Botnet Identification in Randomized DDoS Attacks

Vincenzo Matta, Mario Di Mauro, Maurizio Longo
Department of Information & Electrical Engineering and Applied Mathematics
University of Salerno, Fisciano (SA), Italy
{vmatta, mdimauro, longo}@unisa.it

***Abstract*—Recent variants of Distributed Denial-of-Service (DDoS) attacks leverage the flexibility of application-layer protocols to disguise malicious activities as normal traffic patterns, while concurrently overwhelming the target destination with a large request rate. New countermeasures are necessary, aimed at guaranteeing an early and reliable identification of the compromised network nodes (the *botnet*). In this work we introduce a formal model for the aforementioned class of attacks, and we devise an inference algorithm that estimates the botnet hidden in the network, converging to the *true* solution as time progresses. Notably, the analysis is validated over *real* network traces.**

***Index Terms*—Distributed Denial-of-Service, DDoS, Cyber-Security, Signal Processing for Network Security.**

I. Motivation and Related Work

Communication networks, especially the Internet, emerge more and more as the favorite attackers' land to launch a broad variety of threats. One of the most dangerous attacks is Denial-of-Service (DoS), a kind of *volumetric* attack where the target destination is overwhelmed by a huge number of requests, which eventually lead to the impossibility of serving any of the users. In its most powerful variant, the *Distributed* DoS (DDoS), such requests are produced in parallel by a *botnet*, a large net of ro*bots* acting cooperatively under the supervision of a *botmaster*. The bots may be either malicious users acting consciously, or legitimate users that have been preliminarily infected, (e.g., by warms and/or Trojans).

The anomalous request rate is produced in broad daylight, and, therefore, its detection is not a big concern. The main challenge is instead ascertaining whether the anomaly is caused by an attack and identifying the compromised nodes. Successful DDoS mitigation relies upon an early identification of the botnet, since discriminating legitimate from malicious users would allow the destination to ban the latter, without denying the service to the former.

The literature about DoS attacks is abundant, and we refer the Reader to the survey in [1] as a useful entry-point. The earliest DoS attacks (see, e.g., TCP SYN flooding) were based on specific protocol vulnerabilities, and were characterized by a high-rate, repeated transmission of the same requests from a single user. In this case, the source of the attack could be simply identified by its unusually large rate. In contrast, in a DDoS attack, the huge rate is produced by the botnet as a whole, while the rate of each bot is kept moderate. This notwithstanding, the bots can be still identified at a single-user level, because normal traffic patterns are typically characterized by a certain degree of innovation (for instance, as time progresses, distinct web-pages are likely to be visited), while the repetition scheme implicitly emphasizes the bot character. In fact, several useful inferential strategies have been proposed for such kind of DDoS attacks, see [1] for an excellent summary.

Recently, a novel class of powerful DDoS attacks is emerging, which leverage the many possibilities offered by the application layer, to circumvent the aforementioned repeatability issue [2]–[5]. In such a novel attacks, the bots choose randomly their requests from a set of admissible messages (an *emulation dictionary*), trying so to disguise their traffic patterns as normal ones. With a sufficient variety of messages (e.g., the large number of web-pages accessible in surfing through a certain website), a sufficient degree of variability is assigned to each individual bot's pattern, which prevents from revealing a bot by simple single-user inspection. In this work we first introduce a formal model to represent the aforementioned new class of DDoS attacks, and then try to answer the following fundamental question: *Despite the strong power given to the attacker, is it still possible to consistently unveil the presence of a botnet?*

Unfortunately, the existing inferential strategies are not conceived to face the novel class of DDoS attacks [1]. Some of these strategies might be in principle open to generalization, but, as far as we can tell, ready-to-use solutions to our problem are currently unavailable. Therefore, new inferential solutions are required. To this aim, we shall follow emerging trends in signal processing for network cyber-security, to design universal and/or nonparametric inference strategies — see, e.g., sparsity-aware algorithms for unveiling traffic volume anomalies [6], [7], or solutions to trace clandestine information flows across the network [8]–[12]. Such approaches rely neither on parametric statistical methods (e.g., maximum likelihood, Neyman-Pearson tests), nor on fully data-driven techniques (e.g., distribution-free statistical learning, machine learning), since: the former would require detailed statistical models of the attacks [13]–[15], a condition that is far from being verified in our setting; while the latter would typically lack of performance guarantees, analytical results, physical interpretation, and might require heavy tuning of the algorithms when the parameters change. The methodologies presented in [6]–[12] suggest instead to pursue the following principled approach: $i)$ focus on minimal-and-realistic physical assumptions; $ii)$ build physically-meaningful descriptive indicators arising from the modeling assumptions; $iii)$ develop consequently an inference strategy.

II. The DDoS Attack

Throughout the article we shall imply that: $i)$ the network analyst has access to the message content of the collected traffic patterns; and $ii)$ the mere content of a message does not reveal any information about the nature, legitimate or malicious, of the sender. Moreover, no statistical models are available for the legitimate users' activities.

As an indicator of the *transmission* activity of a given subnet $\mathcal{S}$, we introduce the *empirical* transmission rate at time t, namely,

$$\boxed{\hat{\lambda}_{\mathcal{S}}(t) \triangleq \frac{N_{\mathcal{S}}(t)}{t}} \tag{1}$$

where $N_{\mathcal{S}}(t)$ is the total number of transmissions occurred in $\mathcal{S}$, up to a given time t. Whenever a limiting rate (as t goes to infinity) is meaningfully defined, it will be denoted by $\lambda_{\mathcal{S}}$.

As a second descriptive indicator of the network activity, we define a quantity that relates to the message *content*. We are interested in the *new* messages that are incrementally produced by the users

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

during their activities, namely, in a Message Innovation Rate (MIR). Let $\mathscr{D}_{\mathcal{S}}(t)$ denote the empirical dictionary composed by the *distinct* messages sent, up to time t, by users within $\mathcal{S}$. The *empirical* Message Innovation Rate (MIR) is:

$$\boxed{\hat{\rho}_{\mathcal{S}}(t) \triangleq \frac{|\mathscr{D}_{\mathcal{S}}(t)|}{t}} \tag{2}$$

In particular, when $\hat{\rho}_{\mathcal{S}}(t) \xrightarrow{\text{p}} \rho_{\mathcal{S}}$ (the symbol $\xrightarrow{\text{p}}$ denotes convergence in probability [16]), the limiting value $\rho_{\mathcal{S}}$ will be simply referred to as the MIR of subnet $\mathcal{S}$. We stress that the quantities $\hat{\lambda}_{\mathcal{S}}(t)$ and $\hat{\rho}_{\mathcal{S}}(t)$ refer to a generic subnet $\mathcal{S}$, irrespectively of the nature of the users belonging to $\mathcal{S}$.

Starting from the most recent kind of attacks documented in the literature [2]–[5], we now introduce a formal DDoS model. The botnet has at its disposal a sufficiently rich dictionary wherein it gleans admissible messages to emulate normal patterns. Such emulation dictionary is learned *continually*, namely, its cardinality increases as time progresses, in order to ensure that the bots can sustain a reasonable innovation rate to emulate normal users. The (common) dictionary available at time t to all users in the botnet will be denoted by $\mathscr{E}(t)$. The richness of the emulation dictionary is quantified through the cardinality of the dictionary per unit time, which provides the Emulation Dictionary Rate (EDR):

$$\boxed{\alpha \triangleq \lim_{t\to\infty} \frac{|\mathscr{E}(t)|}{t}} \tag{3}$$

Accordingly, the EDR rules the variability in the emulated traffic patterns (in a precise quantitative way that will be revealed by the forthcoming theorem). We see that our formal model includes as a special case the simplest attack using always the same repeated pattern ($\alpha = 0$). Following [2]–[5], we focus on a *randomized* strategy where, at each time instant t, each botnet member picks, uniformly at random, a message from the available emulation dictionary $\mathscr{E}(t)$. The probability of a particular message is accordingly $1/|\mathscr{E}(t)|$. Thus, for any given subnet $\mathcal{B}$ of the overall botnet, a certain *empirical* dictionary $\mathscr{D}_{\mathcal{B}}(t)$ is constructed at time t. Given the empirical dictionary $\mathscr{D}_{\mathcal{B}}(t)$, at time $t+\tau$ the number of *distinct* messages increases by the number of *distinct* messages not contained in $\mathscr{D}_{\mathcal{B}}(t)$, which have been picked during the interval τ by the bots belonging to $\mathcal{B}$. Now, assume that all bots act synchronously each $1/\lambda$ seconds, and let $d_n = \mathbb{E}[|\mathscr{D}_{\mathcal{B}}(n/\lambda)|]$, and $e_n = |\mathscr{E}(n/\lambda)|$. Neglecting the possibility that two or more bots pick the same message during the same time interval, we have:

$$d_n = d_{n-1} + B\left(1 - \frac{d_{n-1}}{e_n}\right), \tag{4}$$

where $B = |\mathcal{B}|$. Assuming further that a limiting MIR exists, and that we can use the approximation $d_{n-1}/e_n \approx \rho_{\mathcal{B}}/\alpha$, we have, iterating over n and setting $d_0 = 0$:

$$d_n \approx d_{n-1} + B\left(1 - \frac{\rho_{\mathcal{B}}}{\alpha}\right) \Rightarrow d_n \approx nB\left(1 - \frac{\rho_{\mathcal{B}}}{\alpha}\right), \tag{5}$$

which, considering that in the synchronous case the aggregate rate of $\mathcal{B}$ is simply $\lambda_{\mathcal{B}} = B\lambda$, yields:

$$\frac{d_n}{n/\lambda} \overset{n\to\infty}{\longrightarrow} \rho_{\mathcal{B}} = B\lambda\left(1 - \frac{\rho_{\mathcal{B}}}{\alpha}\right) \Rightarrow \rho_{\mathcal{B}} = \frac{\alpha\,\lambda_{\mathcal{B}}}{\alpha + \lambda_{\mathcal{B}}}. \tag{6}$$

The latter intuitive explanation can be in fact made rigorous. More precisely, the forthcoming theorem provides a closed-form expression for the MIR of a botnet, when the transmission schedulings are either synchronous with constant rate, or independent Poisson processes [17]. The corresponding mathematical proof is substantially more involved. Due to space limitations, the necessary technical details (as well as those relevant to Theorem 2) will be reported elsewhere [18], and are available upon request. Let us preliminarily introduce the function $\mathscr{R}(\alpha, \lambda) \triangleq \frac{\alpha\,\lambda}{\alpha+\lambda}$.

Theorem 1 (Botnet MIR). *Consider a botnet $\mathcal{B}_{\text{tot}}$ launching a DDoS attack, where the node transmission policies are either synchronous with constant transmission rate, or independent Poisson processes, with rates λ_u, for $u \in \mathcal{B}_{\text{tot}}$. Consider a subset of the botnet $\mathcal{B} \subset \mathcal{B}_{\text{tot}}$. Let $\mathscr{E}(t)$ be the emulation dictionary available to the botnet, with emulation dictionary rate α, and let $\mathscr{D}_{\mathcal{B}}(t)$ be the empirical dictionary of the subnet $\mathcal{B}$ at time t. Then, the message innovation rate of $\mathcal{B}$ is:*

$$\boxed{\frac{|\mathscr{D}_{\mathcal{B}}(t)|}{t} \xrightarrow{\text{p}} \rho_{\mathcal{B}} = \mathscr{R}(\alpha, \lambda_{\mathcal{B}})} \tag{7}$$

where $\lambda_{\mathcal{B}} = \sum_{u\in\mathcal{B}} \lambda_u$ is the aggregate transmission rate of the considered botnet subset. ■

Exploiting the definition of $\mathscr{R}(\alpha, \lambda)$ and (7), an empirical estimator of α based on the messages in the subnet $\mathcal{S}$ is:

$$\boxed{\hat{\alpha}_{\mathcal{S}}(t) \triangleq \frac{\hat{\lambda}_{\mathcal{S}}(t)\,\hat{\rho}_{\mathcal{S}}(t)}{\hat{\lambda}_{\mathcal{S}}(t) - \hat{\rho}_{\mathcal{S}}(t)}} \tag{8}$$

and, hence, the empirical MIR $\hat{\rho}_{\mathcal{S}}(t)$ can be conveniently expressed as:

$$\boxed{\hat{\rho}_{\mathcal{S}}(t) = \mathscr{R}(\hat{\alpha}_{\mathcal{S}}(t), \hat{\lambda}_{\mathcal{S}}(t))} \tag{9}$$

Note that the latter expression holds irrespectively of the nature of the users in $\mathcal{S}$, even if the interpretation of $\hat{\alpha}_{\mathcal{S}}(t)$ in terms of emulation dictionary does not necessarily hold for normal users.

III. Botnet Identification Condition

The common *emulation* dictionary used by the botnet to launch the attack implies a certain degree of correlation between the *empirical* dictionaries of the bots. In contrast, the empirical dictionaries of two normal users (or of a normal user and a bot) are expected to be weakly correlated, due to the independence of their activities, some partial overlap arising due to common interests, popular web-pages, peculiar website structure, and so on. In our setting, a convenient way to measure the degree of dependence is provided by the empirical message innovation rate in (2). Then, in order to develop a botnet identification algorithm, we can use as reference case (i.e., as *identification threshold*) for a malicious behavior, the MIR corresponding to the activity performed by a botnet. Let us start by considering the simplest case that we must decide whether users 1 and 2 belong to a botnet. Assume for now that the empirical EDRs of the two users obtained through (8) are comparable, namely, that $\hat{\alpha}_1 \approx \hat{\alpha}_2 \approx \hat{\alpha}$, the explicit dependence on t having been suppressed for ease of notation. When both users belong to a botnet, in view of Theorem 1, for t large enough we can write $\hat{\rho}_{\{1,2\}} \approx \mathscr{R}(\hat{\alpha}, \hat{\lambda}_1 + \hat{\lambda}_2) \triangleq \hat{\rho}_{\text{bot}}$. On the other hand, *irrespectively of the nature of the users*, the empirical MIR of the aggregate subnet $\{1,2\}$ is certainly upper bounded by the sum of the individual MIRs, namely, $\hat{\rho}_{\{1,2\}} \le \mathscr{R}(\hat{\alpha}_1, \hat{\lambda}_1) + \mathscr{R}(\hat{\alpha}_2, \hat{\lambda}_2) \approx \mathscr{R}(\hat{\alpha}, \hat{\lambda}_1) + \mathscr{R}(\hat{\alpha}, \hat{\lambda}_2) \triangleq \hat{\rho}_{\text{sum}}$. Therefore, since $\hat{\rho}_{\text{bot}} < \hat{\rho}_{\text{sum}}$, it makes sense to introduce a threshold γ lying between the two points $\hat{\rho}_{\text{bot}}$ and $\hat{\rho}_{\text{sum}}$. Formally, for $\epsilon \in (0,1)$, we set $\gamma = \hat{\rho}_{\text{bot}} + \epsilon(\hat{\rho}_{\text{sum}} - \hat{\rho}_{\text{bot}})$. If the two users belong to a botnet, from Theorem 1 we conclude that the empirical MIR $\hat{\rho}_{\{1,2\}}$ shrinks down to the value $\hat{\rho}_{\text{bot}}$ for sufficiently large t: for *any* $\epsilon > 0$, as time progresses, the empirical MIR will stay sooner (higher ϵ) or later (lower ϵ) *below* the threshold γ, namely, 1 AND 2 are bots $\Rightarrow \hat{\rho}_{\{1,2\}} < \gamma$. Consider now the case that at least one user is normal. Now, were the dictionaries of the two users perfectly disjoint, we should observe that $\hat{\rho}_{\{1,2\}} \approx \hat{\rho}_{\text{sum}} > \gamma$. However, in the (realistic) case that some overlap between the two dictionaries exists, it is also natural to assume that such a dependence is weaker than the dependence pertaining to groups of bots (since the latter choose their messages from one and the same emulation dictionary). Accordingly, we might expect that, when at least one user is normal, for sufficiently small ϵ, the empirical MIR still stays above the threshold, namely, 1 OR 2 are normal $\Rightarrow \hat{\rho}_{\{1,2\}} > \gamma$.

Unfortunately, there is an important complication that has been deliberately neglected so far. According to the above explanation, we need to compare the empirical MIR to the MIR of a *reference* botnet. However, a botnet is characterized by a *common* underlying EDR, corresponding to a unique value of α, while in practice we shall have, especially when at least one user is normal, $\hat{\alpha}_1 \neq \hat{\alpha}_2$. One approach could be that of discarding *ab initio* the botnet hypothesis whenever $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are too dissimilar. However, the qualification of being "too dissimilar" translates into the appearance of some extra tuning parameter, possibly depending on time, which we want definitely to avoid. On the other hand, the naïvest choice of considering as reference value the arithmetic average $1/2(\hat{\alpha}_1 + \hat{\alpha}_2)$ does not work, since it can be shown that, *even for the case of disjoint dictionaries*, the empirical MIR can be *smaller* than that of a botnet with reference EDR given by the arithmetic average.

In order to overcome such difficulties, we now illustrate a systematic way to select a reference value for $\hat{\alpha}$. Let us consider two disjoint subnets $\mathcal{S}_1$ and $\mathcal{S}_2$, with focus on the case that at least one of them is composed only by normal users, with $\hat{\alpha}_{\mathcal{S}_1} \neq \hat{\alpha}_{\mathcal{S}_2}$. Starting from the original traffic patterns, we want to build (fictitiously) new traffic patterns by replacing and reassigning the messages between the subnets $\mathcal{S}_1$ and $\mathcal{S}_2$, in such a way that *i)* the joint MIR of the aggregate subnet $\mathcal{S}_1 \cup \mathcal{S}_2$ is left unchanged and *ii)* the empirical EDRs after replacement and reassignment are equal, namely, $\hat{\alpha}'_{\mathcal{S}_1} = \hat{\alpha}'_{\mathcal{S}_2} = \hat{\alpha}'$. The aforementioned procedure is defined by the following steps.

1. Replacement of repeated messages. The traffic pattern of a subnet $\mathcal{S}$ contains $|\mathscr{D}_{\mathcal{S}}|$ distinct messages, the remaining $N_{\mathcal{S}} - |\mathscr{D}_{\mathcal{S}}|$ ones being repetitions of messages contained in $\mathscr{D}_{\mathcal{S}}$. The first step of the procedure amounts to replacing the $N_{\mathcal{S}} - |\mathscr{D}_{\mathcal{S}}|$ messages by one and the same message contained in $\mathscr{D}_{\mathcal{S}}$. The replacement is applied to both subnets $\mathcal{S}_1$ and $\mathcal{S}_2$. Obviously, replacement leaves unaltered the transmission rates as well as the various MIRs.

2. Reassignment of messages. In order to reach a common $\hat{\alpha}$, some messages from one subnet will be reassigned to the other subnet. We denote by Δ the rate of *distinct* messages of $\mathcal{S}_2$ that are reassigned to $\mathcal{S}_1$, with the convention that a negative Δ corresponds to messages of $\mathcal{S}_1$ that are reassigned to $\mathcal{S}_2$. Obviously, the admissible values for Δ must obey the inequalities: $\Delta \leq \hat{\rho}_{\mathcal{S}_2}$ and $-\Delta \leq \hat{\rho}_{\mathcal{S}_1}$. For instance, if messages from $\mathcal{S}_2$ are reassigned to $\mathcal{S}_1$, the rate of reassigned messages cannot exceed the rate of distinct messages owned by $\mathcal{S}_2$. Note also that the reassignment leaves unaltered the overall MIR, as well as the overall number of transmissions in the *aggregate* network $\mathcal{S}_1 \cup \mathcal{S}_2$. Now, since for sufficiently large t the contribution of a *single* message is irrelevant, we can safely assume that the reassignment does not include the single messages that form the repeated-messages sets. Furthermore, we can assume that it is always possible to reassign messages that do not belong to the intersection of the two empirical dictionaries, since the degree of correlation between the two traffic patterns is low (one of the subnets being normal). In summary, after reassignment, the individual transmission rates and MIRs of subnets $\mathcal{S}_1$ and $\mathcal{S}_2$, are, respectively, $(\hat{\lambda}'_{\mathcal{S}_1}, \hat{\lambda}'_{\mathcal{S}_2}) = (\hat{\lambda}_{\mathcal{S}_1} + \Delta, \hat{\lambda}_{\mathcal{S}_2} - \Delta)$, and $(\hat{\rho}'_{\mathcal{S}_1}, \hat{\rho}'_{\mathcal{S}_2}) = (\hat{\rho}_{\mathcal{S}_1} + \Delta, \hat{\rho}_{\mathcal{S}_2} - \Delta)$.

3. Choice of Δ for the equilibrium condition. In order to get a common reference EDR $\hat{\alpha}'$, we enforce the condition $\hat{\alpha}'_{\mathcal{S}_1} = \hat{\alpha}'_{\mathcal{S}_2} = \hat{\alpha}'$. Using (8), such condition amounts to $\hat{\alpha}' = \hat{\lambda}'_{\mathcal{S}_1}\hat{\rho}'_{\mathcal{S}_1}/(\hat{\lambda}'_{\mathcal{S}_1} - \hat{\rho}'_{\mathcal{S}_1}) = \hat{\lambda}'_{\mathcal{S}_2}\hat{\rho}'_{\mathcal{S}_2}/(\hat{\lambda}'_{\mathcal{S}_2} - \hat{\rho}'_{\mathcal{S}_2})$. In the light of the above explanation, this is equivalent to seek a value $\Delta^\star$ such that:

$$\hat{\alpha}' = \frac{(\hat{\lambda}_{\mathcal{S}_1} + \Delta^\star)(\hat{\rho}_{\mathcal{S}_1} + \Delta^\star)}{\hat{\lambda}_{\mathcal{S}_1} - \hat{\rho}_{\mathcal{S}_1}} = \frac{(\hat{\lambda}_{\mathcal{S}_2} - \Delta^\star)(\hat{\rho}_{\mathcal{S}_2} - \Delta^\star)}{\hat{\lambda}_{\mathcal{S}_2} - \hat{\rho}_{\mathcal{S}_2}}. \tag{10}$$

The explicit formula for $\Delta^\star$ is then found by solving a quadratic equation, and it is possible to show that the solution fulfilling the admissibility conditions $\Delta \leq \hat{\rho}_{\mathcal{S}_2}$ and $-\Delta \leq \hat{\rho}_{\mathcal{S}_1}$ is:

$$\begin{aligned}\Delta^\star &= \frac{\hat{\lambda}_{\mathcal{S}_1}\hat{\lambda}_{\mathcal{S}_2} - \hat{\rho}_{\mathcal{S}_1}\hat{\rho}_{\mathcal{S}_2}}{(\hat{\lambda}_{\mathcal{S}_1} - \hat{\rho}_{\mathcal{S}_1}) - (\hat{\lambda}_{\mathcal{S}_2} - \hat{\rho}_{\mathcal{S}_2})} \\ &- \frac{\sqrt{(\hat{\lambda}_{\mathcal{S}_1} - \hat{\rho}_{\mathcal{S}_1})(\hat{\lambda}_{\mathcal{S}_2} - \hat{\rho}_{\mathcal{S}_2})(\hat{\lambda}_{\mathcal{S}_1} + \hat{\rho}_{\mathcal{S}_2})(\hat{\lambda}_{\mathcal{S}_2} + \hat{\rho}_{\mathcal{S}_1})}}{(\hat{\lambda}_{\mathcal{S}_1} - \hat{\rho}_{\mathcal{S}_1}) - (\hat{\lambda}_{\mathcal{S}_2} - \hat{\rho}_{\mathcal{S}_2})}.\end{aligned} \tag{11}$$

Now, note that:

$$\begin{aligned}\hat{\rho}_{\text{sum}}(\mathcal{S}_1, \mathcal{S}_2) &\triangleq \mathscr{R}(\hat{\alpha}_{\mathcal{S}_1}, \hat{\lambda}_{\mathcal{S}_1}) + \mathscr{R}(\hat{\alpha}_{\mathcal{S}_2}, \hat{\lambda}_{\mathcal{S}_2}) \\ &= [\mathscr{R}(\hat{\alpha}_{\mathcal{S}_1}, \hat{\lambda}_{\mathcal{S}_1}) + \Delta^\star] + [\mathscr{R}(\hat{\alpha}_{\mathcal{S}_2}, \hat{\lambda}_{\mathcal{S}_2}) - \Delta^\star] \\ &\overset{\text{(a)}}{=} \mathscr{R}(\hat{\alpha}', \hat{\lambda}'_{\mathcal{S}_1}) + \mathscr{R}(\hat{\alpha}', \hat{\lambda}'_{\mathcal{S}_2}) \\ &\overset{\text{(b)}}{>} \mathscr{R}(\hat{\alpha}', \hat{\lambda}'_{\mathcal{S}_1} + \hat{\lambda}'_{\mathcal{S}_2}) \\ &\overset{\text{(c)}}{=} \mathscr{R}(\hat{\alpha}', \hat{\lambda}_{\mathcal{S}_1} + \hat{\lambda}_{\mathcal{S}_2}) \triangleq \hat{\rho}_{\text{bot}}(\mathcal{S}_1, \mathcal{S}_2),\end{aligned} \tag{12}$$

where (a) follows by conservation of the sum of MIRs; (b) follows by Theorem 1, since, given a botnet $\mathcal{S}_1 \cup \mathcal{S}_2$, for generic values $\alpha, \lambda_1, \lambda_2 \in \mathbb{R}^+$, the function $\mathscr{R}(\alpha, \lambda_1) + \mathscr{R}(\alpha, \lambda_2)$ can be regarded as the sum of the individual MIRs, while the function $\mathscr{R}(\alpha, \lambda_1 + \lambda_2)$ can be regarded as the MIR of the whole botnet; and (c) follows by conservation of the sum of transmission rates.

Let us switch now to the case that $\mathcal{S}_1$ and $\mathcal{S}_2$ form a botnet. Theorem 1 implies that, for t large enough, $\hat{\alpha}_{\mathcal{S}_1} \approx \hat{\alpha}_{\mathcal{S}_2} \approx \hat{\alpha}' \approx \alpha$. Therefore, in this case the inequality $\hat{\rho}_{\text{sum}}(\mathcal{S}_1, \mathcal{S}_2) > \hat{\rho}_{\text{bot}}(\mathcal{S}_1, \mathcal{S}_2)$ is justified by the approximations: $\hat{\rho}_{\text{sum}}(\mathcal{S}_1, \mathcal{S}_2) \approx \mathscr{R}(\alpha, \lambda_{\mathcal{S}_1}) + \mathscr{R}(\alpha, \lambda_{\mathcal{S}_2})$ and $\hat{\rho}_{\text{bot}}(\mathcal{S}_1, \mathcal{S}_2) \approx \mathscr{R}(\alpha, \lambda_{\mathcal{S}_1} + \lambda_{\mathcal{S}_2})$.

In summary, we have shown that, for *arbitrary transmission schedulings*, as well as for *arbitrary message-picking policies*, the reference EDR value (10) arising from the proposed strategy does *always* provide a lower bound to the sum of individual MIRs (i.e., the perfectly disjoint case). Therefore, it makes sense to introduce an intermediate threshold $\gamma(\mathcal{S}_1, \mathcal{S}_2) = \hat{\rho}_{\text{bot}}(\mathcal{S}_1, \mathcal{S}_2) + \epsilon\,[\hat{\rho}_{\text{sum}}(\mathcal{S}_1, \mathcal{S}_2) - \hat{\rho}_{\text{bot}}(\mathcal{S}_1, \mathcal{S}_2)]$.

We are now ready to characterize the interaction between the empirical MIR and the threshold. When $\mathcal{S}_1$ and $\mathcal{S}_2$ form a botnet, we have, at least for sufficiently large t:

$$\boxed{\hat{\rho}_{\mathcal{S}_1 \cup \mathcal{S}_2} < \gamma(\mathcal{S}_1, \mathcal{S}_1)} \tag{13}$$

When at least one subnet is normal, it is realistic to assume that the degree of dependence is consistently lower than the degree of dependence occurring when both subnets form a botnet. Since the replacement-and-reassignment procedure does preserve the empirical MIR of $\mathcal{S}_1 \cup \mathcal{S}_2$, it makes sense referring to the *new* traffic patterns, and using for the reference botnet an EDR equal to $\hat{\alpha}'$. Otherwise stated, it is reasonable to assume that the empirical MIR, even if not coinciding with the upper bound $\hat{\rho}_{\text{sum}}(\mathcal{S}_1, \mathcal{S}_2)$, is still sufficiently far from the lower bound $\hat{\rho}_{\text{bot}}(\mathcal{S}_1, \mathcal{S}_2)$. The latter observation corresponds, at least for small ϵ, to the following identification condition.

Botnet Identification Condition

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two subnets with $\mathcal{S}_1 \bigcap \mathcal{S}_2 = \emptyset$. *If at least one of the subnets is composed only by normal users*:

$$\boxed{\hat{\rho}_{\mathcal{S}_1 \cup \mathcal{S}_2} \geq \gamma(\mathcal{S}_1, \mathcal{S}_2)} \tag{14}$$

■

The latter condition characterizes the behavior of the MIR when at least one subnet is made of normal users.

IV. BotBuster: The Botnet Identification Test

Given two disjoint subnets, Eqs. (14) and (13) would allow discriminating the situation where both subnets are part of a botnet, from the situation where at least one of them is made of normal users. Such a basic property can be exploited to design the identification algorithm described by the pseudo-code reported in the right column above. Let us examine quickly how the algorithm works. At the beginning of the algorithm, user 1 is initially declared as a bot, namely, $\hat{\mathcal{B}} = \{1\}$. Then, it is checked whether users 1 and 2 form a botnet. If so, $\hat{\mathcal{B}} = \{1, 2\}$, otherwise $\hat{\mathcal{B}} = \{1\}$. Then, it is checked whether the currently estimated botnet $\hat{\mathcal{B}}$ forms a bot with user 3, and so on. At the end of the inner loop, the algorithm ends up with an estimate $\hat{\mathcal{B}}$. If the cardinality of the estimated set is greater than one, it is taken as a current estimate. The procedure restarts by choosing user 2 as initial pivot, and sequentially checking the remaining users as explained before. At the end of the inner loop, the algorithm ends up with another estimate $\hat{\mathcal{B}}$. If the cardinality of the estimated set is greater than one *and* greater than the cardinality of the previously estimated set, then it is taken as a current estimate. Otherwise, the previous estimate is retained. The procedure ends when all users have been scanned as pivots.

With regards to the computational burden, we see that the complexity is $\mathcal{O}(N^2)$. Moreover, in view of its looping structure, the algorithm is open to parallelization, which can be useful when working with large networks. With regards to the inference performance, we see that, under the botnet identification condition in (14), all checks performed by the algorithm will give the right answer, with probability tending to 1 as t goes to infinity. The BotBuster algorithm is accordingly expected to provide a *consistent* estimator of the underlying botnet. To see this, let us introduce a quantitative measure of the inference performance. With reference to a network $\mathcal{N} = \{1, 2, \dots, N\}$, containing a botnet $\mathcal{B}$, and letting $\hat{\mathcal{B}}(t)$ be the botnet estimated at time t by the BotBuster algorithm, we introduce the two performance indices:

$$\eta_{\text{bot}}(t) = \frac{\mathbb{E}[|\hat{\mathcal{B}}(t) \cap \mathcal{B}|]}{|\mathcal{B}|}, \quad \eta_{\text{nor}}(t) = \frac{\mathbb{E}[|\hat{\mathcal{B}}(t) \cap (\mathcal{N} \setminus \mathcal{B})|]}{|\mathcal{N} \setminus \mathcal{B}|}, \tag{15}$$

namely, the expected fraction of *correctly banned users* (i.e., discovered bots), and the expected fraction of incorrectly-banned users (i.e., normal users erroneously declared as bots). Clearly, $\eta_{\text{bot}}(t)$ (resp., $\eta_{\text{nor}}(t)$) is not defined when $\mathcal{B} = \emptyset$ (resp., when $\mathcal{B} = \mathcal{N}$). We would like to see $\eta_{\text{bot}}(t) \to 1$, and $\eta_{\text{nor}}(t) \to 0$ as t goes to infinity. Under the ideal assumption that the condition in (14) is always verified, such requirement is in fact fulfilled, as stated, without proof, in the following theorem.

Algorithm 1: $\hat{\mathcal{B}}_{\text{new}}$=BotBuster

```
$\mathcal{N} = \{1, 2, \dots, N\}$; $\hat{\mathcal{B}}_{\text{new}} = \emptyset$;
for $b_0 \in \mathcal{N}$ do
    $\hat{\mathcal{B}} = \{b_0\}$;
    for $j \in \mathcal{N} \setminus \{b_0\}$ do
        if $\hat{\rho}(\hat{\mathcal{B}} \cup \{j\}) < \gamma(\hat{\mathcal{B}}, \{j\})$ then
            $\hat{\mathcal{B}} = \hat{\mathcal{B}} \bigcup \{j\}$;
        end
    end
    if $|\hat{\mathcal{B}}| > \max(1, |\hat{\mathcal{B}}_{\text{new}}|)$ then
        $\hat{\mathcal{B}}_{\text{new}} = \hat{\mathcal{B}}$;
    end
end
```

Theorem 2 (Consistency of BotBuster). *Consider a network $\mathcal{N} = \{1, 2, \dots, N\}$, containing a botnet $\mathcal{B}$ (the case $\mathcal{B} = \emptyset$ is admitted), launching a randomized DDoS attack. The bots' transmission policies are either synchronous with constant transmission rate, or independent Poisson processes, while the normal users' transmission policies are arbitrary. If condition (14) holds, then, for any finite emulation dictionary rate α, the BotBuster algorithm is consistent, namely,*

$$\boxed{\lim_{t \to \infty} \eta_{\text{bot}}(t) = 1, \qquad \lim_{t \to \infty} \eta_{\text{nor}}(t) = 0} \tag{16}$$

■

Were the condition in (14) exactly verified, extracting the estimated botnet with highest cardinality is unnecessary. However, in real-world applications, assuming that (14) is verified *for all* normal/normal, and botnet/normal interactions, as well as *for all* time epochs, is definitely an *over-idealized* assumption. In practice, spurious clusters of normal users might be erroneously included in the botnet estimated by the algorithm. What is expected to remain true is that the occurrence of such cases is rare and that the spurious clusters are small. Since DDoS with small botnet sizes make little sense, estimated botnets of unreasonably small cardinality should be easily ruled out by the maximum-extraction performed by the algorithm, and the final estimate should contain the true botnet plus, possibly, a small fraction of normal users. Therefore, *even under non-ideal conditions*, it is expected that $\eta_{\text{bot}}(t)$ converges to 1 as time progresses, whereas $\eta_{\text{nor}}(t)$ possibly takes on some relatively small value.

V. Experimental Results

The analysis conducted so far is mainly theoretical, and, hence, relies upon several assumptions. Thus, when dealing with real network traces, and with challenging DDoS attacks, the operational validity of the algorithm BotBuster is not at all obvious. This fact motivates the experimental analysis that we are going to illustrate. A popular e-commerce website has been selected as target destination of the attack. About 20 minutes of (application-layer) traffic have been collected, from 10 students and/or researchers of our laboratory, carrying on their surfing activity almost independently. The streams have been partitioned into 2-minutes chunks.

The DDoS attack has been generated as described in Sec. II. Given the ensemble of distinct messages obtained from the *whole* recorded activity, at epoch t, the emulation dictionary $\mathscr{E}(t)$ is constructed by taking the first $\lfloor e_0 + \alpha t \rfloor$ messages of such ensemble (e_0 is the initial dictionary size and α is the EDR). Independently at each

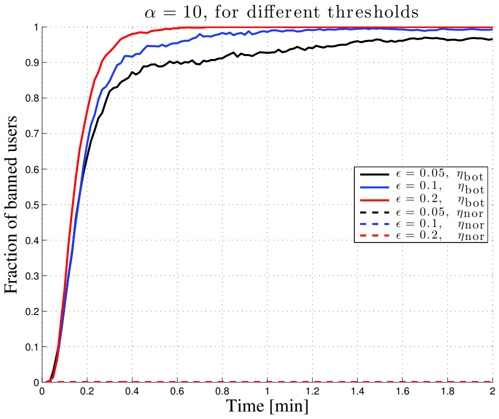

Fig. 1. Fraction of banned users as a function of time, estimated over 100 Monte Carlo runs, for different values of ϵ. The network is made of 10 normal users and 10 bots. Solid curves refer to $\eta_{\rm bot}(t)$, dashed curves to $\eta_{\rm nor}(t)$.

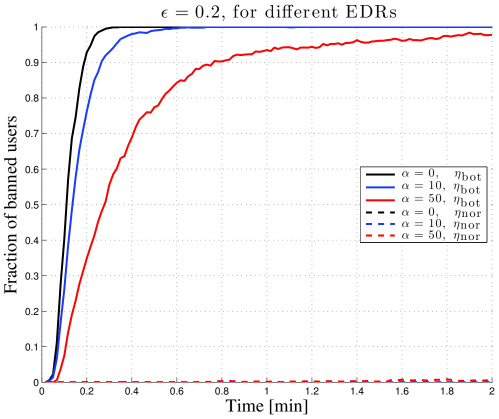

Fig. 2. Fraction of banned users as a function of time, estimated over 100 Monte Carlo runs, for different values of α. The network is made of 10 normal users and 10 bots. Solid curves refer to $\eta_{\rm bot}(t)$, dashed curves to $\eta_{\rm nor}(t)$.

bot, a Poisson time-scheduling is randomly generated and, per each transmission epoch, each bot picks at random from the emulation dictionary available at that transmission epoch.

In Fig. 1 we display the results corresponding to a network with 10 normal users and 10 bots using an emulation dictionary with $e_0 = 100$ and $\alpha = 10$. We remark that such a value is compatible with some of the empirical values $\hat{\alpha}$ estimated over the normal users' traces, which implies that no particular inference could be made by simply examining the single-user behavior. The points of each curve refer to the output of the algorithm taken each 1 seconds, and the performance is averaged over 100 Monte Carlo trials. Per each trial, the trace of a given normal user is chosen at random among its available 2-minutes chunks. The algorithm is run for three values of the threshold parameter $\epsilon \in (0, 1)$. First, we remark that the dashed curves are in practice invisible, implying that the fraction of erroneously banned users $\eta_{\rm nor}$ is almost zero for all the considered values of ϵ. Such an evidence suggests that the spurious-and-small estimated clusters containing normal users can be efficiently ruled out by the fact that the algorithm selects, as a final estimate, only the cluster with maximum size, which is expected to contain only bots. Let us then switch to the analysis of $\eta_{\rm bot}$. The average fraction of correctly identified bots is relatively high ($> 80\%$), even at the beginning of the monitoring activity. Then, $\eta_{\rm bot}$ increases, approaching unity as time progresses, in perfect accordance with Theorem 2. Moreover, the performance increases as ϵ increases from 0.05 to 0.2. In fact, increasing ϵ makes easier staying *below* the threshold, which in turn facilitates the inclusion in the estimated botnet. The excellent performance delivered by the algorithm when the threshold parameter spans the range $(0.05, 0.2)$ reveals that the algorithm is so flexible that a fine tuning of the threshold parameter can be avoided (recall that $0 < \epsilon < 1$, and that ϵ must be small, since otherwise (14) would be clearly violated). We conclude that the choice of the threshold is not critical.

In Fig. 2, the incidence of the EDR on the algorithm performance is examined. We considered $e_0 = 100$, and three values of the EDR α. First, irrespectively of the value of α, $\eta_{\rm nor}$ stays approximately constant at 0, which matches our previous evidences. Switching to $\eta_{\rm bot}$, we see again that all the curves increase as time progresses, eventually approaching unity. With regards to the dependence of performance upon the EDR, we see that the curves corresponding to $\eta_{\rm bot}$ move upward as α is decreased. Indeed, increasing α corresponds to increasing the learning ability (i.e., the power) of the botnet, which in turn corresponds to decreasing the inference performance of the algorithm. We remark that the uppermost curve refers to the degenerate case $\alpha = 0$, which corresponds to the well-documented (and simpler) case where the botnet uses repeatedly the same patterns.

References

[1] N. Hoque, D. Bhattacharyya, and J. Kalita, "Botnet in DDoS attacks: trends and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2242–2270, fourth quarter 2015.

[2] "Layer 7 DDoS — blocking http flood attacks." http://blog.sucuri.net/2014/02/layer-7-ddos-blocking-http-flood-attacks.html.

[3] "Taxonomy of DDoS attacks." http://www.riorey.com/types-of-ddos-attacks/#attack-15.

[4] "Global DDoS threat landscape: assaults resemble advanced persistent threats." https://www.incapsula.com/blog/ddos-global-threat-landscape-report-q2-2015.html.

[5] S. Ferretti and V. Ghini, "Mitigation of random query string DoS via gossip," *Communications in Computer and Information Science*, vol. 285 CCIS, pp. 124–134, 2012.

[6] M. Mardani, G. Mateos, and G. B. Giannakis, "Dynamic anomalography: tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 50–66, Feb. 2013.

[7] M. Mardani and G. B. Giannakis, "Estimating traffic and anomaly maps via network tomography," *IEEE/ACM Trans. Networking*, DOI: 10.1109/TNET.2015.2417809, date of publication, Apr. 2015.

[8] P. Venkitasubramaniam, T. He, and L. Tong, "Anonymous networking amidst eavesdroppers," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, Jun. 2008.

[9] T. He and L. Tong, "Distributed detection of information flows," *IEEE Trans. Inf. Forensics and Security*, vol. 3, no. 3, pp. 390–403, Sep. 2008.

[10] T. He, A. Agaskar, and L. Tong, "Distributed detection of multi-hop information flows with fusion capacity constraints," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3373–3383, Jun. 2010.

[11] J. Kim and L. Tong, "Unsupervised and nonparametric detection of information flows," *Signal Processing*, vol. 92, no. 11, pp. 2577–2593, Nov. 2012.

[12] S. Marano, V. Matta, T. He, and L. Tong, "The embedding capacity of information flows under renewal traffic," *IEEE Trans. Inform. Theory*, vol. 59, no. 3, pp. 1724–1739, Mar. 2013.

[13] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of Byzantine attacks," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 16–29, Jan. 2009.

[14] S. Marano, V. Matta, and P. Willett, "Distributed detection with censoring sensors under physical layer secrecy," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1976–1986, May 2009.

[15] B. Kailkhura, S. Brahma, B. Dulek, Y. S Han, and P. Varshney, "Distributed detection in tree networks: Byzantines and mitigation techniques," *IEEE Trans. Inf. Forensics and Security*, vol. 10, no. 7, pp. 1499–1512, Jul. 2015.

[16] H. Shao, *Mathematical Statistics*, 2nd ed., Springer, 2003.

[17] S. Ross, *Stochastic Processes*, 2nd ed. New York: John Wiley & Sons, Inc., 1996.

[18] V. Matta, M. Di Mauro, and M. Longo, "DDoS attacks with randomized traffic innovation: botnet identification challenges and strategies," *submitted for publication*. Available as arXiv:1606.03986 [cs.IT], Jun. 2016.