Are Covert DDoS Attacks Facing Multi-Feature Detectors Feasible?

Amir Reza Ramtin[1], Don Towsley[1], Philippe Nain[2],
Edmundo de Souza e Silva[3], Daniel S. Menasche[3]
[1]University of Massachusetts at Amherst, USA, [2]INRIA, France, [3]UFRJ, Brazil

ABSTRACT

We state and prove the square root scaling laws for the amount of traffic injected by a covert attacker into a network from a set of homes under the assumption that traffic descriptors follow a multivariate Gaussian distribution. We numerically evaluate the obtained result under realistic settings wherein traffic is collected from real users, leveraging detectors that exploit multiple features. Under such circumstances, we observe that phase transitions predicted by the model still hold.

1. INTRODUCTION

Distributed denial of service attacks are pervasive[1]. The extraordinary amount of traffic generated by those attacks has already disrupted cloud services and critical infrastructures, but the visibility of those attacks usually renders them short lived. From the attacker standpoint, this suggests that the next generation of botnets may attempt to be covert (undetectable).

A covert attack must be executed under the limits of detector capabilities. To achieve covertness, the key insight consists in leveraging a large number of devices, in such a way that the aggregate amount of traffic regularly generated by those devices hides the attack traffic.

Prior art. There is vast literature on machine learning methods for anomaly detection and intrusion detection systems. The literature on scaling laws of covert attacks is much scarcer, and the works on communication with low probability of detection (LPD) typically account for resource constrained detectors. In this work, we point towards novel directions to analyze scaling laws for covert attacks, considering detectors that can account for multiple features while determining the presence of attacks.

Goals. We consider a set of n compromised home networks. In each home network, there is at least one device amenable to be used by the attacker in an attack campaign. We pose the following questions: at what rate can the attack traffic grow, with respect to n, while still remaining covert? How do multiple features impact the scaling laws? To what extent is covertness sensitive to the nature of the distribution models?

Our answers to the above questions involve theoretical and numerical evaluation methods.

[1]https://www.digitalattackmap.com/

Contributions. We begin by establishing a square-root scaling law for the amount of traffic that can be injected into the network with respect to the number of homes, so that the attack is covert (Section 3). The law is derived under the assumption that traffic features follow a multivariate Gaussian distribution, and the achievability of the attack is derived accounting for likelihood ratio test (LRT) detectors. Then, we illustrate the tightness of the scaling law regardless of the distribution type (Section 4). Numerical evaluation allows us to assess the square root law with respect to a multi-feature LRT detector. To that aim, we partnered with a Brazilian ISP to collect regular traffic from its home users. Then, we simulate attacks on top of such traffic and observe that phase transitions predicted by the model still hold (Section 5).

2. PROBLEM SETUP

Consider a collection of n homes. Each home-router continuously measures the traffic during a time slot (slot) and sends this information to an ISP fusion center where detection takes place. We assume that there is an attacker who may or may not launch an attack during a slot. The system administrator (henceforth known as admin) can perform a hypothesis test on observations with the null hypothesis H_0 being that the attacker does not launch an attack and the alternate hypothesis H_1 that he does launch an attack. We are interested in the following question: can the attacker launch an attack without being detected by the admin and, if so, how large can such an attack be? Admin can tolerate some false positives, or cases when the statistical test incorrectly concludes an attack is under way. This rejection of H_0 is known as a false alarm, and, following standard nomenclature, we denote its probability by p_{FA}. Admin's test may also fail to indicate that an attack is taking place. Acceptance of H_0 when it is false is known as a missed detection, and we denote its probability by p_{MD}. Then, the sum $p_{FA} + p_{MD}$ characterizes the necessary trade-off between false alarms and missed detections in the design of a hypothesis test.

Denote by f_0 the pdf of the regular traffic by a home-router in a slot in the absence of an attack (i.e. when H_0 is true) and by f_1 the pdf of traffic by a home-router in a slot in the presence of an attack (i.e. when H_1 is true). When f_0 and f_1 are known to admin, he can construct an optimal statistical hypothesis test (such as the Neyman-Pearson or likelihood ratio test) that minimizes the sum of error probabilities [4, Ch. 13], $p_E = p_{FA} + p_{MD}$. When an attack is launched the adversary targets each home with probability

$q(n)$; in particular, all homes are chosen when $q(n) = 1$. An attack is covert provided that, for any $\epsilon > 0$, the attacker has a strategy for each n such that

$$\liminf_n p_E \geq 1 - \epsilon. \quad (1)$$

3. ACHIEVABILITY

In the following $|\mathbf{M}|$ denotes the determinant of any square matrix $\mathbf{M}$. We recall that a real matrix $\mathbf{M}$ is positive definite if for any non-zero vector $\underline{x}^T \in \mathbb{R}^N$, $\underline{x}^T \mathbf{M} \underline{x} > 0$. Both regular and attack traffic are represented by vectors. More precisely, $\underline{X}_r = (X_{r,1}, \ldots, X_{r,N}) \in \mathbb{R}^N$ characterizes the regular traffic passing from home-router r in a slot and $\underline{Y}_r = (Y_{r,1}, \ldots, Y_{r,N}) \in \mathbb{R}^N$ characterizes the attack traffic passing from home-router r in a slot. We assume that $\underline{X}_1, \ldots, \underline{X}_n$ are iid rvs, that $\underline{Y}_1, \ldots, \underline{Y}_n$ are iid rvs, and that $\underline{X}_r$ is independent of $\underline{Y}_s$ for any r and s.

As another illustration, take $N = 2$ with $X_{r,1}$ the regular traffic at home-router r (resp. $Y_{r,1}$ the attack traffic) counted in *packets* and $X_{r,2}$ the regular traffic at home-router r (resp. $Y_{r,2}$ the attack traffic) counted in *bytes*.

Denote by $f_0(\underline{x})$ and $g(\underline{x}, n)$ the pdfs of $\underline{X}_r$ and $\underline{Y}_r$, respectively. Assume that $\underline{X}_r$ and $\underline{Y}_r$ have multivariate Gaussian distributions with location (mean) parameters $\underline{\mu}_0^T = (\mu_{0,1}, \ldots, \mu_{0,N})$ and $\underline{\mu}_1(n)^T = (\mu_{1,1}(n), \ldots, \mu_{1,N}(n))$, and positive definite covariance matrices Σ_0 and $\Sigma_1(n)$, namely,

$$f_0(\underline{x}) = \frac{e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_0)^T \Sigma_0^{-1} (\underline{x}-\underline{\mu}_0)}}{\sqrt{(2\pi)^N |\Sigma_0|}}, \quad (2)$$

$$g(\underline{x}, n) = \frac{e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_1(n))^T \Sigma_1(n)^{-1} (\underline{x}-\underline{\mu}_1(n))}}{\sqrt{(2\pi)^N |\Sigma_1(n)|}}. \quad (3)$$

The sum of regular and attack traffic, $X_r + Y_r$, at home r when an attack occurs has pdf

$$f_1(\underline{x}, n) = \frac{e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_0-\underline{\mu}_1(n))^T (\Sigma_0+\Sigma_1(n))^{-1} (\underline{x}-\underline{\mu}_0-\underline{\mu}_1(n))}}{\sqrt{(2\pi)^N |\Sigma_0 + \Sigma_1(n)|}}.$$

Under H_0, $\underline{Z}_r = \underline{X}_r$ with pdf f_0. Under H_1, $\underline{Z}_r = \underline{X}_r + \chi_r \underline{Y}_r$ with $q(n) := \mathbb{P}(\chi_r = 1)$, where χ_r is an indicator random variable which equals 1 if home r is selected to actively issue an attack, so that the pdf of $\underline{Z}_r$ is given by

$$h(\underline{x}, n) = (1 - q(n)) f_0(\underline{x}) + q(n) f_1(\underline{x}, n).$$

Denote by $f_0^{(n)}$ (resp. $h^{(n)}$) the joint pdf of $\underline{X}_1, \ldots, \underline{X}_n$ (resp. $\underline{Z}_1, \ldots, \underline{Z}_n$). It is known that the minimum p_E is

$$p_E^\star = 1 - T_V\left(f_0^{(n)}, h^{(n)}\right), \quad (4)$$

with $T_V(u, v) := \int |u(x) - v(x)| dx$ the total variance distance between pdfs u and v [4, Theorem 13.1.1]. The lemma below gives an upper bound on $T_V\left(f_0^{(n)}, h^{(n)}\right)$,

LEMMA 3.1 (UB. ON TOTAL VARIATION DISTANCE). *For all $n \geq 1$, $T_V\left(f_0^{(n)}, h^{(n)}\right) \leq \frac{1}{2}\sqrt{(1 + q(n)^2 C(n))^n - 1}$, where constant $C(n)$, known as the Fisher information constant at origin [3], is given by*

$$C(n) = -1 + \int_{\mathbb{R}^N} \frac{f_1(\underline{x}, n)^2}{f_0(\underline{x})} dx_1 \cdots dx_N.$$

The proof is similar to the proof of Lemma 7.1 in [2]. Together with (1) and (4) this yields the following,

COROLLARY 3.1. *Fix $\epsilon > 0$. If $q(n)\sqrt{C(n)} = \mathcal{O}(1/\sqrt{n})$ then $\limsup_n T_V\left(f_0^{(n)}, h^{(n)}\right) \leq \epsilon$.*

Let $\mathbf{A}(n) = (\Sigma_0 + \Sigma_1(n))^{-1}$ and $\mathbf{B}(n) = \Sigma_0^{-1} - 2\Sigma_0^{-1}(\Sigma_0^{-1} + \Sigma_1(n)^{-1})^{-1}\Sigma_0^{-1}$.

THEOREM 3.1 (ACHIEVABILITY). *The attack is covert if $\exists\, n_0$ such that $\mathbf{B}(n)$ is positive definite for $n \geq n_0$ and if $q(n)\sqrt{C(n)} = \mathcal{O}(1/\sqrt{n})$, where*

$$C(n) = -1 + \frac{|\Sigma_0|\sqrt{|\mathbf{B}(n)^{-1}|}}{|\Sigma_0 + \Sigma_1(n)|} \times e^{-\underline{\mu}_1(n)^T \left(\mathbf{A}(n) - 4\mathbf{A}(n)\mathbf{B}(n)^{-1}\mathbf{A}(n)\right)\underline{\mu}_1(n)}, \quad (5)$$

$\mathbf{B}(n)$ is positive definite if the eigenvalues of $\Sigma_0\Sigma_1(n)^{-1}$ are strictly larger than one.

Sketch of proof. By Corollary 3.1 the attacker is covert if $q(n)\sqrt{C(n)} = \mathcal{O}(1/\sqrt{n})$. On the other hand, when matrix $\mathbf{B}(n)$ is positive definite one can show that constant $C(n)$ is given by (5). The last statement of the proof follows from standard linear algebra arguments. ∎

The proof of Theorem 3.1 holds under the assumption that admin knows the distribution of the attack traffic. If this assumption does not hold then Theorem 3.1 still holds as admin cannot do better with less knowledge.

As a first illustration assume that $N = 1$ (i.e. X_r and Y_r have Gaussian distributions). Introduce $\mathbb{E}[Y_r] = \mu_1(n)$, $\mathrm{var}(\mathrm{X_r}) = \sigma_0^2$, and $\mathrm{var}(\mathrm{Y_r}) = \sigma_1(\mathrm{n})^2$. Then $\mathbf{B}(n) = (\sigma_0^2 - \sigma_1(n)^2)/\sigma_0^2(\sigma_0^2 + \sigma_1(n)^2)$ is strictly positive when $\sigma_1(n)^2 < \sigma_0^2$, and in this case

$$C(n) = -1 + \frac{\sigma_0^2}{\sqrt{\sigma_0^4 - \sigma_1(n)^4}} e^{\frac{\mu_1(n)^2}{\sigma_0^2 - \sigma_1(n)^2}}. \quad (6)$$

By Theorem 3.1 it is easily shown from (6) that the attacker is covert if $\limsup_n \sigma_1(n)^2 < \sigma_0^2$, $\mu_1(n) = \mathcal{O}(1)$, $q(n)\mu_1(n) = \mathcal{O}(1/\sqrt{n})$, and $q(n)\sigma_1(n)^2 = \mathcal{O}(1/\sqrt{n})$.

Another illustration is $N = 2$. Choose (with $\Sigma_1 := \Sigma_1(n)$)

$$\Sigma_i = \begin{pmatrix} a_i^2 & \rho_i a_i b_i \\ \rho_i a_i b_i & b_i^2 \end{pmatrix}, \quad i = 0, 1,$$

with a_0 and b_0 (resp. a_1 and b_1) the standard deviations of $X_{1,1}$ and $X_{1,2}$ (resp. $Y_{1,1}$ and $Y_{1,2}$) respectively, and ρ_0 (resp. ρ_1) the Pearson correlation coefficient of $X_{1,1}, X_{1,2}$ (resp. $Y_{1,1}, Y_{1,2}$). Matrix Σ_i is positive definite iff

$$-1 < \rho_i < 1. \quad (7)$$

The eigenvalues of the (symmetric) matrix $\Sigma_0\Sigma_1^{-1}$ are

$$\frac{-2a_0b_0a_1b_1\rho_0\rho_1 + a_0^2b_1^2 + b_0^2a_1^2 \pm \sqrt{d}}{2a_1^2b_1^2(1 - \rho_1^2)}, \quad (8)$$

with

$$d = -4a_0a_1b_0b_1\rho_0\rho_1(a_0^2b_1^2 + a_1^2b_0^2) + 4a_0^2a_1^2b_0^2b_1^2(\rho_0^2 + \rho_1^2) + (a_0^2b_1^2 - a_1^2b_0^2)^2.$$

Note that $d \geq 0$ since the eigenvalues of a symmetric matrix are all real. We deduce from (8) that both eigenvalues of $\Sigma_0\Sigma_1^{-1}$ are strictly larger than one iff

$$-a_0b_0a_1b_1\rho_0\rho_1 + \frac{1}{2}(a_0^2b_1^2 + b_0^2a_1^2) - \frac{1}{2}\sqrt{d} > a_1^2b_1^2(1 - \rho_1^2). \quad (9)$$

When (7) holds for $i = 0, 1$ and (9) is met Theorem 3.1 implies that the attacker is covert when $q(n) = \mathcal{O}(1/\sqrt{n})$.

4. CONVERSE

In this section we relax the assumption that $\underline{X}_r$ and $\underline{Y}_r$ have multivariate Gaussian distributions. Recall that admin observes $\{\underline{z}_r\}_{r=1}^n$, $\underline{z}_r = (z_{r,1}, \ldots, z_{r,N})$, with $\underline{z}_r$ a realization of $\underline{Z}_r$, which can be reorganized as $\{z_{r,1}\}_{r=1}^n, \ldots, \{z_{r,N}\}_{r=1}^n$, where $\{z_{r,j}\}_{r=1}^n$ contains information about feature j. Admin can therefore detect an attack by investigating separately sequences $\{z_{r,1}\}_{r=1}^n, \ldots, \{z_{r,N}\}_{r=1}^n$. From this observation, the following converse result can be shown,

THEOREM 4.1 (CONVERSE).
If for at least one j ($j = 1, \ldots, N$)

$$0 < \inf_{n\geq 1} \text{var}(Y_{n,j}) \leq \sup_{n\geq 1} \text{var}(X_{n,j}) < \infty, \tag{10}$$

$$\sup_{n\geq 1} \mathbb{E}[|X_{n,j} - \mathbb{E}[X_{n,j}]|^3] < \infty, \tag{11}$$

$$\lim_n \sqrt{n} q(n) \mathbb{E}[Y_{n,j}] = +\infty, \tag{12}$$

$$q(n)(\text{var}(Y_{n,j}) + (1 - q(n))\mathbb{E}[Y_{n,j}])) = \mathcal{O}(1), \tag{13}$$

then the attacker is not covert.

The proof uses a detector of the form $\frac{1}{n}\sum_{r=1}^n z_{r,j} \lessgtr \tau_j$ for feature j and the Berry-Esseen theorem. The detector built in the proof of Theorem 4.1 does not use the attack distribution. However, Theorem 4.1 still holds if admin knows it as this knowledge can only increase the effectiveness of the detector.

When $\underline{X}_r$ and $\underline{Y}_r$ have multivariate Gaussian distributions given in (2)-(3), $\mathbb{E}[X_{n,j}] = \mu_{0,j}$, $\mathbb{E}[Y_{n,j}] = \mu_{1,j}(n)$, $\text{var}(X_{n,j}) = \sigma_{0,j}^2$, and $\text{var}(Y_{n,j}) = \sigma_{1,j}^2$, with $\sigma_{0,j}^2$ (resp. $\sigma_{1,j}(n)^2$) the jth diagonal element of the covariance matrix Σ_0 (resp. $\Sigma_1(n)$). In this case (10) becomes $0 < \inf_{n\geq 1} \sigma_{1,j}(n)^2 < \sigma_{0,j}^2$ and (11) is automatically satisfied.

5. EVALUATION

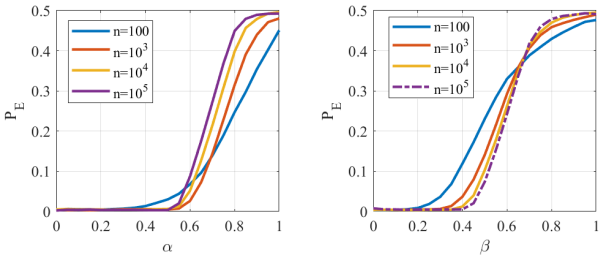
Next, we report numerical results relying on real network traffic. Our goal is to show that the phase transition on the error probability predicted by the model still holds beyond the assumptions considered in the model. In particular, the model considers features sampled from a multivariate Gaussian, whereas we consider real network traffic.

Evaluation setup. We base our results on real data of regular traffic collected from network interfaces of more than 2000 home-routers between June 10th 2018 and August 18th 2018. These routers gather information about network usage. For the purpose of this work, we use packet and byte counts of upstream and downstream traffic. We observe that the joint probability distribution of data (upload and download, byte and packet counts) can be characterized by a mixture of five multivariate Gaussian distributions. Then, we ran a controlled experiment in the lab to estimate the distribution of traffic generated by a typical DDoS attack [5]. The obtained statistics were used as a baseline to generate the synthetic data of attack traffic from a multivariate Gaussian mixture model with four components, where

$$\underline{\mu}_{1,i}(n) = \delta c_i n^{-\alpha} \text{ and } \Sigma_{1,i}(n)^2 = \delta \mathbf{C}_i n^{-\alpha}, \tag{14}$$

with $i \in \{1, 2, 3, 4\}$. When considering an attacker that issues an attack from a fraction of the homes, we let

$$q(n) = c_q n^{-\beta} \text{ and } \alpha = 0. \tag{15}$$



(a) all homes (b) fraction of homes

Figure 1: Phase transition under real traffic.

In the experiments, the number of observations is assumed to be n, i.e., the number of compromised homes.

Phase transitions under real traffic. Figure 1 reports the error probability as a function of the aggressiveness of the attacker. We start from our reference setup, with $p_{FA} = 0.01$ and $\delta = 0.1$, considering an attacker issuing attack from all homes (Figure 1(a)). The average total traffic injected by the attacker is proportional to $n^{1-\alpha}$: as α grows the total attack traffic decreases and the probability of error transitions from 0 to 1. For values of n as small as 1,000, we already observe a sharp phase transition occurring around $\alpha = 0.5$, in agreement with the square root law. Then, we consider a variation obtained after considering attacks from a subset of homes and letting $\delta = 1$, where the fraction of homes $q(n)$ is given by $q(n) = n^{-\beta}$ (Figure 1(b)). As β grows the total attack traffic decreases and the probability of error transitions from 0 to 1, with a phase transition at $\beta = 0.5$, agreeing with the square root law.

To produce Figure 1 we consider a simple binary classifier for attack detection that implements a likelihood ratio test [1, Chapter 9], wherein the likelihood ratio is compared against a threshold τ to determine the class of a given set of traffic samples. Both threshold and probability of error are computed using Monte Carlo methods, where given a target p_{FA}, we obtain the threshold τ for the hypothesis test. Then, we assess the probability of error $p_{FA} + p_{MD}$.

6. CONCLUSION

In this work we extend the square-root scaling laws of DDoS attacks in the realm of covertness [2] to the multi-feature setup. Our results pave the way towards a method for extending and validating scaling laws associated to covert attacks in realistic settings.

7. REFERENCES

[1] D. P. Bertsekas and J. N. Tsitsiklis. *Introduction to Probability*. Athena Scientific, 2008.

[2] B. Jiang, P. Nain, and D. Towsley. Covert cycle stealing in a single FIFO server. *To appear in ACM ToMPECS. Preprint at arXiv:2003.05135*, 2021.

[3] S. Kullback. *Information Theory and Statistics*. Dover Publications, 1968.

[4] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer Science & Business Media, 2006.

[5] G. Mendonça, G. H. A. Santos, E. de Souza e Silva, R. M. M. Leao, D. S. Menasche, and D. Towsley. An Extremely Lightweight Approach for DDoS Detection at Home Gateways. In *2019 IEEE International Conference on Big Data*, pages 5012–5021, 2019.