The Multi-fractal Nature of Worm and Normal Traffic at Individual Source Level*

Yufeng Chen[1], Yabo Dong[1], Dongming Lu[1], and Yunhe Pan[1]

College of Computer Science and Technology,
Zhejiang University,
Hangzhou 310027, P. R. China
{xztcyfnew, dongyb, ldm, panyh}@zju.edu.cn

Abstract. Worms have been becoming a serious threat in web age because worms can cause huge loss due to the fast-spread property. To detect worms effectively, it is important to investigate the characteristics of worm traffic at individual source level. We model worm traffic with the multi-fractal process, and compare the multi-fractal property of worm and normal traffics at individual source level. The results show that the worm traffic possesses less multi-fractal property.

1 Introduction

Data security is very important in web age because we should assure the availability of Internet and web-based information systems. Worms have been becoming a serious threat because worms can spread in short time and cause huge loss. Last year, two notorious worms, the "Blaster" worm [1] and "Welchia" worm [2] infected a lot of computers and the losses are heavy. Early warning is an effective method to prevent the spread of worms. The infectious computers should be located for eradicating worms from systems. Thus, we investigated the traffic characteristics of worm and normal traffic at individual source level to compare the diversity of the traffic characteristics.

The self-similar [3, 4, 5, 6, 7] and multi-fractal [8, 9] models have been proposed to depict the characteristics of network traffic from the angle of traffic engineering [3]. However, because the fractal characteristic is the nature of network traffic, we tried to investigate the fractal characteristics of worm traffic find the abnormity of worm traffic. Because worm detection is a competition against worm propagation, the short-range characters is more important. Thus we payed our attentions to the multi-fractal nature of traffics because multi-fractal model possesses the capability of describing the short-range property of time series as

* This work is supported by a grant from Zhejiang Provincial Natural Science Foundation (No.Y104437), Hubei Provincial Natural Science Foundation (No. 2004ABA018), Science and Technology Program of Hubei Provincial Department of Education (No. 2004D005), and Science and Technology Program of Hubei Provincial Department of Education (No. D200523007).

P. Kantor et al. (Eds.): ISI 2005, LNCS 3495, pp. 505–510, 2005.

well as long-range property. We studied the worm and normal traffic at individual source level, i.e., traffics are generated by individual computers. We found that the worm traffic owns less multi-fractal characteristic than normal traffic does. The data set was collected on a 100Mbps link connecting one dormitory building to our campus network. And as an example of worm traffic, the traffic generated by "Welchia" worm was collected and analyzed.

The paper is organized as follows. The related works are introduced briefly in Sect. 2 and the mathematical background is introduced in Sect. 3. In Sect. 4, we give an overview of our data set. The diversity of multi-fractal properties of normal and worm traffics at individual source level are compared in Sect. 5. In Sect. 6, the conclusions and future work are presented.

2 Related Works

Some models and methods of worm traffic have been proposed, which mostly focus on the aggregated traffic characters of worm and normal traffics. Cowie et al. described the idea of "worm induced traffic diversity" that is the primary cause of the BGP instabilities[10]. However, they just propose the preliminary conclusions without deeper investigation. [11] presented an statistical-based approach that utilizes application specific knowledge of the network services, which is used to find Remote-to-Local attacks. [12] presented the ideas of exploiting the property of self-similarity in network traffic to detect the faulty behavior. [13] took advantage of multi-fractal model to detect fault in network traffic based on the fact that faults in a self-similar traffic destroy the structure at the time points they occur. However, the meaning of errors and faults in [12] and [13] is rather broad, maybe including the abnormal traffic generated by worms.

The proposed statistical models of worm behaviors and fractal models of abnormal traffic are investigated at the aggregated traffic level. However, they can not provide information to identify the infectious computers. Thus, we try to provide the idea for detecting the infectious computers at individual source level from the angle of multi-fractal nature of network traffic.

3 Background

Firstly, we give the basis of multi-fractal analysis briefly. For detailed discussion, please refer to [8, 9, 14]. Consider a probability measure μ on the unit interval $[0, 1]$ and random variables

$$Y_n = \log \mu(I_K^{(n)}) \ , \tag{1}$$

where $I_K^{(n)}$ denotes the partition into 2^n equal subintervals

$$I_K^{(n)} := [k2^{-n}, (k+1)2^{-n}] \ , \tag{2}$$

and K is a random number from $\{0, \ldots, 2^n - 1\}$ with uniform distribution P_n. If the rate function, also called "partition function" or "free energy"

$$\tau(q) := \lim_{n\to\infty} \frac{-1}{n} \log_2 \sum_{k=1}^{2^n} \mu(I_k^{(n)})^q \tag{3}$$

exists and is differentiable on $\Re$, then the double limit

$$f_G(\alpha) := \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log_2 2^n P_n[\alpha(I_K^{(n)}) \in (\alpha - \varepsilon, \alpha + \varepsilon)] \tag{4}$$

with

$$\alpha(I_K^{(n)}) := \frac{-1}{n \log 2} Y_n = \frac{\log \mu(I_K^{(n)})}{\log |I_K^{(n)}|} \tag{5}$$

exists, where $\alpha(I_K^{(n)})$ is termed Holder exponent. And

$$f_G(\alpha) = f_L(\alpha) := \tau^*(\alpha) := \inf_{q\in\Re} (q\alpha - \tau(q)) \ . \tag{6}$$

And f_L is often referred to as Legendre spectrum. Because of the robustness and simplicity of f_L, we will analyze the multi-fractal property of our data series through f_L. And the typical shape of f_L is a $\cap$. The parameter α quantifies the degree of regularity in a point x: here, the measure of an interval $[x, x + \Delta x]$ behaves as $(\Delta x)^\alpha$. In traffic measurements, $(\Delta x)^\alpha$ can be interpreted as the number of packets or bytes in this interval. Consequently, $\alpha < 1$ indicates a burst of events around x "on all levels", while $\alpha > 1$ is found in regions where events occur sparsely. Thus, for a process, if the interval of $\alpha < 1$ is larger, the process possesses more multi-fractal property.

The scaling of "sample moments" can also be studied through the partition function

$$\tau(q) := \liminf_{n\to\infty} \frac{\log S^{(n)}(q)}{-n \log 2} \ , \tag{7}$$

where the partition sum

$$S^{(n)}(q) := \sum_{k=0}^{2^n - 1} |Y((k+1)2^{-n}) - Y(k2^{-n})|^q \ . \tag{8}$$

When we inspect the log-log plots of partition sum against q, if the plot appears to be linear, the observed process is multi-fractal. To get the partition sum and Legendre spectrum of the worm and normal traffic, we make use of a tool named Fraclab [15].

4 Data Set

Our data sets were collected from a 100Mbps link connecting one dormitory building to our campus network, in December 26, 2003, when the "Welchia" worm broke out. The data sets contains information of source and destination

Table 1. Summary of our data set

Item	Value
Packet Counts	25,436,630
Number of Distinct Source Addresses	2,889
Number of Distinct Destination Addresses	300,053
Number of Total Distinct IP Addresses	300,133

IP address, timestamp, packet length. The packet count of the data set is about 26 million. Table 1 shows the summaries of the data sets.

We picked up two packet streams generated by two sources, one for normal traffic and the other for worm traffic. The summaries of the two packet streams are shown in Table 2. The source addresses are renumbered for privacy reasons, which are 2 and 72. The source 2 generates normal traffic, and the source 72 generates worm traffic. The source 72 marked with "worm" exhibits the character of "Welchia" worm because the volume of the corresponding destination addresses is rather vast.

5 Diversity of Multi-fractal Properties at Individual Source Level

We analyze the number of packets of normal and worm traffic generated by individual source with the time unit of 100 milliseconds. And because we want to investigate the characteristics of short term, we choose 4096 observations for each time series, and each observation represents the number of packets sent over the Ethernet by the corresponding source every 100 milliseconds. Fig. 1 shows the plots of the partition sum and Legendre spectrum of the two streams.

Fig. 1 depicts the multi-fractal characteristics of the two packet streams. The top plots show the partition sum against the scale on log-log scale. It's obvious that $S^{(n)}(q)$ of the worm traffic is much different from that of the normal traffic. And the bottom plots show Legendre spectrum $f_L(\alpha)$ against α. And we can learn from the plots that the worm traffic possesses less multi-fractal property because the interval of $\alpha > 1$ is much larger than that of normal traffic, which

Table 2. Summary of the two packet streams. The corresponding destination address means the destination address of the packet whose source address is the corresponding renumbered source address, and the corresponding packet has the similar meaning

Traffic	Normal	Worm
Renumbered Source Address	2	72
Number of Corresponding Destination Addresses	220	129,437
Number of Corresponding Packets	141,092	129,528

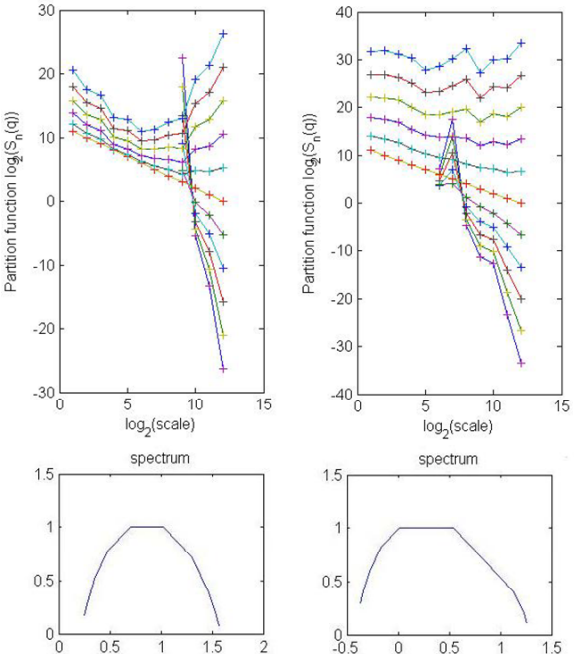

Fig. 1. the plots of partition sum $S^{(n)}(q)$ and Legendre spectrum $f_L(\alpha)$ based on Discrete Wavelet Transform. The number of q ranging from -5 to 5, and the total number of q is 11. The left part is corresponding to the infectious source 72, and the right part is corresponding to the normal source 2

means that the proportion of burstiness of events is less for worm traffic. The reason of less multi-fractal for worm traffic can be interpreted intuitively. When a computer is infected by "Welchia" worm, this source will send ICMP echo request packets continuously, and the traffic generated by this source exhibits less burstiness on all levels.

6 Conclusions and Future Works

In this paper, we presented the idea to investigate the diversity of multi-fractal nature of worm and normal traffics for further application, such as worm detection. And the results show that worm traffic possesses less multi-fractal property than that of normal traffic. The further work includes: 1) investigate the multi-fractal characteristics of traffic of other worms; 2) analyze the factors that influence the multi-fractal property and how they influence; 3) implement the algorithm of detecting infectious sources based on the multi-fractal characteristics of traffic.

Acknowledgements. The authors acknowledge the continuing support from the Ningbo Network Access Point and Networking Center of Zhejiang University.

References

1. CERT, 2003, "CERT advisory CA-2003-20 w32-blaster worm" http://www.cert.org/advisories/CA-2003-20.html
2. CCERT, 2003, "CCERT advisory of security" http://www.ccert.edu.cn/notice/advisory.htm
3. Leland, W.E., Willinger, W., Taqqu, M.S. and Wilson, D.V., 1995, "On the self-similar nature of ethernet traffic", *ACM SIGCOMM Computer Communication Review*, Vol. 25, No. 1, pp. 202-213
4. Willinger, W., Taqqu, M.S., Sherman, R. and Wilson, D.V., 1997, "Self-similarity through high-variability: Statistical analysis of ethernet LAN traffic at the source level", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 1, pp. 71-86
5. Crovella, M.E. and Bestavros, A., 1997, "Self-similarity in world wide web traffic: Evidence and possible causes", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, pp. 835-846
6. Garrett, M.W. and Willinger, W., 1994, "Analysis, modeling and generation of self-similar VBR video traffic", *ACM SIGCOMM Computer Communication Review*, Vol. 24, No. 4, pp. 269-280
7. Paxson, V. and Floyd, S., 1995, "Wide area traffic: The failure of poisson modeling", *IEEE/ACM Trans. Networking*, Vol. 3, No. 3, pp. 226-244
8. Mannersalo, P. and Norros, I., 1997, "Multifractal analysis of real ATM traffic: a first look", http://www.vtt.fi/tte/tte21/cost257/multifaster.ps.gz
9. Riedi, R.H. and Vehel, J.L., 1997, "Multifractal properties of TCP traffic: a numerical study", http://www.stat.rice.edu/∼riedi/Publ/ts.ps.gz
10. Cowie, J., Ogielski, A.T., Premore, B. and Yuan, Y., 2001, "Global routing instabilities triggered by codered II and nimda worm attacks", http://www.renesys.com/projects/bgp_instability
11. Krugel, C., Toth, T. and Kirda, E., 2002, "Service specific anomaly detection for network intrusion detection", In *Proc. of ACM Symposium on Applied Computing*, pp. 201-208, March 2002
12. Schleifer, W. and Mannle, M., 2001, "Online error detection through observation of traffic self-similarity", In *Proc. of IEE on Communications*, Vol. 148, pp. 38-42, February 2001
13. Tang, Y., Luo, X. and Yang, Z., 2002, "Fault detection through multi-fractal nature of traffic", In *Proc. of IEEE on Communications, Circuits and Systems and West Sino Expositions*, Vol. 1, pp. 695-699, June 2002
14. Riedi, R.H., 2002, "Multifractal processes", http://www.stat.rice.edu/∼riedi/Publ/mp.ps.gz
15. FRACTALES group, 2001, "Fraclab: A fractal analysis toolbox for signal and image processing", http://fractales.inria.fr/index.php?page=download_fraclab