2018 ASEE Mid-Atlantic Section Spring Conference: Washington, District of Columbia Apr 6

Analysis of the Low Rate of Denial of Service Attacks Detection by Using Statistical Fisher Methods

Mr. Yasser R Salem, University of the District of Columbia
Dr. Paul Cotae, University of the District of Columbia

Dr. Paul Cotae,Professor of Electrical and Computer Engineering has more than 25 years of experience in the communication field (research and education). He received a Dipl. Ing. and a M.S. degrees in communication and electronic engineering in 1980 from the Technical University of Iassy and a Ph.D. degree in telecommunications from "Politechnica" University of Bucharest, Romania in 1991, and a Master in Applied Mathematics in 1998 from the University of Colorado at Boulder. From 1994 to 1998 he spent four years at the University of Colorado at Colorado Springs and the University of Colorado at Boulder as a Fulbright Scholar and Visiting Associate Professor doing research and teaching in ECE department and APPM department. He served also as a consultant to Navsys Corp., Colorado Springs, in 1997. From 2002 to 2008 he was with the Department of Electrical and Computer engineering at the University of Texas at San Antonio (UTSA). From 1984 to 2001, he was with the Department of Electrical Engineering, Technical University of Iassy, where he conducted research and teaching in the area of digital communications as a Full Professor at the same department. Since 2008, he has been with the University of the District of Columbia as an Associate Professor. His current research interests include multiple access, modulation and coding, mobile communications, and digital communication systems. He has authored or coauthored more than 100 papers in these areas and four books. Dr. Cotae serves as an Associate Editor for IEEE Communication Letters, EURASIP Journal on Signal Processing, Elsevier International Journal of Computers and Electrical Engineering (JCEE), and he has been on the Technical Program Committee and Session chair of the IEEE Conferences GLOBECOM (2003-2011), VTC Spring 2005, 2006 and ICC 2005-2011. He is a Senior Member of IEEE, member of ASEE, member of HKN (Eta Kappa Nu) and SIAM. He is cited in Who's Who in American Education, Who's Who in America, and in Who's Who in the World. He has been nominated two times for the best teacher award (2005 and 2006) in the College of Engineering at UTSA. He is the IEEE Vice Chair Washington Section and IEEE ComSoc chair Whasington DC Section. He is Faculty Fellow for the ONR-ASEE Summer Faculty Research Program 2009-2018.

Analysis of Low Rate of Denial Service Attacks Detection Using Fisher Statistical Method

Yasser Salem[1] y.salem@hotmail.com, **Paul Cotae**[2] pcotae@udc.edu

Department of Electrical and Computer Engineering
School of Engineering and Applied Sciences
University of the District of Columbia
4200 Connecticut Avenue NW,
Washington, DC 20008,USA

Abstract

This paper emphasizes on the cybersecurity for the Low Rate Denial of Service attacks (LR DoS). By exploiting the weaknesses of TCP protocol, attackers send a cracking packet in quasi-periodic fashion in order to overwhelm the server before TCP sender enters retransmission timeout (RTO) phase. Therefore, the attacker causes lost packets for the legitimate users. The attacker synchronizes with the RTO period and in this way the attacker can over control the legitimate user TCP.

In this paper, we focus on the detection of Low Rate Denial of Service attacks. We use the Fisher statistics methods for detection of the LR DoS attacks components. Fisher statistical method is used to calculate the data that is hiding in the process in one time series. In addition, we are going to analyze and simulate different types of flows in a computer network by using MATLAB.

Keywords

g-statistic, Fisher test, peridogram, periodic content, Low Rate Denial of Service attacks.

Introduction

Detection and prevention of denial of service (DoS) attacks, and other traffic anomalies are curial for efficient network[5]. Recently, variants of DoS (low and slow) attacks have been identified and they are hard to detect [6]. An attack tries to attack Transmission Control Protocol (TCP)[7]. Attackers send a cracking packet periodically to occupy a server before TCP sender enters retransmission timeout (RTO), and it causes lost packets of legitimate user. Because of lost packets, TCP sender is going to enter RTO. In LR DoS attack, attacker sends short burst of packets to overflow a router's queue and causes packet loss for users Consequently, TCP source will back off to recover from the congestion and retransmit only after one RTO. The attacker congests the router again at the times of retransmission, then little or no real user traffic can get through the network. The attacker can immediately shut off most legitimate TCP sources even though the rate of attack is low and weak. The TCP operates on longer timescales of retransmission timeout (RTO) where the minimum recommend RTO is 1 second. In each period, the wave has a magnitude of zero except of a one unit time. The wave, also, has a magnitude of a normalized burst. Common to the above, attacks are a large number of compromised machines involved in the attack and approach high-rate transmission of packets towards to hack nodes. An attacker which consists of periodic "on-off bursts" exploits the homogeneity of the TCP's RTO mechanism. In the time domain, this attack can be modeled by a set of three parameters in[7]. When such a burst attack arrives at the link, TCP senders stop transmitting packets and enter timeout state due to packet loss. After minRTO, when a valid sender attempts to retransmit its lost packets, a new burst from the attacker arrives and the sender is forced to re-enter timeout state. Then, the sender is being denied bandwidth. An attack exploits the TCP slow starting mechanism and let users exit timeout state and enter slow start phase after every burst. The burst pulses are exploit the slow start mechanism. The attacker can transmit at a lower rate and it is more easily background traffic in time domain, which is the usual avenue of avoiding detection by defense mechanisms that combat DoS attacks[1].

In this paper, we are going to use one side test in alternative hypothesis in one time series. We take a frequency domain to detect periodicities of Low Rate DOS. The periodicity in the frequency domain provides chances to create a new method for anti-attacks mechanisms. Spectral analysis methods are invaluable tools in the analysis of time series for detecting and testing periodicities. The basic tool that we use in the frequency domain is the periodogram. After noticing a periodogram in [9] (for a single time series) contains a peak, a formal test should be carried out to determine whether this peak is significant or not. We analyze the periodogram from[9] using MATLAB. To test the periodicity, we use the Fisher statistical methods from[1-4, 8-9] for the periodic signal in time series. We apply the Fisher statistics test for more than one-time series and we applied it for the detection in[3-4, 8]. Also, we use periodogram analysis to go deeply in the detection. We want to apply Fisher statistical methods test for a number of samples as in real environment. In addition, we analyze our results using MATLAB. We estimate the variance directly from the time series by using sample variance. We make no assumption about the "time duration" of periodic content embedded in the time series and can detect a malicious flow even when the attack is bursty in nature.

Statistical Model

In this problem, we want to make sure that there is at least a periodicity in one of the time series. We consider a stochastic model to characterize our processing model as:

$$x(t) = \mu + \sum_{l=1}^{k} D_l \cos(2\pi f_l t\Delta t + \theta_l) + \epsilon(t) \qquad (1)$$

We apply it on one time series case, and we consider a statistical method which is real value discrete process in deterministic model in alternative hypothesis that is represented in:

$$x(t) = \sum_{l=1}^{k} D_l \cos(2\pi f_l t\Delta t + \theta_l) \qquad (2)$$

where k≥1 and it is denoting the number of components in time series$x(t)$. The amplitudes D_l of periodical components and their harmonic frequencies f_l are shown in the frequency domain. We use t to describe the time intervals for the measurement in (2) where t= 1, 2, 3,….,N, where N is the parameter form calculated by $N = 2m + 1$. Finding N is required in order to compute the periodogram and Fisher test. In addition, we take our deterministic random phase θ_l in the interval [-π,π]. The problem of detecting the periodical components in a time series (described in (2)) is equivalent to the problem of detecting peaks in a periodogram in[2,9].We focus in detecting a periodicity from (2) in frequency domain and the periodogram that is coordinated at Fourier in[9]. In spectral analysis, therefore, we take Δt as sampling time, and we convert it in to frequency by calculating Nyquist frequency f_N which is going to observe in frequency domain as double side band with the interval $S_N = [-f_N, f_N]$ from[1-2] which can be defined as:

$$f_N = \frac{1}{2\Delta t} \qquad (3)$$

Given a time series that can be regarded as a realization of a discrete sequence $x(1), x(2), x(3), \ldots, x(N)$ described by (2), the periodogram with this sequence in[2,9] is defined in the frequency domain as:

$$S(f) = \frac{\Delta t}{N}\left[\sum_{t=1}^{N} x(t)e^{-j2\pi ft\Delta t}\right]^2 \quad (4)$$

Under alternative hypothesis $H1$, we are testing a single significant periodical content in the whole set of N time series. We can decide if the number of time series group is malicious or not and, if so, we can identify the attacker host. The proposed approach is N-time faster to isolate the suspicious host than test each individual flow coming from all hosts in one time series system.

Fisher Statistical Test

The problem of deciding if a time series is random or periodic can be cast as a statistical decision problem by using hypothesis testing. We focus in one time series periodicity test. So, we apply Fisher g- statistic test to find exact periodicity in[3-4,8]. Fisher exact test identifies the ratio to find the g-statistic:

$$g = \frac{\max_{1 \le i \le N} x_i}{x_1 + x_2 + x_3 + \cdots + x_N} \quad (5)$$

Fisher determines $x_1 + x_2 + x_3 + \cdots + x_N$ as positive coordinates, which are represented as a point in the space with N dimensions[3-4]. Hence the periodogram components are also positive quantities, Fisher shows that the exact distribution by the ratio of the maximum ordinate of the periodogram, and the sum over all periodogram ordinates at the Fourier frequencies. Therefore:

$$g = \frac{\max_{1 \le i \le m} S(f_i)}{S(f_1) + S(f_2) + S(f_3) + \cdots + S(f_m)} \quad (6)$$

Where frequency interval $S_f = [-f, f]$ in[1-2] we assume that the null hypothesis H_0 where exists with no periodicity (no attack) in this frequency domain. We would like to test the alternative hypothesis H_1 when there is exist at least one periodicity (at least one attack) in the frequency domain. From[3-4], we calculate the initial Fisher test in first term by:

$$g_f = 1 - \left(\frac{\alpha}{m}\right)^{\frac{1}{m-1}} \quad (7)$$

If we consider only the first term (which is the dominant term) in summation and solve from[3,4,8] We cannot reject the null hypothesis if the test shows $g < g_f$. It suggests that the observation data are inconsistent with the assumption that in null hypothesis is true, so that hypothesis can have been rejected but this does not mean that the alternative hypothesis can be accepted as true. From[8], we organize the test by:

1) Select the attack frequency interval S_f.
2) Calculate the initial g_f value from (7).
3) Calculate Fisher g-statistic value from (6).
4) Fisher test Decision.

Example

According to our data, we have four different TCP flows: TCP- type1 flow1 and TCP- type3 flow 1, 2 and 3. We have an attacker and 50 normal TCP clients for each flow. The traffic shows attackers pattern for the flows and they are different because attackers interfere with the traffic of normal clients and network characteristics. We use Signal processing and Communication to analyze and show the malicious periodicity. Our majority detects an attacks based on the pattern. Our data includes:

- TCP maximum size: 64000byte.
- The duration: five minutes (300 seconds).
- Burst length (L): Randomly selected in 0.01 sec
- Length duration: 3000.
- Max burst rate: 75000byte/s (bottleneck capacity)
- Period (T): randomly between 0.1sec to 0.5sec
- Set the level of value which corresponds to confidence interval α=0.01.

We find spectral analysis from the MATLAB. We used FFT to find the periodogram of the signal and plot it using periodogram code. on another hand, we calculate time series from the Nyquist frequency to get the Fourier frequency. The test is performed on one side so the negative side is neglected. We start collecting data for interval frequency of $0 \leq f_N \leq 5$ Hz. However, we consider analyzing from 0 to 3 Hz. We want to detect an attacker based on the traffic pattern. In TCP- type 1, we see there are packets sending every certain time period during the five minutes. The maximum rate is 22218 b/s. We take the Fourier function of the signal, so that we can check the periodicity of the signal. Then, we apply the Fisher test in our result from (6), we calculate (7) to check if the condition is true. An attack flow represented in frequency domain, the Fisher test's g-statistic value shows there is an attack with high amplitude values at dome frequency components. From (7), g-value=0.01860, if the test shows $g < g_f$ therefore, there is an attack in the signal. In the spectral domain, we observe three high amplitude values: 6.52e+10 ,6.33e+09 and 5.22e+09. These amplitudes represent periodic attack signals in TCP-type 1.

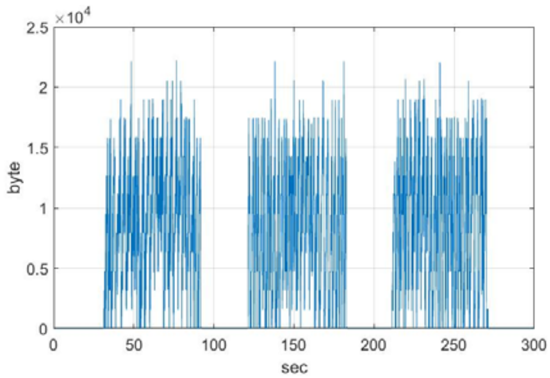

Figure 1: TCP1Flow1 Time Domain

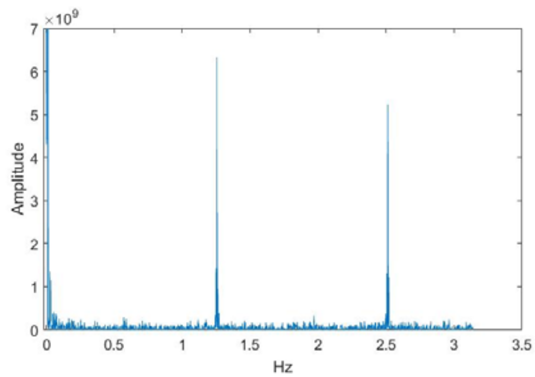

Figure 2: TCP1Flow1 Periodogram

For TCP- type 3, we follow the same procedure above. In flow1, we find the lowest amplitude in time domain equals to 18984. From (6), g-value=0.01890, if the test shows g<gf therefore, there is an attack in the signal. In the spectral domain, we observe three high amplitude frequency components: 5.45e+10 ,4.35e+09 and 3.12e+09.

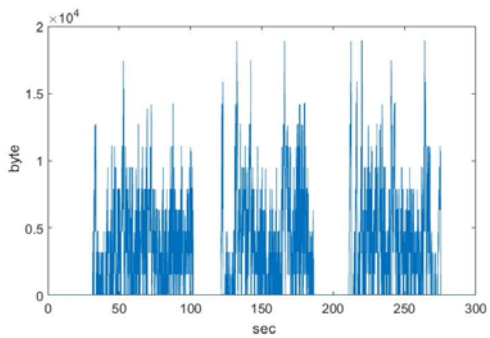

Figure 3: TCP3Flow1 Time Domain

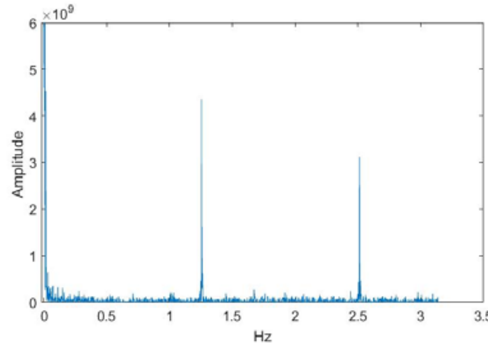

Figure 4: TCP3Flow1 Periodogram

Similarly, in flow 2 and 3, we follow the fisher method to analyze the data. In contrast, we observe the flow3 has the highest amplitude in the spectral domain with g value=0.0200. If the test shows g<gf therefore, there is an attack in the signal. In the spectral domain, we observe three high signals, and they are periodic attack signal for all three TCP types

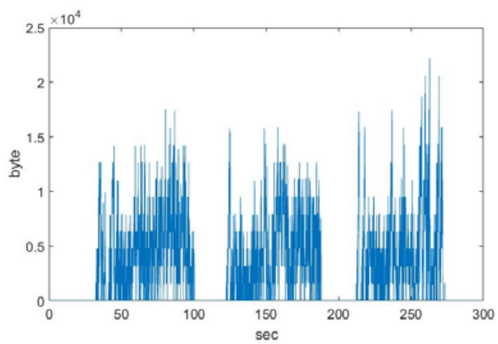

Figure 5: TCP3Flow2 Time Domain

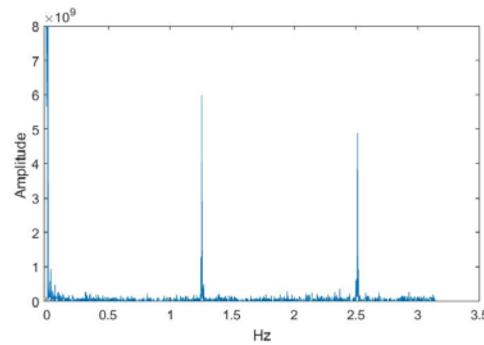

Figure 6: TCP3Flow2 Periodogram

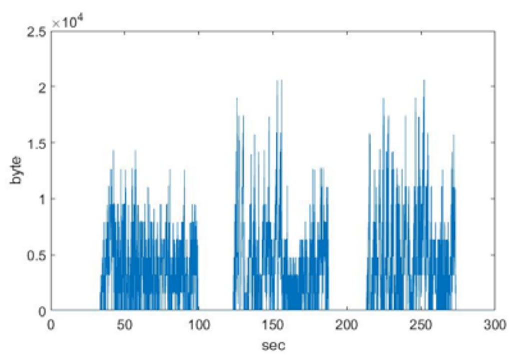

Figure 7: TCP3Flow3 Time Domain

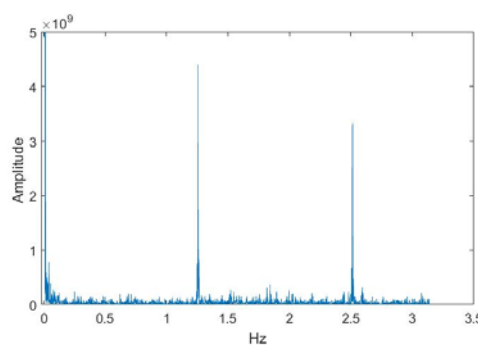

Figure 8: TCP3Flow3 Periodogram

Conclusion

In Low Rate Denial of service attack (DoS), attackers send packet quasiperiodic in order to overwhelm the server, before TCP sender enters retransmission timeout (RTO) phase. Attackers resend packets in the TCP during the RTO duration, so that no data traffic can get through the network. Therefore, attacker synchronizes with the RTO period and in this way the attacker can over control the legitimate TCP users. We have presented practical applications of Fisher periodicity tests for detection of Low-Rate Denial of Service (LR DoS) attacks. Attacks such as based on their periodicities in the frequency domain. We focused on how to examine the signals in the periodogram at TCP flow by using MATLAB application and toolbox. We applied our test in alternative hypothesis in one side time series. We have found the malicious time series with the periodic content. The Fisher g-statistic test has better results for detecting one dominant spectral line. Our results show that those attackers are effectively identified by using Fisher statistical methods for periodic attack.

References

1. Paul Cotae, Mayong Kange, Alexander Velazquez, "Spectral Analysis of Low Rate of Denial of Service Attacks Detection based on Fisher and Siegel Tests", IEEE,ICC, 2016.
2. Paul Cotae, Mayong Kange, Alexander Velazquez, "Multiple Time Series Fisher Periodicity Test for the Detection of the Distributed New Shrew Attacks", IEEEICC,2015.
3. Ronald. A. Fisher,"Test of Significance in Harmonic Analysis", Proceedings of the Royal Society of London, London,1929,pp54-59.
4. A. A. Nowroozi, "Table for Fisher's Test of Significance in Harmonic Analysis", Geophysical Journal of Royal Astronomical Society vol.12, 1967, pp.517-520.
5. Thatte Gautam, Mitra Urbashi, John Heidemann, "Detection of Low-Rate Attacks in Computer Networks", INFOCOM Workshops IEEE, Phoenix, AZ, USA 2008.
6. J. Luo, X. Yang "The NewShrew Attack: A New Type of Low rate TCP-Targeted DoS Attack", IEEE ICC, Australia, 2014, pp.713-718.
7. A. Kuzmanovic, Knightly, E.W.,"Low rate-Targeted Denial of Service Attacks (The Shrew vs. Mice and Elephants", SIGCOMM, Germany, 2003, pp.75-86 , 2003.
8. Fisher,Ronald A Statistical Methods, Experimental Design, and Scientific Inference: A Re-issue of Statistical Methods for Research Workers, The Design of Experiments, and Statistical Methods and Scientific Inference, Oxford University Press,UK, 1993.
9. Bloomfield, P., Fourier Analysis of Time Series - An Introduction, 2nd edition, John Wiley&Sons, Inc, New York, 2000.

Bio

Author[1] Yasser Salem

Yasser Salem received the undergrad in electrical engineering from Yanbu Industrial College. He worked in Saudi Aramco as instrumentation and communications engineer at King Abdullah University of Science and Technology (KAUST) in water plan project. He worked as instrumentation and control engineering at Marafiq Company in Technical service Department. Today, he is continuing his master degree in electrical Engineering at University of the District of Columbia in communication track. Hi is interesting in cybersecurity.

Auther[2] Dr. Paul Cotae

Dr. Paul Cotae, Professor of Electrical and Computer Engineering is the Director of the PhD Program at SEAS and Director of the SEAS Research Center. His research is in Digital

Communication, Information theory, Statistics and Applied Mathematics and Cybersecurity: Anomaly detection, Detection of Low Rate Denial of Service Attacks, Intrusion Detection, Information Visualization. He published more than 140 conference and journal papers, many of them at IEEE level, authored 2 books and coauthored 3 books in the area of digital communications systems. During the AY 2014-2015 he spent his sabbatical at the Center for High Assurance Computer Systems Code 5540, Naval Research Laboratory, Washington DC, 20375. Since 2009 he has been selected every summer as ONR Senior Research Fellows for the ASEE Summer Faculty Research Program at NRL. His research is sponsored by NSF, ONR, AFOSR and USAF. He received in last five years more than $1M for his research from DOD as a sole PI for the following grants: –Army Research Office (ARO) –Award No. W911NF-15-1-0481: "Performance Data-Driven Methods and Tools for Computer Network Defense through Network Science" and Office of Naval Research - Award no. W911NF-11-1-0144 "Information-Driven Blind Doppler Shift Estimation and Compensations Methods for Underwater Wireless Sensor Networks".