

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221284679>

Periodic Behavior in Botnet Command and Control Channels Traffic

Conference Paper · November 2009

DOI: 10.1109/GLOCOM.2009.5426172 · Source: DBLP

CITATIONS

33

READS

522

3 authors, including:



Basil AsSadhan

King Saud University

41 PUBLICATIONS 605 CITATIONS

[SEE PROFILE](#)



Jose M F Moura

Carnegie Mellon University

751 PUBLICATIONS 28,911 CITATIONS

[SEE PROFILE](#)

Periodic Behavior in Botnet Command and Control Channels Traffic*

Basil AsSadhan and José M. F. Moura
Electrical and Computer Engineering Department
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA, USA
bsadhan@ece.cmu.edu, moura@ece.cmu.edu

David Lapsley
BBN Technologies
10 Moulton Street, Cambridge, MA, USA
dlapsley@sonusnet.com

Abstract—A botnet is a large network of bots that are under the control of a bot herder. Botnets have become a significant threat to network communications and applications. Botnets' execution relies on Command and Control (C2) communication channels traffic, which occur prior to the attack activity itself. Therefore, the detection of C2 communication channels traffic enables the detection of the members of a botnet before any target is attacked.

We study the periodic behavior of C2 traffic that is caused by the pre-programmed behavior of bots to check for and download updates every T seconds. We use this periodic behavior of the C2 traffic to detect bots. This involves evaluating the periodogram of traffic in the monitored network. Then applying Walker's large sample test to the maximum ordinate of the periodogram to determine if it is due to a high periodic component in the traffic or not, and, if it is, then it is bot traffic.

We apply the test to a TinyP2P botnet generated by SLINGbot and show a strong periodic behavior in the bots traffic. We study the effect of the period's length and duty cycle of the C2 traffic on the test performance and find that it increases with the increase of the duty cycle and/or the decrease of the period length. We analyze the test's performance in the presence of injected random noise traffic and develop a lower and an upper bounds for the test performance.

Index Terms—Botnet detection, period, duty cycle, SNR, periodogram, Walker large sample test.

I. INTRODUCTION

A botnet is a large network of bots (compromised computers) that are under the control of a bot herder. Botnets have become a significant threat to network communications and applications, as they increase the efficiency of network attacks such as denial-of-service (DoS) attacks, scanning, phishing, E-mail spam, identity theft, click fraud, and espionage. This capability of a botnet is attributed to the large number of hosts that it controls, which ranges from hundreds to thousands that work together in carrying out an attack, as opposed to when only a few number of hosts carry out the attack.

Unlike other types of malware, botnets execution relies on *Command and Control* (C2) communication channels traffic. A bot herder uses these channels to command bots to execute attack activities and to control bots to download the

information and code they need to execute their attacks. We note here that the C2 communication channels traffic occurs prior to the attack activity itself. Therefore, the detection of C2 communication traffic will enable the detection of the members of a botnet before any target is attacked.

Botnet C2 communication traffic, in general, has low volume and is well behaved (i.e., it does not violate network protocol rules). This makes its detection difficult especially if there is only a small number of botnet members in the monitored network. However, in many botnets variants, bots are pre-programmed to check for and download updates every T seconds. This may be due to a number of factors including ease of coordination and ease of programming. Due to this pre-programmed behavior of bots, we observe periodic behavior in C2 communication traffic [1]. We use this observation to detect bots by looking for periodic behavior in the C2 communication traffic in the monitored network.

In this paper, we extract from a bot's C2 communication traffic the packet count sequence to evaluate its periodogram, [2]. Then we apply Walker's large sample test, [3], to the maximum ordinate of the periodogram to determine if it is due to a high periodic component in the traffic or not. The presence of a periodic component in the traffic is a strong sign that the host sending/receiving it is a member of a botnet. We also study the effect of the period's length and duty cycle of the C2 traffic on the test's performance. In addition, we examine how does the test perform in the presence of injected random noise traffic. We note that the use of statistical properties of network traffic to analyze, infer, classify, or detect network activities has been used in previous studies, see for example, [4], [5], [6], [7], [8], [9].

The rest of the paper is organized as follows, Section II explains the motivation of using periodograms to detect the periodic behavior of the C2 botnet traffic. Section III presents Walker's large sample test that is applied to the maximum ordinate in the periodogram to test if it is due to a periodic component or not. In Section IV, we generate the test data traffic, apply the test, and evaluate and analyze the results, and in Section V we give our conclusions.

*This material is partially based on work supported by the United States Air Force under Contract No. FA8750-07-C-0212. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Department of Homeland Security or the U.S. Air Force.

II. PERIODIC BEHAVIOR IN C2 BOTNET TRAFFIC

Periodic behavior arises in botnet C2 traffic due to the pre-programmed behavior in bots. In many botnet variants with different structures and communication protocols, bots frequently contact each other every T seconds to receive commands, update data, and send *keep-alive* messages. This results in a periodic behavior in a host's traffic at the transport port number used by the bot. Therefore, detecting periodic behavior in a host's network traffic is an indication that the host is part of a botnet. We stress that periodic behavior might also arise in other forms of traffic, for example, software updates. We also note that, in other botnet variants, bots can contact each other at arbitrary times, which results in C2 communication that does not exhibit periodic behavior. We point to all these and their effect on the false alarm and detection rates and on the efficiency of the message exchange of a botnet in [1].

In addition to our work in [1], we are aware of a previous study, [10], that exploits the periodic behavior of botnet C2 traffic to detect bots. The autocorrelation function of the host's traffic was computed in the time domain to examine whether the traffic has a periodic component or not. We propose to work in the frequency domain, as it involves less amount of computations, therefore is faster in time. This can be achieved by analyzing the *Power Spectral Density* (PSD) of the C2 traffic to detect its periodic behavior. The PSD can be estimated by taking the Fourier Transform of the autocorrelation function. Alternatively, a PSD can be estimated using periodograms, [2]. The periodogram of a time sequence (signal) provides its power at different frequencies. Periodograms have been used in other areas to detect periodic behavior like biology [11] and geophysics [12]. It has also been used to analyze network traffic, see for example [13].

The periodogram is useful to identify frequency components that possess high power levels. Therefore, the periodogram of a periodic signal will have a high peak at the reciprocal of the fundamental period of the signal when compared to the mean of the periodogram.

To evaluate the periodogram of the traffic of a given host at a given port, we first count the number of packets over a selected aggregation interval to produce a discrete time packet count sequence. The periodogram $P_{xx}[k]$ of a discrete time sequence $x[n]$ is the square magnitude of the Discrete Fourier Transform (DFT) of the signal evaluated by

$$P_{xx}[k] = \frac{1}{N} |X[k]|^2,$$

where

$$X[k] = \sum_{n=0}^{N-1} x[n] \exp\left(\frac{-j2\pi kn}{N}\right)$$

is the N -point DFT.

Figure 1 illustrates the usefulness of the periodogram in detecting periodic behavior. The top-left plot shows a square wave signal with levels 0 and 1, a duty cycle of 20%, a

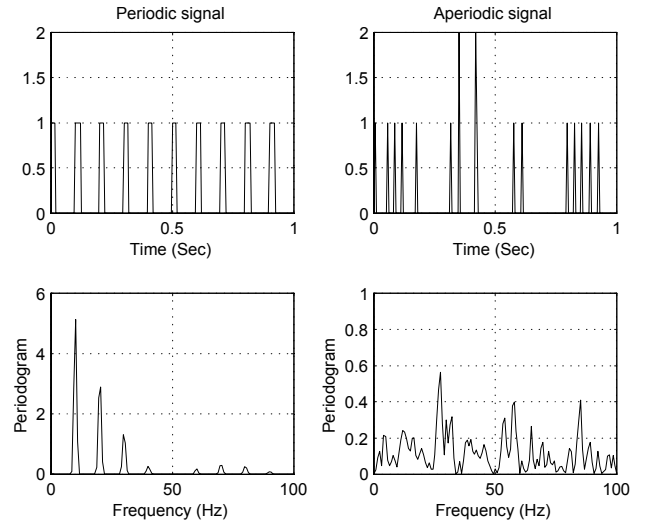


Fig. 1. Left plots show a periodic signal with a period of 0.1 second, and its one sided periodogram that consists of a large peak at the fundamental frequency and smaller peaks at the harmonics. Right plots show an aperiodic signal, and its one sided periodogram that consists of several small peaks.

period of 0.1 second, and a duration of 1 second. The one sided periodogram of this square wave after subtracting its mean (bottom-left) consists of a “large” peak at 10 Hz, smaller peaks at multiples of 10 Hz, and almost zero elsewhere. These smaller peaks represent the harmonic components. The top-right plot of the figure shows an aperiodic Poisson random signal with a mean and variance of 0.16, and a duration of 1 second. The value of the variance of this signal is selected so that it has its power after subtracting its mean is the same as the power of square wave after subtracting its mean, which is 0.16. The one sided periodogram of the aperiodic signal after subtracting its mean (bottom-right) consists of several peaks, however, none of them has a large value when compared to the mean of the periodogram.

We note that we use the standard periodogram rather than the Welch's method of averaged periodogram [2], [14]. This is because in our work we are interested in the detection and estimation of a single periodic component, which is better achieved using the standard periodogram as discussed in [15]. The averaged periodogram is used by others to represent the spectral density of the traffic [13] and to detect a periodic component with nonstationary phase [15].

III. TESTING THE SIGNIFICANCE OF THE PERIODOGRAM PEAK

From the bottom plots of Figure 1, one can see that the periodogram of either a periodic or aperiodic signal will always have a peak. Therefore, we need to be able to determine whether the peak is significant enough to declare it is due to a periodic component with the frequency where the peak is located or not. We use binary hypothesis testing, [16], [17], to achieve this.

Before we set up the two hypotheses, we assume the following; the packet count sequence extracted from the network traffic communication of a given host on a given port has a Poisson distribution. We acknowledge that Poisson statistics may or may not be an accurate model for network traffic at a given port. We use it as a first step to simplify the analysis. More complex models will affect the threshold selection.

Poisson statistics are a good approximation to binomial statistics when the binomial parameter n is large and the binomial parameter p is small¹ [18]. A binomial random variable with a large n and its Poisson approximation (when p is small) can be, based on the central limit theorem, approximated by a Gaussian random variable [18]. Then, the packet count sequence extracted from the network traffic communication for a given host on a given port, after subtracting its mean and normalizing it by its standard deviation, can be treated as having a standard Gaussian distribution (i.e., $N(0, 1)$).

We now set up the null hypothesis H_0 , which is that the packet count sequence $x[n]$ is Gaussian, against the alternative hypothesis that $x[n]$ has a periodic component at some unspecified frequency plus the Gaussian noise. Under H_0 , it can be shown that the ordinates $P_{xx}[k_0]$ of the periodogram of $x[n]$ are independently identically distributed (i.i.d.) [3]. Each $P_{xx}[k_0]$ has a distribution that is proportional to a Chi-Square distribution with two degrees of freedom [3]. Specifically,

$$P_{xx}[k_0]/\sigma_x^2 = \chi_2^2.$$

Since a Chi-Square distribution with two degrees of freedom is equivalent to an exponential distribution with mean 2, it follows that the probability density function of $P_{xx}[k_0]/\sigma_x^2$ is

$$f(x) = \frac{1}{2} \exp(-x/2), \quad 0 \leq x < \infty.$$

Therefore, for $z \geq 0$,

$$\begin{aligned} \Pr [P_{xx}[k_0]/\sigma_x^2 \leq z] &= \int_0^z \frac{1}{2} \exp(-x/2) dx \\ &= 1 - \exp(-z/2). \end{aligned} \quad (1)$$

Since we are interested in the maximum value of the periodogram ordinates, we define the ratio test statistics,

$$\gamma_x = \frac{\max_{0 \leq k \leq m-1} (P_{xx}[k])}{\sigma_x^2}. \quad (2)$$

Since under H_0 the periodogram ordinates $P_{xx}[k_0]$ are i.i.d., then it follows that, for $z \geq 0$,

$$\begin{aligned} \Pr [\gamma_x > z] &= 1 - \Pr [\gamma_x \leq z] \\ &= 1 - \Pr [(P_{xx}[k_0]/\sigma_x^2) \leq z, \text{ all } k_0] \\ &= 1 - (1 - \exp(-z/2))^m, \end{aligned} \quad (3)$$

¹A binomial random variable is defined as the sum of n independently identically distributed (i.i.d.) Bernoulli random variables with probability p .

where m is the number of ordinates at the positive frequencies of the periodogram.

Equations (1)–(3) assume that the variance σ_x^2 is known a priori. However, in practice it is almost never known, and an estimate is used. The variance σ_x^2 can be estimated directly from the time sequence $x[n]$ using the sample variance. But since $x[n]$ might not be always available, it is preferable to estimate σ_x^2 directly from $P_{xx}[k]$. The estimate of σ_x^2 according to [3] can be evaluated by

$$\widehat{\sigma_x^2} = \frac{1}{2m} \sum_{k=0}^{m-1} P_{xx}[k].$$

The quantity $\widehat{\sigma_x^2}$ is an unbiased estimate of σ_x^2 , and we use it in place of σ_x^2 in (2) to define the sample ratio test statistic,

$$g_x^* = \frac{\max_{0 \leq k \leq m-1} (P_{xx}[k])}{\frac{1}{2m} \sum_{k=0}^{m-1} P_{xx}[k]}. \quad (4)$$

When m is large, $\widehat{\sigma_x^2}$ will be a good approximate to σ_x^2 , thus, we can treat the denominator of (4) as σ_x^2 . Then, g_x^* will have the same distribution as γ_x , and asymptotically under H_0 we have, for $z \geq 0$,

$$\Pr [g_x^* > z] \sim 1 - (1 - \exp(-z/2))^m. \quad (5)$$

The asymptotic distribution of g_x^* is the basis of *Walker's large sample test* for $\max(P_{xx}[k])$, page 407 of [3].

Under the alternative hypothesis H_1 , where the signal is periodic, g_x^* will be large. This allows us to use a one-sided test and select the critical region $g_x^* > z_\alpha$, where z_α is selected so the right hand side of (5) is equal to α , which is the false alarm probability of the test. If the calculated value of g_x^* from the sample data is less than z_α , then we accept H_0 , and conclude that $x[n]$ does not have any periodic component. If g_x^* is larger than z_α , then we reject H_0 with a false alarm probability of α and conclude that $x[n]$ has a periodic component. The value of α is selected based on how small we would like the false alarm probability to be.

We note that Fisher has derived an exact test for $\max(P_{xx}[k])$ [3], [11], [19]. However, we prefer to use Walker's test for the following reasons: first, Fisher's test involves using combinatorial coefficients, which are limited in their accuracy as the number of sample points gets large, hence we lose the exactness of the test. Second, even if the number of sample points is not large, evaluating z_α to set the critical region to α is not straight forward. This requires for each calculated g_x^* to evaluate $\Pr [\gamma_x > g_x^*]$ and check if it is smaller than α or not. Third, usually, we are not short of network traffic to get a good estimate $\widehat{\sigma_x^2}$ to use in the denominator of (4).

IV. EXPERIMENTAL SETUP: EVALUATION & ANALYSIS

A. Botnet Setup

We use SLINGbot, [20], (System for Live Investigation of Next Generation bots) to generate examples of ground truth

C2 traffic. SLINGbot generates the C2 communication traffic, which includes downloading bot software, connecting to bot C2 servers, and receiving botnet commands. SLINGbot uses a C2 feature space that consists of five separate dimensions for the functionality of botnets. The five dimensions are: topology, rallying mechanism, communication protocol, control mechanism, and command authentication mechanism.

We use SLINGbot to set a of TinyP2P (Peer-to-Peer) botnet² that consists of five bots and one bot herder. Each of the five bots is pre-programmed to update its data with a different period. The experiment was run for about 35 seconds, and the periods that the bots use are 3, 4, 5, 6, and 7 seconds. We use an aggregation interval of 100 ms to extract the packet count of each bot's traffic on port number 11375, which is the port number used for C2 communication. In the following section, we discuss the results on this dataset.

B. Results

The top plots of Figure 2 show the packet count sequences of two TinyP2P bot C2 traffic. The periods of the traffic for the first bot (top-left) and the second bot (top-right) are 4 and 6 seconds, respectively. The periodic behavior is apparent in both plots. At the beginning of each period each bot checks with its peers to see if there are any updates, and if so it downloads them; after that it becomes silent until the next period starts. The duration of the active time in the period is not affected by the duration of the period in the two bots. Therefore, the duty cycle of the first bot's traffic is higher than the duty cycle of the second bot's traffic; the reference to the duty cycle will be apparent below.

The bottom plots of Figure 2 show the *modified* periodogram³, [2], of the packet count sequences after subtracting their mean and normalizing by their standard deviation. The periodogram of the first bot (bottom-left) consists of a large peak at 273 MHz, which corresponds to a period of 3.7 seconds. This period agrees with the 4-second pre-programmed period of the bot. The periodogram of the first bot also consists of smaller peaks at the multiples of the frequency. The smaller peaks are the harmonic components as discussed in Section II. The same observations apply to the periodogram of the second bot (bottom-right), except that the peak is located at 176 mHz, which corresponds to a period of 5.7 seconds. Again, this period agrees with the 6-second pre-programmed period of the bot.

Duty Cycle: We test the significance of the peak, $\max(P_{xx}[k])$, of each of the two periodograms in Figure 2 by evaluating the ratio g_x^* in (4) for each periodogram. We select the false alarm probability, α , to be equal to 0.1%. This value of α is selected since, in general, it is desired to have a low false alarm rate, in particular, in network anomaly detection systems. The value of α is equated to the right hand

²A TinyP2P botnet is a botnet that uses TinyP2P as its communication protocol [20].

³The modification is achieved by multiplying the sequence by a Hamming window to reduce the level of the side-lobes. We note that windowing comes at the cost of reducing the sharpness of the peak.

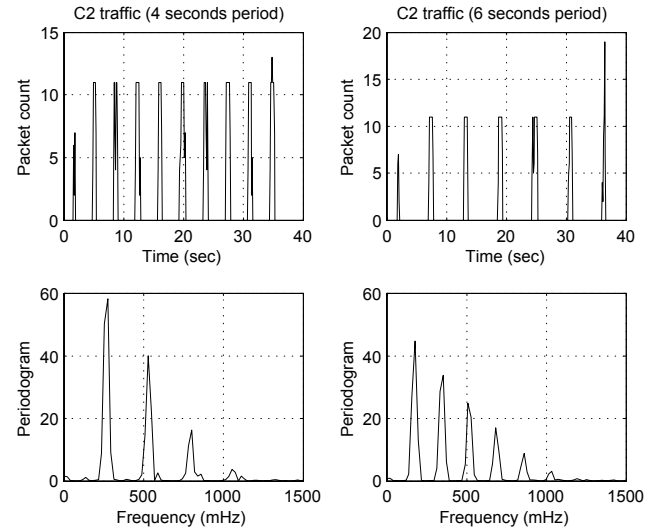


Fig. 2. The top plots show the packet count sequences for the C2 communication channels traffic of two TinyP2P bots with two different periods. The bottom plots show the one sided periodogram of each sequence. The aggregation interval for the packet count is 100 ms.

side of (5), resulting in a threshold value, $z_{0.1\%}$, of 24.9⁴. We find the values of g_x^* for each of the two periodograms to be 101.4 and 77.0; both are larger than $z_{0.1\%}$. Therefore, we reject the null hypothesis and conclude that both sequences have a periodic component. The periodogram of the first bot has a higher value of g_x^* than the second bot because the ratio of the periodogram's peak to the harmonic components is larger, which is due to the larger duty cycle. This is due to the following fact. The Fourier transform of a train of rectangular pulses is a sampled sinc function. For trains of rectangular pulses with a larger duty cycle, the ratio between the main peak and the harmonic components at the Fourier transform is also larger. Hence, for trains of rectangular pulses with larger duty cycle, the main peak is more significant.

Noise: The periodic behavior of the traffic in the two bots is very apparent as confirmed by the fact that the two values of the ratio g_x^* are more than three times larger than the value of $z_{0.1\%}$. The bot herder can attempt to reduce this strong periodic behavior by injecting random noise traffic $w[n]$ to the C2 traffic $x[n]$ to hide its periodic behavior. The total traffic is $y[n] = x[n] + w[n]$, and the resulting ratio g_y^* becomes noisy. The effect of injecting random noise to the packet count on the value g_y^* is illustrated in Figure 3, where a Poisson random sequence is added to the packet count sequences in Figure 2. The variance of the Poisson random sequence σ_w^2 in each case is set so that the *Signal-to-Noise Ratio* (SNR) is -6 dB (i.e., the power of the Poisson sequence is four times the power of the packet count sequence). The periodic behavior of the two packet count sequences is no longer visible in the time domain (top plots). But it is still apparent in the frequency domain

⁴The number of ordinates at the positive frequencies of both periodograms, m , that is used in evaluating $z_{0.1\%}$ is 256.

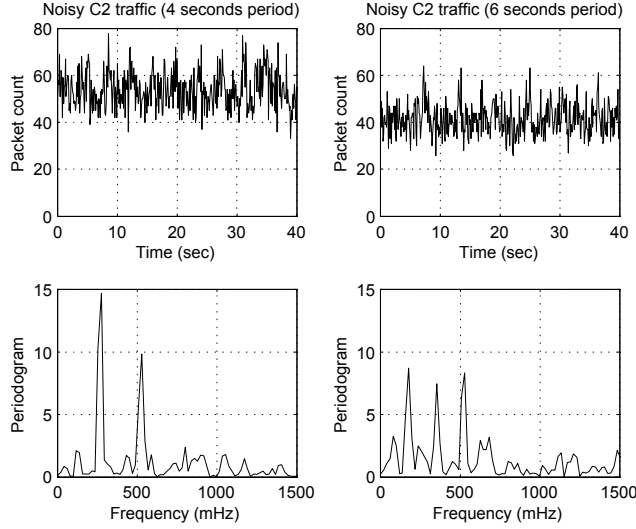


Fig. 3. The packet count sequences and their one sided periodogram for the C2 traffic in Figure 2 with added Poisson noise (SNR = -6 dB).

(bottom plots), where $\max(P_{yy}[k])$ of each periodogram is still located at the frequency of the sequence $x[n]$.

The added noise results in having a g_y^* value of 31.8 and 17.0 for the periodogram of the first and second bots traffic, respectively (we remind the reader that the two corresponding values of g_x^* without adding the noise are 101.4 and 77.0). The g_y^* value of the first bot is still larger than $z_{0.1\%} = 24.9$, hence, the test will declare that the sequence has a periodic component, whereas the g_y^* value of the second bot is smaller than $z_{0.1\%}$, hence, the test will declare that the sequence does not have a periodic component. The first bot resisted the added noise since it had a higher g_x^* value in the absence of the noise. This is because its traffic had a higher duty cycle and a higher number of repeated periods than the second bot's traffic. The detection of the periodic behavior of the second bot's C2 traffic can be achieved at the expense of increasing the false alarm probability α , which will decrease the value of z_α . In addition, the periodic behavior of the C2 traffic can be detected by monitoring longer intervals of the C2 traffic. This will allow observing larger number of periods, which will increase the peak of the traffic's periodogram.

We further study the effect of adding noise to the first bot C2 traffic by varying the level of the noise power, and plotting $E[g_y^*]$ against the inverse of the signal-to-noise ratio, SNR^{-1} . The solid curve in Figure 4 is the average of 2000 runs; it shows the effect of increasing the noise power on reducing the value of $E[g_y^*]$. The value $E[g_y^*]$ when $\text{SNR}^{-1} = 0$ (i.e., in the absence of noise) is the value of $g_x^* = 101.4$. The value $E[g_y^*]$ reaches the value of $z_{0.1\%} = 24.9$ at $\text{SNR}^{-1} = 4.5$.

We develop two bounds for $E[g_y^*]$ that are shown in Figure 4. The dashed curve above the $E[g_y^*]$ curve shows

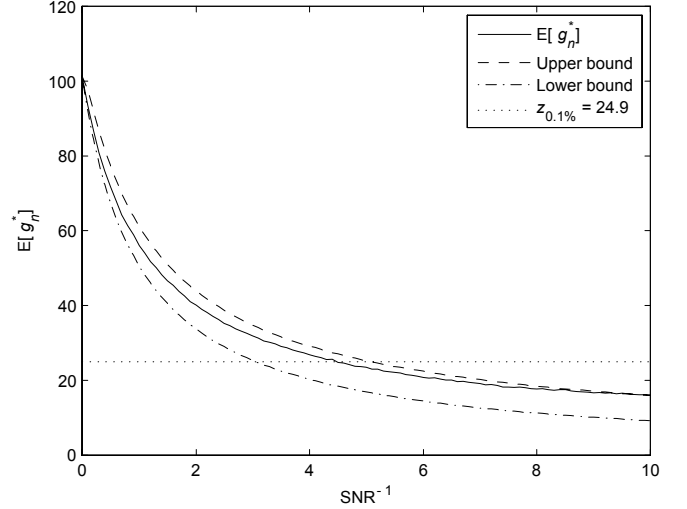


Fig. 4. The effect of adding random noise traffic to the bot's C2 traffic in the left plots of Figure 2 on $E[g_n^*]$.

$$\frac{(\sqrt{g_x^*} + \sqrt{\text{SNR}^{-1}})^2}{1 + \text{SNR}^{-1}}, \quad (6)$$

and the dashed dotted curve below the $E[g_y^*]$ curve shows

$$\frac{g_x^*}{1 + \text{SNR}^{-1}}. \quad (7)$$

The plots of Figure 4 shows that (6) is a tight upper bound, while (7) is a looser lower bound. These two bounds are helpful in determining the level of noise that the periodogram of a bot traffic can sustain for a given false alarm probability.

We now show how we develop the two bounds for $E[g_y^*]$ starting with the upper bound. Instead of bounding the mean of the sample ratio test statistic $E[g_y^*]$, we bound the ratio test statistic $E[\gamma_y]$. Equation (2) shows that γ_y is expressed by $P_{yy}[k]$ and σ_y^2 ⁵. Since the added noise traffic $w[n]$ is generated independently from the C2 traffic $x[n]$, σ_y^2 is equal to $\sigma_x^2 + \sigma_w^2$. We expand $P_{yy}[k]$ as follows:

$$P_{yy}[k] = \frac{1}{N} |Y[k]|^2 = \left| \frac{X[k]}{\sqrt{N}} + \frac{W[k]}{\sqrt{N}} \right|^2.$$

Since $|X + Y| \leq |X| + |Y|$, we have

$$P_{yy}[k] \leq \left[\frac{|X[k]|}{\sqrt{N}} + \frac{|W[k]|}{\sqrt{N}} \right]^2.$$

We note that $|X[k]|/\sqrt{N}$ is $\sqrt{P_{xx}[k]}$, and by taking the expectation of both sides, we get

$$E[P_{yy}[k]] \leq \left[\sqrt{P_{xx}[k]} + \sqrt{\sigma_w^2} \right]^2.$$

⁵Recall that in our analysis we normalize $y[n]$ by subtracting its mean and dividing by its standard deviation. So, in the sequel, $y[n]$, $w[n]$, and $x[n]$ are zero mean.

We can now express $E[\gamma_y]$ as

$$\begin{aligned} E[\gamma_y] &= \frac{\max_{0 \leq k \leq m-1} (E[P_{yy}[k]])}{\sigma_y^2} \\ &\leq \frac{\max_{0 \leq k \leq m-1} \left(\left[\sqrt{P_{xx}[k]} + \sqrt{\sigma_w^2} \right]^2 \right)}{\sigma_x^2 + \sigma_w^2}. \end{aligned}$$

Dividing the numerator and denominator by σ_x^2 we get

$$E[\gamma_y] \leq \frac{\left(\sqrt{\gamma_x} + \sqrt{\text{SNR}^{-1}} \right)^2}{1 + \text{SNR}^{-1}},$$

which corresponds to (6).

To develop the lower bound, we note that $E[P_{yy}[k]] \geq P_{xx}[k]$, therefore we have

$$E[\gamma_y] = \frac{\max_{0 \leq k \leq m-1} (E[P_{yy}[k]])}{\sigma_y^2} \geq \frac{\max_{0 \leq k \leq m-1} (P_{xx}[k])}{\sigma_x^2 + \sigma_w^2}.$$

Dividing the numerator and denominator by σ_x^2 we get

$$E[\gamma_y] \geq \frac{\gamma_x}{1 + \text{SNR}^{-1}},$$

which corresponds to (7).

V. CONCLUSIONS

We study the periodic behavior of botnets' command and control (C2) communication traffic to detect bots. The periodic behavior arises in botnets because bots are pre-programmed to check for and download updates every T seconds. This is true in many botnets variants independent of the structure and communication protocol they use. To test the periodic behavior, we compute the periodogram of the packet count sequence from the C2 communication traffic. We use Walker's large sample test to determine the significance of the maximum ordinate of the periodogram. If it passes the test, then the traffic is determined to have periodic behavior and so determined to be botnet C2 traffic.

We test this procedure by analyzing the C2 traffic of a TinyP2P botnet generated by SLINGbot, and show a strong periodic behavior in the bots C2 traffic in Figure 2. We study the effect of the period's length and duty cycle of the C2 traffic on the test performance and find that it increases with the increase of the duty cycle and/or the decrease of the period length. We analyze the test performance in the presence of injected random noise traffic and develop a lower and an upper bounds for the test performance as shown in Figure 4. The bot herder can attempt to hide the C2 traffic periodic behavior by uniformly randomizing the period within a certain small range (e.g., 4–6 seconds). This can be modeled by a random phase, which will be part of our future work.

ACKNOWLEDGMENT

David Lapsley would like to acknowledge the support from the Cyber Security Program Area of the Command, Control and Interoperability Division within the Science and Technology Directorate of the U.S. Department of Homeland Security.

REFERENCES

- [1] B. AsSadhan, J. M. F. Moura, D. Lapsley, C. Jones, and W. T. Strayer, "Detecting botnets using command and control traffic," in *IEEE International Symposium on Network Computing and Applications (NCA)*, Cambridge, MA, USA, Jul. 9–11, 2009.
- [2] A. Oppenheim, R. Schaffer, and J. Buck, *Discrete-Time Signal Processing*. Prentice-Hall, 1999.
- [3] M. B. Priestley, *Spectral Analysis and Time Series*. Academic Press, 1982.
- [4] R. R. Kompella, S. Singh, and G. Varghese, "On scalable attack detection in the network," in *USENIX/ACM Internet Measurement Conference*, Taormina, Sicily, Italy, Oct. 2004.
- [5] Y. Kim, W. C. Lau, M. C. Chuah, and H. J. Chao, "PacketScore: a statistics-based packet filtering scheme against distributed denial-of-service attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 3, no. 2, pp. 141–155, Apr.-Jun. 2006.
- [6] A. Dainotti, A. Pescapé, and G. Ventre, "Wavelet-based detection of DoS attacks," in *IEEE GLOBECOM*, New Orleans, LA, USA, Nov.27–Dec.1 2006.
- [7] A. Karasaridis, K. Meier-Hellstern, and D. Hoeflin, "Detection of DNS anomalies using flow data analysis," in *IEEE GLOBECOM*, New Orleans, LA, USA, Nov.27–Dec.1 2006.
- [8] A. Dainotti, W. de Donato, A. Pescapé, and P. S. Rossi, "Classification of network traffic via packet-level hidden Markov models," in *IEEE GLOBECOM*, San Francisco, CA, USA, Nov.30–Dec.4 2008.
- [9] Y.-C. Chang, K.-T. Chen, C.-C. Wu, and C.-L. Lei, "Inferring speech activity from encrypted Skype traffic," in *IEEE GLOBECOM*, San Francisco, CA, USA, Nov.30–Dec.4 2008.
- [10] G. Gu, J. Zhang, and W. Lee, "BotSniffer: Detecting botnet command and control channels in network traffic," in *the 15th Annual Network and Distributed System Security Symposium (NDSS'08)*, San Diego, CA, USA, Feb. 10–13, 2008.
- [11] M. Ahdesmlaki, H. Lähdesmlaki, and O. Yli-Harja, "Robust Fisher's test for periodicity detection in noisy biological time series," in *IEEE International Workshop on Genomic Signal Processing and Statistics GENSPS*, Tuusula, Finland, Jun. 2007.
- [12] D. J. Thomson, L. Lanzerotti, L. Medford, C. MacLennan, A. Meloni, and G. Gregori, "Study of tidal periodicities using a transatlantic telecommunications cable," *Geophysical Research Letters*, vol. 13, no. 6, pp. 525–528, 1986.
- [13] C. Partridge, D. Cousins, A. W. Jackson, R. Krishnan, T. Saxena, and W. T. Strayer, "Using signal processing to analyze wireless data traffic," in *ACM Workshop on Wireless Security (WiSe)*, Atlanta, GA, USA, Sep. 2002, pp. 67–76.
- [14] P. D. Welch, "The use of the fast Fourier transform for estimation of spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–74, Aug. 1967.
- [15] H. C. So, Y. T. Chan, Q. Ma, and P. C. Ching, "Comparison of various periodograms for sinusoid detection and frequency estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 945–952, Jul. 1999.
- [16] H. V. Trees, *Detection, Estimation, and Modulation Theory. Part 1*. John Wiley & Sons, 2001.
- [17] K. Fukunaga, *Statistical Pattern Recognition*, 2nd ed. Academic Press, 1990.
- [18] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*. Addison Wesley, 1994.
- [19] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag, 1991.
- [20] A. W. Jackson, D. Lapsley, C. Jones, M. Zatko, C. Golubitsky, and W. T. Strayer, "SLINGbot: A system for live investigation of next generation botnets," in *Cybersecurity Application and Technologies Conference for Homeland Security (CATCH)*, Washington, DC, USA, Mar. 2009.