network traffic generation," in *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, New York, NY, USA, 2004, pp. 68–81, ACM Press.

[22] J. Ridoux, A. Nucci, and D. Veitch, "Seeing the difference in ip traffic: Wireless versus wireline," in *Proceedings of IEEE INFOCOM 2006*, Barcelona, Spain, April 2006.

[23] C. Rolland, J. Ridoux, and B. Baynat, "Litgen, a lightweight traffic generator: application to p2p and mail wireless traffic," in *Proceedings of the Passive and Active Measurement Conference (PAM 2007)*, Louvain-La-Neuve, Belgium, April 2007, vol. 4427 of *LNCS*, pp. 52–62, Springer.

[24] Gianluca Iannaccone, Christopher Diot, Derek McAulley, Andrew Moore, Ian Pratt, and Luigi Rizzo, "The como white paper," Tech. Rep., INTEL, 2004, Page last visited 22/08/05.