ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #24578

Analysis of Distributed Denial of Service Attacks Detection Using Fisher Statistical Method

Mr. Yasser R Salem, Graduate student
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 a Professor of Electrical Engineering. His current research interests include multiple access, modulation and coding, mobile communications, and digital communication systems. He has authored or coauthored more than 150 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 Distributed Denial Service Attacks Detection Using Fisher Statistical Methods

Yasser Salem[1] yasser.salem@udc.edu , **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

Distributed Denial of Service attacks are become one of the most serious security threats to our society. In this paper, we presented Distributed Denial of Service (DDoS) attacks detection in low and slow rate attack scenarios. Attackers generate low rate packets and try congested the TCP. After a user enters in the retransmission timeout (RTO) phase, attackers synchronize with the RTO period and they over-control legitimate users. In this paper, we focus on attacks detection in frequency domain. We tested, analyzed, and simulated different types of flows in a computer network by using MATLAB and Fisher statistical methods for detection of DDoS attack components in multiple times series. This paper intends to present applications of theoretical lectures and lab sessions to solve real engineering problems. It is a groundbreaking work in applying Fisher methods to cybersecurity and gives a new perspective to use MATLAB and other software to preprocess and simulating our data.

Keywords

G-statistic, Fisher test, Periodogram, Low-Rate**,** Distributed Denial of Service attacks.

Introduction

Detection and prevention of attacks are curial for efficient network security performance[5]. The Low Rate Distributed Denial of Service attacks (LR-DDoS) is a major issue in network security[6]. On October 2016, DDoS attacks shut down many servers and participated in DDoS attack by targeting systems operation by Domain Name System (DNS) provider. Attackers made Internet platforms and services disable to myriad swathes of users in Europe and North America. In a DDoS attack, an attacker attempts subversion of the Transmission Control Protocol (TCP) by sending large amount of data packets through networks targeting specific servers. The usual avenue of avoiding detection is by defense mechanisms that combat DDoS attacks[1]. In a LR-DDoS attack, an attacker sends short burst of attack packets in order to avoid detection by devices dedicated to DDoS preventions. This type of attack is classified as a protocol attack and as such this attack focuses on retransmission timeout parameter of the TCP protocol. Attackers make users exit the timeout state and enter slow start phase after every burst. Burst pulses are exploiting the slow start mechanism. The TCP operates on two time scales each of which deal with congestion avoidance in an IP network. When burst packets arrive in to the TCP link, TCP senders stop transmitting packets and enter the timeout state due to packets loss. The shorter time

scale in congestion avoidance deals with the round-trip times (RTT) between node links for milliseconds. The longer timescales of retransmission timeout (RTO) where the minimum recommend RTO is 1 second. Attacke consists of periodic "on-off bursts" exploiting the homogeneity of the TCP's RTO mechanism. Attackers send burst packets to overflow user's router queue and it causes packet loss, so users enter RTT. Attacks continue sending packets to make the TCP enters the RTO so that they can occupy a server. Consequently, TCP source will back off to recover from the congestion and retransmit only after one RTO. Attacks are a large number of compromised machines involved in attack and approach low-rate transmission of packets towards to occupy nodes. When such a burst attack arrives at the bottleneck link, TCP senders stop transmitting packets and enter timeout state due to packet loss. A valid sender attempts to retransmit the lost packets after minRTO. Simultaneously, a new burst arrives and a sender is forced to re-enter timeout state. Then, the sender is being denied bandwidth. Attackers congest the router again at the times of retransmission, then little or no real users traffic can get through the network. Attacker can immediately shut off most legitimate TCP sources even though the rate of attack is low and weak. Attackers can transmit at a lower rate and it is more easily background traffic in time domain. In this paper, we are going to test alternative hypothesis in multiple time series by using frequency domain to detect periodicities of DDOS. Our test is finding at least a malicious. The periodicity in the frequency domain constitutes a new method for anti-attacks mechanisms. The basic tool that we use in the frequency domain is the periodogram. Periodogram is an invaluable tool in the analysis of time series to detect and test periodicities. We analyzed 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 for the detection in[3-4, 8]. After noticing a periodogram in [9] (for multiple time series) a formal test should be carried out to determine whether this peak is significant or not.

Statistical Model

In this paper, we consider a stochastic model to characterize our processing model as:

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

We apply it on multiple time series cases, and we consider values in alternative hypothesis and it represents in:

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

Where $k\geq1$, and it denotes 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 $N = 2m + 1$,so we find N in order to compute the periodogram and Fisher test. In addition, we take our deterministic random phase θ_l in the interval $[-\pi, \pi]$. We extended our model (2) by showing the vector length and we got:

$$X(t) = [X_1(t), X_2(t). X_3(t), \dots, X_N(t)]^T \quad (3)$$

We can decide if the numbers of time series group are malicious or not. The problem of detecting periodical components in a time series (described in (2)) are equivalent the problem for detecting peaks in 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 Fourier frequency f_v which is going to observe in frequency domain with the interval $S_v = [-f_v, f_v]$ from[1-2]. Given a time series that can regard as a realization of a discrete sequence $x(1), x(2), x(3), \dots, x(N)$described by (2), the periodogram with this sequence in2,9 is defined in the frequency domain as:

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

Therefore, the total $S(f)$ is extended as:

$$S(f) = [S_1(f), S_2(f), S_3(f), \dots, S_v(f)]^T \quad (5)$$

Under alternative hypothesis $H1$, we are testing multiple significant periodical contents in whole sets of N time series. We want to make sure there is at least a periodicity attacks in multiple time series.

Fisher Statistical Test

The problem of deciding is time series either randomly or periodic that can cast as a statistical decision problem by using hypothesis testing. 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:

$$G = \frac{\max\limits_{1 \le i \le N} X_t}{x_1 + x_2 + x_3 + \cdots + x_N} \quad (5)$$

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:

$$G = \frac{\max\limits_{1 \le i \le m} S(f_v)}{S(f_1) + S(f_2) + S(f_3) + \cdots + S(f_m)} \quad (6)$$

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. We cannot reject the alternative hypothesis where the test shows $G \le G_f$. It suggests that the data observations are inconsistent with an assumption that in null hypothesis is true, so that hypothesis can have been rejected but this does not mean that the null hypothesis can be accepted as true.

We organize the test by:

1) Select the attack frequency interval S_v.
2) Calculate the initial G_f value from (7).
3) Calculate Fisher G-statistic value from (6).
4) Fisher test Decision.

Example

Our topology, we have three attackers. Each one attempts attacking a server. We applied that each one controls a botnet to do the attacking.
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.
- Period (T): Randomly.
- Data Rate: 1500000byte/sec

We want to detect an attacker based on the traffic pattern. We find frequency domain from the MATLAB. We used the periodogram of the signal and plot it using periodogram code to find the frequency domain; we calculate time series to get the Fourier frequency. Data packets start from zero response. We start collecting interval frequency of $0 \leq f_N \leq 5$ Hz. However, we consider analyzing from 0 to 3 Hz. We applied many toolboxes from MATLAB such as: communications and signal processing toolboxes. From the average traffic TCP that shown in figure 1, three attackers are sending burst packets to a victim. They are sending the low rate packets in the minRTO duration until attackers occupy services.

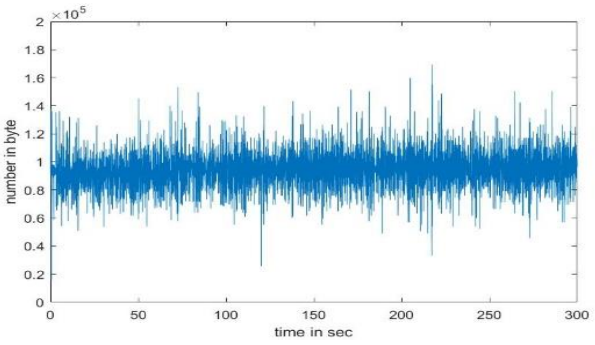

Figure 1: Average traffic TCP Time Domain

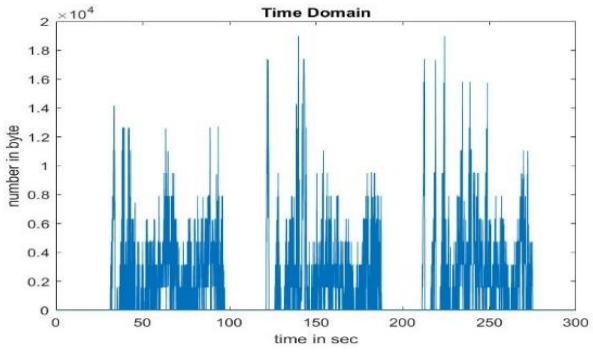

Figure 2: 1st Attacker packets in Time Domain

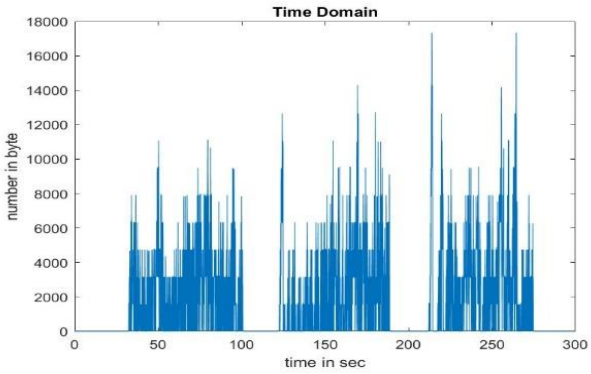

Figure 3: 2nd Attacker packets in Time Domain

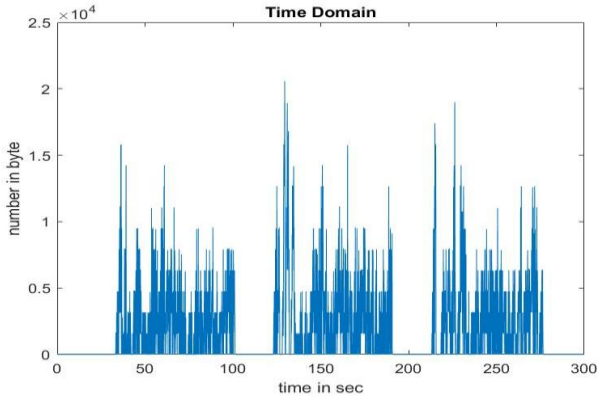

Figure 4: 3rd Attacker Packets in Time domain

For each packet, we studied and applied statistical model, we got X values result for each packet from (2) and applied in (3). We take the Fourier Frequency of the signal, so that we can check the periodicity of the signal. Then, we applied the periodogram in (4). For total we used (5). We apply the Fisher test in our result from (6), we calculate (7) to check if the condition is true. Attack flows represented in frequency domain, the Fisher test's G-statistic value shows there is an attack From (7), the test shows $G \leq G_f$ there is an attack in the signal.

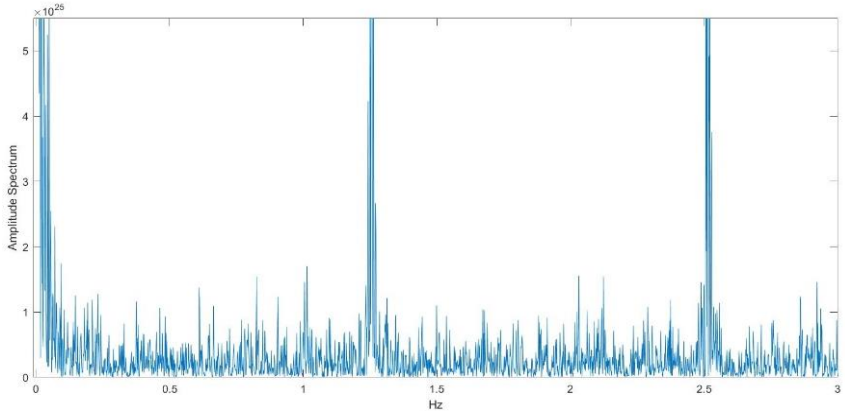

Figure 6: Total S(F) TCP Periodogram

Conclusion

In Distributed Denial of service (DDoS), 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, attackers synchronize 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 Distributed Denial of Service (DDoS) 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. We focus on the detection of DDoS attacks in spectral domain. We study our data using MATLAB and statistical methods. We use the periodogram to find the malicious attack vector. We tested, analyzed, and simulated the different types of flows in a computer network by using MATLAB. Also, we use Fisher statistical methods for detection of the DDoS attack components. The Fisher statistical methods is used to study the data in multiple times series. Then, we test the data by using the Fisher test to find an attacker that is hiding in the process. We applied the MATLAB toolboxes to study the data. We used to apply most communications and signal processing toolboxes to notice differences in the time domain and spectral domain. This paper intends to present the applications of our theoretical class lectures and lab sessions in order to solve real world engineering problems. It is a groundbreaking work in applying fisher statistics methods to cybersecurity and gives a new perspective to use MATLAB and other software to preprocess data and simulating our results.

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. Abhilash Singh, Kausthav Pratim Kalita, Sweta Bhadra “An Efficient Entropy Based Approach for the Detection of DDOS Attack”, International Journal of Scientific Research in Science, Engineering and Technology,2018
6. Lu Zhou, Mingchao Liao, Cao Yuan, Haoyu Zhang “Low-Rate DDoS Attack Detection Using Expectation of Packet Size”, Hindawi, Security and Communication Networks, Volume 2017.
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.
10. Changwang Zhang, Zhiping Cai, Weifeng Chen, Xiapu Luo, Jianping YinBhadra “Flow level detection and filtering of low-rate DDoS”, ElSEVIR Journal, 2012.

Bio

Auther[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. His 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”.