Wavelet-Based Unwanted Traffic Time Series Analysis

Kriangkrai Limthong
Kasetsart University
Bangkok 10900 Thailand
e-mail:g5065373@ku.ac.th

Fukuda Kensuke
National Institute of Informatics
PRESTO JST
Tokyo 101-8430 Japan
e-mail:kensuke@nii.ac.jp

Pirawat Watanapongse
Kasetsart University
Bangkok 10900 Thailand
e-mail:pw@ku.ac.th

Abstract

Identifying traffic anomalies precisely and instantaneously is critical to network stability. Most studies have focused on analyzing unwanted traffic from a Darknet system. However, conventional methods of detecting anomalous activities from these data are not applicable to detection. We apply Discrete Wavelet Transform (DWT) techniques for traffic signal decomposition and examine unknown anomalous activities from unwanted traffic data. Our work focuses on three unwanted traffic packets: TCP SYNs, TCP SYN/ACKs, and UDP packets and on three intervals: 10-ms, 100-ms and 1-s. Furthermore, we discuss the features of this approach and consider some of its possible realizations. Our goal is to reveal properties when wavelet techniques are used to detect network anormalies behavior.

Index Terms—Wavelet Decomposition, Darknet, Time Series,Unwanted Traffic, Anomalies Detection

1. Introduction

Identification of anomalies in network traffic have been extensively research topics in these days. Because they are responsible for network congestion and consumes resources utilization. The forms and causes of anomalies can vary considerably. For example, they can take the from of worms [1], [2], Denial of Service (DoS) attacks [3], port scans [4], flash crowds [5], etc. Additionally, new varieties of anomalies are appearing every day.

One popular method of detecting anomalies in network traffic that are due to some form of attack is to monitor unused network addresses. This method has been called Darknet [6], network telescope [7], internet motion sensor [8], or background radiation monitor [9]. If there were an efficient method for analyzing unwanted traffic, it would become possible to detect anomalies or to assist in identifying their causes. Thus, as a way of detecting and investigating malicious network activity, we need an effective method for analyzing data from Darknet monitoring systems.

In this paper, we present a proficiently method for estimating and analyzing unwanted traffic behaviors that is based on the Discrete Wavelet Transform (DWT), time series analysis, and statistics approach. We focus on three types of connection request and response from an attacker or misconfigured host. The packets are TCP SYNs, TCP SYN/ACKs, and UDP packets. In addition, we analyze and compare the correlation coefficients for three windows sizes (10-ms, 100-ms and 1-s).

The contribution of this paper is twofold. First, we address significance of a parameter such as the window size interval and find a DWT level that is appropriate for detecting anomalies activity from unwanted traffic data. Second, we compare the relevances of the DWT levels for each window size by computing their correlation coefficients. In this work results suggest ways and guidance for researchers who may use anomaly detections based on the DWT method to well understand the macroscopic behavior of unwanted traffic.

The remainder of this paper is structured as follows. Section II gives an overview of related work; we summarize the previous studies on using wavelet-based analysis for anomaly detection and introduce about a Darknet system. Section III describes the wavelet decomposition, measurement data, and our methodology. Section IV presents our results, and Section V discusses them. Finally, we present our conclusions and mention future work in Section VI.

2. Related Work

Anomaly detection and identification techniques have been studied for many years [10]-[14]. Many of the approaches rely on known statistical properties of normal traffic when the observed traffic deviates significantly from the normal behavior. The work of Kim and Reddy in [15] and Kompella et al. in [16] are examples of statistical approaches. The cumulative sum (CUSUM) algorithm to identify traffic changes

978-0-7695-3504-3/08 $25.00 © 2008 IEEE
DOI 10.1109/ICCEE.2008.106

was presented by Tarkakovsky et al. in [17]. Another statistical method, by Dewaele et al. in [18], uses sketch techniques and Gaussian marginal distribution modeling as a means of extracting hidden anomalies from a large-scale packet trace database.

A second line of work focuses on signal processing. The method of Thottan and Ji in [19] detects abrupt changes in signals. Lakhina et al. in [20] propose a solution method based on signal processing and Bayesian models to diagnose anomalies. Both methods detect anomalies behavior by monitoring statistical changes in signals.

A third line of work, such as the work of Barford et al. in [21], focuses on wavelet-based solutions. Carl et al. in [22] apply wavelets transform for detecting change-points in the CUSUM statistic. Hamdi and Boudriga in [23] and Xunyi et al. in [24] devise wavelet techniques for detecting DoS attacks. Lu et al. in [25] study wavelet basis functions that have an important impact on the intrusion detection performance.

Besides, much of the previous Darknet research has been covered in [26]-[28]. The Darknet is a portion of the routed IP address space in which no legitimate traffic exists. Moore et al. in [26] analyze the Code-Red internet worm, by using data from the Darknet system. Yegneswaran et al. in [27] proposed honeynets data analysis for daily network security monitoring. Soto in [28] analyses a range of class B and C subnets with the intent of discovering the types of variability that are characteristic of unwanted traffic from the Darknet.

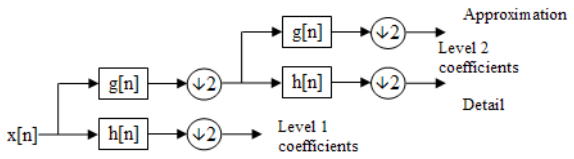

Fig 1. Wavelet decomposition tree

3. Analysis Methodology

This section describes the methods and tools we use to measure and analyze unwanted traffic data.

3.1. Measurement Data

Our data set comprised pcap packet traces captured at the Darknet in Japan for 1 week on Dec. 2007. We investigated three types of unwanted traffic. TCP SYNs packet are mainly generated by flooding, viruses, worms or port scanners. TCP SYN/ACKs packets are known to be backscatter or background radiation of DoS or Distributed DoS (DDoS) attacks, caused by an original packet with a spoofed source address. UDP packets are often due to flooding, viruses, or worms.

3.2. Wavelet Decomposition

Wavelets are basis functions used in representing data or other functions that can achieve good frequency resolution at low frequencies and good time resolution at high frequencies [29]. They are different from Fourier Transform (FT) which are suited to only the study of stationary signals, and Short-Time Fourier Transform (STFT) that localize the Fourier analysis by using a sliding window. The major limitation of STFTs is that they give either a good frequency resolution or a good time resolution depending upon the window size.

The wavelets function are generated from a single basic wavelet $\Psi(t)$, the so-called *mother wavelet*, by scaling and translation.

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j} \right) \qquad (1)$$

Equation (1) is the Discrete Wavelet Transform. J and k are integers. The scale factor s0 > 1 is a fixed dilation step. The translation factor tau0 depends on the dilation step, and the factor square s0 is for energy normalization across different scales.

Our propose is to study the use of multiple-level wavelet decomposition for the time series problem [30]. We need a decomposition that can be computed quickly and easily. Therefore, we used a Haar mother wavelet to decompose the time series signal. Wavelet decomposition is a wavelet transform where the signal is passed through more filters than the DWT. The wavelet decomposition tree is shown in Figure 1. A time series signal x[n] is broken down into g[n] approximation components of low resolution and h[n] detail components of high resolution. In this work, we broke down the signal into lower and high resolution components on 15 levels.

3.3. Implementation

The analysis method was implemented from the GNU Compiler Collection (GCC) 4.2.3 on the Ubuntu 8.0.4 linux platform. We read packet trace files through the libpcap libraries. In wavelet analysis section, the Harr wavelet was adopt for mother of wavelet from GNU Scientific Library (GSL), and it was selected for its simplicity. Finally, our data analysis and presentation use Gnuplot 4.2.2.

4. Results

The first step was to develop an automated method for converting pcap files from the Darknet into time

series data. Figure 2 shows the number of unwanted TCP SYNs packets in time series intervals of 10-ms, 100-ms and 1-s. The data in different intervals were similar. However, the 1-s interval data was finer than other interval times with log scale on the y-axis.

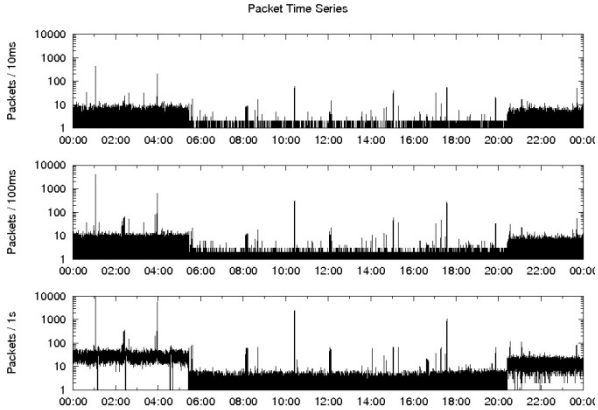

Fig 2. TCP SYNs packet time series in different window size of 10-ms, 100-ms and 1-s on 20-Dec-2007

In the second step, we evaluated the histograms, the proportion of unwanted packet cases for each interval time, and the Complementary Cumulative Distribution Function (CCDF), which is the probability of the number of packets per interval. In Figure 3, we plots the histograms and CCDF of TCP SYNs packets for each interval. The y-axis represents frequency for the histogram function and probability for the CCDF function. The x-axis is the number of packets per interval time.

TABLE I Anomalous Properties

Anomaly Label	Saddr	Daddr	Sport	Dport
Scan Port 1	-	-	-,*	*
Scan Port m	-	*	-,*	-
DoS	-	-	*	-,*
DDoS	*	-	*	-,*
Flash Crowd	*	-	*	-
Worm,Virus	*	*	*	-

Next, we considered the number of anomalies per interval time. Table I lists the anomalous properties used in our analysis, i.e., source IP address, destination IP address, source port, and destination port. (-) sign, that is means the properties are static, whereas (*) sign indicates they vary.

As shown in Figure 4, we plots the number of anomalies from TCP SYNs packets per time interval of 10-ms, 100-ms and 1-s.

Finally, we decomposed the original packet time series into low-frequency and hight-frequency parts. The original packet time series are taken from the Darknet data shown in Figure 1. However, each output has half the frequency band of the input so the frequency resolution has been doubled. The low-frequency content is the most important part. It is what gives the anomalous behavior its identity. Nevertheless, we can determine abrupt changes in the signal from abnormal traffic behavior in high-frequency.

Figure 5 shows the low-frequency of wavelet decomposition of levels 3, 7 and 11. Situated at low level, the details of the average signal are out of sight, but a huge signal was identified. The high level of wavelet decomposition is grainier than the original; it can not be used describe anomalous activities. We found after some experimentation that level 7 with a 1-s interval time was suitable for showing anomalous behavior.

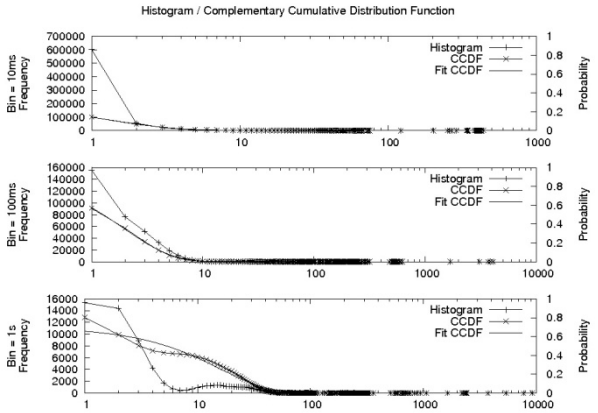

Fig 3. Histogram and Complementary Cumulative Distribution Function of TCP SYNs packets in different window size of 10-ms, 100-ms and 1-s on 20-Dec-2007

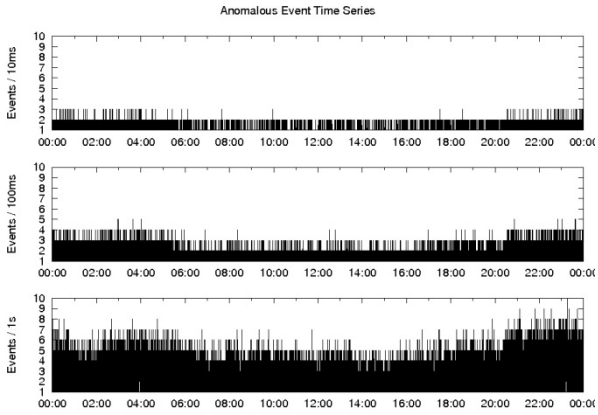

Fig 4. Anomalous event time series of TCP SYNs packets in different window size of 10-ms, 100-ms and 1-s on 20-Dec-2007

5. Discussion

From the time series of packet and anomalous events, we calculated the correlation between one interval of time series and another interval of time series. Table II lists the correlation coefficients of the

time series. The first number in this table lists packet time series correlations and the second number lists anomalous event time series correlations.

In Table II indicates that the 100-ms time series is correlated with 1-s interval time at 0.851, and the anomalous 10-ms time series are correlated with the 100-ms interval time with value 0.443. These are maximum correlation from the packet and event time series.

Moreover, our results show time series and wavelet decomposition at each level. We found that level 7 with a 1-s interval time was appropriate for delineating anomalies in Darknet data. At this level, it takes fewer number of data and a shorter computation time for decomposition signal.

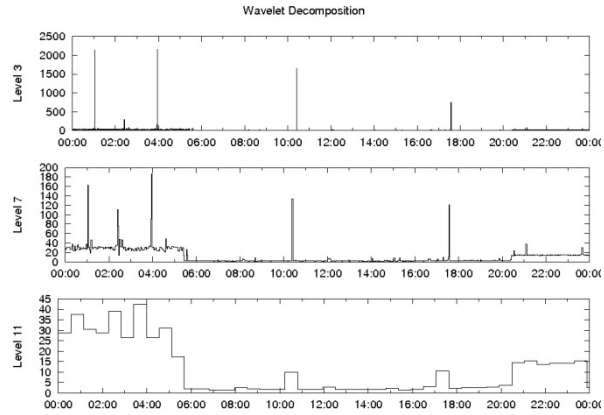

Fig 5. Discrete Wavelet Transfrom levels 3, 7 and 11 of TCP SYNs packets in the window size of 1-s on 20-Dec-2007

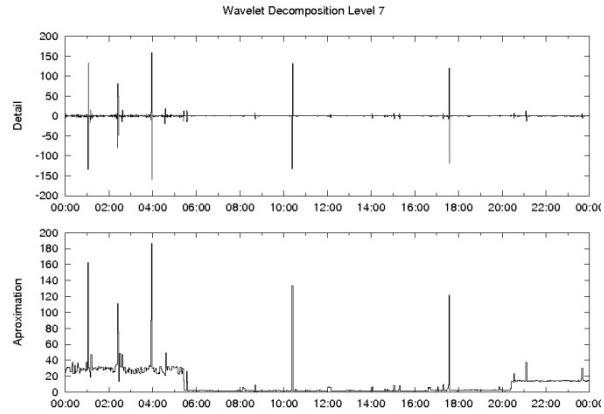

Fig 6. Discrete Wavelet Transform level 7 of TCP SYNs packets in the window size of 1-s on 20-Dec-2007

Figure 6 shows the detail and approximation of level 7 for anomalous behavior from the Darknet data with a 1-s time interval. From detail resolution, we can see an abrupt change caused by irregular activities, and we can obviously see the ordinary behavior from the approximate resolution.

We discovered that signal level 11 with 100-ms and level 15 with 10-ms are strongly correlated with signal level 7 with a 1-s time interval. Note that with a 0.827 correlation coefficient, signal level 11 with a 100-ms time interval can be used to interpret anomalous network traffic, as can signal level 7 with a 1-s time interval.

TABLE II Packet/Event Time Series Correlation of TCP SYNs

Interval Time	10ms	100ms	1s
10ms	1,1	0.613,0.443	0.522,0.155
100ms	0.613,0.443	1,1	0.851,0.365
1s	0.522,0.155	0.851,0.365	1,1

From our results, indicate that TCP SYNs and TCP SYN/ACKs packets regularly have a high wavelet decomposition correlation coefficients. However, UDP packets often have a low correlation coefficient.

6. Conclusions and Future Work

This paper described important measurement issues associated with wavelet-based analysis on unwanted traffic data from the Darknet. We evaluated three types of unwanted traffic packet: TCP SYNs, TCP SYN/ACKs and UDP packets and three different window time intervals: 10-ms, 100-ms and 1-s. We attempted to provide researchers with a general overview of wavelet decomposition and the important details needed to analyze Darknet data.

The results of this paper bring to light the method and fundamental parameters of the wavelet decomposition for anomaly behavior analysis of Darknet data.

Ours in a work-in-progress and we will continue to investigate the suitability of these and other parameters for an efficient anomaly detection method. We believe that our study is a promising step towards the broader goal of developing a practical analysis of unwanted traffic data.

7. Acknowledgments

This work was done during NII international internship program and supported by the National Science and Technology Development Agency (NSTDA) and the Thailand Graduate Institute of Science and Technology (TGIST) under contract number TGIST 01-51-092, and this work was made possible through cooperation between the National Institute of Informatics (NII), Kasetsart University and Thailand Post Co.,Ltd. The authors would like to thank all the anonymous reviewers for their excellent suggestions that have greatly improved the quality of this paper.

8. References

[1] H.-A. Kim and B. Karp, "Autograph: toward automated, distributed worm signature detection," in SSYM'04: Proceedings of the 13th conference on USENIX Security Symposium. Berkeley, CA, USA: USENIX Association, 2004, pp. 19–19.

[2] S. Schechter, J. Jung, and A. W. Berger, "Fast detection of scanning worm infections," in 7th International Symposium on Recent Advances in Intrusion Detection (RAID), French Riviera, France, September 2004.

[3] A. Hussain, J. Heidemann, and C. Papadopoulos, "A framework for classifying denial of service attacks," in SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. New York, NY, USA: ACM, 2003, pp. 99–110.

[4] J. Jung, V. Paxson, A. W. Berger, and H. Balakrishnan, "Fast portscan detection using sequential hypothesis testing," in IEEE Symposium on Security and Privacy 2004, Oakland, CA, May 2004.

[5] J. Jung, B. Krishnamurthy, and M. Rabinovich, "Flash crowds and denial of service attacks: characterization and implications for cdns and web sites," in WWW '02: Proceedings of the 11th international conference on World Wide Web. New York, NY, USA: ACM, 2002, pp. 293–304.

[6] M. Bailey, E. Cooke, F. Jahanian, A. Myrick, and S. Sinha, "Practical darknet measurement," Information Sciences and Systems, 2006 40th Annual Conference on, pp. 1496–1501, March 2006.

[7] D. Moore, C. Shannon, G. M. Voelker, and S. Savage., "Network telescopes," University of California, San Diego, Tech. Rep., July 2004.

[8] M. Bailey, E. Cooke, F. Jahanian, and J. Nazario, "The internet motion sensor - a distributed blackhole monitoring system," in NDSS, 2005.

[9] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson, "Characteristics of internet background radiation," in IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. New York, NY, USA: ACM, 2004, pp. 27–40.

[10] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," Comput.Netw., vol. 51, no. 12, pp. 3448–3470, 2007.

[11] C. Krugel, T. Toth, and E. Kirda, "Service specific anomaly detection for network intrusion detection," in SAC '02: Proceedings of the 2002 ACM symposium on Applied computing. New York, NY, USA: ACM, 2002, pp. 201–208.

[12] J. M. Estevez-Tapiador, P. Garcia-Teodoro, and J. E. Diaz-Verdejo, "Anomaly detection methods in wired networks: a survey and taxonomy," Computer Communications, vol. 27, no. 16, pp. 1569–1584, Oct. 2004.

[13] P. Barford and D. Plonka, "Characteristics of network traffic flow anomalies," in IMW '01: Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement. New York, NY, USA: ACM, 2001, pp. 69–73.

[14] N. Ye, S. Vilbert, and Q. Chen, "Computer intrusion detection through ewma for autocorrelated and uncorrelated data," Reliability, IEEE Transactions on, vol. 52, no. 1, pp. 75–82, March 2003.

[15] S. S. Kim and A. L. N. Reddy, "Statistical techniques for detecting traffic anomalies through packet header data," Networking, IEEE/ACM Transactions on, vol. 16, no. 3, pp. 562–575, June 2008.

[16] R. R. Kompella, S. Singh, and G. Varghese, "On scalable attack detection in the network," IEEE/ACM Trans. Netw., vol. 15, no. 1, pp. 14–25, 2007.

[17] A. Tartakovsky, B. Rozovskii, R. Blazek, and H. Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," Signal Processing, IEEE Transactions on, vol. 54, no. 9, pp. 3372–3382, Sept. 2006.

[18] G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, and K. Cho, "Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedures," in LSAD '07: Proceedings of the 2007 workshop on Large scale attack defense. New York, NY, USA: ACM, 2007, pp. 145–152.

[19] M. Thottan and C. Ji, "Anomaly detection in ip networks," Signal Processing, IEEE Transactions on, vol. 51, no. 8, pp. 2191–2204, Aug. 2003.

[20] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," SIGCOMM Comput. Commun. Rev., vol. 34, no. 4, pp. 219–230, 2004.

[21] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment. New York, NY, USA: ACM, 2002, pp. 71–82.

[22] G. Carl, R. R. Brooks, and S. Rai, "Wavelet based denial-of-service detection," Computers & Security, vol. 25, no. 8, pp. 600–615, Nov. 2006.

[23] M. Hamdi and N. Boudriga, "Detecting denial-of-service attacks using the wavelet transform," Comput. Commun., vol. 30, no. 16, pp. 3203–3213, 2007.

[24] R. Xunyi, W. Ruchuan, and W. Haiyan, "Wavelet analysis method for detection of DDoS attack on the basis of self-similarity," Frontiers of Electrical and Electronic Engineering in China, vol. 2, no. 1, pp. 73–77, March 2007.

[25] W. Lu, M. Tavallaee, and A. A. Ghorbani, "Detecting network anomalies using different wavelet basis functions," cnsr, vol. 0, pp. 149–156, 2008.

[26] D. Moore, C. Shannon, and K. Claffy, "Code-red: a case study on the spread and victims of an internet worm," in IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment. New York, NY, USA: ACM, 2002, pp. 273–284.

[27] V. Yegneswaran, P. Barford, and V. Paxson, "Using honeynets for internet situational awareness," in In Proc. of the ACM/USENIX Fourth Workshop on Hot Topics in Networks, 2005.

[28] P. Soto, "Identifying and modeling unwanted traffic on the internet," Thesis, Massachusetts Institute of Technology, 2006.

[29] S. Mallat, A wavelet tour of signal processing. Academic Press, 1999.

[30] D. B. Percival and A. T. Walden, Wavelet Methods for Time Series Analysis. Cambridge University Press, 2007.