

Network Traffic Anomaly Detection Through Correlation Integrals

Song Luo

School of Computer Science
University of Central Florida
Orlando, Florida 32816-2362
sluo@cs.ucf.edu

Gerald Marin

Department of Computer Science
Florida Institute of Technology
Melbourne, Florida 32901
gmarin@fit.edu

Categories and Subject Descriptors

C.2.0 [Computer Communication Networks]: General –
security and protection

General Terms

Measurement, Experimentation, Security

Keywords

Network traffic template, anomaly detection, intrusion detection

1. EXTENDED ABSTRACT

The Correlation integral was defined by Grassberger and Procaccia [1] as a tool to calculate the fractal dimension of a time series. Previous research has indicated that Internet traffic can be characterized by a fractal dimension that changes slowly over time [2] [3].

Due to the close relationship between the correlation integral and the fractal dimension, it is natural to presume that the correlation integral is also capable of characterizing network traffic. In this paper, we use captured traffic traces to illustrate that one can indeed describe the dynamics of the Internet traffic with a template of correlation integrals. Furthermore, this template can be leveraged to detect abnormal traffic.

For the experiments we use 10-hours of traffic captured on the CS LAN at the University of Central Florida from 8am to 6pm on the Wednesday of February 05, 2003. The traffic traces include more than 30 million TCP packets. A time series $\{x_i\}$, $1 \leq i \leq 36000$, is constructed by counting the TCP packet arrivals for each second.

To obtain the traffic template the time series is divided into 36 non-overlapping segments; thus, each

segment has 1000 seconds of packet arrival data. Correlation integrals, $C(r)$, are computed for each segment by the following formula:

$$C(r) = \frac{1}{N^2} \sum_{i,j=1,i \neq j}^N \theta(r - |x_i - x_j| / D).$$

Here x_i and x_j are elements of the time series, and $|x_i - x_j|$ is the difference between the number of the packet arrivals during the i^{th} and j^{th} seconds. D is the maximum such distance found in a particular segment, r is a real number in $[0,1]$, and θ is the Heaviside function:

$$\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0 \end{cases}.$$

It follows that $C(r)$ measures the probability that the standardized distance $|x_i - x_j|/D$ between any pair of points in the time series is less than r . A vector is constructed for each segment by computing the correlation integrals when r varies from 0.05 to 0.1 with a step size of 0.05. More than 70% of pairs have a distance less than 0.1

For our initial experiments we constructed a traffic template as another correlation integral vector in which each element (fixed r) is the mean of the corresponding elements (the same r) in other vectors of all 36 segments. Figure 1 displays the template of the UCF CS LAN traffic and the 99% confidence intervals.

To verify that the template actually characterizes the network traffic, we compare the template against the correlation integral vector of each segment using the Kolmogorov-Smirnov test. A high p -value indicates a good match, which means the template correctly represents the dynamics of the examined traffic. The tests return p -values above 0.8 for 20 out

of 36 segments. 27 segments obtain p-values above 0.1. Figure 2 depicts the traffic of a typical “normal” segment. The segments with low p-values have either continuously increasing packet volume or extremely high packet arrival rates, and are regarded as abnormal.

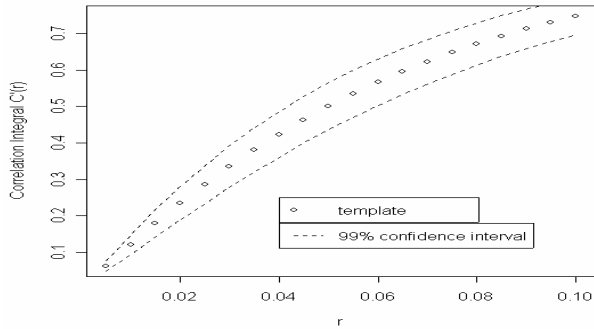


Figure 1: The traffic template constructed by correlation integrals

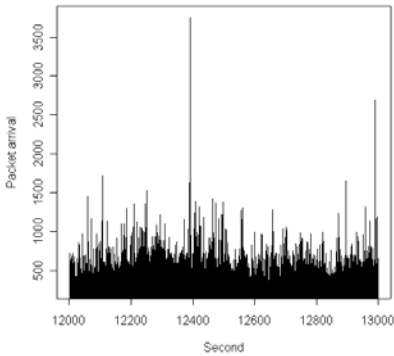


Figure 2: Packet arrivals of typical “normal” pattern

We want to know if an attack, especially a DoS attack, would possibly change the structure of the correlation integral of the normal traffic, and be detected by comparing the structure with the template.

We selected 19 DoS attacks used in the 1999 DARPA Lincoln Lab Intrusion Detection Evaluation experiment [4]. We also selected a period of “normal” traffic (i.e. having high p-values in the previous K-S tests) from the UCF CS data as the background. Because the average packet arrivals of the DARPA Lincoln Lab experiment is much lower than that of the UCF CS data, the attack intensity (packets/second) is rescaled to a higher level before being combined with the UCF CS background. (This

is done by keeping the peak attack ratio to average background ratio the same as the originals.)

We apply a sliding window of 1000 seconds, with a step size of one second, to the background-plus-attack traffic. At each step of the sliding window, its correlation integral vector is computed and compared with the template using the K-S test. A drop in p-value below 0.1 indicates the detection of a DoS attack. Table 1 lists the 19 attacks and how long after the occurrence they are detected.

The results in Table 1 show that the technique successfully detects 18 attacks in less than 10 seconds after the attacks begin. Thus, the technique is promising. The longest detection time is 40 seconds.

Table 1: Results of attack detection experiment

Attack Name	Detection Time(sec.)	Attack Name	Detection Time (sec.)
Fri4.mailbomb	9	Thu4.mailbomb	1
Fri4.smurf	2	Thu4.satan	2
Fri5.back	40	Tue5.back	4
Fri5.neptune	2	Tue5.neptune	2
Mon4.smurf	1	Wed4.mailboubm	1
Mon5.apache-1	2	Wed4.satan	1
Mon5.apache-2	2	Wed4.smurf	2
Mon5.neptune	2	Wed5.apache	1
Mon5.udpstorm	1	Wed5.back-1	1
Wed5.back-2	2		

Thus, the technique may support early intervention. However, clearly more work remains to be done. Future work will consider more realistic approaches for developing the templates (based on previously captured traffic, for example). Once this is done, it will be possible to estimate the pfa (probability of false alarm), perhaps the most critical determinant of effectiveness.

2. REFERENCES

- [1] Perter Grassberger and Itamar Procaccia. Characterization of strange attractors. *Physical Review Letters*, 50(5), 1983
- [2] Taquq MS, Teverovsky V, Willinger W. Is network traffic self-similar or multifractal? *Fractals* 5:63, 1997
- [3] D. Chakraborty, A. Ashir, T. Suganuma, et. al. Self-similar and fractal nature of Internet traffic. *International Journal of Network Management* 14, 119-129, 2004
- [4] Richard Lippmann, Joshua W. Haines, David J. Fried, Jonathan Korba, Kumar Das. The 1999 DARPA Off-Line Intrusion Detection Evaluation. *Draft of paper submitted to Computer Networks*, In Press, 2000.